Probabilistic Treatment Planning with Polynomial Chaos Expansion for Proton Therapy

BY

J.H. SALVERDA

to obtain the degree of Master of Science in Applied Physics,
to be defended publicly on Tuesday June 18, 2019 at 03:00 PM.

Abstract

Radiotherapy is an important treatment type for patients with cancer. An advantage of state of the art proton therapy with respect to traditional photon therapy is the spatial energy deposition of protons, which is characterized by the Bragg peak. Due to this particular course of energy deposition, the tumor can be irradiated more precisely and as a consequence more healthy tissue can be spared. A drawback is that proton therapy is more sensitive to uncertainties like misalignment of the patient, which is referred to as a set-up error. Robust optimization is the current way to account for such uncertainties during which a few discrete error scenarios are included in the planning. Due to the nature of the uncertainties, probabilistic optimization, on the other hand, is more promising, since it can handle many more scenarios with their occurrence probability taken into account as well.

To investigate probabilistic planning, Polynomial Chaos methods were used in this research. Polynomial Chaos can approximate a stochastic response, depending on for example the set-up error, by a series expansion in terms of polynomials. The advantage is that typically the Polynomial Chaos Expansion (PCE) can be evaluated much faster than the stochastic response itself. In this research an investigation is done how PCE can be used with probabilistic treatment planning for proton therapy.

First, an overview is given on how PCE can be used to describe probabilistic objective functions. It has been found that in general it is possible to describe random set-up errors with PCE without the need of considerably additional computational power. Moreover, polynomials in dose are especially suitable when fractionation is taken into account.

Second, a simple one-dimensional geometry consisting of a tumor with a surrounding organ has been considered. The effects of fractionation on the dose distribution has been calculated with PCE and the results were in accordance with earlier research by Unkelbach et al. (2018).

Third, a simplified three-dimensional geometry which mimics a spinal cord with surrounding tumor has been optimized probabilistically with PCE. The expected value of the sum of the quadratic differences between the dose and the prescribed dose has been minimized. The effects of probabilistic optimization became clear from both the nominal and the expected dose distribution. This objective function (i.e. the expected value of a quadratic function of the dose) turned out to be extremely suitable for probabilistic optimization with PCE, however its clinical relevance is less clear.

Last, a real skull base meningioma patient has been considered. As objective function, a percentile of the dose volume parameter has been used with PCE. This function demonstrated a bumpy dependence on the beam intensity and it has been proven that it cannot have a negative gradient. Consequently, a high step size is recommended to approximate the gradient and, if needed, the Hessian. Probabilistic treatment plans were obtained with a single probabilistic objective, with two probabilistic objectives and with nominal constraints and objectives on OARs.

The results have been compared with a robust treatment plan according to the recipes of Ter Haar et al. (2018). It has been found that the probabilistic treatment plan shows a more conformal dose distribution, since it does not exhibit any preferred direction, contrary to the robust treatment plan. As a consequence, the automatic extension of the tumor is smaller. A clear benefit has been demonstrated in favor of the probabilistic treatment plan. Besides that, a major advantage of probabilistic optimization with respect to robust optimization is that it makes it possible to directly steer to a probabilistic goal independent of the patient and the treatment site. The computation time of probabilistic planning (in the order of 1-3 weeks) is unfortunately still too high for clinical practice.

Acknowledgements

Contents

Nomenclature

Symbols

$\delta_{k,l}$ Kronecker delta

Δ_{lev} Increased quadrature level for the single dimensional directions of a Smolyak sparse grid

μ_ξ Expected value of uncertainty parameter ξ

σ_ξ Standard deviation of uncertainty parameter ξ

$\boldsymbol{\xi}$ Uncertainty vector

$\boldsymbol{\xi}^{(s)}$ Realization or random sample of uncertainty vector

$\boldsymbol{\xi}^{(t)}$ Uncertainty vector for fraction t

$\tilde{\boldsymbol{\xi}}$ Combined uncertainty vector for multiple fractions

$\mathcal{B}$ Index set of beamspots

$D_{\alpha\%}$ Dose volume parameter

D Dose deposition matrix

D_{ij} ij-th element of dose deposition matrix

$\boldsymbol{d}$ Dose vector

d_i i-th element of dose vector (i.e. voxel dose)

$\mathrm{E}[\cdot]$ Expected value

$\hat{\boldsymbol{e}}_n$ Unit vector along the n-th dimension

L Grid level

$\mathcal{M}_{\boldsymbol{\xi}}[\cdot]$ Arbitrary statistical measure over uncertainty vector $\boldsymbol{\xi}$

$\mathcal{N}(\mu, \sigma^2)$ Gaussian distribution with mean μ and standard deviation σ

N_{b} Number of beamspots

N_{s} Number of samples

N_v	Number of voxels
O	Polynomial order
$P^i_{t\%}\{\cdot\}$	t-th percentile of the set of elements distinguished by the index i
SR	Set-up robustness
$\mathcal{T}$	Index set of fractions
T	Number of fractions
$\mathcal{V}$	Index set of voxels
$\mathrm{Var}(\cdot)$	Variance
$\boldsymbol{x}$	Beam intensity vector
x_j	j-th element of beam intensity vector (i.e. beamspot intensity)

Abbreviations

CDF	Cumulative distribution function
CT	Computed tomography
CTV	Clinical target volume
DPH	Dose population histogram
DVH	Dose volume histogram
GTV	Gross tumor volume
IMPT	Intensity modulated proton therapy
$LxEyOz$	$L = x$, $\Delta_{lev} = y$ and $O = z$
MR	Magnetic resonance
OAR	Organ-at-risk
PCE	Polynomial Chaos Expansion
PDF	Probability density function
PTV	Planning target volume

Chapter 1

Introduction

Approximately half of all cancer patients undergo radiotherapy treatment during their course of illness [3–5], either for cure or as a palliative treatment to relieve the patients from symptoms such as pain caused by the cancer. In 2008 the mortality numbers from cancer in Europe are estimated to be 1.7 million deaths [6]. Generally, the most common types of treatment for cancer are surgery, chemotherapy and radiotherapy [7]. Each of these has its own advantages and disadvantages. For example, surgery could spare healthy tissue in contrast to radiotherapy. On the other hand, radiotherapy is non-invasive which might be preferable for certain patients. This research will focus on the latter, radiotherapy. Radiation therapy may be given if the patient is too old to undergo surgery or if surgery is technically not possible due to the location of the tumor. Besides, radiotherapy against microscopic extensions of the tumor is common as well.

There are multiple types of radiotherapy. The patient could for example be irradiated with photons, protons and carbon-ions [8]. Nowadays, photon therapy is the conventional way of radiotherapy: many more patients are treated with photons than with protons for example. However, proton therapy is upcoming in the Netherlands. Recently the first three proton therapy centers have been opened in Groningen, Delft and Maastricht and the first patients have already been treated. However, proton therapy is considerably older, it was first proposed by Wilson in 1946 [9, 10].

Unlike photons, protons have a mass and are positively charged. The main difference between photon and proton therapy is their way of depositing their energy to the body. Photons deliver their energy gradually along their trajectory. On the other hand, protons undergo stronger interactions and the energy of a proton beam is released at a more specific depth along its path. This typical curve is called the Bragg peak. Figure 1.1 shows the difference between the energy deposition of a photon and a proton beam. It can be seen that after the Bragg peak of a proton, almost no energy is deposited anymore. On the contrary, after the proximal peak of the photon beam, relatively much energy is still deposited.

The prompt drop after the Bragg peak is in principle one of the advantages of proton therapy with respect to conventional radiotherapy. More healthy tissue can be spared this way. However, this also causes proton therapy to be much more sensitive to uncertainties. Some examples of uncertainties are errors in patient alignment and range errors of the proton beam.

For photon therapy it is common that a bigger volume is irradiated than only the tumor, to account for such errors. However, this approach is not applicable to proton therapy [12]. Currently, so-called *robust treatment planning* is used to account for these kind of errors in
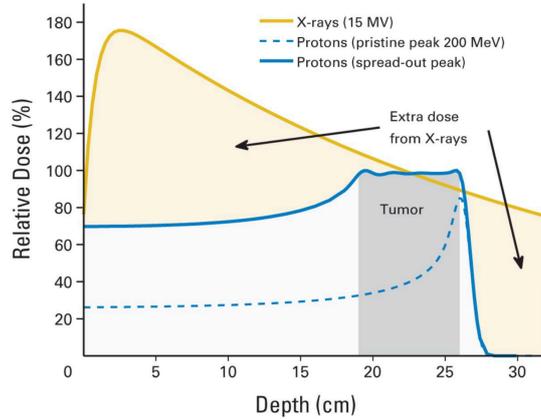
Figure 1.1: Comparison of the energy deposition of photon and proton beams. On the x- and y-axis respectively the depth along the trajectory in the body and the fraction of the deposited dose are shown. The dashed line corresponds to a single Bragg peak of a proton beam. Combining such single Bragg peaks could result in the solid blue line, a spread-out Bragg peak (SOBP). Figure from [11].

proton therapy. It defines a few error scenarios. An error scenarios is a specific realization of the considered uncertainties. Robust treatment planning performs a worst-case optimization over the error scenarios and it minimizes the worst-case of the objective function. This means that for the resulting treatment plan an adequate target coverage is guaranteed in all predefined error scenarios. However, the probability that one of these scenarios actually takes place during treatment is not taken into account and in fact is exactly zero.

So-called *probabilistic treatment planning* tries to overcome these limitations of robust treatment planning and assumes the uncertainties to follow a certain probability density function. An optimization is performed which includes all possible scenarios instead of a few predefined error scenarios. This can be computationally quite expensive and so-called *Polynomial Chaos Expansion* (PCE) can alleviate the computational effort. Polynomial Chaos Expansion is a mathematical tool which serves as a meta-model for a black box model of which the input is stochastic. Van der Voort et al. (2015) and Ter Haar et al. (2018) have used Polynomial Chaos Expansion to describe the dose distribution in a patient. Besides, Van der Voort et al. (2015) devoted one chapter to probabilistic treatment planning in combination with PCE, where the conditional value-at-risk of the LTCP function was considered as objective function.

In this research, the theory, possibilities and applications of probabilistic treatment planning in combination with Polynomial Chaos Expansion for proton therapy are investigated. The structure of this thesis is as follows. In Chapter 2 proton therapy is discussed, both the theory and the application. Chapter 3 describes the working principles of treatment planning. Chapter 4 elaborates on the mathematics involved in Polynomial Chaos Expansion. Subsequently, the possible objective functions which can be used in probabilistic treatment planning in combination with Polynomial Chaos Expansion are described in Chapter 5. Then, the methodology of the applications which are investigated in this research are explained in Chapter 6. Next, the results, discussions, conclusions and recommendations for future research follow.

Chapter 2

Proton therapy

2.1 Radiotherapy

A patient diagnosed with cancer can be treated with radiation therapy. Radiation induces damage in the DNA of tumor cells. This can either be direct on the cellular molecules or indirectly using produced free radicals [5]. Both ways can lead to cell death.

Typically there are two types of radiotherapy, internal and external radiotherapy. The type of radiotherapy which is given to the patients depends on many factors: the type and size of cancer, the patient's health and medical history and other factors. Regarding internal radiotherapy, a radioactive source irradiates the patient from inside the body. Such a treatment can either be given locally of systemic. Locally means that the radioactive source is inserted in or next to the tumor. This is called brachytherapy. For systemic radiotherapy, the source travels through the patient's blood. With external radiotherapy on the other hand, the patient is irradiated by an external machine [14]. This research focuses on external radiotherapy.

Besides the way the patient is irradiated, internally or externally, there are also different types of particles which are used as radiation, for example photons, protons and electrons. Most commonly, photon therapy is used. It consists of electromagnetic waves, which are massless. Besides, photons are charge-less and undergo few interactions. As a consequence, a substantial part of the photon beam passes through the patient. The dose deposition profile of photons has a specific shape. There is a pristine peak in dose deposition, after which the dose deposition gradually decreases. This is shown in Figure 1.1.

Besides photons, also protons are used for radiation therapy. Protons are positively charged particles with a mass. Protons exhibit a very typical dose deposition profile. In contrast to photons, the entrance dose is low and at a certain depth there is a sharp and high peak, which is called the Bragg peak. After the peak, the energy deposition quickly decreases to zero. The depth at which the Bragg peak occurs is dependent on the initial energy of the proton. Since tumors are typically bigger than the width of the Bragg peak in any dimension, multiple energies should be combined to form a so-called *spread-out Bragg peak*. The energy deposition of a proton beam is also shown in Figure 1.1 in Chapter 1. From that figure, it becomes clear that the dose given to normal tissue is lower for protons, partly because of the lower entrance dose and partly because of the distal fall-off of the dose deposition after the Bragg peak.

2.2 Physics of the proton beam

When a proton propagates through matter, it undergoes different kind of interactions with electrons or nuclei. In this section, the most dominant interactions are briefly discussed.

The inelastic Coulombic interaction of a proton with an electron frequently occurs. It is a major cause for the decrease in kinetic energy of the proton. As the rest mass of a proton is 1832 times greater than the rest mass of an electron, most protons are not deflected by this interaction [15]. A schematic representation of this interaction is shown in Figure 2.1a. The elastic Coulombic interaction of a proton with a nucleus is caused by the repulsive force between the positively charged proton and nucleus. This force will deflect the trajectory of the proton. A schematic representation of this interaction is shown in Figure 2.1b. The non-elastic nuclear reactions with a nucleus are less frequent [15]. The proton enters the nucleus and subsequently a gamma ray, a proton, a neutron or another ion might be emitted. A schematic representation of this interaction is shown in Figure 2.1c.

Thus, these interactions cause a proton to be slowed down, to be deflected or to be absorbed. Whether an interaction actually occurs is not deterministic but stochastic. The path of each proton is therefore different. As a consequence, in proton therapy it is more common to consider proton beams instead of single protons. The range of the beam is defined as the depth at which half of the protons have no kinetic energy anymore [15]. The stopping power of the proton beam, defined as the energy loss per unit length, increases for lower energies [16]. This relation is approximately proportional to the inverse of the square of the velocity [15, 17], explaining the Bragg peak.

The range and dose deposition profile of a proton beam (i.e. the Bragg curve) can be computed with the stopping power. The stopping power itself also depends on the physical properties of the absorbing material: the electron density and the effective atomic numbers. To calculate the stopping power a conversion needs to be made from Hounsfield units of a *Computed Tomography* (CT) scan to units of stopping power. In this conversion many types of uncertainties can arise [18]. For example, research has shown that variations in the tissue composition lead to an increase of the uncertainty in the computed stopping power [19]. Calculating the stopping power with a mono-energetic CT scan is also an important source of uncertainty [20].
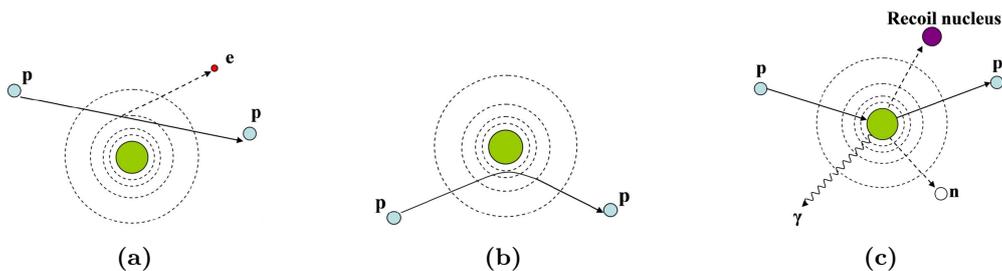
Figure 2.1: Dominant types of interactions of a proton with matter. **(a)** Inelastic Coulombic scattering. **(b)** Elastic Coulombic scattering. **(c)** Non-elastic nuclear reactions. Figures from [15].

2.3 Intensity modulated proton therapy

This research focusses on so-called *intensity modulated proton therapy* (IMPT). It is based upon the pencil beam scanning technique. A pencil beam is a narrow unscattered beam of protons

which can be aimed at any location within the tumor. Since protons are positively charged particles, their direction and kinetic energy can be controlled by electric and magnetic fields. Each pencil beam, or beamspot, can be characterized by its radiation angle, lateral position and energy. For intensity modulated proton therapy, a weight is assigned to each pencil beam, also called the intensity. This weight is proportional to the number of protons in the beam and therefore also to the deposited dose. These pencil beams and their optimal weights can deliver a dose distribution which is conformal to the tumor shape and has the prescribed dose levels. The essence of IMPT is finding the optimal weights of the beamspots.

2.3.1 Gantry

The patient is irradiated within the so-called gantry. This machine is able to irradiate the tumor from multiple angles. The couch, on which the patient lies, as well as the gantry itself is able to rotate. Every possible radiation angle can be defined by the combination of the couch and gantry angle. Besides the angle, also the lateral position of the beam can be varied.

2.4 Skull base meningioma

In this research, the focus will be on a patient with a skull base meningioma, which can be either benign or malignant [21]. Typically, it is not possible to completely remove such a tumor with surgery. Radiotherapy is a common treatment for patients with skull base meningioma, either alone or after subtotal resection [22]. The survival rate after 10 years is approximately 70% [22, 23]. Usually, the treatment is given in multiple fractions, such that healthy tissue has the opportunity to repair itself as much as possible between the fractions. The number of fractions typically ranges from 25 to 30 [24].

Due to the characteristics of proton therapy as described in this chapter, proton therapy is preferred over photon therapy for this type of tumor [25, 26].

2.5 Uncertainties

As discussed in Section 2.2, certain uncertainties could arise when the stopping power is computed. This leads to uncertainties in the calculated range of the proton beam. Different types of uncertainties within proton therapy are treated in this section.

2.5.1 Set-up error

A set-up error is due to a misalignment of the patient, either during the CT scan or during the treatment. A set-up error could be split in two components, a systematic component and a random component. The systematic component is constant for all fractions of a treatment. The random component differs between fractions. Figure 2.2 shows a schematic representation of the components of a set-up error. Two realizations of a systematic set-up error with each three realizations of a random set-up error are shown.

Systematic set-up error

For treatment planning a CT scan is made of the patient. When the patient is misaligned during this CT scan, this error will cause a shift in the finally deposited dose, since the treatment plan
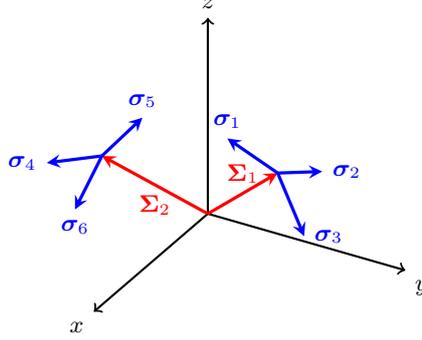
Figure 2.2: A set-up error can be decomposed into a systematic and a random component. Two realizations of a systematic error are shown, each with three realizations of a random error. Note that in this figure $\boldsymbol{\Sigma}_i$ and $\boldsymbol{\sigma}_i$ represent realizations instead of standard deviations of respectively the systematic and random set-up error.

is made according to that particular CT scan. This is called a systematic set-up error, since it is present and constant for all fractions of the treatment. Another cause for a systematic set-up error is the motion of the skin with respect to the organs [27].

A systematic set-up error can be denoted by $\boldsymbol{\xi}$ and generally has three components ξ_1, ξ_2 and ξ_3 corresponding to the three dimensions of Euclidean space, measured in millimeter (mm). These components are stochastic variables and are assumed to have a certain probability density function (PDF). For this research these components are assumed to be Gaussian distributed with zero mean and zero mutual correlation, which is in accordance with clinical data [28]. This can be expressed as in Equation (2.1). Here, $\mathrm{diag}(\boldsymbol{k})$ denotes the diagonal matrix with elements of the vector $\boldsymbol{k}$ on the main diagonal.

$$\boldsymbol{\xi} = (\xi_1, \xi_2, \xi_3)^T \sim \mathcal{N}\left(\mathbf{0}, \mathrm{diag}\left(\Sigma_x^2, \Sigma_y^2, \Sigma_z^2\right)^T\right) \tag{2.1}$$

Random set-up error

In contrast to systematic errors, random errors are not constant over fractions but differ. This error arises if the patient during a fraction is not positioned the way it was during the CT scan. A random error can also be denoted by $\boldsymbol{\xi}$ and also has three components in Euclidean space. Analogously to systematic errors, these components of the random error are also assumed to be described by Gaussian distributions with zero mean and zero correlation, as shown in Equation (2.2).

$$\boldsymbol{\xi} = (\xi_1, \xi_2, \xi_3)^T \sim \mathcal{N}\left(\mathbf{0}, \mathrm{diag}\left(\sigma_x^2, \sigma_y^2, \sigma_z^2\right)^T\right) \tag{2.2}$$

Although the same variable $\boldsymbol{\xi}$ is used for random errors and systematic errors, they are inherently different. In general, $\boldsymbol{\xi}$ is used to denote any uncertainty.

2.5.2 Range error

When the computed range of the proton beam does not coincide with the actual range, one can speak of a range error. There are multiple causes for range errors. It could be due to the conversion from Hounsfield units to stopping power, but wrong calibration of the machine could

give rise to range errors as well. A distinction is made between absolute and relative range errors. Here only relative range errors are described.

Relative

The Hounsfield scale resulting from a CT scan can be used directly for photon therapy. However, for proton therapy a conversion to stopping power is required, as already mentioned. This conversion is not exact, but is based upon algorithms or Monte Carlo simulations [29]. Both introduce uncertainties. The relative range error originates from this conversion of Hounsfield scale of the CT scan to proton stopping power. This type of range error is systematic, since it is constant over fractions. A consequence of the relative range error is that the computed range of the proton beam is an over- or underestimate of the real range.

2.5.3   Delineation error

A CT image gives the physician an insight on where the tumor and the organs are situated in the patient's body. On every slice of the CT scan another part of the body is visible. Before the patient gets radiotherapy, the tumor and other organs need to be delineated by a physician.

The tumor visible on the CT scan is called the *gross tumor volume* (GTV). Due to finite resolution and sometimes low contrast of CT scans, the delineated contour of the GTV demonstrates inter-observer variation. This is a type of delineation error. Research has assessed the inter-observer variability of the delineation of the GTV for brain tumors [30].

Tumors may be delineated on *magnetic resonance* (MR) scans if these are not visible on CT scans. The geometric distortion of MR images is another cause of delineation uncertainty.

The *clinical target volume* (CTV) contains the visible tumor or the GTV including some margins around to include possible microscopic extensions of the tumor, which are not visible on the scans. As one might expect, due to interpretation differences among physicians, the delineated contour of the CTV is physician specific. This is also a type of delineation error.

Several studies have been done to quantify this phenomenon for breast cancer [31–33]. In these articles both inter- and intra-observer variation are examined quantitatively.

When a patient gets a second CT scan, so-called auto-segmentation software can be used to automatically delineate the tumor, based on the delineation done on the first CT scan. The CTV delineated by the software might also differ from one delineated by a physician. This is another cause of variation in tumor contours [34].

2.6   Handling uncertainties

As discussed, uncertainties such as set-up errors and range errors can give rise to a shift or distortion of the desired dose profile.

Conventional treatment planning does not account for possible uncertainties. The plan is made with the assumption that the patient will be at exactly the right position during treatment and the range of the radiation particles will be exactly as calculated. However, if any error occurs, the treatment plan will not be robust against this uncertainty, which means that the treatment goals, such as a high dose in the CTV, are not achieved.

To account for uncertainties in photon therapy, a bigger volume than the CTV is irradiated [27]. It is based upon the idea that the CTV will be covered by this bigger volume, regard-
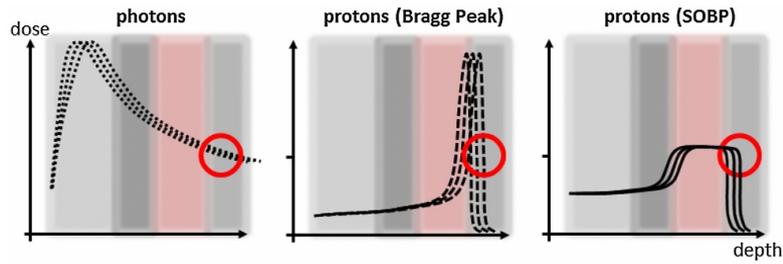
Figure 2.3: The impact of a range error on the dose deposition profile for a photon beam and for a proton beam. Figure from [36].

less of any errors. This bigger volume is called the *planning target volume* (PTV). It entails the CTV including some margin.

Research has shown that handling uncertainties using the PTV approach does not result in proper treatment plans with IMPT [35], since the energy deposition of protons highly depends on the density variations inside the body. Besides, possible range errors have a bigger impact on the dose deposition for protons than for photons. This is caused by the steeper slope of the dose deposition along the depth. This effect of a range shift on the dose deposition for photons and protons is shown in Figure 2.3.

Chapter 3

Treatment planning

In this chapter all important aspects related to treatment planning are discussed. First an overview is given of the important treatment parameters, which is followed by an elaboration on treatment planning itself.

3.1 Clinical treatment parameters

Treatment parameters are characteristics of a treatment plan. They can for example give information about the desired dose or about the actual given dose.

3.1.1 Beam intensity vector

The patient can be irradiated from multiple angles, multiple positions and with multiple energies. Each unique combination of angle, position and energy is called a beamspot, characterized by the index $j \in \mathcal{B} = \{1, ..., N_\mathrm{b}\}$ with N_b the number of beamspots. The beam intensity vector contains the intensities corresponding to the different beamspots, measured in Monitor Unit (MU):

$$\boldsymbol{x} = (x_1, ..., x_{N_\mathrm{b}})^T \ . \tag{3.1}$$

The elements of the beam intensity vector are physically constrained to be non-negative [37], since extracting energy from the body is not possible with radiotherapy. A beamspot is also referred to as a pencil beam.

3.1.2 Uncertainty vector

Uncertainties like the set-up error and the range error are denoted by the uncertainty vector $\boldsymbol{\xi}$. The number of elements of this vector depends on the uncertainties which are considered. For example, when the range error and the three-dimensional systematic set-up error are taken into account, the number of elements of this vector equals four.

Regarding terminology, the scenario in which no errors occur ($\boldsymbol{\xi} = \boldsymbol{0}$) is referred to as the nominal scenario.

3.1.3 Dose deposition matrix

One can calculate the contribution of a beamspot to the deposited dose in a certain voxel. This can be represented by a matrix D with elements D_{ij}. Here, $i \in \mathcal{V} = \{1, ..., N_v\}$ corresponds to a certain voxel and $j \in \mathcal{B} = \{1, ..., N_b\}$ corresponds to the beamspot, with N_v and N_b the number of voxels and beamspots respectively. Thus, D_{ij} is the contribution of beamspot j to the dose received by voxel i, measured in Gray per Monitor Unit $(\text{Gy} \cdot \text{MU}^{-1})$.

$$D = [D_{ij} : i \in \mathcal{V}, j \in \mathcal{B}] \tag{3.2}$$

The elements of the dose deposition matrix are non-negative. The dose deposition matrix is the variable where the uncertainty in set-up and range is introduced. After all, the contribution of a beamspot to the dose in the different voxels changes if for example the patient is shifted. Therefore, a more correct notation would be $D = D(\boldsymbol{\xi})$.

3.1.4 Dose distribution

The dose or dose distribution, measured in Gray (Gy), received by a certain structure is the matrix-vector product of the dose deposition matrix and the beam intensity vector:

$$\boldsymbol{d}(\boldsymbol{x}, \boldsymbol{\xi}) = D(\boldsymbol{\xi})\boldsymbol{x} \quad \text{or} \quad d_i(\boldsymbol{x}, \boldsymbol{\xi}) = \sum_{j \in \mathcal{B}} D_{ij}(\boldsymbol{\xi})x_j . \tag{3.3}$$

3.1.5 Prescribed dose

The prescribed dose represents the dose which is desired in a certain voxel or structure. As one might expect, the prescribed dose in the tumor is usually high, while the prescribed dose in organs, which need to be spared, is low. A dose can either be prescribed to a structure as a whole or to specific voxels within. The former is denoted by:

$$d^{\text{p}} , \tag{3.4}$$

while the latter is denoted by:
$$\boldsymbol{d}^{\text{p}} = \left(d_1^{\text{p}}, ..., d_{N_v}^{\text{p}}\right)^T . \tag{3.5}$$

3.1.6 Dose volume parameters

To examine the quality of a treatment plan, it might be desirable to have information about which volume fraction of a structure receives a certain dose. This kind of information is contained in the parameter $D_{\alpha\%}$, with $\alpha \in [0, 100]$. It is defined as the maximum dose that is received by at least $\alpha\%$ of the volume. In other words: the maximum lower bound of the dose received by $\alpha\%$ of the volume. The formula for the parameter $D_{\alpha\%}$ is as follows:

$$D_{\alpha\%} = P_{(100-\alpha)\%}^i \{d_i : i \in \mathcal{V}\} , \tag{3.6}$$

where the operator $P_{t\%}^i \{\cdot\}$ is used to denote the t-th percentile of the set of values distinguished by the index i. For convenience, the dependence of $D_{\alpha\%}$ and d_i on the beam intensity vector $\boldsymbol{x}$ and uncertainty vector $\boldsymbol{\xi}$ is omitted.

3.1.7 Dose volume histogram

The dose volume histogram (DVH) is a graphical representation of the dose volume parameters. It summarizes the three-dimensional dose distribution in a two-dimensional graph. Spatial information of the dose distribution in the structure is not contained in the graph. It is a plot of the volume fraction $\alpha\%$ against the dose volume parameter $D_{\alpha\%}$. It can be interpreted as a cumulative histogram of the dose in the voxels d_i. Figure 3.1 shows an example of a dose volume histogram, for the GTV and the right hippocampus. Ideally, the shape of the DVH for the tumor would be a horizontal line with a sudden drop at a high dose. On the other hand, the ideal shape of the DVH of a non-tumorous organ would be opposite, a steep fall-off followed by a horizontal tail.
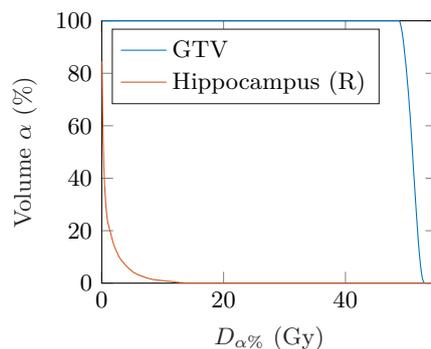
Figure 3.1: Dose volume histogram for the GTV and the right hippocampus.

3.1.8 Dose population histogram

The dose population histogram (DPH) is a curve which gives information about the probability distribution of a parameter. It is the same type of graph as a dose volume histogram, but on the vertical axis the probability or the population is presented instead of the volume. A data point (x, y) on the curve of a DPH means that the parameter of interest equals at least x with a probability of y. It can be interpreted as a cumulative distribution function, which is mirrored with respect to the horizontal line with cumulative probability equal to 0.5.

A population could refer to a set of different patients, but in this research different realizations of the systematic error are meant.

3.2 Treatment planning

This section provides information about all the necessary steps of treatment planning, from the start to the end.

3.2.1 CT scan and delineation

First, a CT scan is made of the patient. Typically only a part of the body is imaged. A requirement for this CT scan is that all parts which receive dose are visible on the scan. For a meningioma patient, this will be the head and neck. All organs of interest are to be delineated by a physician.

Obviously this includes the tumor visible on the scan, the GTV. Besides, from clinical experience physicians are aware of possible microsocopic extensions of the tumor, which are not visible on the image. These expected microscopic extensions are site specific. The GTV including expected microscopic extensions are delineated as well, forming the CTV. Typically, for meningioma patients the CTV coincides with the GTV [38, 39].

Besides, so-called *organs-at-risk* (OARs) are to be delineated. These include all organs, which are to be spared from radiation. Two types of organs-at-risk can be distinguished, serial and parallel organs [40]. Serial organs lose functionality when even a small volume receives a high dose. Thus, the potential risk for serial organs is dominated by the highest dose. Examples are the spinal cord and optic nerves. On the other hand, for parallel organs a local dose within the organ only affects the functionality of that part of the organ. For parallel organs, the risk is dominated by the average dose in the entire organ. Examples are the lungs and liver.

It should be noted that the CT scan is made up of images of adjacent cross-sectional slices of the body. The tumor and organs are delineated on these two-dimensional images. Algorithms exist that combine these two-dimensional delineated images to construct a three-dimensional delineated organ.

3.2.2 Automated treatment planning using the iCycle algorithm

Erasmus Medical Center uses the in-house developed treatment planning software environment called Erasmus-iCycle in Matlab. The main functionality is the multi-criteria optimization according to a predefined wish-list. Besides, it offers a range of other features, of which some examples are briefly explained in this section.

Wish-list

The wish-list is the actual configuration of the optimization algorithm. It contains all dose related requirements of the treatment and needs to be defined by the physician or medical physicist. iCycle distinguishes two types of criteria. A constraint is a criterion that should always be met, while an objective is a criterion which has to be reached as close as possible. Constraints and objectives can be set for structures individually. For example, a constraint could be set on the minimum dose in the GTV, while an objective could be set on the mean dose in a certain organ-at-risk. Furthermore, a priority has to be assigned to each objective, which indicates the order in which the objectives will be optimized. If a higher priority is assigned to an objective, there is a higher chance the goal will be reached. An example of a simplified wish-list is shown in Table 3.1. Regarding the constraints, minimize and maximize indicate whether the limit is respectively an upper or lower bound. The application in which the wish-lists for iCycle are defined is called Lucy.

Such a wishlist is specific for the tumor site, but is not patient-specific. As a consequence, the constraints and objectives of a wish-list are different for different types of tumors, while they are similar for different patients with the same type of tumor [41].

In general, maximizing a function is equivalent to minimizing the function multiplied by -1 [42, 43]. iCycle makes use of this logic. From now on, when a statement is made about minimizing a certain function, this also applies to maximizing the opposite.

14

Table 3.1: An example of a simplified wish-list for a meningioma patient. Note the difference between constraints and prioritized objectives. Here, $\bar{d}(\boldsymbol{x}, \boldsymbol{0})$ denotes the mean dose.

Constraints		
Volume	Type	Limit
GTV	Maximize $\min_i d_i(\boldsymbol{x}, \boldsymbol{0})$	$0.95 \cdot 50.4$ Gy
Optic Nerve (L)	Minimize $\max_i d_i(\boldsymbol{x}, \boldsymbol{0})$	50 Gy
Optic Nerve (R)	Minimize $\max_i d_i(\boldsymbol{x}, \boldsymbol{0})$	50 Gy

Objectives			
Priority	Volume	Type	Goal
1	Brainstem	Minimize $\bar{d}(\boldsymbol{x}, \boldsymbol{0})$	1 Gy
2	Cerebellum	Minimize $\bar{d}(\boldsymbol{x}, \boldsymbol{0})$	1 Gy

Multi-criteria optimization

iCycle uses a multi-criteria optimization approach. The optimization consists of two phases. In the first phase, objectives are optimized in decreasing order of priority. First, the first objective is optimized, within the prescribed constraints. Once the assigned goal has been reached, this objective is set into a constraint. This constraint is slightly relaxed with respect to the reached value of the objective, in order to leave room for optimization of other objectives. Subsequently, the second objective is optimized. However, for this optimization there is one more constraint which should be met. This extra constraint originates from the first objective. Once the second objective has reached its goal, this objective is set into a constraint as well. These steps are repeated until all objectives have been optimized. It might be clear that an objective with a lower priority is less likely to reach its goal, since more constraints are to be met. In the second phase, all objectives which could have been optimized in the first phase further than its initial goal, are further optimized [41]. iCycle requires the gradient and Hessian matrix of the objectives and constraints to determine the proper direction of optimization in beam intensity space. These need to be provided either as an exact representation or as an approximation using finite differencing.

Beam direction selection

iCycle is able to automatically select the optimal beams (i.e. beam directions). This is an iterative process, in which beams are added and the least contributing beams are omitted [41]. It is also possible to prescribe the beam directions manually. This information is often known from clinical experience. Besides, it saves computation time. For this research the beam directions are prescribed manually.

Beamspot resampling

The beamspots corresponding to a certain beam direction can be selected automatically in a similar way. An iterative process entails the addition of new spots and removal of the least contributing spots [44, 45]. This process is called pencil beam resampling or beamspot resampling.

Voxel reduction

Another feature of iCycle is called voxel reduction. It randomly picks a representative subset of the voxels of a certain structure for computational reasons. This saves memory and computation time. This feature can be regulated per organ.

3.2.3 Conventional treatment planning

Conventional treatment planning has already been discussed briefly in Chapter 2. In this section a more detailed explanation is given.

Conventional treatment planning does not account for possible errors to occur. This means that uncertainties in for example the set-up (i.e. alignment of the patient) or the proton range are not considered in the treatment planning.

Consider an objective function $f(\boldsymbol{x}, \boldsymbol{\xi})$. First of all, this function depends on the beam intensity vector $\boldsymbol{x}$, but it also depends on the uncertainties $\boldsymbol{\xi}$ via the dose deposition matrix $D(\boldsymbol{\xi})$. When this objective function is minimized in conventional treatment planning, mathematically this could be expressed as follows:

$$\underset{\boldsymbol{x}}{\text{minimize}}\ f(\boldsymbol{x}, \boldsymbol{0})\,. \tag{3.7}$$

Note the absence of the uncertainty dependence due to the substitution of $\boldsymbol{\xi} = \boldsymbol{0}$. The resulting treatment plan is guaranteed to satisfy the objective only if no error actually occurs, which is statistically never the case.

Conventional treatment planning is possible with iCycle.

3.2.4 Robust treatment planning

Robust treatment planning does account for possible errors. A finite number of predefined scenarios are considered. A scenario is a realization of the possible errors that could occur. In the minimax approach to robust optimization, a worst case optimization is performed. This means that the worst case value of a certain objective is minimized [2, 13, 46], within the possibly imposed constraints. This can be expressed mathematically as follows:

$$\underset{\boldsymbol{x}}{\text{minimize}}\ \underset{s}{\max}\left\{f(\boldsymbol{x}, \boldsymbol{\xi}^{(s)}) : s \in \mathcal{S}\right\}\,, \tag{3.8}$$

where $f(\boldsymbol{x}, \boldsymbol{\xi})$ is an arbitrary objective function and where $\mathcal{S}$ denotes the set of considered scenarios.

Satisfaction of the objective is guaranteed when one of the scenarios in $\mathcal{S}$ actually occurs. In other words: the resulting treatment plan is robust against these scenarios. However, none of these scenarios will statistically ever happen. While it can be verified how an obtained robust treatment plan performs for other scenarios or in the probabilistic sense, a robust treatment plan is inherently robust solely against the finite number of predefined scenarios in $\mathcal{S}$.

Robust treatment planning is the current way of handling errors in proton therapy. This method has been implemented in iCycle. Since iCycle uses a multi-criteria optimization, it performs the minimax worst case optimization objective-wise: the worst case of the current objective is minimized within the worst case of the constraints. For the systematic set-up errors, iCycle typically considers seven scenarios, which have to be defined by the so-called *set-up robustness* (*SR*) setting. Table 3.2 shows the considered scenarios for a robust treatment plan in iCycle.

Table 3.2: The considered scenarios for robust treatment planning in iCycle, when only systematic set-up errors are accounted for.

Scenario s	x-shift ξ_1 (mm)	y-shift ξ_2 (mm)	z-shift ξ_3 (mm)
1	0	0	0
2	SR	0	0
3	$-SR$	0	0
4	0	SR	0
5	0	$-SR$	0
6	0	0	SR
7	0	0	$-SR$

3.2.5 Probabilistic treatment planning

Probabilistic treatment planning differs from robust treatment planning. First of all, each uncertainty parameter is assumed to be distributed according to a probability density function (PDF). In principle, this could be any distribution. However, only continuous distributions are realistic. After all, a discrete distribution does not allow a continuous spectrum of scenarios. As discussed in Chapter 2, Gaussian distributed uncertainty inputs are assumed, with zero correlation. This is in accordance with literature [47, 48]. Second, in contrast to robust treatment planning, all possible scenarios are accounted for in the treatment plan. For any continuous distribution, this would be an infinite number of scenarios.

Besides the assumed probability density function, a choice should be made regarding what kind of statistical measure is desired to be optimized. Any statistical measure is possible. Some examples are listed below:

- Expected value,

- Conditional expected value,

- Expected value plus or minus a multiple of the standard deviation,

- Mode,

- Median,

- Percentile.

Each of these statistical measures can be expressed mathematically. The operator $\mathcal{M}_{\boldsymbol{\xi}}[\cdot]$ is used to denote an arbitrary statistical measure over the uncertainty vector $\boldsymbol{\xi}$. For probabilistic treatment planning, the minimization problem would be described as follows:

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

where $f(\cdot)$ is an arbitrary function and $g(\boldsymbol{x}, \boldsymbol{\xi})$) is a stochastic variable at least depending on the uncertainty vector $\boldsymbol{\xi}$. The dose or the minimum dose are examples of such a stochastic variable.

Depending on the statistical measure which is minimized for, one can make probabilistic statements about the resulting probabilistic treatment plan. For example: the tumor coverage is within certain limits with a 95% probability or the optic nerve is expected to be spared. Of course only if a feasible solution exists and the optimization converges, the resulting treatment plan is guaranteed to satisfy such objectives.

iCycle doesn't offer any features for probabilistic treatment planning. Implementing this functionality in iCycle is a major part of this research.

As already discussed, the dependence on the uncertainty parameters comes in via the dose deposition matrix. The so-called *dose engine* of iCycle is able to compute the dose deposition matrix corresponding to a certain set-up and range error. This is computationally expensive. As expressed in Equation (3.9) for probabilistic treatment planning, a statistical measure of the stochastic variable is needed. Since no explicit analytical expression is available for the dose deposition matrix as function of the uncertainty, the statistical measure cannot be computed analytically. Using Monte Carlo methods is one solution to the problem. However, approximating statistical measures, e.g. an expected value, with Monte Carlo takes a lot of time, since many samples of the dose deposition matrix are required from the dose engine. Polynomial Chaos Expansion (PCE) can be used to bypass this problem. It is a mathematical tool which constructs a meta-model of a black box problem depending on an uncertainty parameter.

Chapter 4

Polynomial Chaos Expansion

This chapter covers the needed theory about Polynomial Chaos Expansion. Substantial parts are based on literature [2, 13, 49–51]. In this research, the openGPC package is used for the construction and evaluating of PCEs [52, 53]. It has been coded by Z. Perkó in Matlab.

4.1 Probability theory

In order to describe uncertainties properly, an overview of probability theory is given. A probability space is defined as $(\Omega, \mathcal{F}, P)$. A random outcome within the sample space Ω is denoted by $\theta \in \Omega$. Such an outcome can be described by a vector $\boldsymbol{\xi}$, with N the number of dimensions:

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

The set of corresponding events is denoted by $\mathcal{F}$, which is a σ-algebra on Ω. The probability measure P is a function which maps the set of events on their probability to occur, $P : \mathcal{F} \to [0, 1]$. For this research, the quantities of interest are real valued stochastic responses $R(\theta)$:

$$R(\theta) : \Omega \to \mathbb{R}. \tag{4.2}$$

Usually, the stochastic responses $R(\theta)$ are restricted to square-integrable functions, i.e. to the L_2 space which is defined as in Equation (4.3).

$$L_2(\Omega, P) = \{R(\theta) : [R(\theta) : \Omega \to \mathbb{R}] \wedge [\langle R, R \rangle < \infty]\} \tag{4.3}$$

The inner product in the L_2 space is defined in Equation (4.4).

$$\langle Q, R \rangle = \int_{\Omega} Q(\theta) R(\theta) \, \mathrm{d}P(\theta) = \int_{\mathcal{D}} Q(\boldsymbol{\xi}(\theta)) \, R(\boldsymbol{\xi}(\theta)) \, p_{\boldsymbol{\xi}}(\boldsymbol{\xi}) \, \mathrm{d}\boldsymbol{\xi} \tag{4.4}$$

Here, $\mathcal{D}$ is the domain of random variables, which describe the sample space Ω. From now on, the dependence of $\boldsymbol{\xi}$ on θ is omitted, because it suffices to describe the random outcomes with only the random variable $\boldsymbol{\xi}$. In this research, mutually independent random variables ξ_j are assumed. The corresponding probability density function (PDF) is denoted by $p_{\xi_j}(\xi_j)$. The joint probability density function of the random variable $\boldsymbol{\xi}$ is the product of the individual PDFs, since they are independent:

$$p_{\boldsymbol{\xi}}(\boldsymbol{\xi}) = \prod_{j=1}^{N} p_{\xi_j}(\xi_j). \tag{4.5}$$

Some quantities of interest of a stochastic response are the mean and the variance, which are denoted by μ_R and $\text{Var}(R)$ respectively. The mean of a stochastic response is defined as the expected value of this response: approximately the average outcome of many samples. The variance describes to which extent these outcomes differ from each other. The mean can be defined as:

$$\mu_R = \text{E}[R] = \int_{\mathcal{D}} R(\boldsymbol{\xi}) p_{\boldsymbol{\xi}}(\boldsymbol{\xi}) \, \text{d}\boldsymbol{\xi} \,, \tag{4.6}$$

and the variance as:

$$\text{Var}(R) = \text{E}\left[(R - \text{E}[R])^2 \right] = \int_{\mathcal{D}} \left(R(\boldsymbol{\xi}) - \mu_R \right)^2 p_{\boldsymbol{\xi}}(\boldsymbol{\xi}) \, \text{d}\boldsymbol{\xi} \,. \tag{4.7}$$

A more common measure to describe the spread is the standard deviation $\sigma_R = \sqrt{\text{Var}(R)}$.

In the following chapters, when the domain of integration is not explicitly stated, one may assume the whole domain $\mathcal{D}$ is meant.

4.1.1 Additivity of Gaussian distribution

In this research all random variables ξ_j are assumed to be Gaussian distributed. Therefore, this section provides some properties of such a distribution.

In general, the PDF of a Gaussian distributed variable X, is shown in Equation (4.8).

$$p_X(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{1}{2}\left(\frac{x-\mu_X}{\sigma_X}\right)^2} \tag{4.8}$$

Let $X \sim \mathcal{N}(\mu_X, \sigma_X^2)$ and $Y \sim \mathcal{N}(\mu_Y, \sigma_Y^2)$. Every Gaussian distributed stochastic variable can be formed from a standard Gaussian variable, $W \sim \mathcal{N}(0,1)$:

$$X = \mu_X + \sigma_X W \,. \tag{4.9}$$

When one considers the sum of two Gaussian distributed variables $Z = X + Y$, this obtained variable has the following interesting properties:

$$Z \sim \mathcal{N}(\mu_Z, \sigma_Z^2) \,, \tag{4.10}$$
$$\mu_Z = \mu_X + \mu_Y \,, \tag{4.11}$$
$$\sigma_Z^2 = \sigma_X^2 + \sigma_Y^2 \,. \tag{4.12}$$

4.2 Polynomial Chaos Expansion

The basic principle of Polynomial Chaos Expansion (PCE) is to describe the dependence of the stochastic response $R(\boldsymbol{\xi})$ on the uncertainty vector $\boldsymbol{\xi}$ as a series expansion with expansion coefficients r_k and basis vectors $\Psi_k(\boldsymbol{\xi})$, as shown in Equation (4.13).

$$R(\boldsymbol{\xi}) = \sum_{k=0}^{\infty} r_k \Psi_k(\boldsymbol{\xi}) \tag{4.13}$$

4.2.1   Basis vectors

The basis vectors $\Psi_k(\boldsymbol{\xi})$ are products of univariate polynomials $\phi_{j,\gamma_{kj}}$, each belonging to the random variable ξ_j. The multi index γ_{kj} is the polynomial order of the j-th polynomial for the basis vector $\Psi_k(\boldsymbol{\xi})$. This is represented by Equation (4.14).

$$\Psi_k(\boldsymbol{\xi}) = \prod_{j=1}^{N} \phi_{j,\gamma_{kj}}(\xi_j) \tag{4.14}$$

Not every family of polynomials can be chosen. For convergence to $R(\boldsymbol{\xi})$, orthogonality of the polynomials is recommended. A property of orthogonal polynomials is that the inner product of two polynomials is only non-zero when the order is equal, as stated in Equation (4.15), where $h_{j,n}$ represents the norm of the j-th polynomial of the n-th order.

$$\int \phi_{j,m}(\xi_j)\phi_{j,n}(\xi_j)p_{\xi_j}(\xi_j)\,\mathrm{d}\xi_j = h_{j,n}^2\delta_{m,n} \quad \forall j \in \{1,...,N\} \tag{4.15}$$

The Kronecker delta is denoted by $\delta_{k,l}$. The orthogonality of the basis vectors can be derived from the orthogonality of the univariate polynomials, as shown in Equations (4.16) to (4.20). Equation (4.17) uses the fact that the stochastic variables ξ_j are independent. Regarding Equation (4.19), $\prod_{j=1}^{N}\delta_{\gamma_{kj},\gamma_{lj}} = \delta_{k,l}$ holds because every basis vector Ψ_k is unique. The norm of the basis vector is defined as $h_k = \prod_{j=1}^{N} h_{j,\gamma_{kj}}$.

$$\int \Psi_k(\boldsymbol{\xi})\Psi_l(\boldsymbol{\xi})p_{\boldsymbol{\xi}}(\boldsymbol{\xi})\,\mathrm{d}\boldsymbol{\xi} = \int p_{\boldsymbol{\xi}}(\boldsymbol{\xi})\prod_{j=1}^{N}\phi_{j,\gamma_{kj}}(\xi_j)\prod_{j'=1}^{N}\phi_{j',\gamma_{lj'}}(\xi_{j'})\,\mathrm{d}\boldsymbol{\xi} \tag{4.16}$$

$$= \prod_{j=1}^{N}\int \phi_{j,\gamma_{kj}}(\xi_j)\phi_{j,\gamma_{lj}}(\xi_j)p_{\xi_j}(\xi_j)\,\mathrm{d}\xi_j \tag{4.17}$$

$$= \prod_{j=1}^{N} h_{j,\gamma_{kj}}^2 \delta_{\gamma_{kj},\gamma_{lj}} \tag{4.18}$$

$$= \delta_{k,l}\prod_{j=1}^{N} h_{j,\gamma_{kj}}^2 \tag{4.19}$$

$$= \delta_{k,l}h_k^2 \tag{4.20}$$

Both the left-hand side and the right-hand side of Equation (4.13) can be multiplied by $\Psi_l(\boldsymbol{\xi})p_{\boldsymbol{\xi}}(\boldsymbol{\xi})$ and be integrated over $\boldsymbol{\xi}$. When using the orthogonality of Equation (4.20), one obtains an expression for the coefficients as in Equation (4.21).

$$r_k = \frac{\langle R(\boldsymbol{\xi}), \Psi_k(\boldsymbol{\xi})\rangle}{\langle \Psi_k(\boldsymbol{\xi}), \Psi_k(\boldsymbol{\xi})\rangle} = \frac{\int R(\boldsymbol{\xi})\Psi_k(\boldsymbol{\xi})p_{\boldsymbol{\xi}}(\boldsymbol{\xi})\,\mathrm{d}\boldsymbol{\xi}}{\int \Psi_k(\boldsymbol{\xi})\Psi_k(\boldsymbol{\xi})p_{\boldsymbol{\xi}}(\boldsymbol{\xi})\,\mathrm{d}\boldsymbol{\xi}} \tag{4.21}$$

The expansion, described by Equations (4.13) and (4.21), is still an exact representation of the stochastic response, under the assumption that $R(\boldsymbol{\xi})$ is continuous and integrable. However, to be useful in practice, the expansion is truncated to include only a finite number $P+1$ of basis vectors:

$$R(\boldsymbol{\xi}) \approx \sum_{k=0}^{P} r_k\Psi_k(\boldsymbol{\xi})\,. \tag{4.22}$$

The multi index γ_{kj} is used to define which orders to include in the expansion. The set of o-th order basis vectors is defined as follows:

$$\Gamma_o = \left\{ \prod_{j=1}^{N} \phi_{j,\gamma_{kj}} : \sum_{j=1}^{N} \gamma_{kj} = o \right\} . \tag{4.23}$$

The full O-th order set of basis vectors is defined as:

$$\Gamma(O) = \bigcup_{o \in \{0,1,...,O\}} \Gamma_o = \left\{ \prod_{j=1}^{N} \phi_{j,\gamma_{kj}} : \sum_{j=1}^{N} \gamma_{kj} \leq O \right\} . \tag{4.24}$$

The number of basis vectors included in a full O-th order basis set is given by Equation (4.25).

$$P + 1 = \frac{(N+O)!}{N!O!} \tag{4.25}$$

4.2.2 Mean and variance

The mean and variance of a Polynomial Chaos Expansion can be calculated easily. An expression for the mean of a PCE is derived in Equations (4.26) to (4.30).

$$\mu_R = \int R(\boldsymbol{\xi}) p_{\boldsymbol{\xi}}(\boldsymbol{\xi}) \, \mathrm{d}\boldsymbol{\xi} \tag{4.26}$$

$$= \int \sum_{k=0}^{\infty} r_k \Psi_k(\boldsymbol{\xi}) p_{\boldsymbol{\xi}}(\boldsymbol{\xi}) \, \mathrm{d}\boldsymbol{\xi} \tag{4.27}$$

$$= \sum_{k=0}^{\infty} r_k \int \Psi_0(\boldsymbol{\xi}) \Psi_k(\boldsymbol{\xi}) p_{\boldsymbol{\xi}}(\boldsymbol{\xi}) \, \mathrm{d}\boldsymbol{\xi} \tag{4.28}$$

$$= \sum_{k=0}^{\infty} r_k \langle \Psi_0(\boldsymbol{\xi}), \Psi_k(\boldsymbol{\xi}) \rangle \tag{4.29}$$

$$= r_0 \tag{4.30}$$

An expression for the variance of a PCE is derived in Equations (4.31) to (4.39).

$$\sigma_R^2 = \int \left(R(\boldsymbol{\xi}) - \mu_R \right)^2 p_{\boldsymbol{\xi}}(\boldsymbol{\xi}) \, \mathrm{d}\boldsymbol{\xi} \tag{4.31}$$

$$= \int R(\boldsymbol{\xi})^2 p_{\boldsymbol{\xi}}(\boldsymbol{\xi}) \, \mathrm{d}\boldsymbol{\xi} - 2 \int \mu_R R(\boldsymbol{\xi}) p_{\boldsymbol{\xi}}(\boldsymbol{\xi}) \, \mathrm{d}\boldsymbol{\xi} + \int \mu_R^2 p_{\boldsymbol{\xi}}(\boldsymbol{\xi}) \, \mathrm{d}\boldsymbol{\xi} \tag{4.32}$$

$$= \int \sum_{k=0}^{\infty} r_k \Psi_k(\boldsymbol{\xi}) \sum_{k'=0}^{\infty} r_{k'} \Psi_{k'}(\boldsymbol{\xi}) p_{\boldsymbol{\xi}}(\boldsymbol{\xi}) \, \mathrm{d}\boldsymbol{\xi} - 2\mu_R \int \sum_{k=0}^{\infty} r_k \Psi_k(\boldsymbol{\xi}) p_{\boldsymbol{\xi}}(\boldsymbol{\xi}) \, \mathrm{d}\boldsymbol{\xi} + \mu_R^2 \int p_{\boldsymbol{\xi}}(\boldsymbol{\xi}) \, \mathrm{d}\boldsymbol{\xi} \tag{4.33}$$

$$= \sum_{k=0}^{\infty} \sum_{k'=0}^{\infty} r_k r_{k'} \int \Psi_k(\boldsymbol{\xi}) \Psi_{k'}(\boldsymbol{\xi}) p_{\boldsymbol{\xi}}(\boldsymbol{\xi}) \, \mathrm{d}\boldsymbol{\xi} - 2\mu_R \sum_{k=0}^{\infty} r_k \int \Psi_0 \Psi_k(\boldsymbol{\xi}) p_{\boldsymbol{\xi}}(\boldsymbol{\xi}) \, \mathrm{d}\boldsymbol{\xi} + \mu_R^2 \tag{4.34}$$

$$= \sum_{k=0}^{\infty} \sum_{k'=0}^{\infty} r_k r_{k'} \langle \Psi_k(\boldsymbol{\xi}), \Psi_{k'}(\boldsymbol{\xi}) \rangle - 2\mu_R \sum_{k=0}^{\infty} r_k \langle \Psi_0, \Psi_k(\boldsymbol{\xi}) \rangle + \mu_R^2 \tag{4.35}$$

$$= \sum_{k=0}^{\infty} \sum_{k'=0}^{\infty} r_k r_{k'} h_k^2 \delta_{k,k'} - 2\mu_R^2 + \mu_R^2 \tag{4.36}$$

$$= \sum_{k=0}^{\infty} r_k^2 h_k^2 - \mu_R^2 \tag{4.37}$$

$$= \sum_{k=0}^{\infty} r_k^2 h_k^2 - r_0^2 h_0^2 \tag{4.38}$$

$$= \sum_{k=1}^{\infty} r_k^2 h_k^2 \tag{4.39}$$

When a PCE is truncated to include only $P+1$ basis vectors as in Equation (4.22), the mean is not affected. However, the truncation does influence the variance:

$$\sigma_R \approx \sum_{k=1}^{P} r_k^2 h_k^2. \tag{4.40}$$

4.2.3   Wiener-Askey scheme

As stated in Section 4.2.1, orthogonal pylonomials are recommended for convergence of the Polynomial Chaos Expansion to the stochastic response of which a PCE is constructed. Moreover, fast convergence is guaranteed only for certain sets of polynomials. This optimal choice of the polynomials is provided by the Wiener-Askey scheme [54]. These sets of polynomials are orthogonal with respect to some common PDFs. According to the Wiener-Askey scheme, when using the polynomials corresponding to the suspected underlying PDF of the random variable ξ_j, faster convergence is expected. For example, probabilists' Hermite polynomials are orthogonal with respect to the Gaussian distribution and faster convergence is expected when $\boldsymbol{\xi}$ is Gaussian distributed. A part of the Wiener-Askey scheme is shown in Table 4.1.

Hermite polynomials

As already stated in Section 4.1.1, the random variables ξ_j are assumed to be Gaussian distributed. The Wiener-Askey scheme prescribes to use probabilists' Hermite polynomials. These

can be defined in multiple ways, for example by a recurrence relation as in Equations (4.41).

$$
\begin{aligned}
He_0(\xi_j) &= 1 \\
He_1(\xi_j) &= \xi_j \\
He_{k+1}(\xi_j) &= xHe_k(\xi_j) - kHe_{k-1}(\xi_j) \quad \forall k \in \mathbb{N}
\end{aligned}
\tag{4.41}
$$

They can also be calculated explicitly as in Equation (4.42).

$$
He_k(\xi_j) = (-1)^k e^{\frac{\xi_j^2}{2}} \frac{d^k}{d\xi_j^k} e^{-\frac{\xi_j^2}{2}} \quad \forall k \in \mathbb{N}
\tag{4.42}
$$

The first six probabilists' Hermite polynomials are shown in Table 4.2 and Figure 4.1.
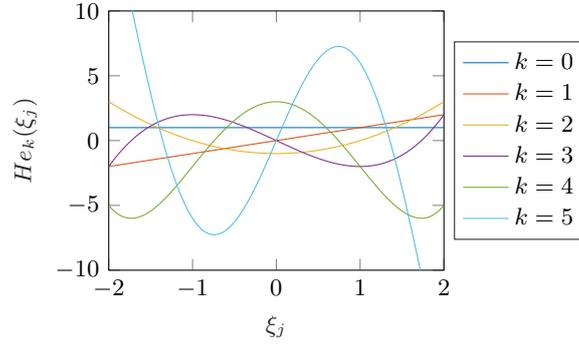
Figure 4.1: The first six probabilists' Hermite polynomials.

4.3 Numerical integration

Typically, no explicit expression is known for the stochastic response. It is assumed to be a black box, to which an input $\boldsymbol{\xi}$ is provided in order to obtain an output $R(\boldsymbol{\xi})$. Therefore, exact calculation of the integral in the numerator of Equation (4.21) is not possible. Hence, numerical integration is used to approximate this integral. Equation (4.21) can also be written as Equation (4.43), due to the mutually independent random variables ξ_j.

$$
r_k = \frac{\int \dots \int R(\boldsymbol{\xi}) \prod_{j=1}^{N} \phi_{j,\gamma_{k_j}}(\xi_j) p_{\boldsymbol{\xi}}(\boldsymbol{\xi}) \, d\xi_1 \dots d\xi_N}{\int \Psi_k(\boldsymbol{\xi}) \Psi_k(\boldsymbol{\xi}) p_{\boldsymbol{\xi}}(\boldsymbol{\xi}) \, d\boldsymbol{\xi}}
\tag{4.43}
$$

Table 4.1: The Wiener-Askey scheme provides the polynomial for a Polynomial Chaos Expansion with fast convergence [54]. Here, $a, b \in \mathbb{R}$ holds.

Distribution	Askey Polynomial	Supported domain
Beta	Jacobi	$[a, b]$
Gamma	Laguerre	$[0, \infty)$
Gaussian	Probabilists' Hermite	$(-\infty, \infty)$
Uniform	Legendre	$[a, b]$

The denominator equals the square of the norm of the k-th order polynomial. Section 4.3.1 covers a method to approximate one-dimensional integrals for the calculation of the numerator of Equation (4.43). Section 4.3.2 covers the extension to multi-dimensional integrals.

4.3.1 Quadrature

Quadratures can be used to approximate one-dimensional integrals [2, 13, 53]. This section focuses on one-dimensional integrals of the form of Equation (4.44), with $f(\xi)$ some function depending on the single variable ξ. It is approximated with a finite sum of the products of the function value at $\xi_{lev}^{(i)} \in [a, b]$ and weight $w_{lev}^{(i)}$.

$$I^{(1)}f = \int_a^b f(\xi)p_\xi(\xi)\,\mathrm{d}\xi \approx Q_{lev}^{(1)} = \sum_{i=1}^{n_{lev}} f\left(\xi_{lev}^{(i)}\right) w_{lev}^{(i)} \tag{4.44}$$

Here, the superscript (1) denotes the dimensionality of the integral, the superscript i distinguishes the quadrature points from each other, lev indicates the quadrature level. The number of quadrature points $\xi_{lev}^{(i)}$ is denoted by n_{lev}. A higher quadrature level means more quadrature points, which is beneficial for the accuracy of the approximation, but is on the other hand computationally more intensive. As one might expect, the quadrature points and weights depend on the probability density function and quadrature rule that is chosen. Multiple quadrature rules exist to approximate such an integral. Two properties are important to keep in mind when choosing a certain rule: the accuracy and so-called *nestedness*. Nestedness is to which extent quadrature points recur for different levels. For example, full nestedness means that every quadrature point of a certain level lev also belongs to a higher level $lev + 1$. Zero nestedness means that different levels don't have common quadrature points. Two examples of quadrature rules are the Gauss-Hermite rule and the Clenshaw-Curtis rule. The Guass-Hermite rule demonstrates low nestedness, as only the origin ($\xi = 0$) recurs at each level. Figure 4.2 shows the Clenshaw-Curtis and the Gauss-Hermite quadrature points. The former are suitable for an integral with the standard Gaussian distribution as weight function. The latter are constructed on an interval $[-3, 3]$.

The Gaussian quadrature rule is used throughout this thesis. The quadrature levels are chosen such that $n_{lev} = 2 \cdot lev - 1$.

Gaussian quadratures are especially suitable for integrating polynomials: they demonstrate exactness up to polynomial order $2n_{lev} - 1 = 4 \cdot lev - 3$ [55].

Table 4.2: The first six probabilists' Hermite polynomials.

Order	Expression
$He_0(\xi_j)$	1
$He_1(\xi_j)$	ξ_j
$He_2(\xi_j)$	$\xi_j^2 - 1$
$He_3(\xi_j)$	$\xi_j^3 - 3\xi_j$
$He_4(\xi_j)$	$\xi_j^4 - 6\xi_j^2 + 3$
$He_5(\xi_j)$	$\xi_j^5 - 10\xi_j^3 + 15\xi_j$
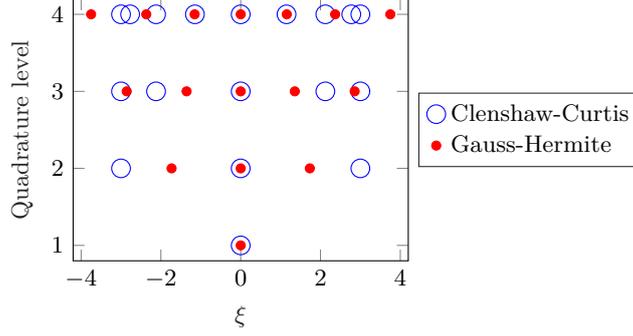
Figure 4.2: Clenshaw-Curtis and Gauss-Hermite quadrature points for several quadrature levels. The Clenshaw-Curtis quadrature points are constructed on the interval $[-3, 3]$ and are fully nested. The Gauss-Hermite quadrature points are constructed for a standard Gaussian distribution as weight function.

4.3.2 Cubature

In Section 4.3.1, one-dimensional numerical integration with quadratures has been treated. However, in this research all problems are multidimensional. So-called cubatures are the extension of quadratures to multiple dimensions. Such cubatures can be derived by tensorization of the single dimensional quadratures, as shown in Equation (4.45).

$$Q_{\boldsymbol{lev}}^{(N)} f = \left(Q_{lev_1}^{(1)} \otimes ... \otimes Q_{lev_N}^{(1)} \right) f \tag{4.45}$$

Here, lev_j denotes the quadrature level along the j-th dimension. These levels lev_j are not necessarily the same for each dimension. For example, the grid $(2, 2, ..., 2)$ means that quadrature level 2 is assigned to each dimension. When the grid $(3, 2, ..., 2)$ is used, quadrature level 2 is assigned to each dimension except from the first dimension, in which a higher accuracy could be required. For $N = 3$, the notation Perm$(3, 2, 2)$ is used to denote the set of all permutations of such grids and therefore it contains all grids with a single dimension with a quadrature level 3, while the others have a quadrature level of 2. As a consequence, Perm$(3, 2, 2) = \{(3, 2, 2), (2, 3, 2), (2, 2, 3)\}$ holds.

The multidimensional integral of a function $f(\boldsymbol{\xi})$ with weight function $p_{\boldsymbol{\xi}}(\boldsymbol{\xi})$ can be approximated with a full N-dimensional cubature grid as follows:

$$I^{(N)} f = \int ... \int f(\boldsymbol{\xi}) p_{\boldsymbol{\xi}}(\boldsymbol{\xi}) \, \mathrm{d}\xi_1 ... \mathrm{d}\xi_N \tag{4.46}$$

$$\approx Q_{\boldsymbol{lev}}^{(N)} f \tag{4.47}$$

$$= \left(Q_{lev_1}^{(1)} \otimes ... \otimes Q_{lev_N}^{(1)} \right) f \tag{4.48}$$

$$= \sum_{i_1=1}^{n_{lev_1}} ... \sum_{i_N=1}^{n_{lev_N}} f \left(\xi_{1,lev_1}^{(i_1)}, ..., \xi_{N,lev_N}^{(i_N)} \right) w_{lev_1}^{(i_1)} ... w_{lev_N}^{(i_N)} \tag{4.49}$$

$$= \sum_{i=1}^{n} f(\boldsymbol{\xi}^{(i)}) w^{(i)} . \tag{4.50}$$

The N sums over the quadrature points of Equation (4.49) are replaced by a single sum over $n = \prod_{j=1}^{N} n_{lev_j}$ cubature points in Equation (4.50), where a cubature point is denoted by $\boldsymbol{\xi}^{(i)}$

and where the weight $w^{(i)}$ is the product of the weights of the individual quadrature points. This full N-dimensional cubature grid leads to many function evaluations, namely $n = \prod_{j=1}^{N} n_{lev_j}$. An example of a three-dimensional full grid with $\boldsymbol{lev} = (4,4,4)^T$ is shown in Figure 4.3a.

4.3.3 Smolyak sparse grids

In order to bypass the need of many function evaluations, sparse grids can be used. They are based upon the so-called *sparsity-of-effects principle*, which means that lower order terms are usually dominant in most models [56, 57]. This means that not all cubature points are equally important. Cubature points corresponding to high quadrature levels in many dimensions are omitted for this purpose. In this way a sparse grid arises. There exist multiple types of sparse grids [58, 59]. In this research Smolyak sparse grids are used. These are based on difference formulas instead of the quadratures. The following relation is used:

$$\Delta_{lev}^{(1)} = Q_{lev}^{(1)} - Q_{lev-1}^{(1)} \quad \text{with} \quad Q_0^{(1)} f = 0 \,. \tag{4.51}$$

From this recurrence relation, one can derive an expression for $Q_{lev}^{(1)}$ in terms of differences:

$$Q_{lev}^{(1)} = \sum_{l=1}^{lev} \Delta_l^{(1)} \,. \tag{4.52}$$

Substituting Equation (4.52) into Equation (4.48) yields:

$$Q_{\boldsymbol{lev}}^{(N)} f = \sum_{l_1}^{lev_1} \dots \sum_{l_N}^{lev_N} \left(\Delta_{l_1}^{(1)} \otimes \dots \otimes \Delta_{l_N}^{(1)} \right) f \tag{4.53}$$

$$= \sum_{\boldsymbol{l} \in \mathcal{I}_{\text{Full}}(G)} \Delta_{\boldsymbol{l}}^{(N)} f \,, \tag{4.54}$$

with

$$\mathcal{I}_{\text{Full}}(\boldsymbol{lev}) = \{ \boldsymbol{l} : l_j \leq lev_j \quad \forall j \in \{1, ..., N\} \} \,. \tag{4.55}$$

Here, $\boldsymbol{l} = (l_1, ..., l_N)^T$ is a multi index with elements l_j indicating the quadrature level of the difference term in the j-th dimension.

A Smolyak sparse grid is created by modifying the set of multi indices over which $\boldsymbol{l}$ is summed in Equation (4.54). For a Smolyak sparse grid this set is represented by $\mathcal{I}_{\text{Sm}}(L)$, where L indicates the so-called *grid level* [2, 13, 49].

$$\mathcal{I}_{\text{Sm}}(L) = \left\{ \boldsymbol{l} : \sum_{j=1}^{N} l_j \leq L + N - 1 \right\} \tag{4.56}$$

Evidently, a higher grid level L means that more quadrature levels are included. On the other hand, this also leads to more function evaluations. It may not be obvious what Equation (4.56) actually means. One starts with the grid $(1, ..., 1)$. Subsequently, there are $L - 1$ levels which are to be divided among the N dimensions. For example, for $L = 3$ and $N = 3$ the Smolyak sparse grid consists of the grids $\text{Perm}(3, 1, 1) = \{(3, 1, 1), (1, 3, 1), (1, 1, 3)\}$ and $\text{Perm}(2, 2, 1) = \{(2, 2, 1), (2, 1, 2), (1, 2, 2)\}$. The number of function evaluations, equal to the number of cubature points, are summarized in Table 4.3. Algorithms exist to compute all cubature points efficiently. An example of a three-dimensional Smolyak sparse grid with $L = 4$ is shown in Figure 4.3b.

(a) Full grid, 343 cubature points

(b) Smolyak sparse grid, 105 cubature points
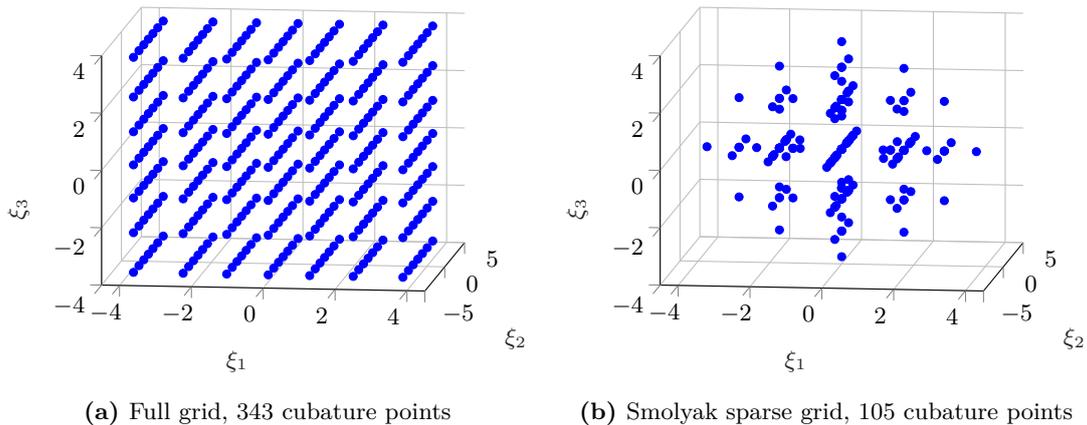
Figure 4.3: Three-dimensional grids. **(a)** Full grid with $\boldsymbol{lev} = (4,4,4)^T$. **(b)** Smolyak sparse grid with $L = 4$.

4.3.4 Extended Smolyak sparse grids

A property of a Smolyak sparse grid is that the maximum quadrature level for a single dimension is equal to the chosen grid level L. In general, the single dimensional grids are $\mathrm{Perm}(L, 1, ..., 1)$. One might wish a more accurate approximation of the single dimensional terms. This might be desirable when single dimensional terms are considered important, but a higher grid level is computationally too costly. With so-called *extended Smolyak sparse grids*, it is possible to increase the accuracy in the single dimensional directions and simultaneously keep the build-up of cubature points within limits. Suppose a Smolyak sparse grid with grid level L is created. This means that the single dimensional grids are $\mathrm{Perm}(L, 1, ..., 1)$. Let the increase in quadrature level in the single dimensional directions be denoted by Δ_{lev}. Now, the quadrature levels in the single dimensional directions can be increased. This yields the grids $\mathrm{Perm}(L + \Delta_{lev}, 1, ..., 1)$, which replace the grids $\mathrm{Perm}(L, 1, ..., 1)$. A higher accuracy is expected, while the number of extra cubature points is equal to only $2\Delta_{lev}N$. The set of $\boldsymbol{l}$ over which has to be summed for an extended Smolyak sparse grid is denoted by $\mathcal{I}_{\mathrm{ExtSm}}(L, \Delta_{lev})$:

$$\mathcal{I}_{\mathrm{ExtSm}}(L, \Delta_{lev}) = \mathcal{I}_{\mathrm{Sm}}(L) \cup \mathrm{Perm}(L + \Delta_{lev}, 1, ..., 1) \setminus \mathrm{Perm}(L, 1, ..., 1). \qquad (4.57)$$

Table 4.3: The number of cubature points for different grid levels and numbers of dimensions when a Smolyak sparse grid with Gauss-Hermite quadrature rule is used [2, 13].

	Grid level L			
N	2	3	4	5
1	3	7	13	21
2	5	17	45	101
3	7	31	105	297
4	9	49	201	1341

4.4 Hyperbolic trimming

In Section 4.3.3 it has been discussed how full grids are transformed into sparse grids by leaving out certain cubature points. The same methodology could be applied to the domain of basis vectors. A way to do this is called hyperbolic trimming [60]. Let the q-quasi-norm be defined as in Equation (4.58) with $q \in (0, 1]$ [56].

$$||\gamma_k||_q = \left(\sum_{j=1}^{N} \gamma_{kj}^q \right)^{\frac{1}{q}} \tag{4.58}$$

Hyperbolic trimming of the Polynomial Chaos basis vectors means that only those vectors remain which meet the requirement of Equation (4.59).

$$||\gamma_k||_q \leq O \tag{4.59}$$

It could be verified that $q = 1$ holds for a full O-th order basis set. Figure 4.4 visualises the effect of hyperbolic trimming on the PC basis set for various values of q. The blue dots represent the basis vectors which are contained in the two-dimensional Polynomial Chaos basis set.
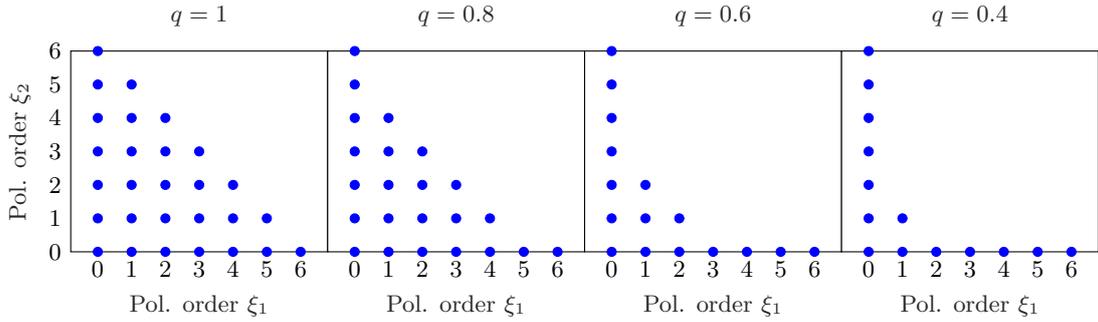
Figure 4.4: Polynomial Chaos basis set with $O = 6$ for different values of the hyperbolic trim factor q. The blue dots show which basis vectors are to be included in the expansion. The polynomial order of ξ_1 and ξ_2 is presented on the horizontal and vertical axis respectively.

4.5 Notation

Throughout this thesis, the abbreviation $LxEyOz$ is used to summarize the main settings of a PCE. It indicates that a Smolyak sparse grid is used with grid level $L = x$ and extended single dimensional level $\Delta_{lev} = y$. A full Polynomial Chaos basis set is used with $O = z$.

Chapter 5

Probabilistic functions with PCE

In treatment planning, objective functions and constraints are used to steer to a solution which meets all clinical wishes, at least to a desired extent. Several possible functions are treated in this chapter, that could serve both as objective function or as constraint in the optimization to obtain a probabilistic treatment plan.

5.1   General overview

Any function which depends on the uncertainty parameters can be used for probabilistic treatment planning. This comes down to all functions depending on the dose, since the dose depends on the dose deposition matrix and the dose deposition matrix depends on the uncertainty parameters. Let $g(\boldsymbol{x}, \boldsymbol{\xi})$ denote such a function.

As already discussed in Chapter 3, a clinical decision should be made what kind of statistical measure is desirable to optimize for. The objective function which is either minimized or maximized is $f\left(\mathcal{M}_{\boldsymbol{\xi}}\left[g(\boldsymbol{x}, \boldsymbol{\xi})\right]\right)$, with $f(\cdot)$ any function. Some examples of statistical measures are listed below:

- Expected value:
  $\int_{\mathcal{D}} g(\boldsymbol{x}, \boldsymbol{\xi}) p_{\boldsymbol{\xi}}(\boldsymbol{\xi}) \, \mathrm{d}\boldsymbol{\xi}$,

- Conditional expected value:
  $\left(\int_{\mathcal{D}_1} p_{\boldsymbol{\xi}}(\boldsymbol{\xi}) \, \mathrm{d}\boldsymbol{\xi}\right)^{-1} \int_{\mathcal{D}_1} g(\boldsymbol{x}, \boldsymbol{\xi}) p_{\boldsymbol{\xi}}(\boldsymbol{\xi}) \, \mathrm{d}\boldsymbol{\xi}$ with $\mathcal{D}_1 \subset \mathcal{D}$,

- Expected value plus or minus a multiple of the standard deviation:
  $\int_{\mathcal{D}} g(\boldsymbol{x}, \boldsymbol{\xi}) p_{\boldsymbol{\xi}}(\boldsymbol{\xi}) \, \mathrm{d}\boldsymbol{\xi} \pm a \int_{\mathcal{D}} \left(g(\boldsymbol{x}, \boldsymbol{\xi}) - \mu_g\right)^2 p_{\boldsymbol{\xi}}(\boldsymbol{\xi}) \, \mathrm{d}\boldsymbol{\xi}$ with $a \in \mathbb{R}$,

- Mode:
  The function value $g(\boldsymbol{x}, \boldsymbol{\xi})$ corresponding to the global maximum of its probability density function $p_g(g)$: $\arg\max_g p_g(g)$,

- Median:
  $P_{50\%}^s \left\{g(\boldsymbol{x}, \boldsymbol{\xi}^{(s)}) : s \in \{1, ..., N_\mathrm{s}\}\right\}$ with N_s samples,

- Percentile:
  $P_{\beta\%}^s \left\{g(\boldsymbol{x}, \boldsymbol{\xi}^{(s)}) : s \in \{1, ..., N_\mathrm{s}\}\right\}$ with N_s samples and $\beta \in [0, 100]$.

Since the explicit dependence of the dose deposition matrix on the uncertainty vector is not known, these statistical measures cannot be calculated analytically. Monte Carlo simulation using samples of the dose deposition matrix from the dose engine is computationally very expensive, as described in Chapter 3. Polynomial Chaos Expansion can alleviate the computational effort of these statistical measures, since sampling from a PCE is faster than sampling dose deposition matrices from the dose engine. Besides, one could easily calculate the expected value and the standard deviation from the coefficients of the PCE, as described in Chapter 4. It is not always evident of what object a PCE is to be constructed. One could for example construct a PCE of the function $g(\boldsymbol{x}, \boldsymbol{\xi})$ as a whole. The advantage is that the computation of the statistical measures is very fast, once the PCE has been constructed. The disadvantage is that the beam intensity vector is contained in the PCE, which means that a PCE has to be constructed for every iteration of the optimization. Besides, the calculation of the gradient and/or Hessian matrix, either analytically or by finite differencing, is typically more complicated. On the other hand, a PCE can be constructed of a part of the function $g(\boldsymbol{x}, \boldsymbol{\xi})$, which depends on the uncertainty parameter.

Some examples of probabilistic objective functions which can be used with Polynomial Chaos Expansion are discussed throughout this chapter. Since most optimization algorithms require a gradient and possibly a Hessian matrix to be provided, these derivations are also included. Note that for the following examples that consider a minimum, for example a minimum dose, the same applies to a maximum.

5.2   Expected quadratic dose

The quadratic objective function is the sum over the voxels $\mathcal{V}$ of the squares of the differences between the dose and the prescribed dose $\boldsymbol{d}^{\mathrm{p}}$. It might be desirable to minimize the function f, which is the expected value of the quadratic objective function:

$$f(\boldsymbol{x}) = \int \sum_{i \in \mathcal{V}} w_i \left(\sum_{j \in \mathcal{B}} D_{ij}(\boldsymbol{\xi}) x_j - d_i^{\mathrm{p}} \right)^2 p_{\boldsymbol{\xi}}(\boldsymbol{\xi}) \, \mathrm{d}\boldsymbol{\xi} \tag{5.1}$$

$$= \int \sum_{i \in \mathcal{V}} w_i \left(\sum_{j \in \mathcal{B}} \sum_{j' \in \mathcal{B}} D_{ij} D_{ij'} x_j x_{j'} - 2 d_i^{\mathrm{p}} \sum_{j \in \mathcal{B}} D_{ij} x_j + (d_i^{\mathrm{p}})^2 \right) p_{\boldsymbol{\xi}}(\boldsymbol{\xi}) \, \mathrm{d}\boldsymbol{\xi} \tag{5.2}$$

$$= \int \left(\sum_{j \in \mathcal{B}} \sum_{j' \in \mathcal{B}} \left[\sum_{i \in \mathcal{V}} w_i D_{ij} D_{ij'} \right] x_j x_{j'} - 2 \sum_{j \in \mathcal{B}} \left[\sum_{i \in \mathcal{V}} w_i d_i^{\mathrm{p}} D_{ij} \right] x_j + \sum_{i \in \mathcal{V}} w_i (d_i^{\mathrm{p}})^2 \right) p_{\boldsymbol{\xi}}(\boldsymbol{\xi}) \, \mathrm{d}\boldsymbol{\xi} \tag{5.3}$$

A PCE can be constructed of the two parts in square brackets of Equation (5.3):

$$\sum_{i \in \mathcal{V}} w_i D_{ij}(\boldsymbol{\xi}) D_{ij'}(\boldsymbol{\xi}) = \sum_{k=0}^{\infty} R_{jj'}^{(k)} \Psi_k(\boldsymbol{\xi}) \tag{5.4}$$

$$\sum_{i \in \mathcal{V}} w_i d_i^{\mathrm{p}} D_{ij} = \sum_{k=0}^{\infty} r_j^{(k)} \Psi_k(\boldsymbol{\xi}) \tag{5.5}$$

The matrix $R^{(k)}$ contains the Polynomial Chaos coefficients of the k-th order corresponding to the first part of Equation (5.3). This matrix is symmetric, so only half of the elements need

to be computed. The vector $\boldsymbol{r}^{(k)}$ contains the Polynomial Chaos coefficients of the k-th order corresponding to the second term of Equation (5.3). The size of $R^{(k)}$ and $\boldsymbol{r}^{(k)}$ is respectively the number of beamspots squared and the number of beamspots. Substituting these Polynomial Chaos Expansions into Equation (5.3), yields:

$$
f(\boldsymbol{x}) = \int \left(\sum_{j \in \mathcal{B}} \sum_{j' \in \mathcal{B}} \left[\sum_{k=0}^{\infty} R_{jj'}^{(k)} \Psi_k(\boldsymbol{\xi}) \right] x_j x_{j'} - 2 \sum_{j \in \mathcal{B}} \left[\sum_{k=0}^{\infty} r_j^{(k)} \Psi_k(\boldsymbol{\xi}) \right] x_j + \sum_{i \in \mathcal{V}} w_i (d_i^{\mathrm{p}})^2 \right) p_{\boldsymbol{\xi}}(\boldsymbol{\xi}) \, \mathrm{d}\boldsymbol{\xi}
$$
(5.6)

$$
= h_0^2 \sum_{j \in \mathcal{B}} \sum_{j' \in \mathcal{B}} R_{jj'}^{(0)} x_j x_{j'} - 2h_0^2 \sum_{j \in \mathcal{B}} r_j^{(0)} x_j + \sum_{i \in \mathcal{V}} w_i (d_i^{\mathrm{p}})^2
$$
(5.7)

with h_0^2 representing the norm of the zero-th basis vector. The derivative of Equation (5.7) is derived in Equation (5.8). The symmetry of the matrix $R^{(0)}$ is used.

$$
\frac{\partial}{\partial x_n} f(\boldsymbol{x}) = h_0^2 \sum_{j \in \mathcal{B}} R_{jn}^{(0)} x_j + h_0^2 \sum_{j' \in \mathcal{B}} R_{nj'}^{(0)} x_{j'} - 2r_n^{(0)} = 2h_0^2 \sum_{j \in \mathcal{B}} R_{nj}^{(0)} x_j - 2r_n^{(0)}
$$
(5.8)

The elements of the Hessian matrix are expressed in Equation (5.9).

$$
\frac{\partial^2}{\partial x_m \partial x_n} f(\boldsymbol{x}) = 2h_0^2 R_{mn}^{(0)}
$$
(5.9)

Note that only the zero-th order coefficient of the Polynomial Chaos Expansions are needed for the objective function, its gradient and its Hessian matrix.

5.3 Minimum expected dose

It might be desirable to include the minimum expected dose in a structure $\mathcal{V}$ in the optimization. In this case the function f is used:

$$
f(\boldsymbol{x}) = \min_{i \in \mathcal{V}} \int \sum_{j \in \mathcal{B}} D_{ij}(\boldsymbol{\xi}) x_j p_{\boldsymbol{\xi}}(\boldsymbol{\xi}) \, \mathrm{d}\boldsymbol{\xi} = \min_{i \in \mathcal{V}} \sum_{j \in \mathcal{B}} x_j \int D_{ij}(\boldsymbol{\xi}) p_{\boldsymbol{\xi}}(\boldsymbol{\xi}) \, \mathrm{d}\boldsymbol{\xi}
$$
(5.10)

A PCE can be constructed of the elements D_{ij} of the dose deposition matrix.

$$
D_{ij}(\boldsymbol{\xi}) = \sum_{k=0}^{\infty} R_{ij}^{(k)} \Psi_k(\boldsymbol{\xi})
$$
(5.11)

The matrix $R^{(k)}$ contains the Polynomial Chaos coefficients of the k-th order. Note that this is a different matrix than the one used in Section 5.2. Substituting the PCE of Equation (5.11) into Equation (5.10), yields:

$$
f(\boldsymbol{x}) = \min_{i \in \mathcal{V}} \sum_{j \in \mathcal{B}} x_j \int \sum_{k=0}^{\infty} R_{ij}^{(k)} \Psi_k(\boldsymbol{\xi}) p_{\boldsymbol{\xi}}(\boldsymbol{\xi}) \, \mathrm{d}\boldsymbol{\xi} = \min_{i \in \mathcal{V}} h_0^2 \sum_{j \in \mathcal{B}} R_{ij}^{(0)} x_j
$$
(5.12)

In general, for a certain $\boldsymbol{x}$ the minimum of a set of arbitrary functions $g_i(\boldsymbol{x})$ equals the lowest function value for that particular $\boldsymbol{x}$. Moreover, the min- and max-function are continuous,

$\lim_{\boldsymbol{x}\to\boldsymbol{x}_0}\left(\min_{i\in\mathcal{V}} g_i(\boldsymbol{x})\right) = \min_{i\in\mathcal{V}} g_i(\boldsymbol{x}_0)$. The derivatives with respect to the beamspot intensities are not defined on a finite number of single points. However, they are defined on the intermediate intervals.

$$\min_{i\in\mathcal{V}} g_i(\boldsymbol{x}) = g_{l(\boldsymbol{x})}(\boldsymbol{x}) \quad \text{with} \quad l(\boldsymbol{x})\in\mathcal{V} : g_l(\boldsymbol{x})\le g_i(\boldsymbol{x}) \quad \forall i\in\mathcal{V} \tag{5.13}$$

$$\frac{\partial}{\partial x_n}\min_{i\in\mathcal{V}} g_i(\boldsymbol{x}) = \begin{cases} \frac{\partial}{\partial x_n} g_{l(\boldsymbol{x})}(x) & \text{if} \quad \exists l(\boldsymbol{x})\in\mathcal{V} : g_l(\boldsymbol{x}) < g_i(\boldsymbol{x}) \quad \forall i\in\mathcal{V}\setminus\{l(\boldsymbol{x})\} \\ \text{undefined} & \text{otherwise} \end{cases} \tag{5.14}$$

If these properties of the min- and max-function are applied to the original problem of Equation (5.12), this yields Equation (5.15)

$$\frac{\partial}{\partial x_n} f(\boldsymbol{x}) = \begin{cases} h_0^2 R_{l(\boldsymbol{x})n}^{(0)} & \text{if} \quad \exists l(\boldsymbol{x})\in\mathcal{V} : \sum_j R_{lj}^{(0)} x_j < \sum_j R_{ij}^{(0)} x_j \quad \forall i\in\mathcal{V}\setminus\{l(\boldsymbol{x})\} \\ \text{undefined} & \text{otherwise} \end{cases} \tag{5.15}$$

The Hessian matrix equals zero for all elements if the condition from Equation (5.15) is satisfied.

5.4 Expected minimum dose

It might be desirable to include the expected value of the minimum dose in a structure $\mathcal{V}$ in the optimization. In this case the function f is used:

$$f(\boldsymbol{x}) = \int \mathrm{d}\boldsymbol{\xi}\, p_{\boldsymbol{\xi}}(\boldsymbol{\xi}) \min_{i\in\mathcal{V}} \sum_{j\in\mathcal{B}} D_{ij}(\boldsymbol{\xi}) x_j \tag{5.16}$$

A PCE can be constructed of the minimum dose in the structure.

$$\min_{i\in\mathcal{V}} \sum_{j\in\mathcal{B}} D_{ij}(\boldsymbol{\xi}) x_j \approx \sum_{k=0}^{\infty} r^{(k)}(\boldsymbol{x}) \Psi_k(\boldsymbol{\xi}) \tag{5.17}$$

The scalar $r^{(k)}$ equals the Polynomial Chaos coefficient of the k-th order. Substituting this Polynomial Chaos Expansion in Equation (5.16), yields:

$$f(\boldsymbol{x}) = \int \mathrm{d}\boldsymbol{\xi}\, p_{\boldsymbol{\xi}}(\boldsymbol{\xi}) \sum_{k=0}^{\infty} r^{(k)}(\boldsymbol{x}) \Psi_k(\boldsymbol{\xi}) \tag{5.18}$$

$$= h_0^2 r^{(0)}(\boldsymbol{x}) \tag{5.19}$$

The derivative of this function is not that obvious. According to the Leibniz rule, the order of integration and differentiation can be interchanged if both the integrand and the derivative of the integrand are continuous [61]. However, the derivative with respect to x_n of the min-function is not continuous, as demonstrated in Equation (5.14). Although validity is therefore

not guaranteed, the Leibniz rule is nevertheless used:

$$\frac{\partial}{\partial x_n} f(\boldsymbol{x}) = \frac{\partial}{\partial x_n} \int \mathrm{d}\boldsymbol{\xi}\, p_{\boldsymbol{\xi}}(\boldsymbol{\xi}) \min_{i \in \mathcal{V}} \sum_{j \in \mathcal{B}} D_{ij}(\boldsymbol{\xi}) x_j \tag{5.20}$$

$$= \int \mathrm{d}\boldsymbol{\xi}\, p_{\boldsymbol{\xi}}(\boldsymbol{\xi}) \frac{\partial}{\partial x_n} \min_{i \in \mathcal{V}} \sum_{j \in \mathcal{B}} D_{ij}(\boldsymbol{\xi}) x_j \tag{5.21}$$

$$= \int \mathrm{d}\boldsymbol{\xi}\, p_{\boldsymbol{\xi}}(\boldsymbol{\xi}) \frac{\partial}{\partial x_n} \sum_{j \in \mathcal{B}} D_{l(\boldsymbol{x},\boldsymbol{\xi})j}(\boldsymbol{\xi}) x_j \tag{5.22}$$

$$= \int \mathrm{d}\boldsymbol{\xi}\, p_{\boldsymbol{\xi}}(\boldsymbol{\xi}) D_{l(\boldsymbol{x},\boldsymbol{\xi})n}(\boldsymbol{\xi})\,, \tag{5.23}$$

with $l(\boldsymbol{x},\boldsymbol{\xi}) \in \mathcal{V}$ such that $\sum_j D_{lj}(\boldsymbol{\xi}) x_j < \sum_j D_{ij}(\boldsymbol{\xi}) x_j$ for all $i \in \mathcal{V} \setminus \{l(\boldsymbol{x},\boldsymbol{\xi})\}$. The domain of the integral of Equation (5.23) might also contain single points where the derivative of the min-function is not defined, as demonstrated in Equation (5.14). Therefore, the integrand of Equation (5.23) could exhibit jump discontinuities. This integral can be approximated with numerical integration using a Gaussian quadrature rule. The Hessian matrix equals zero for all elements if the derivative of the min-function is defined.

It would also be possible to approximate the gradient with a finite difference scheme.

5.5 Expected mean dose

It might be desirable to include the mean expected dose in a structure $\mathcal{V}$ in the optimization. In this case the function f is used:

$$f(\boldsymbol{x}) = \int \frac{1}{N_{\mathrm{v}}} \sum_{i \in \mathcal{V}} \sum_{j \in \mathcal{B}} D_{ij}(\boldsymbol{\xi}) x_j p_{\boldsymbol{\xi}}(\boldsymbol{\xi})\, \mathrm{d}\boldsymbol{\xi} \tag{5.24}$$

$$= \frac{1}{N_{\mathrm{v}}} \sum_{j \in \mathcal{B}} x_j \int \sum_{i \in \mathcal{V}} D_{ij}(\boldsymbol{\xi}) p_{\boldsymbol{\xi}}(\boldsymbol{\xi})\, \mathrm{d}\boldsymbol{\xi} \tag{5.25}$$

A PCE can be constructed of the sum of the columns of the dose deposition matrix.

$$\sum_{i \in \mathcal{V}} D_{ij}(\boldsymbol{\xi}) = \sum_{k=0}^{\infty} r_j^{(k)} \Psi_k(\boldsymbol{\xi}) \tag{5.26}$$

The vector $\boldsymbol{r}^{(k)}$ equals the Polynomial Chaos coefficient of the k-th order. Substituting this Polynomial Chaos Expansion in Equation (5.25), yields:

$$f(\boldsymbol{x}) = \frac{1}{N_{\mathrm{v}}} \sum_{j \in \mathcal{B}} x_j \int \sum_{k=0}^{\infty} r_j^{(k)} \Psi_k(\boldsymbol{\xi}) p_{\boldsymbol{\xi}}(\boldsymbol{\xi})\, \mathrm{d}\boldsymbol{\xi} \tag{5.27}$$

$$= \frac{1}{N_{\mathrm{v}}} h_0^2 \sum_{j \in \mathcal{B}} r_j^{(0)} x_j \tag{5.28}$$

The derivative of the expected value of the mean of the dose in the structure is straightforward.

$$\frac{\partial}{\partial x_n} f(\boldsymbol{x}) = \frac{1}{N_{\mathrm{v}}} h_0^2 r_n^{(0)} \tag{5.29}$$

The Hessian matrix equals zero for all elements.

35

5.6 Mean expected dose

One might also be interested in the mean of the expected value of the dose in a structure $\mathcal{V}$. This equals the expected mean dose of Equation (5.28). The derivative is also equivalent to the derivative of the expected mean dose of Equation (5.29). The Hessian matrix equals zero as well.

5.7 $\beta\%$-probabilistic dose volume parameter

It might be desirable to include dose restrictions to a certain fraction of a structure $\mathcal{V}$ in the optimization. A dose parameter such as $D_{\alpha\%}(\boldsymbol{\xi})$ suits for this purpose. This is the maximum dose which at least $\alpha\%$ of the structure receives. A $\beta\%$-probabilistic $D_{\alpha\%}$ is defined as the maximum $D_{\alpha\%}$ that is given at least with a $\beta\%$ probability. Mathematically it would be described as the $(100-\beta)$-th percentile of the dose volume parameter $D_{\alpha\%}$, i.e. $P^s_{(100-\beta)\%}\left\{P^i_{(100-\alpha)\%}\left\{d_i(\boldsymbol{\xi}^{(s)})\right\}\right\}$ with $\boldsymbol{\xi}^{(s)}$ random samples of the uncertainty vector.

Some interesting and useful parameters are the 95%-probabilistic $D_{98\%}(\boldsymbol{\xi})$, which is abbreviated as the probabilistic near-minimum dose. Another example is the 5%-probabilistic $D_{2\%}(\boldsymbol{\xi})$, which is abbreviated as the probabilistic near-maximum dose.

Because the $\beta\%$-probabilistic $D_{\alpha\%}$ is calculated using percentiles, it is not possible to express it analytically.

5.7.1 PCE of dose deposition matrix

First, PCEs of the elements $D_{ij}(\boldsymbol{\xi})$ of the dose deposition matrix are constructed, as in Equation (5.11).

$$D_{ij}(\boldsymbol{\xi}) = \sum_{k=0}^{\infty} R_{ij}^{(k)} \Psi_k(\boldsymbol{\xi}) \tag{5.30}$$

Subsequently, N_s samples from these Polynomial Chaos Expansions are drawn randomly. These are denoted by $D_{ij}^{(s)} \in \{D_{ij}^{(1)}, ..., D_{ij}^{(N_s)}\}$. The near-minimum dose of the structure $\mathcal{V}$ is calculated for every realization s as follows:

$$D_{\alpha\%}^{(s)} = P^i_{(100-\alpha)\%}\left\{\sum_{j\in\mathcal{B}} D_{ij}^{(s)} x_j\right\}. \tag{5.31}$$

The $\beta\%$-probabilistic $D_{\alpha\%}$, $f(\boldsymbol{x})$, is determined by taking the $(100-\beta)$-th percentile of the set $\left\{D_{\alpha\%}^{(s)} : s \in \{1, ..., N_s\}\right\}$.

$$f(\boldsymbol{x}) = P^s_{(100-\beta)\%}\left\{D_{\alpha\%}^{(s)}\right\} = P^s_{(100-\beta)\%}\left\{P^i_{(100-\alpha)\%}\left\{\sum_{j\in\mathcal{B}} D_{ij}^{(s)} x_j\right\}\right\} \tag{5.32}$$

As no explicit analytical expression is present for the $\beta\%$-probabilistic $D_{\alpha\%}$, the same holds for the derivative with respect to the beam intensity vector. The components of the gradient of the $\beta\%$-probabilistic $D_{\alpha\%}$ are approximated using finite differences. The step size is denoted by h, while the unit vector along the n-th element of the beam intensity vector is denoted by $\hat{\boldsymbol{e}}_n$. Equation (5.33) shows a forward finite difference scheme:

$$\frac{\partial}{\partial x_n} f(\boldsymbol{x}) \approx \frac{f(\boldsymbol{x} + h\hat{\boldsymbol{e}}_n) - f(\boldsymbol{x})}{h}, \tag{5.33}$$

which yields:

$$\frac{\partial}{\partial x_n} f(\boldsymbol{x}) \approx \frac{P^s_{(100-\beta)\%} \left\{ P^i_{(100-\alpha)\%} \left\{ \sum_{j \in \mathcal{B}} D^{(s)}_{ij} x_j + D^{(s)}_{in} h \right\} \right\}}{h} - $$
$$\frac{P^s_{(100-\beta)\%} \left\{ P^i_{(100-\alpha)\%} \left\{ \sum_{j \in \mathcal{B}} D^{(s)}_{ij} x_j \right\} \right\}}{h} . \quad (5.34)$$

5.7.2 PCE of dose volume parameter

The $\beta\%$-probabilistic $D_{\alpha\%}$ could also be computed in a different way, but again using Polynomial Chaos Expansion. Instead of constructing a PCE of the dose deposition matrix and sampling from this PCE, it is also possible to construct a PCE of the $D_{\alpha\%}$ itself, as shown in Equation (5.35).

$$D_{\alpha\%}(\boldsymbol{x}, \boldsymbol{\xi}) = \sum_{k=0}^{\infty} r^{(k)}(\boldsymbol{x}) \Psi_k(\boldsymbol{\xi}) \quad (5.35)$$

The scalar $r^{(k)}(\boldsymbol{x})$ equals the Polynomial Chaos coefficient of the k-th order. Subsequently, N_s samples from this Polynomial Chaos Expansion are calculated. These are denoted by $D^{(s)}_{\alpha\%} \in \{D^{(1)}_{\alpha\%}, ..., D^{(N_s)}_{\alpha\%}\}$. The $\beta\%$-probabilistic $D_{\alpha\%}$, $f(\boldsymbol{x})$, is determined by taking the $(100 - \beta)$-th percentile of the set $\left\{ D^{(s)}_{\alpha\%} : s \in \{1, ..., N_s\} \right\}$.

$$f(\boldsymbol{x}) = P^s_{(100-\beta)\%} \left\{ D^{(s)}_{\alpha\%} \right\} \quad (5.36)$$

Because no explicit analytical expression exists for the $D_{\alpha\%}$, the gradient and Hessian are approximated with a forward finite difference scheme:

$$\frac{\partial}{\partial x_n} f(\boldsymbol{x}) \approx \frac{f(\boldsymbol{x} + h \hat{\boldsymbol{e}}_n) - f(\boldsymbol{x})}{h} . \quad (5.37)$$

The Hessian matrix can be calculated similarly.

It should be noted that the parameter $D_{\alpha\%}$ depends on the beam intensity vector and therefore the same holds for the coefficients $r^{(k)}(\boldsymbol{x})$. However, the actual dependence of $r^{(k)}(\boldsymbol{x})$ on $\boldsymbol{x}$ is unknown. Consequently, a PCE needs to be constructed for every $\boldsymbol{x}$ which is contained in the finite difference scheme of the gradient and Hessian. For a forward difference scheme, $N_b + 1$ PCEs are to be constructed for the gradient, with N_b the number of elements of the beam intensity vector. If the Hessian is needed as well, this number increases to $\frac{1}{2} N_b^2 + \frac{3}{2} N_b + 1$ PCEs [62].

5.8 $\beta\%$-probabilistic minimum dose

It might be desirable to include the $\beta\%$-probabilistic minimum dose in a structure $\mathcal{V}$ in the optimization. The $\beta\%$-probabilistic minimum dose is the maximum dose in a structure that is given at least with a $(100 - \beta)\%$ probability. Because the $\beta\%$-probabilistic minimum dose is calculated using the percentile function, it is not possible to express the probabilistic minimum dose analytically.

First, PCEs of the elements $D_{ij}(\boldsymbol{\xi})$ of the dose deposition matrix are constructed.

$$D_{ij}(\boldsymbol{\xi}) = \sum_{k=0}^{\infty} R^{(k)}_{ij} \Psi_k(\boldsymbol{\xi}) \quad (5.38)$$

Subsequently, N_s samples from this Polynomial Chaos Expansion are calculated. These are denoted by $D_{ij}^{(s)} \in \{D_{ij}^{(1)}, ..., D_{ij}^{(N_\text{s})}\}$. The minimum dose of the structure $\mathcal{V}$ is calculated for every realization s. The probabilistic minimum dose, $f(\boldsymbol{x})$, is determined by taking the $(100 - \beta)$-th percentile of the set $\left\{\min_{i \in \mathcal{V}} \sum_{j \in \mathcal{B}} D_{ij}^{(s)} x_j : s \in \{1, ..., N_\text{s}\}\right\}$.

$$f(\boldsymbol{x}) = P_{(100-\beta)\%}^s \left\{\min_{i \in \mathcal{V}} \sum_{j \in \mathcal{B}} D_{ij}^{(s)} x_j\right\} \tag{5.39}$$

As no explicit analytical expression is present for the probabilistic minimum dose, this also holds for the derivative with respect to the beam intensity vector. The components of the gradient of the probabilistic minimum dose are approximated using finite differences. The scalar h and the unit vector $\hat{\boldsymbol{e}}_n$ are defined in the same way as in Section 5.7.

$$\frac{\partial}{\partial x_n} f(\boldsymbol{x}) \approx \frac{f(\boldsymbol{x} + h\hat{\boldsymbol{e}}_n) - f(\boldsymbol{x})}{h} \tag{5.40}$$

$$= \frac{P_{(100-\beta)\%}^s \left\{\min_{i \in \mathcal{V}} \left(\sum_{j \in \mathcal{B}} D_{ij}^{(s)} x_j + D_{in}^{(s)} h\right)\right\} - P_{(100-\beta)\%}^s \left\{\min_{i \in \mathcal{V}} \sum_{j \in \mathcal{B}} D_{ij}^{(s)} x_j\right\}}{h}$$

$$\tag{5.41}$$

The Hessian matrix can be calculated similarly.

Alternatively, one could also construct a PCE of the minimum dose itself. Subsequently, the $(100 - \beta)$-th percentile from random samples from this PCE should be calculated.

5.9   Handling multiple fractions

A treatment plan could consist of multiple fractions, periodically delivered to a patient. This section describes how to handle these fractions properly with Polynomial Chaos Expansion. In this section, $\boldsymbol{\xi}^{(t)}$ is used to denote the fractional uncertainty vector, contrary to previous sections where $\boldsymbol{\xi}^{(s)}$ was used to denote a random sample of the uncertainty vector.

5.9.1   General

Consider a tumor, which is to be irradiated, subject to a set of errors $\boldsymbol{\xi} = (\xi_1, ..., \xi_N)$. Consider a function $g(\boldsymbol{x}, D(\boldsymbol{\xi}))$. The dependence on the dose deposition matrix is explicitly stated for clarification. One could define an objective function $f(\mathcal{M}_{\boldsymbol{\xi}}\left[g(\boldsymbol{x}, D(\boldsymbol{\xi}))\right])$, which is a function of any statistical measure of $g(\boldsymbol{x}, D(\boldsymbol{\xi}))$. For the optimization of this treatment consisting of only a single fraction, this objective function is to be minimized. The mathematical problem is summarized in Equation (5.42).

$$\underset{\boldsymbol{x}}{\text{minimize}} \; f\left(\mathcal{M}_{\boldsymbol{\xi}}\left[g(\boldsymbol{x}, D(\boldsymbol{\xi}))\right]\right) \tag{5.42}$$

When the same problem is to be optimized for multiple fractions, the mathematical formulation changes slightly, if the considered errors are not systematic but differ among the fractions. Consider a set of fractions $\mathcal{T}$, distinguished by the index $t \in \mathcal{T} = \{1, ..., T\}$. Each fraction is subject to a different and independent error $\boldsymbol{\xi}^{(t)} = \left(\xi_1^{(t)}, ..., \xi_N^{(t)}\right)^T$. These set-up errors lead to

different dose deposition matrices $D(\boldsymbol{\xi}^{(t)})$. The dose $\boldsymbol{d}$, which is delivered in the end, is the sum of the doses of the fractions $\boldsymbol{d}^{(t)} = \frac{1}{T}D(\boldsymbol{\xi}^{(t)})\boldsymbol{x}$.

$$d_i = \sum_{t\in\mathcal{T}} d_i^{(t)} = \sum_{t\in\mathcal{T}} \frac{1}{T} \sum_{j\in\mathcal{B}} D_{ij}(\boldsymbol{\xi}^{(t)})x_j = \sum_{j\in\mathcal{B}} \left(\frac{1}{T} \sum_{t\in\mathcal{T}} D_{ij}(\boldsymbol{\xi}^{(t)}) \right) x_j \tag{5.43}$$

The finally delivered dose equals the matrix-vector product of the mean of the dose deposition matrices and the beam intensity vector. This *effective dose deposition matrix* $\tilde{D}_{ij}(\tilde{\boldsymbol{\xi}})$ is defined in Equation (5.44):

$$\tilde{D}_{ij}(\tilde{\boldsymbol{\xi}}) = \frac{1}{T} \sum_{t\in\mathcal{T}} D_{ij}(\boldsymbol{\xi}^{(t)}), \tag{5.44}$$

where $\tilde{\boldsymbol{\xi}}$ is the combined random error for the whole treatment and is formed by a vertical concatenation, in this thesis denoted with $\|$, of the random errors of the individual fractions:

$$\tilde{\boldsymbol{\xi}} = \mathop{\|}_{t\in\mathcal{T}} \boldsymbol{\xi}^{(t)}. \tag{5.45}$$

Note that the combined random error is a vector of size $T \cdot N$. Substituting the effective dose deposition matrix into $g(\boldsymbol{x}, D(\boldsymbol{\xi}))$ and subsequently calculating the statistical measure over all fractions, yields the adjusted objective function $f\left(\mathcal{M}_{\tilde{\boldsymbol{\xi}}}\left[g(\boldsymbol{x}, \tilde{D}(\tilde{\boldsymbol{\xi}}))\right]\right)$. Equation (5.46) describes the mathematical formulation of this optimization problem for a treatment consisting of multiple fractions.

$$\underset{\boldsymbol{x}}{\text{minimize }} f\left(\mathcal{M}_{\tilde{\boldsymbol{\xi}}}\left[g(\boldsymbol{x}, \tilde{D}(\tilde{\boldsymbol{\xi}}))\right]\right) \tag{5.46}$$

Considering the expected value instead of an arbitrary statistical measure, yields Equation (5.47).

$$\underset{\boldsymbol{x}}{\text{minimize }} f\left(\int g(\boldsymbol{x}, \tilde{D}(\tilde{\boldsymbol{\xi}}))p_{\tilde{\boldsymbol{\xi}}}(\tilde{\boldsymbol{\xi}})\,\mathrm{d}\tilde{\boldsymbol{\xi}}\right) =$$
$$\underset{\boldsymbol{x}}{\text{minimize }} f\left(\int ... \int g(\boldsymbol{x}, \tilde{D}(\boldsymbol{\xi}^{(1)}, ..., \boldsymbol{\xi}^{(T)}))p_{\boldsymbol{\xi}^{(1)}}(\boldsymbol{\xi}^{(1)})...p_{\boldsymbol{\xi}^{(T)}}(\boldsymbol{\xi}^{(T)})\,\mathrm{d}\boldsymbol{\xi}^{(1)}...\mathrm{d}\boldsymbol{\xi}^{(T)}\right) \tag{5.47}$$

As Equations (5.46) and (5.47) show, more uncertainty parameters are introduced when fractionating is taken into account. Such statistical measures can be calculated with PCE in different ways. One could construct a PCE of the fractional dose deposition matrix $D(\boldsymbol{\xi}^{(t)})$ or of the fractional dose $\boldsymbol{d}^{(t)}(\boldsymbol{\xi}^{(t)})$. A drawback is that if sampling is required for the computation of the statistical measure, one would need T times more samples than for a single fraction treatment. After all, for each sample of the finally delivered dose $\boldsymbol{d}$, T samples are needed of either $D(\boldsymbol{\xi}^{(t)})$ or $\boldsymbol{d}^{(t)}(\boldsymbol{\xi}^{(t)})$. Since the dose deposition matrix contains usually many elements (i.e. $N_{\mathrm{v}} \cdot N_{\mathrm{b}}$), sampling would be computationally expensive in terms of memory.

Instead of constructing a PCE of a fractional quantity, one could also construct a PCE of $g(\boldsymbol{x}, \tilde{D}(\tilde{\boldsymbol{\xi}}))$ as a whole. In that case, it should be noted that for the random error of each fraction N separate stochastic variables should be assigned for the PCE. The number of dimensions of the PCE therefore increases with the number of fractions and equals $T \cdot N$. So calculating the Polynomial Chaos coefficients will require more cubature points, as quantified in Table 4.3.

However, the mutual order of the fractions is physically irrelevant. This means that it does not matter in which order the different fractions are received by the patient. This also has a consequence on the mathematics involving the PCE. When multiple fractions are considered, the set of cubature points contains multiple permutations of the same combination. Since the

order of the fractions is irrelevant, different permutations of a single cubature point yield the same stochastic response. Consequently, less function evaluations are needed than described in Table 4.3.

As an example, consider a treatment consisting of two fractions (i.e. $T = 2$) subject to a three-dimensional set-up error (i.e. $N = 3$). Suppose two set-up errors for a single fraction, $\boldsymbol{m} \in \mathbb{R}^3$ and $\boldsymbol{n} \in \mathbb{R}^3$, with $\boldsymbol{m} \neq \boldsymbol{n}$. Since the order of the fractions is physically irrelevant, the cubature points $\boldsymbol{m}||\boldsymbol{n}$ and $\boldsymbol{n}||\boldsymbol{m}$ yield the same effective dose deposition matrix. Therefore, the stochastic response of these two cubature points is equal.

The number of function evaluations increases for higher number of fractions, however, for $T \geq L - 1$, it remains constant, as shown in Table 5.1. This constant increases with the grid level and the number of dimensions. A mathematical explanation is provided in Appendix B. For $L = 5$ and $N = 3$, the number of function evaluations is 780 for $T \geq L - 1$. For $L = 5$ and $N = 4$, the number of function evaluations is 1607 for $T \geq L - 1$.

Table 5.1: The number of required function evaluations (N_{fval}) and the number of cubature points (N_{cub}) against the number of fractions (T) for a PCE with $L = 5$ and $\Delta_{lev} = 1$, considering a three-dimensional random error ($N = 3$). For $T \geq L - 1$, the number of function evaluations is constant.

T	$T \cdot N$	N_{cub}	N_{fval}
1	3	303	303
2	6	2393	669
3	9	9079	765
4	12	24465	780
5	15	53951	780

The number of cubature points and function evaluations does not necessarily increase for every function of which a PCE is constructed when considering multiple fractions. Section 5.9.2 treats an example of a function for which the number of dimensions of the PCE does not increase with the number of fractions considered.

Infinite fractions

For the hypothetical case of infinite fractions, the uncertainty in the delivered dose reduces to zero. The finally delivered dose equals the expected value of the dose distribution, $\mathrm{E}[\boldsymbol{d}] = \mathrm{E}[D(\xi)]\boldsymbol{x}$. After all, the expected value of a stochastic variable is the mean of an infinite number of samples. The mathematical formulation of the optimization reduces to Equation (5.48).

$$\underset{\boldsymbol{x}}{\text{minimize}} \; f\left(g\left(\boldsymbol{x}, \mathrm{E}[D(\boldsymbol{\xi})]\right)\right) \tag{5.48}$$

5.9.2 Expected quadratic dose

In Section 5.2, the expected value of the quadratic objective function has been treated. This function is particularly suitable when treatment planning is done for multiple fractions. For this example, a one-dimensional tumor with an evidently one-dimensional set-up error (i.e. $N = 1$) is considered.

Substituting Equation (5.44) into (5.3), yields:

$$f(\boldsymbol{x}) = \int \left(\frac{1}{T^2} \sum_{j \in \mathcal{B}} \sum_{j' \in \mathcal{B}} x_j x_{j'} \sum_{t \in \mathcal{T}} \sum_{t' \in \mathcal{T}} \left[\sum_{i \in \mathcal{V}} w_i D_{ij}(\xi^{(t)}) D_{ij'}(\xi^{(t')}) \right] - \right.$$
$$\left. \frac{2}{T} \sum_{j \in \mathcal{B}} x_j \sum_{t \in \mathcal{T}} \left[\sum_{i \in \mathcal{V}} w_i d_i^{\mathrm{p}} D_{ij}(\xi^{(t)}) \right] + \sum_{i \in \mathcal{V}} w_i (d_i^{\mathrm{p}})^2 \right) p_{\tilde{\boldsymbol{\xi}}}(\tilde{\boldsymbol{\xi}}) \, \mathrm{d}\tilde{\boldsymbol{\xi}} \quad (5.49)$$

Some rearrangements yield:

$$f(\boldsymbol{x}) = \frac{1}{T^2} \sum_{j \in \mathcal{B}} \sum_{j' \in \mathcal{B}} x_j x_{j'} \sum_{t \in \mathcal{T}} \sum_{t' \in \mathcal{T}} \left[\int \sum_{i \in \mathcal{V}} w_i D_{ij}(\xi^{(t)}) D_{ij'}(\xi^{(t')}) p_{\tilde{\boldsymbol{\xi}}}(\tilde{\boldsymbol{\xi}}) \, \mathrm{d}\tilde{\boldsymbol{\xi}} \right] - $$
$$\frac{2}{T} \sum_{j \in \mathcal{B}} x_j \sum_{t \in \mathcal{T}} \left[\int \sum_{i \in \mathcal{V}} w_i d_i^{\mathrm{p}} D_{ij}(\xi_t) p_{\tilde{\boldsymbol{\xi}}}(\tilde{\boldsymbol{\xi}}) \, \mathrm{d}\tilde{\boldsymbol{\xi}} \right] + \sum_{i \in \mathcal{V}} w_i (d_i^{\mathrm{p}})^2 \quad (5.50)$$

It should be noted that the dose deposition matrices of the fractions, $\{D(\xi^{(t)}) : t \in \mathcal{T}\}$, are uncorrelated and mutually independent. However, they depend each on a Gaussian distribution with the same mean and variance. Therefore, the following holds for $\alpha_{jj'tt'}$, the first term in square brackets of Equation (5.50):

$$\alpha_{jj'tt'} = \int \sum_{i \in \mathcal{V}} w_i D_{ij}(\xi^{(t)}) D_{ij'}(\xi^{(t')}) p_{\tilde{\boldsymbol{\xi}}}(\tilde{\boldsymbol{\xi}}) \, \mathrm{d}\tilde{\boldsymbol{\xi}} \quad (5.51)$$

$$= \int \int \sum_{i \in \mathcal{V}} w_i D_{ij}(\xi^{(t)}) D_{ij'}(\xi^{(t')}) p_{\xi^{(t)}}(\xi^{(t)}) p_{\xi^{(t')}}(\xi^{(t')}) \, \mathrm{d}\xi^{(t)} \mathrm{d}\xi^{(t')} \prod_{l \neq t, t'} \int p_{\xi^{(l)}}(\xi^{(l)}) \, \mathrm{d}\xi^{(l)} \quad (5.52)$$

$$= \int \int \sum_{i \in \mathcal{V}} w_i D_{ij}(\xi^{(t)}) D_{ij'}(\xi^{(t')}) p_{\xi^{(t)}}(\xi^{(t)}) p_{\xi^{(t')}}(\xi^{(t')}) \, \mathrm{d}\xi^{(t)} \mathrm{d}\xi^{(t')} \quad (5.53)$$

$$= \begin{cases} C_{jj'}^{(1)} & \forall t, t' \in \mathcal{T} : t = t' \\ C_{jj'}^{(2)} & \forall t, t' \in \mathcal{T} : t \neq t' \end{cases} \quad (5.54)$$

Similar rearrangements can be applied to β_{jt}, the second term in square brackets of Equation (5.50):

$$\beta_{jt} = \int \sum_{i \in \mathcal{V}} w_i d_i^{\mathrm{p}} D_{ij}(\xi^{(t)}) p_{\tilde{\boldsymbol{\xi}}}(\tilde{\boldsymbol{\xi}}) \, \mathrm{d}\tilde{\boldsymbol{\xi}} \quad (5.55)$$

$$= \int \sum_{i \in \mathcal{V}} w_i d_i^{\mathrm{p}} D_{ij}(\xi^{(t)}) p_{\xi^{(t)}}(\xi^{(t)}) \, \mathrm{d}\xi^{(t)} \prod_{l \neq t} \int p_{\xi^{(l)}}(\xi^{(l)}) \, \mathrm{d}\xi^{(l)} \quad (5.56)$$

$$= \int \sum_{i \in \mathcal{V}} w_i d_i^{\mathrm{p}} D_{ij}(\xi^{(t)}) p_{\xi^{(t)}}(\xi^{(t)}) \, \mathrm{d}\xi^{(t)} \quad (5.57)$$

$$= c_j \quad \forall t \in \mathcal{T} \quad (5.58)$$

One could expand the sums over $\mathcal{T}$ of Equation (5.50), into which Equations (5.54) and (5.58) can be substituted. This results in Equations (5.59) and (5.60).

$$f(\boldsymbol{x}) = \frac{1}{T^2} \sum_{j \in \mathcal{B}} \sum_{j' \in \mathcal{B}} x_j x_{j'} \left(T C_{jj'}^{(1)} + (T^2 - T) C_{jj'}^{(2)} \right) - \frac{2}{T} \sum_{j \in \mathcal{B}} x_j T c_j + \sum_{i \in \mathcal{V}} w_i (d_i^{\mathrm{p}})^2 \qquad (5.59)$$

$$= \sum_{j \in \mathcal{B}} \sum_{j' \in \mathcal{B}} x_j x_{j'} \left(\frac{1}{T} C_{jj'}^{(1)} + \left(1 - \frac{1}{T} \right) C_{jj'}^{(2)} \right) - 2 \sum_{j \in \mathcal{B}} x_j c_j + \sum_{i \in \mathcal{V}} w_i (d_i^{\mathrm{p}})^2 \qquad (5.60)$$

Polynomial Chaos Expansion can be used to approximate the elements of the matrices $C_{jj'}^{(1)}$ and $C_{jj'}^{(2)}$. For this purpose, a PCE of the integrand of Equation (5.53) is to be constructed for both two scenarios, $t = t'$ and $t \neq t'$. For the former scenario the dimensionality of the PCE is reduced to 1, while for the latter it is reduced to 2. Subsequently, the elements of the matrices $C_{jj'}^{(1)}$ and $C_{jj'}^{(2)}$ can be approximated with the zero-th coefficients of the computed PCEs.

Analogously, the elements of the vector c_j can be approximated with the zero-th order PCE coefficients of the integrand of Equation (5.57). This PCE is only single dimensional.

For the hypothetical case of infinite fractions, the limit of Equation (5.60) to infinity is to be taken. The cross terms $C_{jj'}^{(2)}$ disappear. This result can also be obtained with Equation (5.48).

Similar derivations apply to objective functions which are expected values of a polynomial of degree p in dose. Analogously, the dimensionality of the PCE can be reduced to $p \cdot N$.

Regarding Equation (5.53), it would also be possible to change the order of integration and summation. Then, the two-dimensional integral can be split up into two identical one-dimensional integrals. If PCE would be used to approximate these integrals, the dimensionality would indeed further reduce to one, or in general to N. However, the number of PCEs to be calculated could increase. In general, for handling multiple fractions with the expected value of a polynomial in dose of degree p, the choice whether to split up the p-dimensional integral into p identical one-dimensional integrals is a trade-off which depends on the the number of dimensions N, the degree p and the ratio between the number of beamspots and the number of voxels. For example, if $N = 1$ and $N_{\mathrm{v}} > N_{\mathrm{b}}$, the number of required PCEs increases when one would split up the integral in p identical one-dimensional integrals. On the other hand, the number of cubature points would be considerably less.

5.9.3 Examples: reduction of variables

Examples 5.1 and 5.2 demonstrate that generally it is not possible to replace the random errors of the different fractions by one variable which describes all.

Example 5.1. *Suppose that the dose deposition in each voxel is linear with the uncertainty,* $D_{ij}(\xi) = a_{ij}\xi + b_{ij}$. *Let the uncertainty be Gaussian distributed:* $\xi^{(t)} \sim \mathcal{N}(\mu_\xi, \sigma_\xi^2)$ *for all* $t \in \mathcal{T}$.

$$\int \frac{1}{T} \sum_{t \in \mathcal{T}} D_{ij}(\xi^{(t)}) p_{\xi^{(t)}}(\xi^{(t)}) \, \mathrm{d}\xi^{(t)} = \frac{1}{T} \sum_{t \in \mathcal{T}} \int D_{ij}(\xi^{(t)}) p_{\xi^{(t)}}(\xi^{(t)}) \, \mathrm{d}\xi^{(t)} \qquad (5.61)$$

$$= \frac{1}{T} \sum_{t \in \mathcal{T}} \int (a_{ij}\xi^{(t)} + b_{ij}) p_{\xi^{(t)}}(\xi^{(t)}) \, \mathrm{d}\xi^{(t)} \qquad (5.62)$$

$$= a_{ij}\mu_\xi + b_{ij} \qquad (5.63)$$

Let $\bar{\xi} = \frac{1}{T} \sum_{t \in \mathcal{T}} \xi^{(t)}$ and as a consequence $\bar{\xi} \sim \mathcal{N}(\mu_\xi, \frac{\sigma_\xi^2}{T})$.

$$\int D_{ij}(\bar{\xi}) p_{\bar{\xi}}(\bar{\xi}) \, \mathrm{d}\bar{\xi} = \int (a_{ij}\bar{\xi} + b_{ij}) p_{\bar{\xi}}(\bar{\xi}) \, \mathrm{d}\bar{\xi} \tag{5.64}$$

$$= a_{ij}\mu_\xi + b_{ij} \tag{5.65}$$

Equations (5.61) and (5.64) yield the same answer. Therefore, for a linear dependence of the fractional dose deposition on the uncertainty parameter, it is possible to combine the fractional uncertainty parameters into a mean uncertainty parameter $\bar{\xi}$.

Example 5.2. *Suppose a non-linear dependence of the dose deposition in each voxel on the uncertainty, for example $D_{ij}(\xi) = \xi^2$. Let the uncertainty be Gaussian distributed: $\xi^{(t)} \sim \mathcal{N}(\mu_\xi, \sigma_\xi^2)$ for all $t \in \mathcal{T}$.*

$$\int \frac{1}{T} \sum_{t \in \mathcal{T}} D_{ij}(\xi^{(t)}) p_{\xi^{(t)}}(\xi^{(t)}) \, \mathrm{d}\xi_t = \frac{1}{T} \sum_{t \in \mathcal{T}} \int D_{ij}(\xi^{(t)}) p_{\xi^{(t)}}(\xi^{(t)}) \, \mathrm{d}\xi^{(t)} \tag{5.66}$$

$$= \frac{1}{T} \sum_{t \in \mathcal{T}} \int (\xi^{(t)})^2 p_{\xi^{(t)}}(\xi^{(t)}) \, \mathrm{d}\xi^{(t)} \tag{5.67}$$

$$= \mu_\xi^2 + \sigma_\xi^2 \tag{5.68}$$

Let $\bar{\xi} = \frac{1}{T} \sum_{t \in \mathcal{T}} \xi^{(t)}$ and as a consequence $\bar{\xi} \sim \mathcal{N}(\mu_\xi, \frac{\sigma_\xi^2}{T})$.

$$\int D_{ij}(\bar{\xi}) p_{\bar{\xi}}(\bar{\xi}) \, \mathrm{d}\bar{\xi} = \int \bar{\xi}^2 p_{\bar{\xi}}(\bar{\xi}) \, \mathrm{d}\bar{\xi} \tag{5.69}$$

$$= \mu_\xi^2 + \frac{\sigma_\xi^2}{T} \tag{5.70}$$

Equations (5.66) and (5.69) yield a different answer. Therefore, in general, for a non-linear or unknown dependence of the fractional dose deposition on the uncertainty parameter, it is not possible to combine the fractional uncertainty parameters into a mean uncertainty parameter $\bar{\xi}$.

Despite this not being an actual proof, these two examples show that combining uncertainty variables of different fractions into one variable, which describes the mean of the uncertainties over the fractions, is only possible when the dose deposition is linearly dependent on the uncertainty. However, obviously this is not the case, due to the Gaussian lateral spread-out of the beams.

5.9.4 Evaluating a treatment plan

For evaluating a treatment plan consisting of T fractions subject to a random error consisting of N dimensions, it is not possible to construct a PCE with only N variables such that one could draw one sample and obtain information about the treatment.

However, earlier research showed that considering only one fraction and therefore including only N variables in the PCE for the dose, sampling T times from this PCE allows to obtain information about the treatment as a whole [2, 13]. In fact, it should be noted that this way comes down to the construction of a PCE of a fractional quantity instead of a treatment quantity. An example of a fractional quantity is the fractional dose and an example of a treatment quantity is the treatment dose. Obtaining information about a treatment quantity with PCE by considering only N variables, is only possible when the relation between the treatment quantity and the fractional quantity is known.

For example: the relation between the treatment dose and the individual fractional doses is known (i.e. sum) and therefore multiple samples could be drawn from a PCE of the fractional dose and summing them yields the treatment dose. On the other hand, the relation between the treatment dose volume parameter and the fractional dose volume parameter is unknown and therefore it is not possible to obtain information about the treatment dose volume parameter by sampling from a PCE of the fractional dose volume parameter.

Chapter 6

Applications

This chapter describes the methodologies for three applications treated in this research.

6.1 Multiple fractions in 1D

Recently, research has shown the consequences of incorporating a random one-dimensional set-up error probabilistically when minimizing the expected value of a quadratic function in dose [1]. It is a one-dimensional problem with a tumor surrounded by non-tumorous tissue.

This problem is reworked in Matlab in order to investigate whether similar results can be obtained using Polynomial Chaos Expansion. A one-dimensional, discrete voxel-grid is constructed. The domain ranges from -60 to 60 mm and the voxel width is 0.5 mm. The tumor is located between -30 mm and 30 mm.

The tumor is irradiated by $N_{\mathrm{b}} = 36$ equidistantly spaced beamspots, with 2 mm distance. For this problem, the lateral spread-out of the proton beamspots is assumed to be Gaussian with a standard deviation of 3 mm. Since a discrete voxel-grid is used, dose conservation needs to be imposed in each voxel. This means that the mean dose in each voxel is assigned to the center of that voxel. The mean dose is approximated using the Gauss-Legendre quadrature rule with 2 points [63].

This one-dimensonal set-up error is assumed to be Gaussian distributed with zero mean and a standard deviation equal to $\sigma_x = 3$ mm. These settings will also be provided to all PCEs which are to be constructed for this problem.

The expected value of the quadratic differences between the dose and the prescribed dose is minimized for a fractionated treatment, as described in Section 5.9.2. First, a Polynomial Chaos Expansion is to be constructed of the integrand of Equation (5.53). Two cases are distinguished; the integrand depending on either one or two fractions. In both cases, the integrand depends on the indices $j, j' \in \mathcal{B}$ and is invariant under interchange of these indices. This results in $\frac{1}{2}N_{\mathrm{b}}^2 + \frac{1}{2}N_{\mathrm{b}} = 666$ two-dimensional PCEs and an equal number of one-dimensional PCEs. The integral is approximated with the zero-th coefficients of the PCEs. Another option would be to split the integral of Equation (5.53) into two identical one-dimensional integrals. Then, a PCE of the dose deposition matrix could be calculated, which would require more PCEs, i.e. $N_{\mathrm{v}} \cdot N_{\mathrm{b}} = 8676$.

Second, a Polynomial Chaos Expansion is to be constructed of the integrand of Equation (5.57). This leads to $N_{\mathrm{b}} = 36$ one-dimensional PCEs. Again, the integral is approximated with the zero-th coefficients of the PCEs.

The Matlab optimizer *fmincon* is used to find the beam intensity vector which minimizes Equation (5.60). Some arguments can be provided to this optimizer. The starting point of the optimizer is set to be the zero vector. The beam intensity vector is restricted to be non-negative.

After the optimization, the obtained solution can be evaluated, for example as the nominal dose, the expected dose or the variance of the dose.

Expected dose

The expected dose equals the matrix-vector product of the expected value of the dose deposition matrix and the beam intensity vector. The expected value of the dose deposition matrix is approximated with the zero-th order coefficients of Polynomial Chaos Expansions of the elements of the dose deposition matrix.

Variance

The variance of the finally delivered dose in each voxel is proportional to the variance of the dose of each fraction, as derived in Equation (6.1). Use of the definition of the variance and some rearrangements, yield the sum over an integral and the expected dose squared.

$$\mathrm{Var}\left(d_i\right) = \mathrm{Var}\left(\sum_{t \in \mathcal{T}} d_i^{(t)}\right) = \sum_{t \in \mathcal{T}} \mathrm{Var}\left(d_i^{(t)}\right) =$$

$$\frac{1}{T^2} \sum_{t \in \mathcal{T}} \left(\sum_{j \in \mathcal{B}} \sum_{j' \in \mathcal{B}} x_j x_{j'} \int D_{ij}(\xi^{(t)}) D_{ij'}(\xi^{(t)}) p_{\xi^{(t)}}(\xi^{(t)}) \, \mathrm{d}\xi^{(t)} - \left(\sum_{j \in \mathcal{B}} \mathrm{E}[D_{ij}(\xi^{(t)})] x_j\right)^2\right) \quad (6.1)$$

Subsequently, a PCE is to be constructed of the integrand of Equation (6.1). Equation (6.2) shows that the integral is approximated with the zero-th order coefficients of this PCE, which is denoted by $R_{ijj'}^{(0)}$.

$$\mathrm{Var}\left(d_i\right) \approx \frac{1}{T} \left(\sum_{j \in \mathcal{B}} \sum_{j' \in \mathcal{B}} x_j x_{j'} R_{ijj'}^{(0)} - \left(\sum_{j \in \mathcal{B}} \mathrm{E}[D_{ij}] x_j\right)^2\right) \quad (6.2)$$

Note again the invariance under interchange of the indices j, j', which means that almost half of the coefficients does not need to be calculated.

6.2 Simplified 3D geometry

This section describes a model that mimics the case of a tumor at the abdominal side of the spinal cord in a body. The spinal cord is considered to be an organ-at-risk (OAR). The three-dimensional geometry is simplified. A cross sectional slice is shown in Figure 6.1. The tumor and the OAR are shown in red and blue respectively.

A discrete voxel-grid is constructed of the geometry. Each voxel is a cube with sides of 2 mm. The resulted number of voxels is $N_{\mathrm{v}} = 5022$.
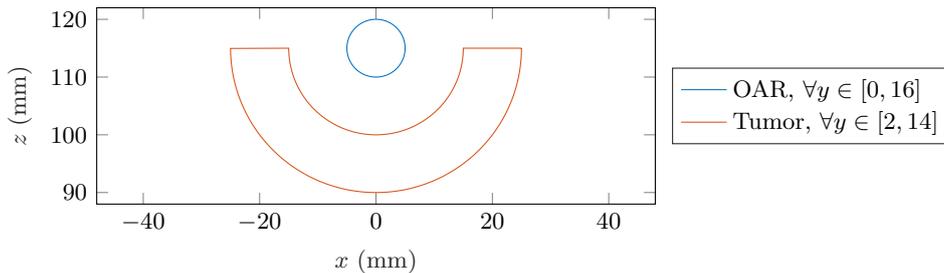
Figure 6.1: Cross-section of simplified geometry of a tumor with an OAR and surrounding tissue. This cross section looks the same for all $y \in [2, 14]$. Outside this interval, only the OAR is present.

The geometry is irradiated by a grid of equidistant beamspots with multiple energies. They enter the body at $z = 0$ and propagate in positive z-direction. This grid is defined in Equation (6.3).

$$x_j = -36 + 3j \quad \forall j \in \{1, 2, ..., 23\} \tag{6.3}$$

$$y_i = -1 + 3i \quad \forall i \in \{1, 2, ..., 5\} \tag{6.4}$$

$$z_k = 87 + 3k \quad \forall k \in \{1, 2, ..., 11\} \tag{6.5}$$

Here, (x_i, y_j) is the lateral position of the beamspot and z_k is the depth at which the beam has deposited all its energy. So the set of beamspots $\mathcal{B} = \{(x_i, y_j, z_k)\}$ consists of $N_b = 23 \cdot 5 \cdot 11 = 1265$ beamspots.

The energy deposition profile along the depth z of the proton beamspot is calculated according the so-called *pencil beam algorithm* [64, 65]. A uniform density is assumed throughout the geometry. The range straggling of the proton beamspot, as described by Bortfeld (1997), is assumed to be $\sigma = 3$ mm, while the accuracy parameter is assumed to be $\epsilon = 0.01$. The pencil beam algorithm calculates the integral depth dose for a certain beamspot for various depths.

The lateral spread-out of the proton beamspots is assumed to be Gaussian distributed. The standard deviation is modelled to be depth dependent: it increases for higher depth. The standard deviation of the lateral spread-out at entrance of the body is 3 mm. Since a discrete voxel-grid is used, dose conservation needs to be imposed in each voxel. This means that the mean dose in each voxel is assigned to the center of that voxel. The mean dose is approximated using the three-dimensional Gauss-Legendre quadrature rule with 8 points [63].

A systematic set-up error ξ in the horizontal direction (x) is considered. The deposited dose is allocated proportionally to the corresponding voxels, in presence of a variation in set-up. The error is assumed to be Gaussian distributed with zero mean and $\Sigma_x = 3$ mm, so $\xi \sim \mathcal{N}(0, 3^2)$.

For this problem, the expected value of the sum of the squares of the difference between the voxel dose and the prescribed dose is minimized. Section 5.2 describes how Polynomial Chaos Expansion can be used to solve this problem. Polynomial Chaos Expansions are to be constructed of the two terms in square brackets of Equation (5.3). The first term results $\frac{1}{2}N_b^2 + \frac{1}{2}N_b = 800745$ PCEs, taken into account the symmetry of the object. The second term results in $N_b = 1265$ different PCEs. Since the expected value is approximated, only the zero-th order coefficients suffice. Subsequently, the expected value described by Equation (5.7) is minimized with the Matlab optimizer fmincon.

The optimizer requires a gradient or an approximation for the gradient. If no gradient is provided by the user, the optimizer approximates the gradient itself with a finite difference

47

scheme. For this problem, the gradient, as expressed in Equation (5.8), is provided to the optimizer. This alleviates the computational cost and decreases the uncertainty.

After the optimization, the obtained solution can be evaluated, for example as the nominal dose, the expected dose or the variance of the dose.

Expected dose

The expected dose equals the matrix-vector product of the expected value of the dose deposition matrix and the beam intensity vector. The expected value of the dose deposition matrix is approximated with the zero-th order coefficients of Polynomial Chaos Expansions of the elements of the dose deposition matrix.

6.3   Patient in iCycle

For a real demonstration of the applicability of probabilistic treatment planning, a skull base meningioma patient is considered.

iCycle, the in-house treatment planning software of Erasmus Medical Center, is used. There was no existing compatibility of iCycle with the openGPC software. This link has been made ourselves. This section describes the methodology for obtaining a probabilistic treatment plan. For comparison, robust treatment plans have also been made. That methodology is not described here, but follows the same methodology as described by Ter Haar et al. (2018).

First of all, the CT data and the delineated organs are loaded into iCycle. The CT data is required for the computation of the stopping power units. The delineated organs define all the structures for which constraints or objectives can be set.

Second, a wish-list is to be defined in Lucy. This application lets the user set the constraints and objectives for the different structures. Different wish-lists have been used. They are discussed in Chapter 7.

The voxel reduction feature, as described in Chapter 3, is turned off for all structures, for which a probabilistic constraint or objective is included in the wish-list. This feature introduces more uncertainty.

The beamspot resampling feature, as described in Chapter 3, is turned off as well, for both the probabilistic and the robust treatment plans. Ideally, this feature would be turned on. However, this research focuses on handling probabilistic constraints or objectives with Polynomial Chaos Expansion. The construction of the PCE requires samples of the dose deposition matrix corresponding to the cubature points. If beamspot resampling would be turned on and the set of beamspots would differ among the iterations, computation of the samples of the dose deposition matrix by the dose engine would be computationally quite expensive. Instead, the same set of beamspots is used for the treatment planning.

Since beamspot resampling is turned off, the set of beamspots which is to be used for the probabilistic treatment plans, has to be selected manually. These beamspots are selected by running a full nominal wish-list, as shown in Table A.2 in Appendix A, with the beamspot resampling feature turned on. The beamspots which result from this optimization, are used as input for the probabilistic treatment plans. In order to include enough beamspots that cover the GTV in (almost) all possible scenarios in the probabilistic treatment plan, the GTV is expanded with a ring of 4 mm in the full nominal wish-list. The assumption is made here that a certain set-up error of the body leads to an equally big shift of the beamspot within the body. Due

to the heterogeneous density of the body, this is actually not true. However, assuming such an equal relation between the set-up error and the shift of the Bragg peak, guarantees the GTV expanded with a ring of 4 mm to cover the beamspots with a probability of 99.9%, when a three-dimensional set-up error is considered with standard deviation equal to 1 mm, which is reasonable according to literature [66].

6.3.1 Implementation of $\beta\%$-probabilistic $D_{\alpha\%}$ into iCycle

Multiple probabilistic functions have been implemented in iCycle.

Here, only the function for the $\beta\%$-probabilistic $D_{\alpha\%}$ is discussed. The output of the function is the $\beta\%$-probabilistic $D_{\alpha\%}$ and optionally its gradient and its Hessian. Its input is the beam intensity vector and some parameters which will be explained later in this section.

The pseudocode of Algorithm 6.1 summarizes the implemented function in iCycle which calculates the $\beta\%$-probabilistic $D_{\alpha\%}$ and if needed the gradient and Hessian. This section provides a walkthrough of this implemented function.

Algorithm 6.1 Calculate $\beta\%$-probabilistic $D_{\alpha\%}$

1: $X = [\boldsymbol{x}_0]$
2: **if** Gradient **then**
3:     Expand matrix X with beam intensity vectors
4: **if** Hessian **then**
5:     Expand matrix X with beam intensity vectors
6: **function** OPENGPC
7:     Generate cubature points $\{\boldsymbol{\xi}_k\}$
8:     **function** COMPUTERESPONSES
9:         **for** all $\boldsymbol{\xi} \in \{\boldsymbol{\xi}_k\}$ **do**
10:            Compute $(100 - \alpha)$-th percentile of $\sum_j D_{ij}(\boldsymbol{\xi})X_{jl}$ over index i
11:     Calculate PCE coefficients
12: Sample from all PCEs
13: Compute $(100 - \beta)$-th percentile
14: **if** Gradient **then**
15:     Calculate gradient
16: **if** Hessian **then**
17:     Calculate Hessian matrix

Since a PCE is constructed of $D_{\alpha\%}$ as a whole, the dependence on the beam intensity vector is not contained in the PCE anymore. As a consequence, for all beam intensity vectors, for which a function evaluation is required in the approximation gradient and/or Hessian by finite differencing, a separate PCE is to be constructed. A set is constructed of all beam intensity vectors for which a PCE of the $D_{\alpha\%}$ is to be made. The size of this set depends on whether the gradient and/or Hessian is required as output.

Subsequently, Polynomial Chaos Expansions of $D_{\alpha\%}$ are constructed for all beam intensity vectors contained in the just discussed set. For this purpose, a matrix is constructed, which is a horizontal concatenation of all beam intensity vectors. The number of columns equals the number of beam intensity vectors. If solely the gradient is required and a forward difference scheme is used to approximate it, the different beam intensity vectors are contained within matrix X:

$$X = \begin{bmatrix} \boldsymbol{x}_0 & \cdots & \boldsymbol{x}_0 \end{bmatrix} + \begin{bmatrix} \boldsymbol{0} & \mathrm{diag}(h,...,h)^T \end{bmatrix} . \tag{6.6}$$

Subsequently, a parallelized loop is executed over the different cubature points, as line 9 and 10 of Algorithm 6.1 show. In each iteration of this loop, the matrix product of the dose deposition matrix corresponding to the current cubature point and the matrix X is calculated and the $(100 - \alpha)$-th percentile is calculated of this matrix along the voxel dimension (i.e. index i). This resulting row vector is stored for each iteration of the loop in a matrix. This matrix contains the the responses to the different cubature points along the first dimension. The number of columns equals the number of beam intensity vectors for which a PCE is needed. The coefficients of the PCEs are calculated with these responses.

Now we have obtained PCEs of the $D_{\alpha\%}$ for the different beam intensity vectors. For each PCE a number of samples is drawn. The number of samples can be provided as a variable to the function. The $(100 - \beta)$-th percentile is calculated for the sample set. This yields the $\beta\%$-probabilistic $D_{\alpha\%}$ for all different beam intensity vectors.

Subsequently, the gradient and Hessian matrix are calculated using finite differencing in a computationally efficient way with parallelized loops. It turns out that the gradient of the $\beta\%$-probabilistic $D_{\alpha\%}$ always has to be non-negative, which is proven in Section 6.3.2. In case the finite differencing still yields a negative derivative, an option has been implemented, which enforces requirements on the gradient and Hessian. Four options regarding monotonicity enforcement can be distinguished:

1. No requirements are enforced,

2. Each negative element of the gradient is set to zero:

$$\frac{\partial f}{\partial x_n} = \frac{\max\{f(\boldsymbol{x}_0 + h\hat{\boldsymbol{e}}_n), f(\boldsymbol{x})\} - f(\boldsymbol{x})}{h} ,$$

3. Each negative element of the gradient is set to zero and these adjustments are also embraced in the Hessian:

$$\frac{\partial f}{\partial x_n} = \frac{\max\{f(\boldsymbol{x}_0 + h\hat{\boldsymbol{e}}_n), f(\boldsymbol{x})\} - f(\boldsymbol{x})}{h} ,$$

$$\frac{\partial^2 f}{\partial x_m \partial x_n} = \frac{1}{h^2} \left(f(\boldsymbol{x}_0 + h\hat{\boldsymbol{e}}_m + h\hat{\boldsymbol{e}}_n) - \max\{f(\boldsymbol{x}_0 + h\hat{\boldsymbol{e}}_m), f(\boldsymbol{x}_0)\} - \right.$$
$$\left. \max\{f(\boldsymbol{x}_0 + h\hat{\boldsymbol{e}}_n), f(\boldsymbol{x}_0)\} + f(\boldsymbol{x}_0) \right) ,$$

4. Positive monotonicity is enforced everywhere:

$$\frac{\partial f}{\partial x_n} = \frac{\max\{f(\boldsymbol{x}_0 + h\hat{\boldsymbol{e}}_n), f(\boldsymbol{x})\} - f(\boldsymbol{x})}{h} ,$$

$$\frac{\partial^2 f}{\partial x_m \partial x_n} = \frac{1}{h^2} \left(\max\{f(\boldsymbol{x}_0 + h\hat{\boldsymbol{e}}_m + h\hat{\boldsymbol{e}}_n), f(\boldsymbol{x}_0 + h\hat{\boldsymbol{e}}_m), f(\boldsymbol{x}_0 + h\hat{\boldsymbol{e}}_n), f(\boldsymbol{x}_0)\} - \right.$$
$$\left. \max\{f(\boldsymbol{x}_0 + h\hat{\boldsymbol{e}}_m), f(\boldsymbol{x}_0)\} - \max\{f(\boldsymbol{x}_0 + h\hat{\boldsymbol{e}}_n), f(\boldsymbol{x}_0)\} + f(\boldsymbol{x}_0) \right) .$$

Regarding the sampling, two options have been implemented: either deterministic sampling or non-deterministic sampling. With deterministic sampling, the same samples of the uncertainty vectors are used every iteration of the optimization.

Every time the implemented function, summarized in Algorithm 6.1, gets called, the same dose deposition matrices corresponding to the cubature points are needed for the computation of the PCEs. This set of dose deposition matrices can be provided to the function to avoid that the dose engine has to calculate these every time the function is called.

Linear constraints and objectives, like a minimization of the maximum dose as shown in Equation (6.7), are often included in wish-lists.

$$\underset{\boldsymbol{x}}{\text{minimize}}\ \underset{i}{\max}\ d_i(\boldsymbol{x}, \boldsymbol{0}) \tag{6.7}$$

To solve such problems, iCycle introduces an extra decision variable t. The new decision vector becomes $\boldsymbol{y} = (x_1, ... x_{N_\mathrm{b}}, t)^T$. A differently formulated but equivalent optimization problem follows, as shown in Equation (6.8).

$$\begin{aligned} \underset{\boldsymbol{y}}{\text{minimize}}\ \begin{bmatrix} 0 & \cdots & 0 & 1 \end{bmatrix} \boldsymbol{y} \\ \text{subject to}\ \begin{bmatrix} D(\boldsymbol{0}) & -\boldsymbol{1} \end{bmatrix} \boldsymbol{y} < 0 \end{aligned} \tag{6.8}$$

Here, $-\boldsymbol{1}$ is a vector with N_v elements, each equal to -1. A functionality has been implemented into Algorithm 6.1, which guarantees proper working when extra decision variables are added for linear constraints and objectives. This functionality is needed when the maximum dose in a structure and the $\beta\%$-probabilistic $D_{\alpha\%}$ in a structure are minimized simultaneously.

The function is coded such that the parameters α and β can be provided in the wish-list in Lucy.

6.3.2 Non-negative gradient $\beta\%$-probabilistic $D_{\alpha\%}$

If the intensity of a beamspot is increased, one would expect that the total deposited dose increases as well. Therefore, a negative gradient of a DVH-like parameter is intuitively not natural. This is stated in Theorem 6.1 and is followed by a mathematical proof.

Theorem 6.1. *Let $d(\boldsymbol{r}, \boldsymbol{x}, \boldsymbol{\xi}) = \sum_j D_j(\boldsymbol{r}, \boldsymbol{\xi}) x_j$ be the dose with $x_j \geq 0$ for all j and $D_j(\boldsymbol{r}, \boldsymbol{\xi}) \geq 0$ for all $j, \boldsymbol{r}, \boldsymbol{\xi}$. The parameter $D_{\alpha\%}(\boldsymbol{x}, \boldsymbol{\xi})$ is non-decreasing, in other words $\frac{\partial D_{\alpha\%}(\boldsymbol{x}, \boldsymbol{\xi})}{\partial x_j}$ is non-negative for all $j, \boldsymbol{x}, \boldsymbol{\xi}$.*

Proof. Let $\boldsymbol{x}_0$ be an arbitrary non-negative beam intensity vector and let k denote an arbitrary element of this vector with corresponding unit vector $\hat{\boldsymbol{e}}_k$. Let $V = \int_{d(\boldsymbol{r}, \boldsymbol{x}, \boldsymbol{\xi}) < \infty} \mathrm{d}^3\boldsymbol{r}$ be the volume of the considered structure. For all $\alpha \in [0, 100]$, the dose volume parameters are defined as:

$$D_{\alpha\%}(\boldsymbol{x}, \boldsymbol{\xi}) = \inf \left\{ D \in \mathbb{R} : \int_{d(\boldsymbol{r}, \boldsymbol{x}, \boldsymbol{\xi}) \leq D} \mathrm{d}^3\boldsymbol{r} \geq \left(1 - \frac{\alpha}{100}\right) \cdot V \right\}.$$

Since $\boldsymbol{x}_0 + h\hat{\boldsymbol{e}}_k \geq \boldsymbol{x}_0$ for any $h > 0$, $d(\boldsymbol{r}, \boldsymbol{x}_0 + h\hat{\boldsymbol{e}}_k, \boldsymbol{\xi}) \geq d(\boldsymbol{r}, \boldsymbol{x}_0, \boldsymbol{\xi})$ for all $\boldsymbol{r}, \boldsymbol{\xi}$. Therefore, the following holds for all $\boldsymbol{\xi}$:

$$\left\{ D \in \mathbb{R} : \int_{d(\boldsymbol{r}, \boldsymbol{x}_0 + h\hat{\boldsymbol{e}}_k, \boldsymbol{\xi}) \leq D} \mathrm{d}^3\boldsymbol{r} \geq \left(1 - \frac{\alpha}{100}\right) \cdot V \right\} \subseteq$$

$$\left\{ D \in \mathbb{R} : \int_{d(\boldsymbol{r}, \boldsymbol{x}_0, \boldsymbol{\xi}) \leq D} \mathrm{d}^3\boldsymbol{r} \geq \left(1 - \frac{\alpha}{100}\right) \cdot V \right\} .$$

Consequently, $D_{\alpha\%}(\boldsymbol{x}_0 + h\hat{\boldsymbol{e}}_k, \boldsymbol{\xi}) \geq D_{\alpha\%}(\boldsymbol{x}_0, \boldsymbol{\xi})$ for any $h > 0$ and for all $\boldsymbol{\xi}$. Therefore, $\frac{\partial D_\alpha(\boldsymbol{x}, \boldsymbol{\xi})}{\partial x_j} \geq 0$ for all $j, \boldsymbol{x}, \boldsymbol{\xi}$. $\qquad\square$

For Theorem 6.1 and its proof, a continuous dose profile $d(\boldsymbol{r}, \boldsymbol{x}, \boldsymbol{\xi})$ and a continuous dose deposition profile $D_j(\boldsymbol{r}, \boldsymbol{\xi})$ are assumed. However, in clinical practice such a continuous dose profile is not known. Theorem 6.1 also applies to a discretized voxel grid with corresponding dose distribution $d_i(\boldsymbol{x}, \boldsymbol{\xi})$ and dose deposition matrix $D_{ij}(\boldsymbol{\xi})$. Its proof is analogous.

A positive gradient of the dose volume parameters has been proven now. A similar approach can proof Theorem 6.2, a non-negative gradient of the probabilistic dose volume parameters.

Theorem 6.2. *Let $D_{\alpha\%}(\boldsymbol{x}, \boldsymbol{\xi})$ be the dose volume parameter as defined in the proof of Theorem 6.1. The $\beta\%$-probabilistic $D_{\alpha\%}(\boldsymbol{x}, \boldsymbol{\xi})$ is denoted by $f_\beta(\boldsymbol{x})$. $f_\beta(\boldsymbol{x})$ is non-decreasing, in other words $\frac{\partial f_\beta(\boldsymbol{x})}{\partial x_j}$ is non-negative for all $j, \boldsymbol{x}$.*

Proof. Let $\boldsymbol{x}_0$ be an arbitrary non-negative beam intensity vector and let k denote an arbitrary element of this vector with corresponding unit vector $\hat{\boldsymbol{e}}_k$. For all $\beta \in [0, 100]$, the $\beta\%$-probabilistic $D_{\alpha\%}(\boldsymbol{x}, \boldsymbol{\xi})$ is defined as:

$$f_\beta(\boldsymbol{x}) = \inf \left\{ D \in \mathbb{R} : \int_{D_{\alpha\%}(\boldsymbol{x}, \boldsymbol{\xi}) \leq D} p_{\boldsymbol{\xi}}(\boldsymbol{\xi}) \, \mathrm{d}\boldsymbol{\xi} \geq \left(1 - \frac{\beta}{100}\right) \right\} .$$

Theorem 6.1 states that $D_{\alpha\%}(\boldsymbol{x}_0 + h\hat{\boldsymbol{e}}_k, \boldsymbol{\xi}) \geq D_{\alpha\%}(\boldsymbol{x}_0, \boldsymbol{\xi})$ for any $h > 0$ and for all $\boldsymbol{\xi}$. Therefore, the following holds:

$$\left\{ D \in \mathbb{R} : \int_{D_{\alpha\%}(\boldsymbol{x}_0 + h\hat{\boldsymbol{e}}_k, \boldsymbol{\xi}) \leq D} p_{\boldsymbol{\xi}}(\boldsymbol{\xi}) \, \mathrm{d}\boldsymbol{\xi} \geq \left(1 - \frac{\beta}{100}\right) \right\} \subseteq$$

$$\left\{ D \in \mathbb{R} : \int_{D_{\alpha\%}(\boldsymbol{x}_0, \boldsymbol{\xi}) \leq D} p_{\boldsymbol{\xi}}(\boldsymbol{\xi}) \, \mathrm{d}\boldsymbol{\xi} \geq \left(1 - \frac{\beta}{100}\right) \right\} .$$

Consequently, $f_\beta(\boldsymbol{x}_0 + h\hat{\boldsymbol{e}}_k) \geq f_\beta(\boldsymbol{x})$ for any $h > 0$. Therefore, $\frac{\partial f_\beta(\boldsymbol{x})}{\partial x_j} \geq 0$ for all $j, \boldsymbol{x}$. $\quad\square$

For computational reasons, the $\beta\%$-probabilistic $D_{\alpha\%}(\boldsymbol{x}, \boldsymbol{\xi})$ is not calculated as in Theorem 6.2 and its proof. Instead, a finite set of samples are drawn from a Polynomial Chaos Expansion of $D_{\alpha\%}(\boldsymbol{x}, \boldsymbol{\xi})$. In that case, Theorem 6.2 is still valid. Its proof is analogous.

Chapter 7

Results

In this chapter the results are presented of the three applications discussed in Chapter 6.

7.1 Multiple fractions in 1D

The example, addressed by Unkelbach et al. (2018), has been reworked, as described in Section 6.1. Similar results have been found.

The prescribed dose in the tumor is $d^{\mathrm{p}} = 1$. Elsewhere the prescribed dose equals zero. The weights w_i assigned to the voxels inside and outside the tumor are respectively 1 and 0.01.

For all PCEs involved in this problem, the grid level is $L = 8$. This is rather conservative, especially since solely the zero-th coefficients are required. However, as this problem is small, it does not influence the computation time considerably.

Figure 7.1 shows the results for different number of fractions. The tumor is presented as the shaded area in the graphs. In blue, the nominal dose is shown, which corresponds to a random error equal to zero for every fraction. The expected dose distribution is presented in red. The standard deviation for the dose in every voxel is shown in green.
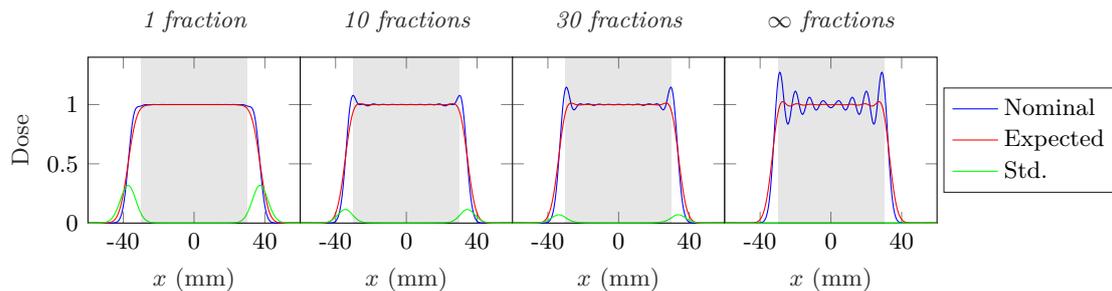
Figure 7.1: Probabilistic treatment plans for different number of fractions. The expected value of a quadratic function in dose has been optimized. The prescribed dose in the tumor, the shaded area ($-30 \leq x \leq 30$), is 1. In blue, the nominal dose is shown, which corresponds to a random error equal to zero for every fraction. The expected dose distribution is presented in red. The standard deviation for the dose in every voxel is shown in green. The standard deviation of the horizontal Gaussian distributed error is $\sigma_x = 3$ mm. This example has been reworked from the problem described by Unkelbach et al. (2018) and the results are the same.

Figure 7.1 shows that the standard deviation of the dose in each voxel decreases for higher number of fractions. This can be explained as follows. The finally delivered dose can be interpreted as an estimate of the expected dose, the so-called *sample mean*. The standard deviation of the sample mean is inversely proportional with the square root of the number of samples, which corresponds to the number of fractions in this problem. As expected, the standard deviation of the dose decreases to zero for an infinite number of fractions. This is also explainable physically. If the number of fractions goes to infinity, the random set-up errors of the fractions can be neglected.

In Figure 7.1 it can be observed that the nominal dose expands around the tumor for a single fraction treatment. This resembles a conventional treatment plan, for which a PTV is used to be defined. However, there is a difference, since for probabilistic treatment planning a PTV does not need to be defined manually, as demonstrated in this section. Instead the automatic extension of the tumor results from the optimization of the CTV itself including the set-up uncertainty. A single fraction treatment is equivalent to a treatment subject to a systematic error. If only one fraction is considered, a random set-up error causes the nominal dose distribution to shift equally. As a consequence, the standard deviation of the dose is higher in a neighbourhood where the nominal dose shows a steeper slope. The size of this neighbourhood depends on the standard deviation of the random error. After all, bigger shifts of the dose distribution are likely, if the standard deviation of the random error is higher. Figure 7.1 demonstrates this observation for a single fraction treatment: the standard deviation is maximal where the magnitude of the gradient is highest.

For a higher finite number of fractions of a probabilistic treatment plan, the fall-off of the nominal dose at the edge of the tumor is steeper: the regions outside the tumor receive less dose. As a consequence, the edge could be underdosed for fractions with a large set-up error. However, the average dose of all fractions is optimized. So, to compensate the possible underdosage at the edge for some fractions, *horns* arise near the edge. These horns become more visible for higher number of fractions. This phenomenon has also been described and simulated in earlier research [1, 67, 68].

As the number of fractions increases, the horns at the edge are emphasized more. These horns trigger a *negative horn* towards the interior. Subsequently, such a reduction in dose triggers a new horn towards the interior of the tumor. This oscillatory behaviour is more prominent as the number of fractions increase. This phenomenon has also been described by Lof, Lind, and Brahme (1995).

The computation time is recorded as well and is presented in Table 7.1.

Table 7.1: Computation time for one-dimensional treatment planning for the optimization of the expected value of a quadratic function in dose. A finite and infinite number of fractions are distinguished. Both the time needed for construction of the PCE and for optimization have been recorded.

Finite T		Infinite T	
PCE	Optimization	PCE	Optimization
4.5 s	2.2 s	0.6 s	2.2 s

7.2 Simplified 3D geometry

This section presents the results of the three-dimensional simplified geometry representing a tumor at the abdominal side of the spinal cord, as described in Section 6.2. The expected value of the sum of the quadratic differences between the dose and the prescribed dose is minimized, as described in Section 5.2.

For this problem, the prescribed dose in the tumor is set to $d^{\mathrm{p}} = 60$ Gy. Elsewhere, the prescribed dose is set to 0 Gy. The weight w_i assigned to the voxels of the tumor and the organ-at-risk is set to 1, while the weight assigned to the voxels outside the tumor and the OAR is set to $5 \cdot 10^{-3}$.

For all PCEs involved in this problem, the grid level is set to $L = 5$. This is reasonable, especially since only the zero-th coefficients of the PCEs are needed.

Figures 7.2a and 7.2b shows the nominal dose distribution and the expected dose distribution for a cross-sectional view ($y = 8$ mm) of the geometry. The tumor and the organ-at-risk are delineated with a white line. The shape of the tumor and OAR in Figure 7.2 is different from the one in Figure 6.1, due to the finite resolution of the voxel-grid. A conformal target coverage is achieved and the organ-at-risk is spared.

Figure 7.2c shows a conventional treatment plan. For this plan the sum of the quadratic differences of the dose and the prescribed dose in the voxels is minimized, without any uncertainty taken into account ($\xi = 0$). When one compares this conventional treatment plan with the nominal dose distribution of the probabilistic treatment plan (Figure 7.2a), the automatic expansion of the tumor in the probabilistic treatment plan is notable, as also described in Section 7.1. This means that no planning target volume had to be defined and instead, the tumor has been expanded automatically, as a consequence of probabilistic optimization.

Besides, the nominal dose distribution of the probabilistic treatment plan shows that the automatic expansion of the tumor reaches less far at the interior edge of the tumor than at the exterior edge of the tumor. This is a consequence of the higher weight assigned to the voxels of the OAR. The steeper fall-off of the dose at the interior edge of the tumor, could cause underdosage of the tumor when a large set-up error is present. To compensate possible underdosage here, a slightly higher dose is received just inside the tumor near the interior edge. Figure 7.3 shows the same dose plots as Figure 7.2, however, the dose scale is smaller such that small differences in dose in the tumor are more clearly visible.

The expected dose distribution is more homogeneous than the nominal dose distribution, as expected. After all, averaging occurs for the expected value, which smoothens out the heterogeneities in the dose distribution.

The weights w_i assigned to the voxels of the different structures (i.e. tumor, OAR or outside both) can be adjusted to obtain the desired treatment plan. As an example, if the weight of the voxels outside the tumor and the OAR is decreased, the automatic expansion of the tumor reaches further, since a high dose outside the tumor is less penalized. If the weight of the organ-at-risk voxels is increased, the automatic expansion of the tumor at the interior side of the tumor reaches less far, since a high dose in the OAR is penalized more. Since the objective function is linear with respect to the weights of the voxels, only the ratio between the weights has importance.

Figure 7.4 shows the dose volume histogram of the expected dose distribution in the tumor of both the probabilistic and the conventional treatment plan. It can be observed that the expected dose of the probabilistic treatment plan is more homogeneous and is closer to the prescribed dose of 60 Gy than the expected dose the conventional treatment plan. This is explainable, since for the conventional treatment plan no set-up errors have been taken into account while for the

(a) Nominal dose of probabilistic treatment plan

(b) Expected dose of probabilistic treatment plan

(c) Nominal dose of cobventional treatment plan
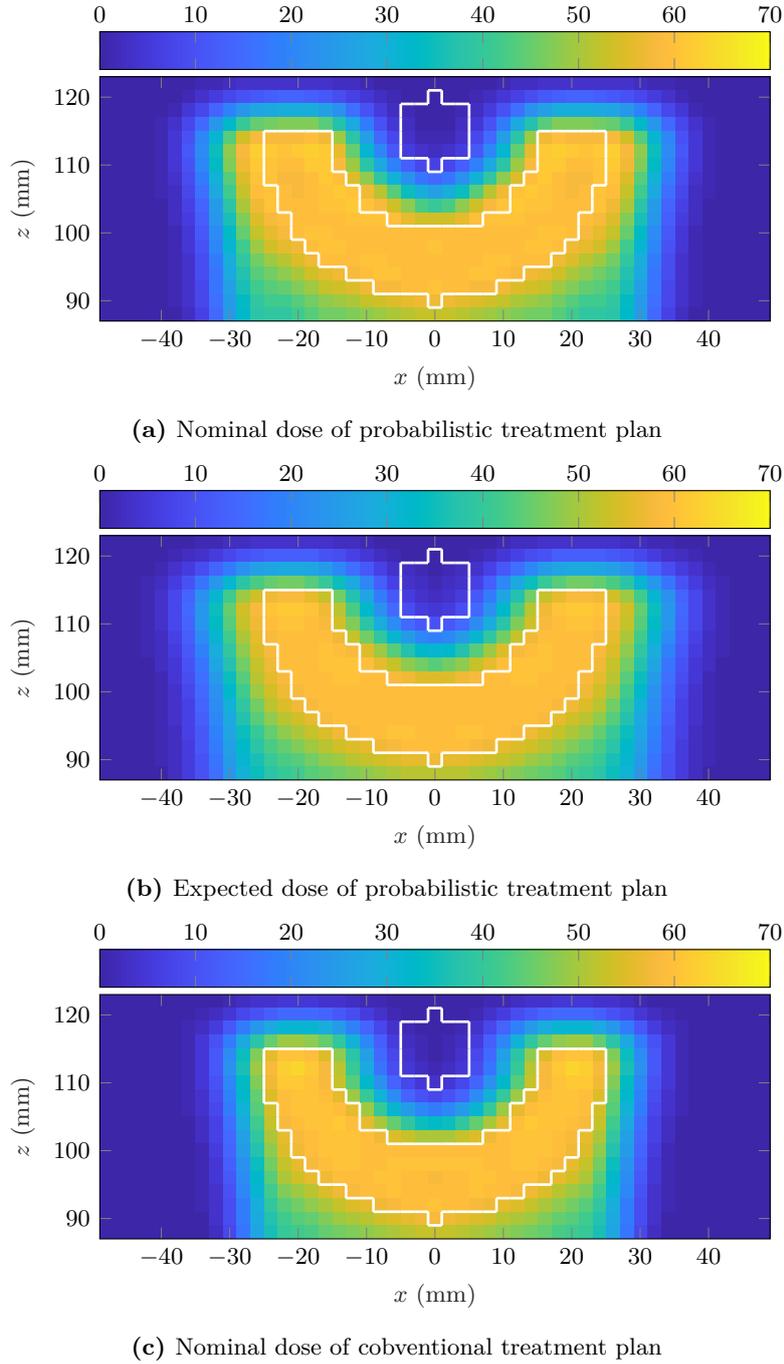
Figure 7.2: Cross-sectional views of the plane described by $y = 8$ mm. The prescribed dose in the tumor is 60 Gy. The weights w_i assigned to the voxels of the tumor and OAR are set to 1, while the weights of the voxels outside both are set to $5 \cdot 10^{-3}$. **(a,b)** Nominal and expected dose of a probabilistic treatment plan for which the sum of the quadratic differences of the voxel dose and the prescribed dose has been minimized. **(c)** Nominal dose of a conventional treatment plan for which the sum of the quadratic differences of the voxel dose and the prescribed dose has been minimized.

(a) Nominal dose of probabilistic treatment plan

(b) Expected dose of probabilistic treatment plan

(c) Nominal dose of cobventional treatment plan
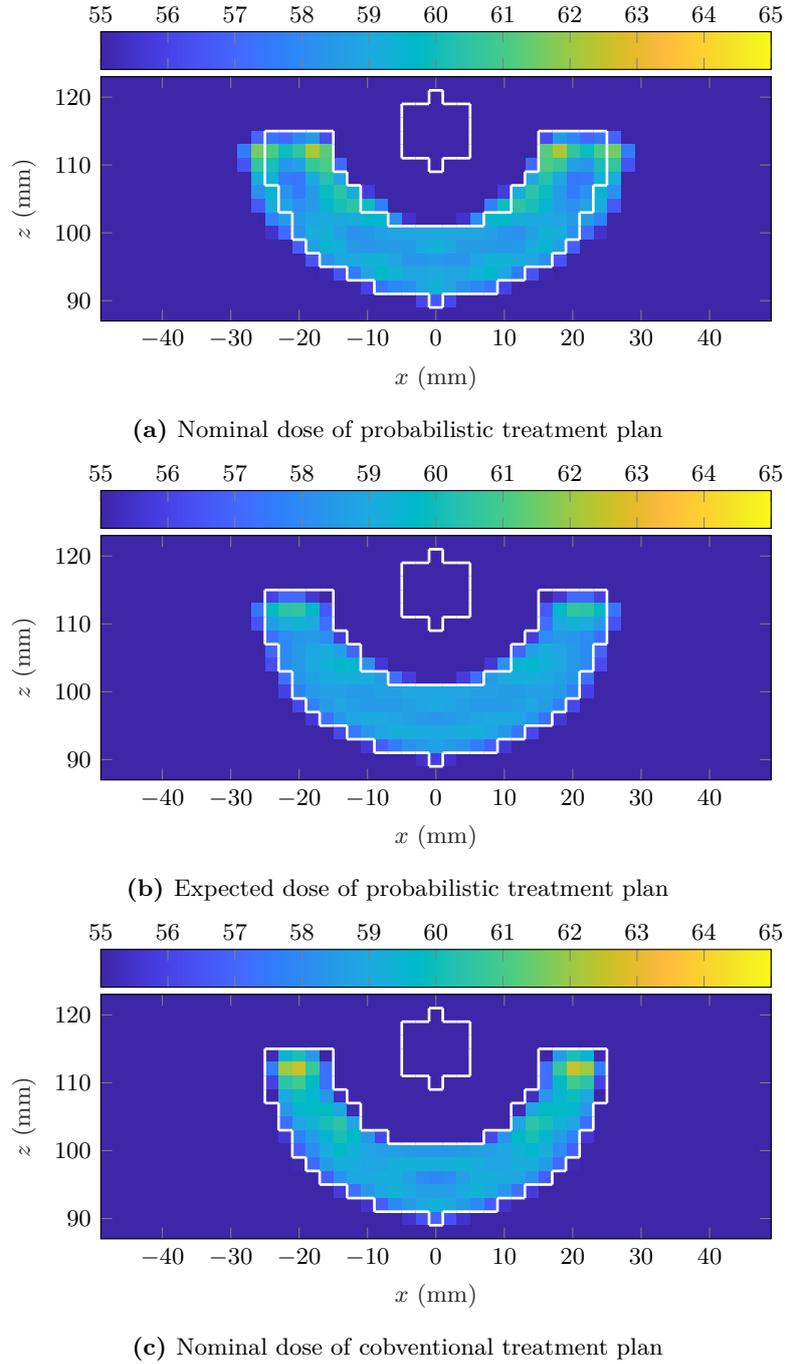
Figure 7.3: Same as Figure 7.2, but with other dose scale. Cross-sectional views of the plane described by $y = 8$ mm. The prescribed dose in the tumor is 60 Gy. The weights w_i assigned to the voxels of the tumor and OAR are set to 1, while the weights of the voxels outside both are set to $5 \cdot 10^{-3}$. **(a,b)** Nominal and expected dose of a probabilistic treatment plan for which the sum of the quadratic differences of the voxel dose and the prescribed dose has been minimized. **(c)** Nominal dose of a conventional treatment plan for which the sum of the quadratic differences of the voxel dose and the prescribed dose has been minimized.

probabilistic treatment plan the difference between the expected dose and the prescribed dose has been minimized [1].
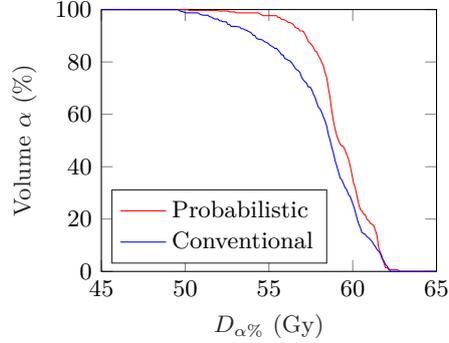
Figure 7.4: Dose volume histogram of the expected dose distribution in the tumor of the probabilistic and the conventional treatment plan.

The expected value of the sum of the quadratic differences of the dose and the prescribed dose in each voxel is technically a suitable function for probabilistic treatment planning. First of all, the function is convex with respect to the beam intensity vector $\boldsymbol{x}$. A convex function is easier to minimize, since the obtained minimum is always the global minimum, if the feasible region is a convex set [69]. Besides, since the function is smooth and an analytical expression is known, the function is differentiable as well. This speeds up the optimization.

However, the expected value of the sum of the quadratic differences of the dose and the prescribed dose in each voxel is a function with little clinical relevance. After all, the function value itself does not have a one-to-one relation with the spatial dose distribution and therefore it does not have an easy to interpret clinical meaning.

The computation time is recorded as well and is presented in Table 7.2.

Table 7.2: Computation time for three-dimensional treatment planning when optimizing the expected value of a quadratic function in dose. Both the time needed for construction of the PCE and for optimization have been recorded. These computations have been run on the cluster of Erasmus MC, with 6 processors and 50 GB of memory reserved. The computation time of the conventional treatment plan was 351.7 s.

PCE	Optimization
21.8 s	219.9 s

7.3   Patient in iCycle

The results of a simplified one-dimensional and three-dimensional problem have been described in previous sections. This section elaborates on the results following from data of a real skull base meningioma patient.

7.3.1 95%-probabilistic $D_{98\%}$ using PCE

In Chapter 5, two methods have been described how PCE can be used for treatment planning with the probabilistic near-minimum dose, either via a PCE of the dose deposition matrix or via a PCE of the near-minimum dose. This section presents the results.

Polynomial Chaos Expansion can be used for treatment planning with the probabilistic near-minimum dose, as described in Section 5.7. For this purpose, a Polynomial Chaos Expansion could be constructed of the near-minimum dose. The quality of the approximation of the PCE has to be validated for different grid levels and polynomial orders. For this purpose, three different Polynomial Chaos Expansions are constructed of the same dose volume parameter $D_{98\%}$. The beam intensity vector which is used resulted from the conventional treatment plan, of which the wish-list is described in Table A.2 in Appendix A.

First, a PCE of the dose deposition matrix has been constructed with $L = 4$, $\Delta_{lev} = 1$ and $O = 5$. For this PCE, 111 cubature points are involved, which means that 111 realizations of the dose deposition matrix are to be calculated by the dose engine. Subsequently, a Polynomial Chaos Expansion is constructed of $D_{98\%}$ with $L = 5$, $\Delta_{lev} = 1$ and $O = 6$. For this PCE, 303 cubature points are involved. The realizations of the dose deposition matrix corresponding to these 303 cubature points are calculated with the PCE of the dose deposition matrix instead of the dose engine. Figure 7.5a shows the dose volume parameter $D_{98\%}$ as a function of the uncertainty vector, either calculated using the dose engine or using a PCE.

Second, a PCE of the $D_{98\%}$ itself has been constructed directly with $L = 5$, $\Delta_{lev} = 1$ and $O = 6$. For this PCE, 303 cubature points are involved, which requires 303 realizations of the dose deposition matrix from the dose engine. Figure 7.5b shows the dependence on the uncertainty vector.

Third, another PCE of the $D_{98\%}$ itself has been constructed with $L = 6$, $\Delta_{lev} = 1$ and $O = 7$. For this PCE, 741 cubature points are involved, which requires an equal number of realizations of the dose deposition matrix from the dose engine. Figure 7.5c shows the dependence on the uncertainty vector.

The second PCE ($L5E1O6$), shown in Figure 7.5b, is preferred. If the first and second PCE (Figures 7.5a and 7.5b) are compared, we observe that the error between the PCE approximation and the exact $D_{98\%}$ is small for both PCEs. The major difference is the required number of realizations of the dose deposition matrix from the dose engine. These samples need to be computed only once for every patient. So the impact on the computation time of the optimization of a full wish-list is considerably small. Besides, there is not much improvement of the accuracy of the third PCE with respect to the second PCE.

In general, the $\beta\%$-probabilistic $D_{\alpha\%}$ cannot be calculated analytically, since the probability density function of the $D_{\alpha\%}$ is unknown. Therefore, a finite number of samples is drawn from the Polynomial Chaos Expansion of $D_{\alpha\%}$. The $(100 - \beta)$-th percentile of this sample set is an estimate for the $\beta\%$-probabilistic $D_{\alpha\%}$. This estimate itself is a stochastic variable with a non-zero variance. This variance is expected to depend on the sample size. For this reason, the influence of the sample size on the the spread of the estimate for the $\beta\%$-probabilistic $D_{\alpha\%}$ is investigated.

Figure 7.6 shows ten realizations of the empirical cumulative distribution function of the $D_{98\%}$ for different number of samples N_s. The value of an estimate for the 95%-probabilistic $D_{98\%}$ is the intersection of a realization and the horizontal line characterized by a cumulative probability equal to 0.05. The purple dot in Figure 7.6 is a realization of the 95%-probabilistic $D_{98\%}$ with $N_s = 10^7$. This value can be interpreted as a benchmark. The spread is defined as the

(a) $L5E1O6$ using PCE of D ($L4E1O5$)

(b) $L5E1O6$

(c) $L6E1O7$
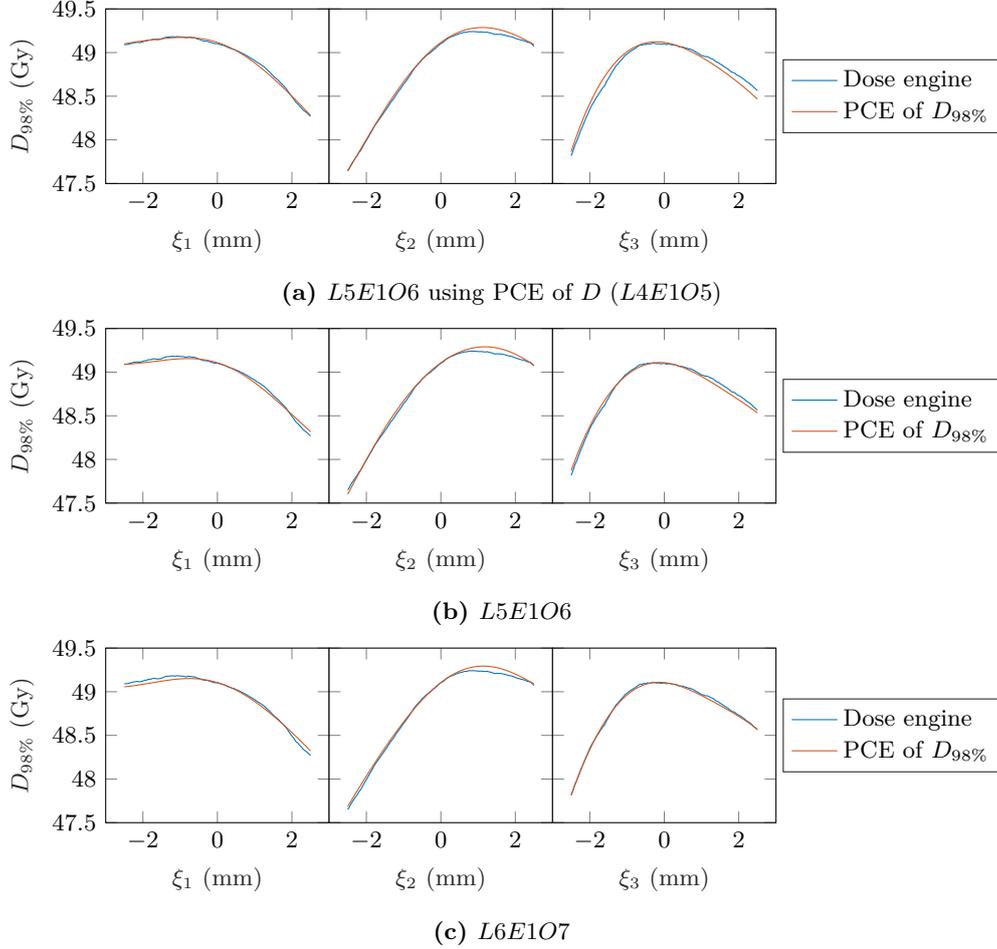
Figure 7.5: The dose volume parameter $D_{98\%}$ of the GTV as a function of ξ_1, ξ_2 and ξ_3, which correspond to the systematic set-up error in the x-, y- and z-direction respectively. The beam intensity vector resulted from a conventional optimization of the wish-list from Table A.2. **(a)** $L = 5$, $\Delta_{lev} = 1$ and $O = 6$ using samples of a lower order PCE of the dose deposition matrix ($L = 4$, $\Delta_{lev} = 1$ and $O = 5$). **(b)** $L = 5$, $\Delta_{lev} = 1$ and $O = 6$. **(c)** $L = 6$, $\Delta_{lev} = 1$ and $O = 7$. Based on these results, $L5E1O6$ was chosen and used in this research.

maximum difference between the realizations of the 95%-probabilistic $D_{98\%}$. In accordance with expectation, the spread of the 95%-probabilistic $D_{98\%}$ appears to decrease for a higher number of samples. For $N_{\mathrm{s}} = 5 \cdot 10^5$ samples, the spread is below 0.01 Gy. The choice of the number of samples is a trade-off between computation time and uncertainty. $N_{\mathrm{s}} = 5 \cdot 10^5$ is preferred and is used throughout this thesis.
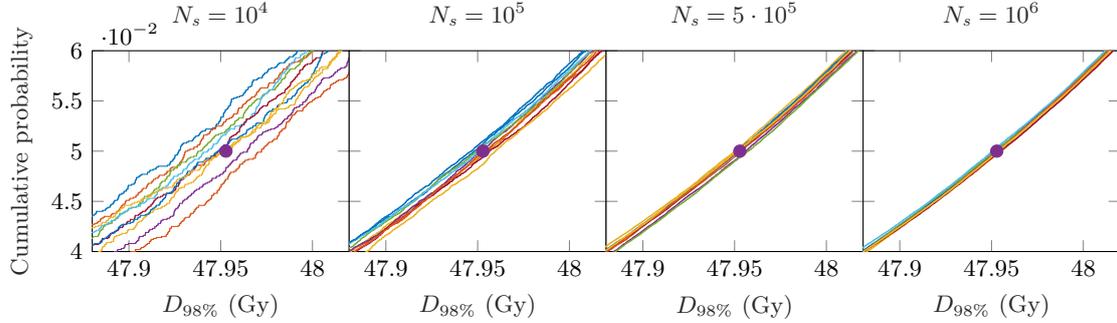
Figure 7.6: Ten realizations of the empirical cumulative distribution function of the $D_{98\%}$ for different number of samples N_{s}. These samples are obtained by sampling from a Polynomial Chaos Expansion ($L5E1O6$) of the $D_{98\%}$ parameter. The beam intensity vector resulted from a conventional treatment plan of which the wish-list is described in Table A.2. The purple dot is a realization of the fifth percentile for $N_{\mathrm{s}} = 10^7$ samples. Based on these results, a sample size of $N_{\mathrm{s}} = 5 \cdot 10^5$ was chosen and used in this research.

Since our approximation of the 95%-probabilistic $D_{98\%}$ itself is a stochastic variable, its derivative with respect to the beam intensity vector is stochastic as well. Figure 7.7 shows ten realizations of the first 30 elements of the gradient of the 95%-probabilistic $D_{98\%}$, calculated with a forward finite difference scheme for different step sizes h. The values of the elements of the gradient are connected for clarification. It appears that the spread decreases for higher step sizes.

Figure 7.8 shows the 95%-probabilistic $D_{98\%}$ as a function of the 19-th element of the beam intensity vector for different scales. The horizontal axis represents the difference with respect to the initial intensity of the 19-th beamspot. Figure 7.8c shows that the 95%-probabilistic $D_{98\%}$ appears to be almost linear with the 19-th beamspot on a larger scale. However, on a smaller scale a bumpy behaviour is observed.

For a part of the domain shown in Figure 7.8c, the 95%-probabilistic $D_{98\%}$ has a negative derivative. However, Theorem 6.2 in Chapter 6 states that the gradient of a probabilistic dose volume parameter always has to be non-negative. Such graphs have also been analyzed for a higher number of samples ($N = 10^7$). It turns out that this bumpy behaviour with negative slopes of the 95%-probabilistic $D_{98\%}$ does not reduce considerably. The reason why negative slopes are observed in contrary to Theorem 6.1 might be that a PCE is used to approximate the $D_{98\%}$. Small errors could cause the probabilistic $D_{98\%}$ to decrease.

The part of the curve of Figure 7.8a with a negative slope corresponds to the dominantly present negative value of the 19-th element of the gradient for a step size of $h = 0.1$, in Figure 7.7. A positive gradient is observed for higher step sizes, as one would expect from the graphs of Figure 7.8.

Whether the dependence of the 95%-probabilistic $D_{98\%}$ on the beam intensity is close to linear for all beamspots, needs to be checked. For this purpose, a linear fit of the functional
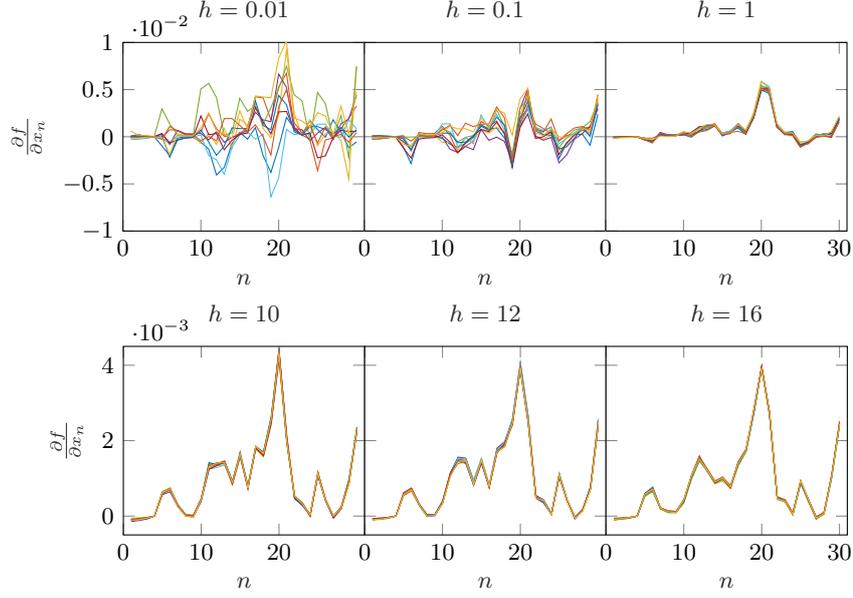
Figure 7.7: Ten realizations of the first 30 elements of the gradient of the 95%-probabilistic $D_{98\%}$ for different step sizes. The elements of the gradient are connected for clarification.

dependence is constructed for all beamspots. Besides, a cubic fit of the functional dependence is constructed. The so-called R^2-value is a measure for the successfulness of the fit in explaining the variation in the data [70]. This parameter can take up values between 0 and 1, where the latter corresponds to a good fit. It turns out that the R^2-value of the linear fit is above 0.75 for 211 of the 235 beamspots, which corresponds to almost 90%. Hence, for 24 out of 235 beamspots the R^2-value is below 0.75. Figure 7.9 is a histogram of the linear coefficients of the fits for all beamspots. It turns out that the linear coefficients of the beamspots with a R^2-value below 0.75, are below $2 \cdot 10^{-4}$, which correspond to the first two bins of Figure 7.9. A slope of $2 \cdot 10^{-4}$ is assumed to be negligible small.
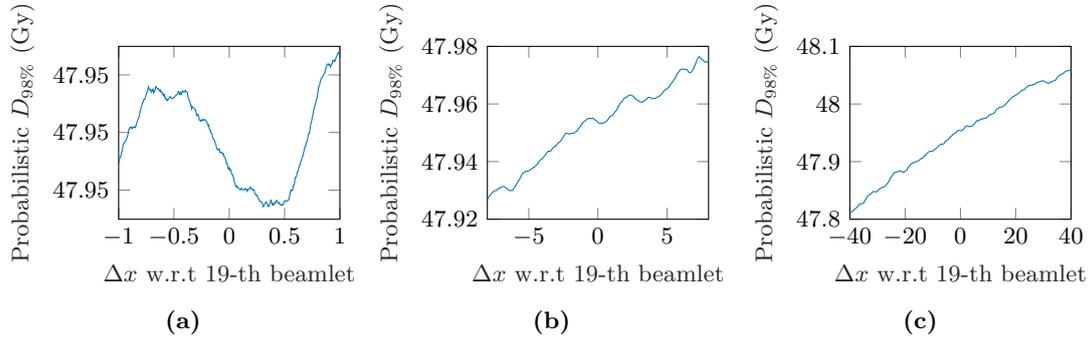
Figure 7.8: Dependence of the 95%-probabilistic $D_{98\%}$ on the intensity of the 19-th beamspot. These graphs are obtained by sampling from a Polynomial Chaos Expansion ($L5E1O6$) of the $D_{98\%}$ parameter with $N_s = 5 \cdot 10^5$.

Besides, the linear coefficient of all beamspots for which the R^2-value of the cubic fit increases with more than 30% with respect to the R^2-value of the linear fit, is below $2 \cdot 10^{-4}$. So, all beamspots for which there is room for improvement of a higher order fit, show a negligible small slope.
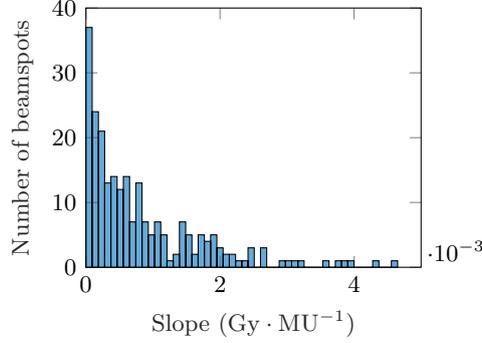
Figure 7.9: Histogram of the linear coefficients of the linear fit of the dependence of the 95%-probabilistic $D_{98\%}$ on the intensity of the different beamspots. The total number of counts is the number of elements of the beam intensity vector, which equals $N_b = 235$.

An example of a linear fit with a coefficient above 0.75 is shown in Figure 7.10a. The variations are well described by the fit. On the other hand, Figure 7.10b shows a linear fit with a coefficient below 0.75. It is obvious that the fit doesn't describe the variations of the probabilistic $D_{98\%}$ well. However, the variations are negligible small and are solely caused by noise.

| (a) | (b) |
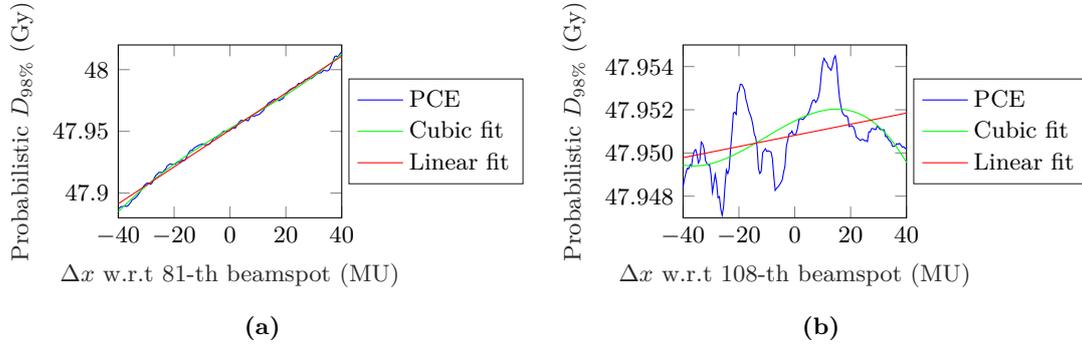
Figure 7.10: Linear fits of the functional dependence of the 95%-probabilistic $D_{98\%}$ for two elements of the beam intensity vector. **(a)** R^2-value above 0.75. **(b)** R^2-value below 0.75.

The bumpy behaviour on a small scale of the 95%-probabilistic $D_{98\%}$ justifies a big step size for approximating the gradient and Hessian with a finite difference scheme. Since the probabilistic $D_{98\%}$ is almost linear with respect to the beam intensity, it is justifiable to use a forward difference scheme instead of a central difference scheme. Besides, the forward difference scheme requires less function evaluations, as discussed in Chapter 5.

Computation time

The computation time of one iteration, as described by Algorithm 6.1, has been recorded and is summarized in Table 7.3. It turns out that the computation time increases substantially

if the Hessian is required. If the Hessian matrix is desired, the construction of the PCEs is computationally most expensive, followed by the calculation of the 5-th percentile of the samples.

Algorithm 6.1 describes a parallelized loop over the cubature points. The responses corresponding to the cubature points can also be computed with a parallelized loop over the beam intensity vectors for which a PCE of the $D_{98\%}$ is to be constructed. This increases the computation time slightly. For higher grid levels, a parallelized loop over the cubature points alleviates the computational effort considerably.

It has also been investigated whether it is computationally beneficial to compute the 95%-probabilistic $D_{98\%}$ and its gradient and Hessian using a PCE of the dose deposition matrix, as described in Section 5.7.1. This takes more than 50 hours for solely $N_s = 10^4$ samples. Due to the size of the dose deposition matrix of the GTV ($N_v = 11781$ and $N_b = 235$), storing many samples simultaneously in the memory is not possible. Therefore, the samples need to be drawn in batches.

7.3.2 Simple wish-lists

This section describes probabilistic treatment plans resulting from a wish-list with either one or two probabilistic objectives.

One probabilistic objective

First, a rather basic wish-list (Table 7.4) has been run in iCycle. As first objective, the so-called *logarithmic tumor control probability* function, as defined in Equation 7.1, is minimized for GTV without any uncertainties taken into account (i.e. $\boldsymbol{\xi} = \mathbf{0}$).

$$\text{LTCP}(\boldsymbol{x}, \boldsymbol{\xi}) = \frac{1}{N_v} \sum_{i \in \mathcal{V}} \exp\left(-\gamma \left(d_i(\boldsymbol{x}) - d^{\text{p}}\right)\right) \tag{7.1}$$

This LTCP function is convex and the minimum is easily found [71]. The sensitivity parameter γ is set to 0.5 and the prescribed dose d^{p} is set to 50.4 Gy. This objective is imposed to drive the optimization efficiently in the proper direction. To prevent the optimization problem to be unbounded, a constraint is set on the maximum nominal dose. Subsequently, the 95%-probabilistic $D_{98\%}$ is maximized to be above $0.95 \cdot 50.4$ Gy. This wish-list, as described in Table 7.4, has been run with multiple combinations of settings.

First of all, iCycle offers a so-called *ForceNonConvex* functionality which could improve convergence of non-convex objective functions. This functionality has a positive effect on the convergence.

Table 7.3: Computation of one iteration of the 95%-probabilistic $D_{98\%}$ by sampling from a PCE (*L5E1O6*) of the $D_{98\%}$, with $N_s = 5 \cdot 10^5$. A distinction is made whether the gradient and Hessian are required. These computations have been run on the cluster of Erasmus MC, with 16 processors and 120 GB of memory reserved.

Function value	Gradient	Hessian	Time (s)
•			6
•	•		21
•	•	•	943

Table 7.4: Simple wish-list for a probabilistic treatment plan with one probabilistic objective on the GTV.

Constraints		
Volume	Type	Limit
GTV	Minimize $\max_i d_i(\boldsymbol{x}, \mathbf{0})$	$1.07 \cdot 50.4$ Gy

Objectives			
Priority	Volume	Type	Goal
1	GTV	Minimize $\mathrm{LTCP}(\boldsymbol{x}, \mathbf{0})$	1
2	GTV	Maximize $P_{5\%}^s\{D_{98\%}(\boldsymbol{x}, \boldsymbol{\xi}^{(s)}\}$	$0.95 \cdot 50.4$ Gy

Besides, the four options regarding the monotonicity enforcement of the probabilistic dose volume parameter as introduced in Section 6.3.1, have been tested. For higher step sizes ($h = 10$ and $h = 16$), the optimization converged for all four options. The computation time of the optimization for which no monotonicity was enforced is fastest. With a smaller step size $h = 1$, the optimization did not converge for all four options. It seems that for a small step size, even when monotonicity is enforced, the bumpy behaviour of the probabilistic 95%-probabilistic $D_{\alpha\%}$ drives the optimization in the wrong direction.

It has also been tested whether the optimization succeeds when the Hessian matrix is manually set to a zero matrix or an identity matrix. Neither leads to convergence. iCycle also offers the so-called *UseHessApprox* functionality, which bypasses the need for the computation of the Hessian with PCEs. Instead, iCycle uses the gradient to approximate the Hessian. As one could expect from Table 7.3, this improves the computation time considerably. For higher step sizes ($h = 10$ and $h = 16$), the optimization converges. However, when this functionality is combined with monotonicity enforcement of the gradient, the optimization does not converge.

Figure 7.6 shows that for $N_s = 5 \cdot 10^5$ there is still some spread in the realizations of the 95%-probabilistic $D_{98\%}$. To get rid of this noise, every iteration of the 95%-probabilistic $D_{98\%}$, described in Algorithm 6.1, the same samples of the uncertainty vector $\boldsymbol{\xi}$ has been used to calculate the samples of the $D_{\alpha\%}$ with the PCE. This is referred to as *deterministic sampling*.

Table 7.5 summarizes the settings which have been used to compute the probabilistic dose volume parameter for the results of probabilistic treatment planning described in Sections 7.3.2 and 7.3.3.

Table 7.5: Settings of the probabilistic dose volume parameter applicable to all results described in Sections 7.3.2 and 7.3.3.

Name	Setting
PCE	*L5E1O6*
ForceNonConvex	enabled
UseHessApprox	disabled
Monotonicity enforcement	disabled
Deterministic sampling	enabled
Number of samples N_s	$5 \cdot 10^5$
Step size h	16

Figure 7.11 shows the nominal scenario of the treatment plan which resulted from the optimization with the wish-list defined in Table 7.4, with $N_\mathrm{s} = 5 \cdot 10^5$, $h = 16$, ForceNonConvex functionality enabled, UseHessApprox disabled and without monotonicity enforcement. A conformal target coverage is observed. Since the fifth percentile of the near-minimum dose is maximized, an automatic expansion of the GTV is visible, which ensures that the GTV receives the prescribed dose in 95% of the possible scenarios.
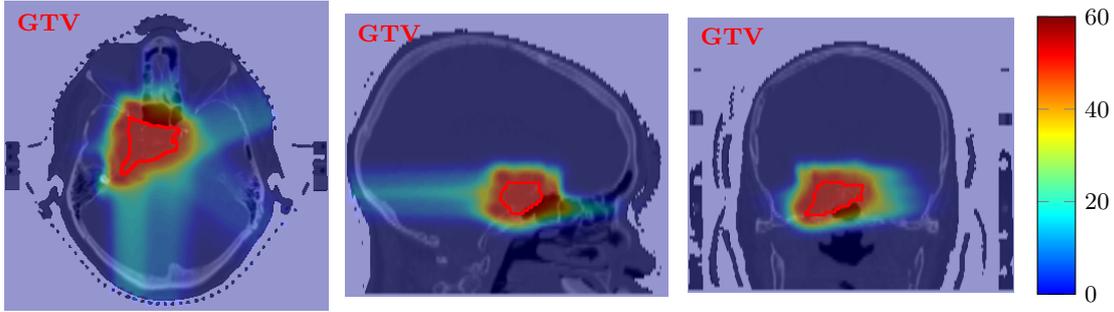
Figure 7.11: Probabilistic treatment plan obtained with the wish-list described in Table 7.4. An axial, sagittal and coronal cross-section through the center of mass of the GTV in the nominal scenario are shown. The following settings are used for the probabilistic dose volume parameter: $N_\mathrm{s} = 5 \cdot 10^5$, $h = 16$, ForceNonConvex enabled, UseHessApprox disabled and no monotonicity enforcement.

Figure 7.12 shows the probabilistic dose volume histogram of the probabilistic treatment plan of which the wish-list is described in Table 7.4. The bandwidth corresponds to a 90% confidence interval, as the 5-th and 95-th percentile of the dose volume parameter are calculated for every α separately. Thus, the probability that a realization of the $D_{\alpha\%}$ for a specific volume α falls within the bandwidth is 90%. No statement can be made about the probability that a realization of the dose volume histogram itself falls within the entire bandwidth for all α. It is observed that for lower α the bandwidth increases, since only the nominal dose is constrained. After all, from the wish-list we know that a high dose, i.e. above $1.07 \cdot 50.4$ Gy, in the GTV is not penalized in any other scenario than the nominal scenario.
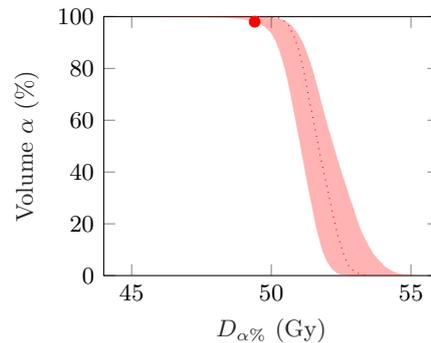
Figure 7.12: Probabilistic DVH corresponding to the probabilistic treatment plan of which the wish-list is described in Table 7.4. The left and right edge of the bandwith correspond to respectively the 5-th and 95-th percentile of samples from a PCE of the dose volume parameter. The dotted curve shows the nominal scenario and the dot is the value which has been optimized.

The computation time for the treatment plan corresponding to Table 7.4 was 30 hours. The number of iterations needed to maximize the probabilistic objective was 84. The optimization has been performed on the cluster of Erasmus MC, with 6 processors and 50 GB memory reserved.

Two probabilistic objectives

Table 7.6 shows almost the same wish-list as Table 7.4 but with two probabilistic objectives. Besides a probabilistic lower bound on the dose in the GTV, also a probabilistic upper bound is imposed. A constraint on the nominal dose is kept to prevent the optimization of the LTCP function to be unbounded.

Table 7.6: Simple wish-list for a probabilistic treatment plan with two probabilistic objectives on the GTV.

Constraints			
Volume	Type		Limit
GTV	Minimize $\max_i d_i(\boldsymbol{x}, \boldsymbol{0})$		$1.07 \cdot 50.4$ Gy
Objectives			
Priority	Volume	Type	Goal
1	GTV	Minimize $\mathrm{LTCP}(\boldsymbol{x}, \boldsymbol{0})$	1
2	GTV	Maximize $P_{5\%}^s\{D_{98\%}(\boldsymbol{x}, \boldsymbol{\xi}^{(s)})\}$	$0.95 \cdot 50.4$ Gy
3	GTV	Minimize $P_{95\%}^s\{D_{2\%}(\boldsymbol{x}, \boldsymbol{\xi}^{(s)})\}$	$1.07 \cdot 50.4$ Gy

Figure 7.13 shows an axial, sagittal and coronal cross-section through the center of mass of the GTV in the nominal scenario. These dose distributions do not show any remarkable visual differences with respect to those corresponding to a treatment plan with one probabilistic objective (Figure 7.11).
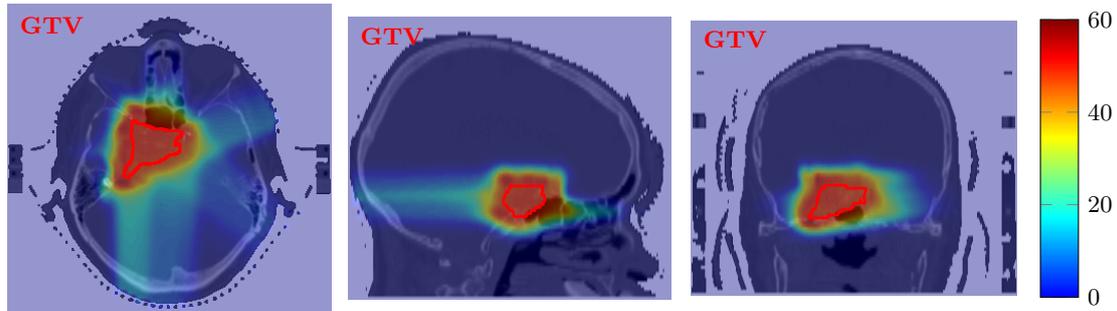
Figure 7.13: Probabilistic treatment plan obtained with the wish-list described in Table 7.6. An axial, sagittal and coronal cross-section through the center of mass of the GTV in the nominal scenario are shown. The following settings are used for the probabilistic dose volume parameter: $N_s = 5 \cdot 10^5$, $h = 16$, ForceNonConvex enabled, UseHessApprox disabled and no monotonicity enforcement.

However, the probabilistic dose volume histogram is considerably different. Figure 7.14 shows the probabilistic DVH of the treatment plan with two probabilistic objectives imposed. It is observed that the bandwidth for lower α has decreased, as expected. It can be concluded that by imposing both a probabilistic lower bound and a probabilistic upper bound on the dose in the

GTV, a more homogeneous dose distribution can be achieved in probabilistic sense, for which the size of the bandwidth is a measure.
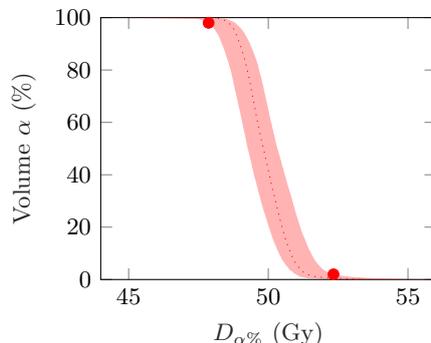
Figure 7.14: Probabilistic DVH corresponding to the probabilistic treatment plan of which the wish-list is described in Table 7.6. The left and right edge of the bandwith correspond to respectively the 5-th and 95-th percentile of the dose volume parameter. The dotted curve shows the nominal scenario and the dots are the values which have been optimized.

The computation time for the treatment plan corresponding to Table 7.6 was 136 hours, which is more than 5 days. The number of iterations needed to optimize the probabilistic $D_{98\%}$ and $D_{2\%}$ was respectively 91 and 64. The optimization has been performed on the cluster of Erasmus MC, with 8 processors and 60 GB memory reserved.

7.3.3 Extended wish-list

Realistic wish-lists also take into account organs-at-risk. Therefore, an extended wish-list has been run with constraints and objectives not only on the GTV but also on the optic nerves and the brainstem. For this wish-list, the 98%-probabilistic $D_{98\%}$ has been maximized, as shown in Table 7.7. Two cases are distinguished, each with a different goal for the maximum nominal dose of the brainstem.

A robust treatment plan functions as a benchmark. This treatment plan resulted from the wish-list described in Table A.1 in Appendix A. The robustness recipe from Ter Haar et al. (2018) has been used which imposes the 98%-probabilistic $D_{98\%}$ to be above 0.95·50.4 Gy by maximizing the nominal dose robustly with $SR = 3$ mm. The resulting treatment plan is scaled such that the 98%-probabilistic $D_{98\%}$ equals the value reached in the probabilistic treatment plan. Again, two cases are distinguished, each with a different goal for the maximum nominal dose of the brainstem.

Case 1

Figure 7.15 shows an axial, sagittal and coronal cross-section through the center of mass of the GTV in the nominal scenario of the probabilistic and robust treatment plan, which resulted from the wish-lists (case 1) described in Table 7.7 and A.1 respectively.

A rather conformal target coverage is visible in the probabilistic treatment plan. Again, the automatic extension of the GTV is observed. Besides, the considered organ-at-risks appear to be spared in the cross-sections of Figure 7.15a. However, single cross-sections do not contain all information about the dose distribution in the entire organ.

Table 7.7: Extended wish-lists for probabilistic treatment plans with one probabilistic objective on the GTV. Nominal constraints and objectives are imposed on some OARs. Two cases are distinguished, each with a different goal for the maximum nominal dose of the brainstem. The resulting treatment plans are compared with robust treatment plans from the wish-lists described in Table A.1.

		Constraints	
		Limit	
Volume	Type	Case 1	Case 2
GTV	Minimize $\max_i d_i(\boldsymbol{x}, \boldsymbol{0})$	$1.07 \cdot 50.4$ Gy	$1.07 \cdot 50.4$ Gy
Optic Nerve (L)	Minimize $\max_i d_i(\boldsymbol{x}, \boldsymbol{0})$	50 Gy	50 Gy
Optic Nerve (R)	Minimize $\max_i d_i(\boldsymbol{x}, \boldsymbol{0})$	50 Gy	50 Gy

			Objectives	
			Goal	
Priority	Volume	Type	Case 1	Case 2
1	GTV	Minimize LTCP$(\boldsymbol{x}, \boldsymbol{0})$	1	1
2	GTV	Maximize $P^s_{2\%}\{D_{98\%}(\boldsymbol{x}, \boldsymbol{\xi}^{(s)})\}$	$0.95 \cdot 50.4$ Gy	$0.95 \cdot 50.4$ Gy
2	Brainstem	Minimize $\max_i d_i(\boldsymbol{x}, \boldsymbol{0})$	50 Gy	46 Gy

Figure 7.16 shows the probabilistic dose volume histograms for the organs considered in the wish-lists of the probabilistic and robust treatment plan. The horizontal bandwidth corresponds to a confidence interval of 96% of the $D_{\alpha\%}$. As imposed, the left edges of the bandwidth of the $D_{98\%}$ of the probabilistic and robust plan intersect. A slightly steeper fall-of of the DVH is observed for GTV of the probabilistic plan. For both plans, the DVH of the left optic nerve show a similar course. The right optic nerve receives less of the lower dose spectrum in the probabilistic treatment plan. On the other hand, the brainstem receives in the probabilistic treatment more dose. Figure 7.17 shows that indeed more dose in the spectrum around 20 Gy is received by the brainstem in the probabilistic treatment plan, which is also observed in the DVH of Figure 7.16. An explanation might be that a probabilistic optimization equally weights all directions in which a set-up error could occur. This is in contrast to a robust optimization which only considers axial set-up errors. Since in this geometry axial displacements of the GTV have considerably less effect on the brainstem than non-axial displacements, the brainstem receives less dose in the robust treatment plan.

In general, it is expected that a specific level of the 98%-probabilistic $D_{98\%}$ is more easily achieved with probabilistic treatment planning than with robust treatment planning, since the whole spectrum of scenarios is taken into account, which can be interpreted as more degrees of freedom for the optimization. On the other hand, if the OARs are not planned probabilistically or robustly, the robust plan might spare an OAR more than a probabilistic plan does if that OAR is situated along a non-axial direction of the GTV.

Besides, the probabilistic treatment plan of Figure 7.15a shows a more conformal target coverage than the robust treatment plan of Figure 7.15b. The automatic extension of the GTV follows the same shape as the GTV itself in the probabilistic plan, while in the robust treatment plan a larger volume receives a high dose. This can be explained as follows. For a probabilistically optimized objective, all possible scenarios are taken into account, while for the robust treatment plan only a finite number of scenarios are taken into account. As a consequence, the probabilistic

(a) Probabilistic

(b) Robust
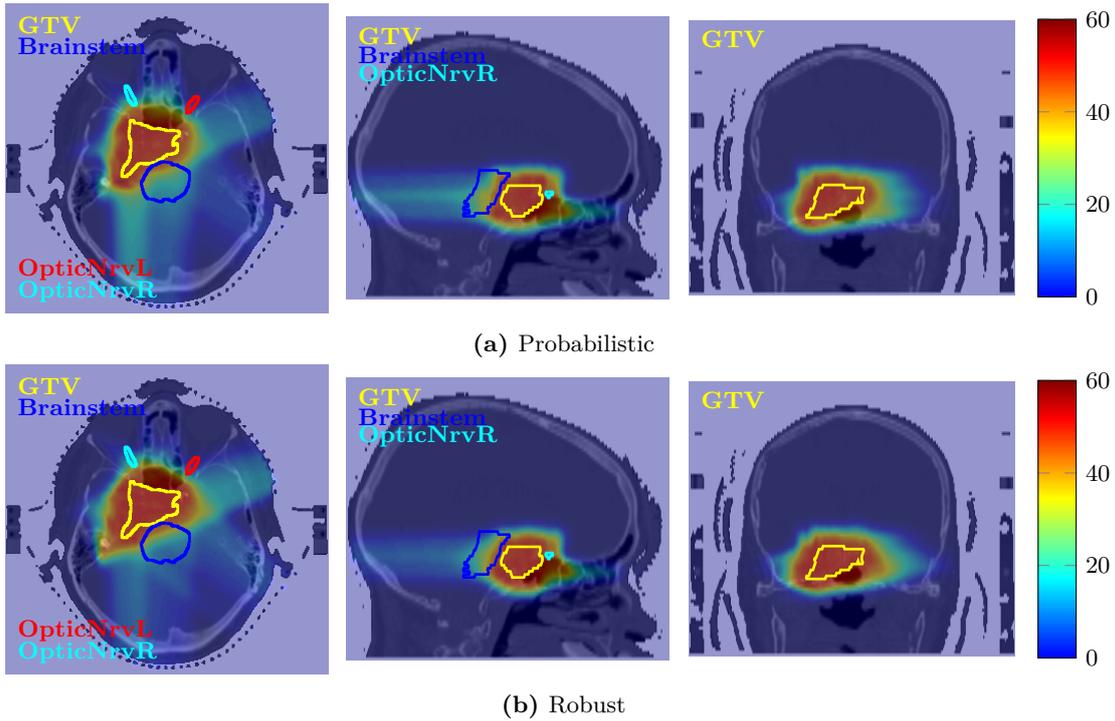
Figure 7.15: An axial, sagittal and coronal cross-section through the center of mass of the GTV in the nominal scenario are shown. **(a)** Probabilistic treatment plan from the wish-list (case 1) described in Table 7.7, with the following settings for the probabilistic dose volume parameter: $N_\mathrm{s} = 5 \cdot 10^5$, $h = 16$, ForceNonConvex enabled, UseHessApprox disabled and no monotonicity enforcement. **(b)** Robust treatment plan from the wish-list (case 1) described in Table A.1, with the set-up robustness set to $SR = 3$ mm.

automatic extension of the tumor does not demonstrate a preferred direction, contrary to a robust automatic extension. Moreover, the weighing of the scenarios is different. The probabilistic dose volume histograms of rings around the GTV (Figure 7.18) show the more conformal target coverage of the probabilistic treatment plan. Example 7.1 on page 76 shows intuitively why a larger volume receives a high dose in a robust treatment plan.

From the probabilistic DVH of the right optic nerve in Figure 7.16, it is not immediately clear whether the total dose received is higher for the probabilistic or the robust treatment plan. After all, in a probabilistic sense, the probabilistic treatment plan receives more dose in the high dose spectrum (i.e. above 30 Gy), while the robust treatment plan receives more dose in the low dose spectrum (i.e. below 30 Gy). The total dose received in a structure cannot directly be read from a dose volume histogram. The mean dose is one-to-one related to this quantity and is therefore a suitable measure for the total dose. Figure 7.19 shows the dose population histogram, which has been introduced in Section 3.1.8. For the GTV the $D_{98\%}$ is shown, while for the other organs the mean dose is shown. It can be observed that the expected mean dose in the brainstem is lower for the robust treatment plan, while the expected mean dose in the right optic nerve is lower for the probabilistic treatment plan. The structure called *elsewhere* is the volume of the brain in which dose is deposited outside the four considered organs. Although it is
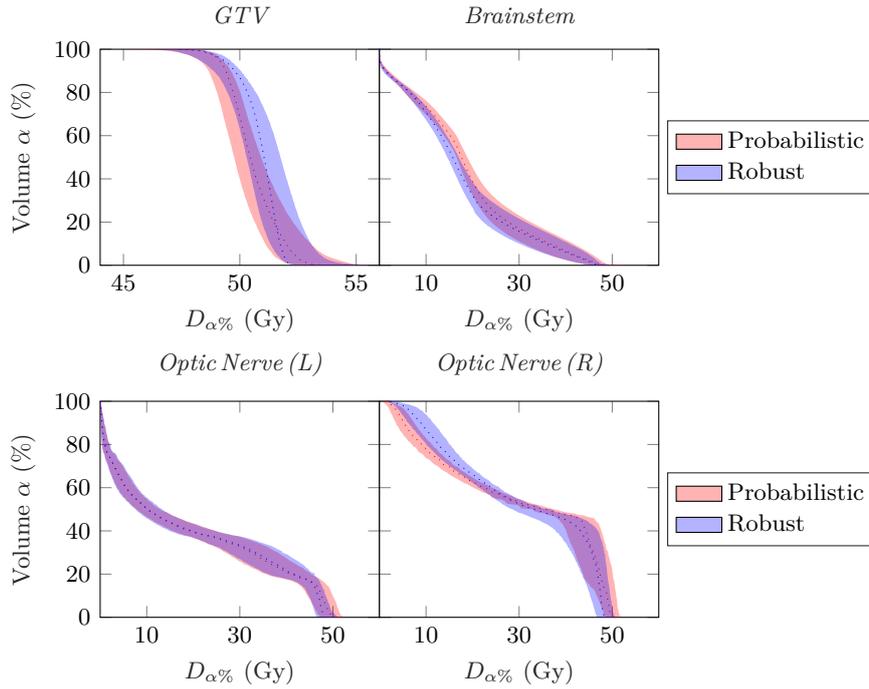
Figure 7.16: Probabilistic DVH corresponding to the probabilistic and robust treatment plan of which the wish-lists (case 1) are described in respectively Table 7.7 and Table A.1. The left and right edge of the bandwith correspond to respectively the 2-nd and 98-th percentile of the dose volume parameter. The dotted curve shows the nominal scenario.

a slight difference, the mean dose in this structure is lower for the probabilistic treatment plan. Besides, it applies to a much bigger structure.

The computation time for the probabilistic treatment plan corresponding to Table 7.7 (case 1) was 115 hours, which is more than 4 days. The number of iterations needed to optimize the probabilistic objective was 222. The optimization has been performed on the cluster of Erasmus MC, with 8 processors and 60 GB memory reserved.

71

(a) Probabilistic   **(b)** Robust
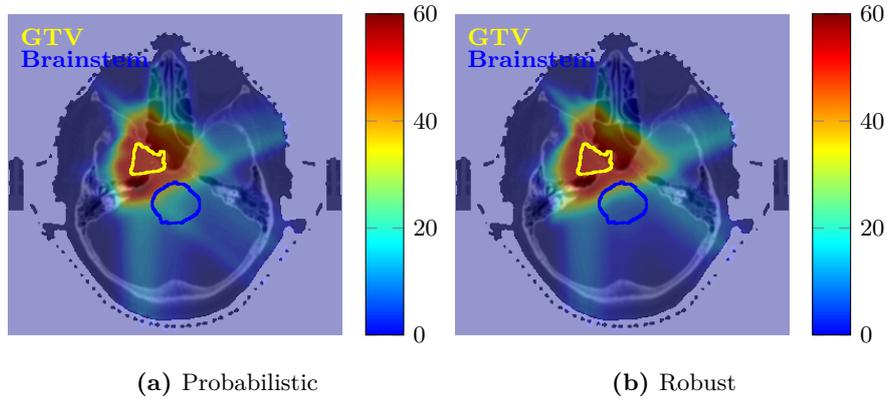
Figure 7.17: Axial cross-section of the probabilistic and robust treatment plan corresponding to the wish-lists (case 1) from respectively Table 7.7 and Table A.1.
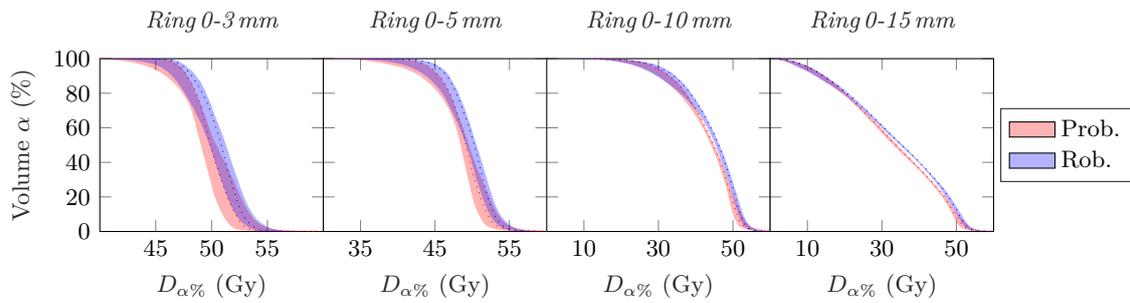
Figure 7.18: Probabilistic DVH corresponding to the probabilistic and robust treatment plan of which the wish-lists (case 1) are described in respectively Table 7.7 and Table A.1. Different rings around the GTV are considered. The left and right edge of the bandwith correspond to respectively the 2-nd and 98-th percentile of the dose volume parameter. The dotted curve shows the nominal scenario.
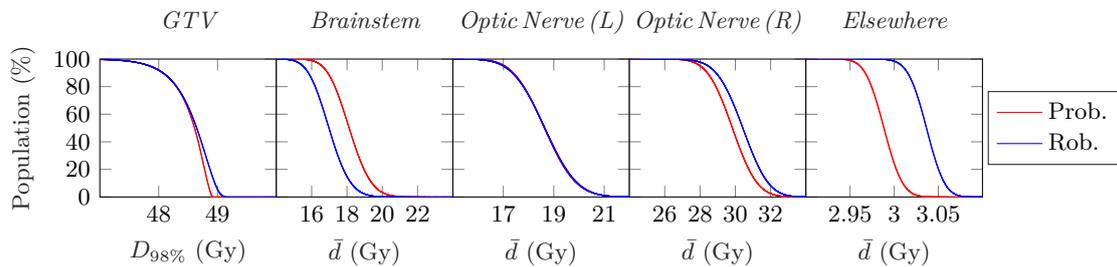
Figure 7.19: Dose population histogram of different organs and outside these organs corresponding to the probabilistic (red line) and robust (blue line) treatment plan of which the wish-lists (case 1) are described in respectively Table 7.7 and Table A.1. For the GTV the $D_{98\%}$ is considered and for the other organs the mean dose $\bar{d}$. Note that the scale of the horizontal axis is different from that of Figure 7.16.

72

Case 2

Figure 7.20 shows an axial, sagittal and coronal cross-section through the center of mass of the GTV in the nominal scenario of the probabilistic and robust treatment plan, which resulted from the wish-lists (case 2) described in Table 7.7 and Table A.1 respectively.

(a) Probabilistic

(b) Robust
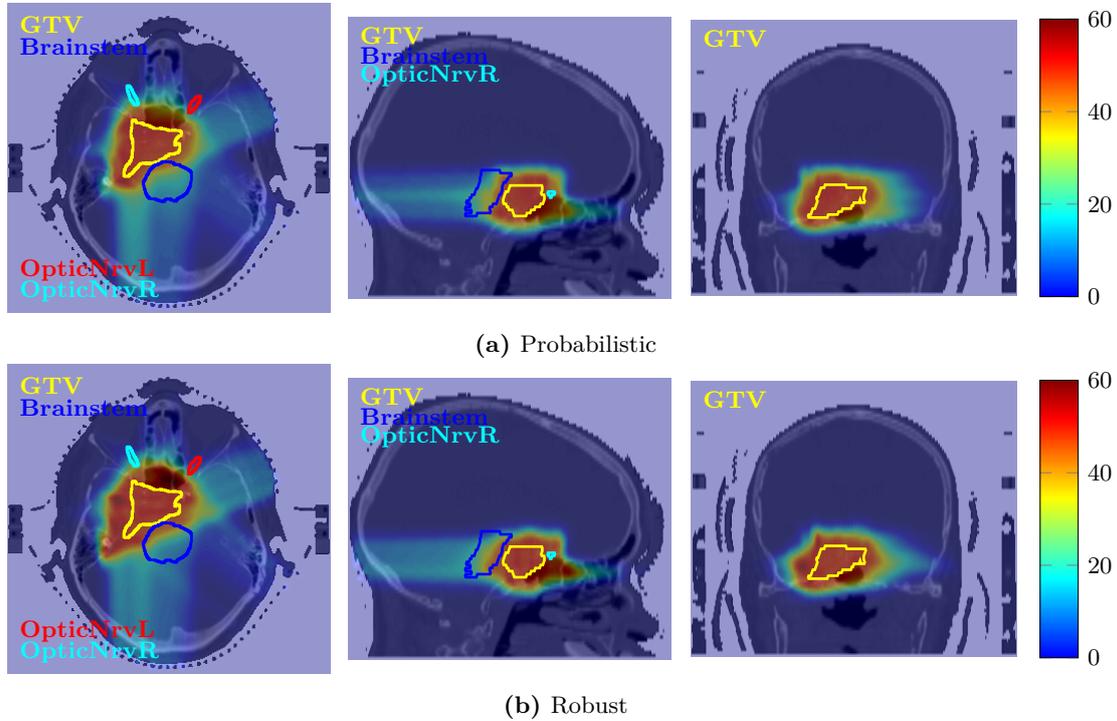
Figure 7.20: An axial, sagittal and coronal cross-section through the center of mass of the GTV in the nominal scenario are shown. **(a)** Probabilistic treatment plan from the wish-list (case 2) described in Table 7.7, with the following settings for the probabilistic dose volume parameter: $N_s = 5 \cdot 10^5$, $h = 16$, ForceNonConvex enabled, UseHessApprox disabled and no monotonicity enforcement. **(b)** Robust treatment plan from the wish-list (case 2) described in Table A.1, with the set-up robustness set to $SR = 3$ mm.

Again, the probabilistic treatment plan shows a more conformal target coverage than the robust treatment plan, which results in a smaller automatic extension of the GTV. This can also be observed in the probabilistic dose volume histograms of rings around the GTV, shown in Figure 7.21.

Figure 7.22 shows the probabilistic dose volume histograms of the GTV and the OARs included in the wish-lists. For the probabilistic and robust treatment plan, the left edge of the bandwidth of the $D_{98\%}$ of the GTV coincides. The brainstem receives less dose in the probabilistic treatment plan, contrary to case 1 (see Figure 7.16). The left optic nerve receives less dose as well in the probabilistic treatment plan, while for case 1 (Figure 7.16) the dose volume histograms of the right optic nerve for the probabilistic and robust treatment plans look the same.

Figure 7.19 shows dose population histograms. For the GTV the $D_{98\%}$ is shown, while for the other organs the mean dose is shown. The 98%-probabilistic $D_{98\%}$ of the GTV is equally high for the probabilistic and robust treatment plan, while the expected value of the $D_{98\%}$ is
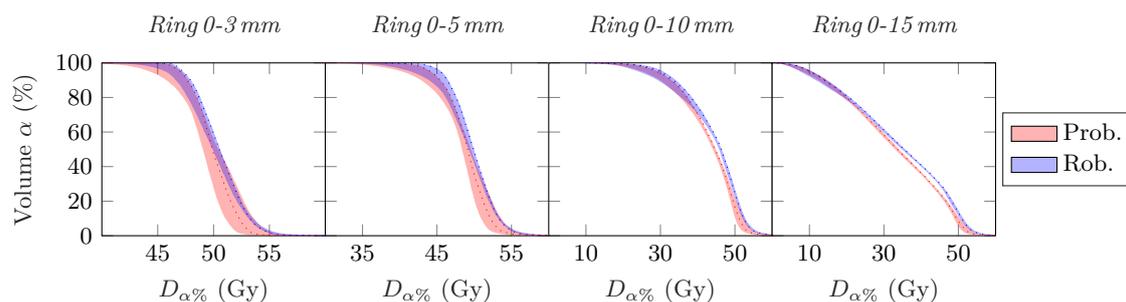
Figure 7.21: Probabilistic DVH corresponding to the probabilistic and robust treatment plan of which the wish-lists (case 2) are described in respectively Table 7.7 and Table A.1. Different rings around the GTV are considered. The left and right edge of the bandwith correspond to respectively the 2-nd and 98-th percentile of the dose volume parameter. The dotted curve shows the nominal scenario.

higher for the probabilistic treatment plan. Furthermore, the mean dose in every organ-at-risk is lower for the probabilistic treatment plan. The structure called *elsewhere* is again the volume of the brain in which dose is deposited outside the four considered organs. In this structure the total deposited dose is also lower for the probabilistic treatment plan.

The probabilistic dose volume histograms and the dose population histograms show that the probabilistic treatment plan outperforms the robust treatment plan for case 2 of the extended wish-lists. Apparently, the probabilistic treatment plan shows more improvement in probabilistic sense than the robust treatment plan when a lower goal on the maximum nominal dose in the brainstem is imposed.

The computation time for the probabilistic treatment plan corresponding to Table 7.7 (case 2) was approximately 3 weeks. The number of iterations needed to optimize the probabilistic objective was 222. The optimization has been performed on the cluster of Erasmus MC, with 8 processors and 60 GB memory reserved. This optimization entailed the whole optimization of the wish-list of case 1 from Table 7.7 with extra iterations to lower the maximum nominal dose in the brainstem. The part of the optimization that is common with the optimization of the wish-list of case 1, took 10 days, which is twice the computation time of case 1. This difference in computation time might be due to different hardware used in the computing cluster or due to a varying number of jobs on the computing cluster, which may cause a variable performance.
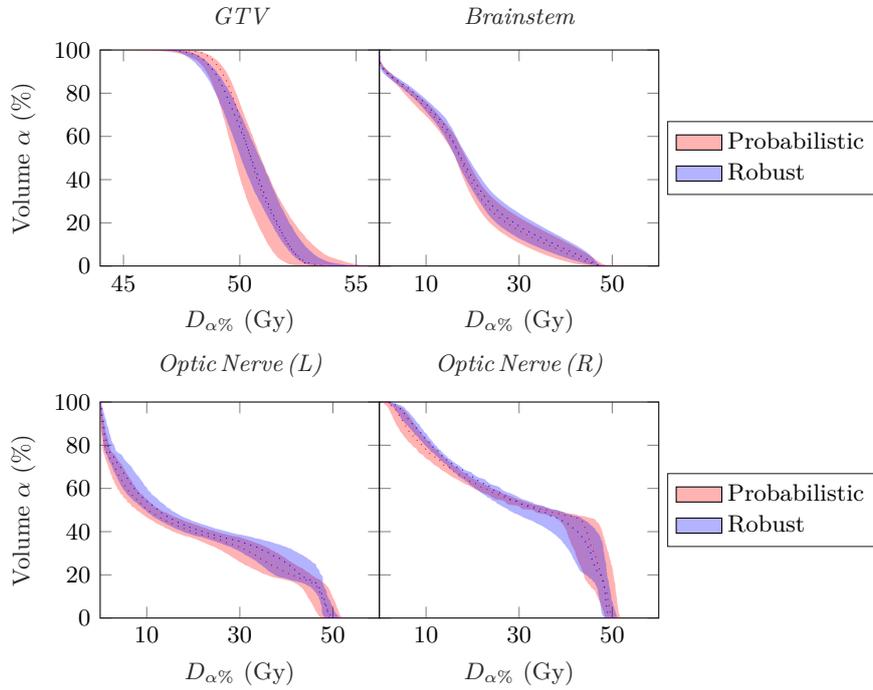
Figure 7.22: Probabilistic DVH corresponding to the probabilistic and robust treatment plan of which the wish-lists (case 2) are described in respectively Table 7.7 and Table A.1. The left and right edge of the bandwith correspond to respectively the 2-nd and 98-th percentile of the dose volume parameter. The dotted curve shows the nominal scenario.
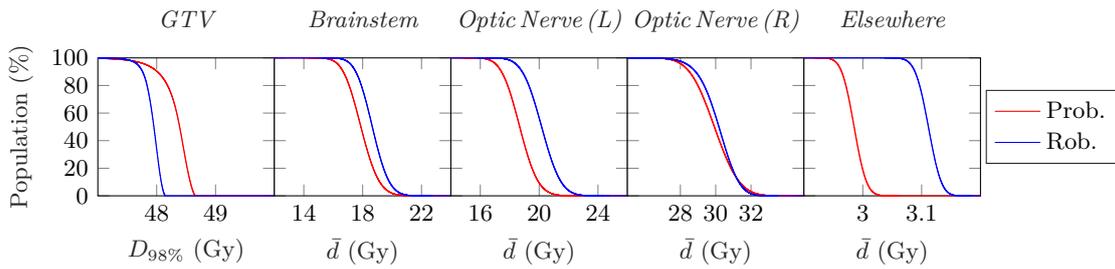
Figure 7.23: Dose population histogram of different organs and outside these organs corresponding to the probabilistic (red line) and robust (blue line) treatment plan of which the wish-lists (case 2) are described in respectively Table 7.7 and Table A.1. For the GTV the $D_{98\%}$ is considered and for the other organs the mean dose $\bar{d}$. Note that the scale of the horizontal axis is different from that of Figure 7.16.

Example: difference probabilistic and robust optimization

For both cases of the extended wish-list, the volume which receives a high dose of the probabilistic plan is smaller than that of the robust plan. Example 7.1 shows an intuitive explanation.

Example 7.1. *This three-dimensional example shows that the automatic extension of the GTV is bigger for a robust treatment plan than for a probabilistic treatment plan. The assumption is made that a set-up error of the patient can be interpreted as an equal shift of nominal dose distribution. In reality, this is not true mainly because of two reasons. First, a shift does not affect the dose distribution if the beam direction is parallel to the set-up error. Second, even if the beam direction is perpendicular to the set-up error, a set-up error does not cause a simple translation of the nominal dose distribution due to heterogeneities in density.*

Consider a spherical GTV with radius r_1 subject to a Gaussian set-up error with zero mean. A larger volume is irradiated when a 95%-probabilistic minimum dose is required at a certain level. To achieve this level, the smallest irradiated area would be a sphere with radius $r_2 > r_1$. The extension, characterized by $r_2 - r_1$, depends on the required confidence interval. For a confidence interval of 95% with a three-dimensional set-up error with standard deviation of 1 mm, the extension would be $r_2 - r_1 = 2.795$ mm.

A probabilistic optimization considers all possible scenarios in probability space and therefore is able to irradiate the whole extension. As a consequence, the automatic extension of a probabilistic treatment plan would ideally be the smallest irradiated volume corresponding to a 2.795 mm extension, which is presented as the red shaded area in Figure 7.24.

On the other hand, a robust optimization considers only a small number of scenarios, such as the scenarios highlighted in blue in Figure 7.24. It is assumed that the set-up robustness SR is high enough that the objective can be reached. In general, these scenarios do not cover the entire ring characterized by $r_2 - r_1 = 2.795$ mm. Therefore, parts within this ring ($\mathcal{A}_1$) are not irradiated, which require parts outside the ring ($\mathcal{A}_2$) to be irradiated in order to meet the 95% probability, as shown in Figure 7.24. Since the magnitude of the set-up errors associated with $\mathcal{A}_2$ is bigger than the magnitude of the set-up errors associated with $\mathcal{A}_1$, the probability density of the set-up errors which are covered by $\mathcal{A}_2$ is lower than the probability density of the set-up errors covered by $\mathcal{A}_1$ due to the Gaussian error with zero mean. To guarantee a target coverage with 95% probability, the volume of $\mathcal{A}_2$ is necessarily bigger than the volume of $\mathcal{A}_1$.

Concluding, under these assumptions, the automatic extension of the GTV is bigger for a robust treatment plan than for a probabilistic treatment plan if the 95%-probabilistic minimum dose is required at a certain level. As a matter of fact these assumptions are not valid, then this example serves as an intuitive explanation.
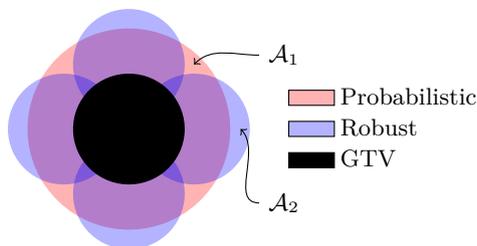
Figure 7.24: A cross-sectional view of the automatic extension of the GTV for a probabilistic and a robust treatment plan.

Chapter 8

Discussions and conclusions

In this chapter, the main findings and conclusions of this research are summarized and a reflection on these is provided.

Methodology results

First of all, a methodology has been provided which shows that a variety of objective functions can be used with Polynomial Chaos Expansion for probabilistic treatment planning. Any statistical measure of a function depending on the uncertainty vector is suitable. Per type of objective function, a specific choice should be made of what part of the function a PCE should be constructed. This could either be a *global* PCE, which means that a PCE is constructed of the probabilistic function as a whole. In general, the construction of such a PCE and sampling from this PCE is computationally less expensive, but on the other hand a PCE is to be constructed in every iteration. One could also construct a more *local* PCE, for example of the dose deposition matrix. Since this matrix generally has many elements, the construction and evaluation of the PCE is computationally more expensive, but on the other hand it is not necessary to construct a PCE every iteration. It is also possible to construct a PCE of a part of the probabilistic function which is in between the two extrema of a global and a local PCE. Besides considerations regarding required computational power, approximation errors of the PCE are also important for the choice of which part of the function a PCE is to be constructed. If the part for which a PCE is constructed also includes the beam intensity vector, a PCE is to be constructed every iteration and the approximation error is different every iteration, which is not beneficial for the optimization. This type of noise can be prevented by constructing a PCE for parts of the objective function which do not depend on the beam intensity.

Second, we have demonstrated that it is possible to probabilistically handle multiple fractions, each subject to an independent random error, with PCE. It turns out that the computational cost does not increase proportional with the number of fractions. In general, more variables are to be assigned for the PCE, to be specific the number of fractions times the number of dimensions, i.e. $T \cdot N$. More variables within the PCE require more cubature points for a certain desirable grid level. It has been found that the set of cubature points contains different permutations of the same combination of random errors amongst the fractions. These different permutations of a certain combination result in the same stochastic response. As a consequence, the number of function evaluations required for the construction of the PCE is less than the number of cubature points. If the number of fractions is greater than or equal to the grid level minus one

(i.e. $T \geq L-1$), the number of function evaluations does not increase anymore for higher number of fractions. Although a considerable number of cubature points is not needed, these are still calculated by the openGPC package, which is computationally expensive.

Third, the expected value of polynomials in dose of degree p turns out to be an objective function which is extremely suitable for handling multiple fractions probabilistically with PCE. For such functions, it is not necessary to assign $T \cdot N$ variables to the PCE. The dimensionality of the PCE can be reduced to $p \cdot N$ or even further to N. This choice depends on many parameters: the degree of the polynomial, the number of dimensions of the random error and the ratio between the number of beamspots and the number of voxels.

Demonstrative results

For the simplified three-dimensional geometry of the spinal cord with surrounding tumor subject to a one-dimensional systematic set-up error, it has been shown that the expected value of the quadratic differences between the dose and the prescribed dose is a suitable objective function for probabilistic treatment planning, since it is convex and analytical which is beneficial for optimization. Polynomial Chaos Expansion is extremely useful to describe this function. A smart choice can be made of which parts a PCE is to be constructed such that less PCEs are required and no PCEs have to be reconstructed every iteration. This resulted in a probabilistic treatment plan which did not require much computational power. The plan shows perfect OAR sparing and conformal target coverage. As expected, the expected dose distribution is more homogeneous than the nominal dose distribution. Weights which could be assigned to each voxel can be adjusted to prioritize for example target coverage or OAR sparing.

For the real skull base meningioma patient, it has been shown that probabilistic treatment planning leads to treatment plans which show a high level of conformal target coverage, enough OAR sparing and minimal dose to surrounding tissue. These plans were obtained by maximizing the probabilistic $D_{98\%}$ in the GTV. It has been shown that the probabilistic treatment plan outperforms the robust treatment plan.

- The probabilistic $D_{98\%}$ is clinically a perfect measure for the GTV to optimize and PCE can be used to compute it. For this purpose, sampling is required, either from a PCE of the dose deposition matrix or from a PCE of the $D_{98\%}$ itself. Sampling from a PCE of the dose deposition matrix turned out to require too much memory. Polynomial Chaos Expansion is able to describe the $D_{98\%}$ well with $L = 5$, $\Delta_{lev} = 1$ and $O = 6$, which requires 303 realizations of the dose deposition matrix from the dose engine. Calculating the probabilistic $D_{98\%}$ including its gradient and Hessian takes about 16 minutes with $N_\mathrm{s} = 5 \cdot 10^5$ samples.

- The probabilistic $D_{98\%}$ is a non-convex function. Besides, calculating it with PCE, results in a bumpy functional dependence on the beam intensity, which is expected to be caused by both the sampling and the approximation of the $D_{98\%}$ with the PCE. As a consequence, the probabilistic $D_{98\%}$ is hard to optimize. However, it has been observed that on a larger scale the 95%-probabilistic $D_{98\%}$ is close to linear with respect to the beam intensity. Moreover, it has been proven that a probabilistic dose volume parameter cannot have a negative gradient. These two findings justify to use a forward finite difference scheme with a high step size to approximate the gradient and the Hessian matrix. A step size of $h = 16$ is used. Besides, it has been observed that enabling the ForceNonConvex and disabling the

UseHessApprox functionality of iCycle yield the best convergence. Monotonicity enforcement of the probabilistic dose volume parameter does not improve the convergence.

- Imposing a logarithmic tumor control probability function on the GTV in the nominal scenario as first objective is a good method to drive the optimization efficiently in the proper direction. A constraint on the maximum nominal dose is required to prevent the optimization to be unbounded. As second objective, the probabilistic $D_{98\%}$ can be imposed. The optimization corresponding to this wish-list converged. The treatment plan shows an automatic extension of the GTV. We have shown that it is also possible to impose a second probabilistic objective on the GTV which minimizes the 5%-probabilistic $D_{2\%}$. This results in a steeper fall-of of the dose volume histogram and therefore in a more homogeneous dose distribution in a probabilistic sense.

- Constraints and objectives on the organs-at-risk in the nominal scenario have been added to the wish-list containing an objective on the 98%-probabilistic $D_{98\%}$ of the GTV. The resulting treatment plan have been compared to a robust treatment plan following the robustness recipes of Ter Haar et al. (2018). The probabilistic treatment plan shows a more conformal GTV coverage. The volume which receives a high dose is smaller than in the robust treatment plan, while the 98%-probabilistic $D_{98\%}$ is equal for both treatment plans. This is due to the fact that a robust treatment plan has preferred directions for the automatic extension of the GTV, which depend on the scenarios. This can cause a difference in the dose distribution of nearby OARs, in favor of either the probabilistic or the robust treatment plan. The overall mean dose is however lower for the probabilistic treatment plan.

- If iCycle reaches a certain value for an objective, this value is set into a constraint and is slightly relaxed (e.g. by 1%), in order to leave room for the optimization of the next objective. Such a relative relaxation in a probabilistic optimization does not have the same impact on the dose distribution as a similar relaxation in a robust optimization. A relaxation of 1% in terms of 98%-probabilistic $D_{98\%}$ differs from a relaxation of 1% in terms of minimum voxel dose over seven scenarios. The actual influence of these relaxations on the treatment plans has not been investigated.

Clinical results

The results of this research have shown that one of the major clinical advantages of probabilistic optimization with respect to robust optimization is that it is possible to directly impose a probabilistic goal, for example that the GTV coverage is adequate with 98% probability. The robustness recipe prescribes robustness settings, from which one may expect a treatment plan which meets the same probabilistic goal for the GTV, but which is not guaranteed. Besides, robustness recipes are site and population specific, while it is expected that probabilistic treatment planning can be applied to any treatment site and to any patient. Furthermore, probabilistic treatment planning has more potential, since probabilistic objectives could be imposed not only on the GTV but on any structure, while robustness recipes are focused on the GTV coverage only.

The results have also shown that probabilistic treatment planning with PCE cannot be used yet in clinical practice, since the computation time is still too long. For the extended wish-list which included a probabilistic objective on the GTV, nominal constraints on the optic nerves and a nominal objective on the brainstem, the computation time was 115 hours which is too long in clinical practice. The optimization of the extended wish-list with a lower goal on the maximum nominal dose in the brainstem took even longer. In Chapter 9 a recommendation is made on how to alleviate the required computational effort.

Chapter 9

Future research

In this chapter, recommendations for future research are provided.

Fractionation with PCE

As a first recommendation regarding fractionation with PCE, a functionality has to be implemented in the openGPC package which bypasses the need for computation of all cubature points when only some of them are required. As discussed, it has been found (see Table 5.1) that the number of function evaluations N_{fval} remains constant for $T > L - 1$ when considering multiple fractions with PCE. However, the number of cubature points N_{cub} rapidly increases. For higher number of fractions, it is computationally very expensive to run the openGPC package, since all cubature points are calculated. However, only N_{fval} cubature points are actually needed. Thus, in general it is possible to consider a higher number of fractions with Polynomial Chaos Expansion, but the openGPC package is not yet suitable for it.

Besides, it has to be confirmed whether the accuracy of the PCE changes when the polynomial order and also the grid level are kept the same when considering multiple fractions. It is expected that it is not required to increase these, due to the symmetry among the fractions and corresponding set-up errors.

Alleviation of computational cost

The computation time of a probabilistic optimization is still too high, as posed in Chapter 8. The voxel reduction functionality of iCycle has been disabled for the GTV for the probabilistic optimizations, therefore all voxels of the GTV were considered. It is recommended to enable this functionality. Then iCycle picks a representative subset of voxels of the GTV, which could considerably reduce the required memory and therefore also the computation time of an iteration. The subset of voxels does not change over iterations and therefore no adjustments to Algorithm 6.1 or to iCycle have to be made. However, the introduced uncertainty of reducing the number of voxels on the final probabilistic treatment plan still has to be quantified.

As stated in Section 7.3.1, the calculation of the $(100 - \beta)$-th percentile takes up a large part of the computation time of Algorithm 6.1. It is recommended to investigate whether the computational cost associated with $(100 - \beta)$-th percentile can be lowered without significant loss of accuracy. Other methods exist to estimate a percentile. For example, a considerably smaller number of samples can be drawn with the Wilks approach to obtain a conservatively estimated

bound [72]. Furthermore, one could approximate a percentile with $\mu + k \cdot \sigma$ with $k \in \mathbb{R}$ depending on the assumed or fitted probability density function of the dose volume parameter $D_{\alpha\%}(\boldsymbol{\xi})$.

Improvement of probabilistic treatment plan

For the real patient data in iCycle, a probabilistic objective was set only on the GTV. It is recommended to include probabilistic constraints or objectives on every organ. Ideally, these criteria would be in terms of percentiles in probability space. After all, such criteria are clinically more relevant than criteria in terms of for example expected value. However, depending on the object of which a PCE is constructed, this can be computationally more expensive, since sampling is required. A first next step could be to minimize the maximum expected dose instead of the maximum nominal dose in the brainstem and the optic nerves. This does not require any additional implementation in iCycle. The nominal dose matrix only has to be substituted in iCycle by the expected dose matrix for all desired organs. By doing so, iCycle's approach to solve linear objectives, as described by Equation 6.8, can be used.

It is recommended to include other uncertainties as well. This research only considered probabilistic optimization with set-up errors, which can be interpreted as a translational shift of the beams. However, one might also want to consider rotational set-up errors of the body, which can be interpreted as a rotational shift of the beams. Range errors are also common to consider. Both rotational set-up errors and range errors can be included easily in probabilistic optimization, since such errors are also introduced in the dose deposition matrix, just like translational set-up errors. This is not necessarily the case for delineation errors. Such an error is inherently different and it has to be investigated what a proper way is to describe such errors and how to take them into account in probabilistic optimization.

To get a better understanding of the differences in terms of OAR sparing between a probabilistic and robust treatment plan, it is recommended to investigate in detail what the influence of the axial preferred directions of robust optimization on the dose distribution in nearby OARs is and above all what the differences are with respect to a probabilistic optimization. For this purpose, a simplified small three-dimensional geometry with a tumor and an organ can be used. The position of the OAR with respect to the tumor can be varied in order to be able to qualitatively describe the differences in dose distribution.

In this research, the beamspots used in the probabilistic optimizations resulted from a conventional treatment plan. Such an approach turned out to work well for this patient, but this is not guaranteed for all patients. After all, it highly depends on the chosen beam directions and the heterogeneities in density the body. As a possible solution, it is recommended to implement the functionality of beamspot resampling in combination with the probabilistic dose volume parameter function in iCycle. This would imply that every iteration a different set of samples of the dose deposition matrix needs to be provided to the probabilistic dose volume parameter function (described by Algorithm 6.1). Beamspot resampling may improve the convergence of the probabilistic optimization, since more degrees of freedom are involved. On the other hand, this may increase the computation cost.

A high step size has been used to approximate numerically the gradient and Hessian of the probabilistic dose volume parameter. If the optimization approaches the optimum, it might be that this step size is too big that it does not measure differences on a smaller scale. Therefore, it might be interesting to investigate whether it is beneficial to use adaptive step sizes for the last few iterations, possibly with enforcement of monotonicity for the gradient and Hessian. A slightly better treatment plan may result.

Bibliography

[1] J. Unkelbach et al. "Robust radiotherapy planning". In: *Physics in Medicine & Biology* 63.22 (2018), 22TR02. DOI: `10.1088/1361-6560/aae659`.

[2] C. ter Haar et al. "Robustness Recipes for Proton Therapy". Master thesis. Delft University of Technology, 2018.

[3] A. C. Begg, F. A. Stewart, and C. Vens. "Strategies to improve radiotherapy with targeted drugs". In: *Nature Reviews Cancer* 11 (2011), pp. 239–253.

[4] G. Delaney et al. "The role of radiotherapy in cancer treatment". In: *Cancer* 104.6 (2005), pp. 1129–1137. DOI: `10.1002/cncr.21324`.

[5] R. Baskar et al. "Biological response of cancer cells to radiation treatment". In: *Frontiers in Molecular Biosciences* 1 (2014), p. 24. ISSN: 2296-889X. DOI: `10.3389/fmolb.2014.00024`.

[6] J. Ferlay, D. Parkin, and E. Steliarova-Foucher. "Estimates of cancer incidence and mortality in Europe in 2008". In: *European Journal of Cancer* 46.4 (2010), pp. 765 –781. ISSN: 0959-8049. DOI: `10.1016/j.ejca.2009.12.014`.

[7] A. C. Society. *Common types of cancer treatment.* Mar. 2019. URL: `https://www.cancer.org/treatment/understanding-your-diagnosis/after-diagnosis/common-cancer-treatments.html`.

[8] J. P. Grutters et al. "Comparison of the effectiveness of radiotherapy with photons, protons and carbon-ions for non-small cell lung cancer: A meta-analysis". In: *Radiotherapy and Oncology* 95.1 (2010), pp. 32 –40. ISSN: 0167-8140. DOI: `10.1016/j.radonc.2009.08.003`.

[9] R. R. Wilson. "Radiological Use of Fast Protons". In: *Radiology* 47.5 (1946). PMID: 20274616, pp. 487–491. DOI: `10.1148/47.5.487`.

[10] X. Tian et al. "The evolution of proton beam therapy: Current and future status". In: *Mol Clin Oncol* 8.1 (Jan. 2018), pp. 15–21. ISSN: 2049-9450. DOI: `10.3892/mco.2017.1499`.

[11] T. Mitin and A. L. Zietman. "Promise and Pitfalls of Heavy-Particle Therapy". In: *Journal of Clinical Oncology* 32.26 (2014). PMID: 25113772, pp. 2855–2863. DOI: `10.1200/JCO.2014.55.1945`.

[12] W. Liu et al. "Robust optimization of intensity modulated proton therapy". In: *Medical Physics* 39.2 (2012), pp. 1079–1091. DOI: `10.1118/1.3679340`.

[13] S. van der Voort et al. "Application of polynomial chaos in proton therapy". Master thesis. Delft University of Technology, 2015.

[14] N. C. Institute. *Radiation Therapy to Treat Cancer.* Mar. 2019. URL: `https://www.cancer.gov/about-cancer/treatment/types/radiation-therapy`.

[15] W. D. Newhauser and R. Zhang. "The physics of proton therapy". In: *Physics in Medicine and Biology* 60.8 (Mar. 2015), R155–R209. DOI: 10.1088/0031-9155/60/8/r155.

[16] W. Ulmer. "Theoretical aspects of energy-range relations, stopping power and energy straggling of protons". In: *Radiation Physics and Chemistry* 76.7 (2007), pp. 1089 –1107. ISSN: 0969-806X. DOI: 10.1016/j.radphyschem.2007.02.083.

[17] Y. Hatano, Y. Katsumura, and A Mozumder. *Charged particle and photon interactions with matter: recent advances, applications, and interfaces.* CRC Press, 2010, p. 328.

[18] M. Yang et al. "Comprehensive analysis of proton range uncertainties related to patient stopping-power-ratio estimation using the stoichiometric calibration". In: *Physics in Medicine and Biology* 57.13 (June 2012), pp. 4095–4115. DOI: 10.1088/0031-9155/57/13/4095.

[19] M Yang et al. "Theoretical variance analysis of single- and dual-energy computed tomography methods for calculating proton stopping power ratios of biological tissues". In: *Physics in Medicine and Biology* 55.5 (Feb. 2010), pp. 1343–1362. DOI: 10.1088/0031-9155/55/5/006.

[20] M. Saito and S. Sagara. "Simplified derivation of stopping power ratio in the human body from dual-energy CT data". In: *Medical Physics* 44.8 (2017), pp. 4179–4187. DOI: 10.1002/mp.12386.

[21] J. Rockhill, M. Mrugala, and M. C. Chamberlain. "Intracranial meningiomas: an overview of diagnosis and treatment". In: *Neurosurgical Focus FOC* 23.4 (2007), E1.

[22] W. M. Mendenhall et al. "Radiotherapy alone or after subtotal resection for benign skull base meningiomas". In: *Cancer* 98.7 (2003), pp. 1473–1482. DOI: 10.1002/cncr.11645.

[23] C. Nutting et al. "Radiotherapy in the treatment of benign meningioma of the skull base". In: *Journal of Neurosurgery* 90.5 (1999), pp. 823 –827.

[24] F. Y. Moraes and C. Chung. "Radiation for skull base meningiomas: review of the literature on the approach to radiotherapy". In: *Chinese Clinical Oncology* 6.1 (2017). ISSN: 2304-3873.

[25] I. Madani et al. "Dose-painting intensity-modulated proton therapy for intermediate- and high-risk meningioma". In: *Radiation Oncology* 10.1 (2015), p. 72. ISSN: 1748-717X. DOI: 10.1186/s13014-015-0384-x.

[26] F. J. Vernimmen et al. "Stereotactic proton beam therapy of skull base meningiomas". In: *International Journal of Radiation Oncology*Biology*Physics* 49.1 (2001), pp. 99 –105. ISSN: 0360-3016. DOI: 10.1016/S0360-3016(00)01457-7.

[27] M. van Herk. "Errors and margins in radiotherapy". In: *Seminars in Radiation Oncology* 14.1 (2004). High-Precision Radiation Therapy of Moving Targets, pp. 52 –64. ISSN: 1053-4296. DOI: 10.1053/j.semradonc.2003.10.003.

[28] V. Rudat et al. "Combined error of patient positioning variability and prostate motion uncertainty in 3D conformal radiotherapy of localized prostate cancer". In: *International Journal of Radiation Oncology*Biology*Physics* 35.5 (1996), pp. 1027 –1034. ISSN: 0360-3016. DOI: 10.1016/0360-3016(96)00204-0.

[29] H. Jiang, J. Seco, and H. Paganetti. "Effects of Hounsfield number conversion on CT based proton Monte Carlo dose calculations". In: *Medical Physics* 34.4 (2007), pp. 1439–1449. DOI: 10.1118/1.2715481.

[30] C. Weltens et al. "Interobserver variations in gross tumor volume delineation of brain tumors on computed tomography and impact of magnetic resonance imaging". In: *Radiotherapy and Oncology* 60.1 (2001), pp. 49 –59. ISSN: 0167-8140. DOI: 10.1016/S0167-8140(01)00371-1.

[31] C. W. Hurkmans et al. "Variability in target volume delineation on CT scans of the breast". In: *International Journal of Radiation Oncology*Biology*Physics* 50.5 (2001), pp. 1366 – 1372. ISSN: 0360-3016. DOI: 10.1016/S0360-3016(01)01635-2.

[32] R. P. Petersen et al. "Target Volume Delineation for Partial Breast Radiotherapy Planning: Clinical Characteristics Associated with Low Interobserver Concordance". In: *International Journal of Radiation Oncology*Biology*Physics* 69.1 (2007), pp. 41 –48. ISSN: 0360-3016. DOI: 10.1016/j.ijrobp.2007.01.070.

[33] A. M. van Mourik et al. "Multiinstitutional study on target volume delineation variation in breast radiotherapy in the presence of guidelines". In: *Radiotherapy and Oncology* 94.3 (2010), pp. 286 –291. ISSN: 0167-8140. DOI: 10.1016/j.radonc.2010.01.009.

[34] M. Prastawa et al. "Automatic brain tumor segmentation by subject specific modification of atlas priors1". In: *Academic Radiology* 10.12 (2003), pp. 1341 –1348. ISSN: 1076-6332. DOI: 10.1016/S1076-6332(03)00506-3.

[35] A. Fredriksson, A. Forsgren, and B. Hårdemark. "Minimax optimization for handling range and setup uncertainties in proton therapy". In: *Medical Physics* 38.3 (2011), pp. 1672–1684. DOI: 10.1118/1.3556559.

[36] A. C. Knopf and A. Lomax. "In vivoproton range verification: a review". In: *Physics in Medicine and Biology* 58.15 (2013), R131–R160. DOI: 10.1088/0031-9155/58/15/r131.

[37] S. V. Spirou and C. S. Chui. "A gradient inverse planning algorithm with dose-volume constraints". In: *Medical Physics* 25.3 (1998), pp. 321–333. DOI: 10.1118/1.598202.

[38] R. G. Wiggenraad et al. "Stereotactic Radiotherapy of Intracranial Tumors: A Comparison of Intensity-Modulated Radiotherapy and Dynamic Conformal Arc". In: *International Journal of Radiation Oncology*Biology*Physics* 74.4 (2009), pp. 1018 –1026. ISSN: 0360-3016. DOI: 10.1016/j.ijrobp.2008.09.057.

[39] F. Martin et al. "Fractionated stereotactic radiotherapy of benign skull-base tumors: a dosimetric comparison of volumetric modulated arc therapy with Rapidarc versus noncoplanar dynamic arcs". In: *Radiation Oncology* 11.1 (2016), p. 58. ISSN: 1748-717X. DOI: 10.1186/s13014-016-0632-8.

[40] D. S. Chang et al. "Normal Tissue Radiation Responses". In: *Basic Radiotherapy Physics and Biology*. Cham: Springer International Publishing, 2014, pp. 265–275. ISBN: 978-3-319-06841-1. DOI: 10.1007/978-3-319-06841-1_26.

[41] S. Breedveld et al. "iCycle: Integrated, multicriterial beam angle, and profile optimization for generation of coplanar and noncoplanar IMRT plans". In: *Medical Physics* 39.2 (2012), pp. 951–963. DOI: 10.1118/1.3676689.

[42] G. Owen. *Discrete mathematics and game theory*. Vol. 22. Springer Science & Business Media, 1999, p. 29.

[43] MathWorks. *Maximizing an Objective*. Apr. 2019. URL: https://nl.mathworks.com/help/optim/ug/maximizing-an-objective.html.

[44] S van de Water et al. "Improved efficiency of multi-criteria IMPT treatment planning using iterative resampling of randomly placed pencil beams". In: *Physics in Medicine and Biology* 58.19 (2013), pp. 6969–6983. DOI: 10.1088/0031-9155/58/19/6969.

[45] S. van de Water et al. "Shortening Delivery Times of Intensity Modulated Proton Therapy by Reducing Proton Energy Layers During Treatment Plan Optimization". In: *International Journal of Radiation Oncology*Biology*Physics* 92.2 (2015), pp. 460 –468. ISSN: 0360-3016. DOI: 10.1016/j.ijrobp.2015.01.031.

[46] S. van der Voort et al. "Robustness Recipes for Minimax Robust Optimization in Intensity Modulated Proton Therapy for Oropharyngeal Cancer Patients". In: *International Journal of Radiation Oncology*Biology*Physics* 95.1 (2016). Particle Therapy Special Edition, pp. 163 –170. ISSN: 0360-3016. DOI: 10.1016/j.ijrobp.2016.02.035.

[47] J. Unkelbach, T. C. Y. Chan, and T. Bortfeld. "Accounting for range uncertainties in the optimization of intensity modulated proton therapy". In: *Physics in Medicine & Biology* 52.10 (2007), p. 2755.

[48] J. Unkelbach et al. "Reducing the sensitivity of IMPT treatment plans to setup errors and range uncertainties via probabilistic treatment planning". In: *Medical Physics* 36.1 (2009), pp. 149–163. DOI: 10.1118/1.3021139.

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

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

[51] D. Xiu. *Numerical methods for stochastic computations: a spectral method approach.* Princeton university press, 2010.

[52] Z. Perkó et al. "Fast and accurate sensitivity analysis of IMPT treatment plans using Polynomial Chaos Expansion". In: *Physics in Medicine & Biology* 61.12 (2016), p. 4646.

[53] Z. Perkó et al. "Grid and basis adaptive polynomial chaos techniques for sensitivity and uncertainty analysis". In: *Journal of Computational Physics* 260 (2014), pp. 54 –84. ISSN: 0021-9991. DOI: 10.1016/j.jcp.2013.12.025.

[54] D. Xiu and G. E. Karniadakis. "The Wiener–Askey polynomial chaos for stochastic differential equations". In: *SIAM journal on scientific computing* 24.2 (2002), pp. 619–644.

[55] L. Trefethen. "Is Gauss Quadrature Better than Clenshaw–Curtis?" In: *SIAM Review* 50.1 (2008), pp. 67–87. DOI: 10.1137/060659831.

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

[57] D. C. Montgomery. *Design and analysis of experiments.* John wiley & sons, 2017.

[58] M. Griebel, M. Schneider, and C. Zenger. *A combination technique for the solution of sparse grid problems.* Citeseer, 1990.

[59] S. A. Smoljak. "Quadrature and interpolation formulae on tensor products of certain function classes". In: *Dokl. Akad. Nauk SSSR* 148 (1963), pp. 1042–1045. ISSN: 0002-3264.

[60] G. Blatman and B. Sudret. "An adaptive algorithm to build up sparse polynomial chaos expansions for stochastic finite element analysis". In: *Probabilistic Engineering Mechanics* 25.2 (2010), pp. 183 –197. ISSN: 0266-8920. DOI: 10.1016/j.probengmech.2009.10.003.

[61] W. Kaplan. *Advanced Calculus*. Addison-Wesley higher mathematics. Addison-Wesley, 2002. ISBN: 9780201799378.

[62] J. E. Dennis and R. B. Schnabel. *Numerical methods for unconstrained optimization and nonlinear equations*. English. Includes index. Englewood Cliffs, N.J. : Prentice-Hall, 1983. ISBN: 0136272169.

[63] M. Abramowitz and I. A. Stegun. *Handbook of mathematical functions: with formulas, graphs, and mathematical tables*. Vol. 55. Courier Corporation, 1965.

[64] T. Bortfeld. "An analytical approximation of the Bragg curve for therapeutic proton beams". In: *Medical Physics* 24.12 (1997), pp. 2024–2033. DOI: 10.1118/1.598116.

[65] "Special Functions". In: *Table of Integrals, Series, and Products (Eighth Edition)*. Ed. by D. Zwillinger et al. Eighth Edition. Boston: Academic Press, 2014. Chap. 9, pp. 1014 –1059. ISBN: 978-0-12-384933-5. DOI: 10.1016/B978-0-12-384933-5.00009-6.

[66] Y. K. Oh et al. "Assessment of setup uncertainties for various tumor sites when using daily CBCT for more than 2200 VMAT treatments". In: *Journal of Applied Clinical Medical Physics* 15.2 (2014), pp. 85–99. DOI: 10.1120/jacmp.v15i2.4418.

[67] J. Lof, B. K. Lind, and A. Brahme. "Optimal radiation beam profiles considering the stochastic process of patient positioning in fractionated radiation therapy". In: *Inverse Problems* 11.6 (1995), pp. 1189–1209. DOI: 10.1088/0266-5611/11/6/005.

[68] M. G. Witte et al. "Beyond the margin recipe: the probability of correct target dosage and tumor control in the presence of a dose limiting structure". In: *Physics in Medicine & Biology* 62.19 (Sept. 2017), pp. 7874–7888. DOI: 10.1088/1361-6560/aa87fe.

[69] T. Hu, V. Klee, and D. Larman. "Optimization of Globally Convex Functions". In: *SIAM Journal on Control and Optimization* 27.5 (1989), pp. 1026–1047. DOI: 10.1137/0327055.

[70] MathWorks. *Evaluating Goodness of Fit*. Apr. 2019. URL: https://nl.mathworks.com/help/curvefit/evaluating-goodness-of-fit.html.

[71] M. Alber and R. Reemtsen. "Intensity modulated radiotherapy treatment planning by use of a barrier-penalty multiplier method". In: *Optimization Methods and Software* 22.3 (2007), pp. 391–411. DOI: 10.1080/10556780600604940.

[72] J. P. Hessling. "Sampling Variance and Bias of the Wilks Conservative Estimate of Confidence Intervals". In: *Nuclear Science and Engineering* 184.3 (2016), pp. 388–399. DOI: 10.13182/NSE16-8.

Appendix A

Other wish-lists

Table A.1: Extended wish-lists for a robust treatment plan with scenarios $s \in \{1, ..., 7\}$ from Table 3.2 with $SR = 3$ mm. Two cases are distinguished, each with a different goal for the maximum nominal dose of the brainstem. The resulting treatment plans are compared with probabilistic treatment plans from the wish-lists described in Table 7.7. The results are presented in Section 7.3.3.

Constraints				
			Limit	
Volume	Type		Case 1	Case 2
GTV	Minimize $\max_i d_i(\boldsymbol{x}, \mathbf{0})$		$1.07 \cdot 50.4$ Gy	$1.07 \cdot 50.4$ Gy
Optic Nerve (L)	Minimize $\max_i d_i(\boldsymbol{x}, \mathbf{0})$		50 Gy	50 Gy
Optic Nerve (R)	Minimize $\max_i d_i(\boldsymbol{x}, \mathbf{0})$		50 Gy	50 Gy

Objectives				
			Goal	
Priority	Volume	Type	Case 1	Case 2
1	GTV	Minimize $\mathrm{LTCP}(\boldsymbol{x}, \mathbf{0})$	1	1
2	GTV	Maximize $\min_s \min_i d_i(\boldsymbol{x}, \boldsymbol{\xi}^{(s)})$	$0.95 \cdot 50.4$ Gy	$0.95 \cdot 50.4$ Gy
2	Brainstem	Minimize $\max_i d_i(\boldsymbol{x}, \mathbf{0})$	50 Gy	46 Gy

Table A.2: Wish-list for a conventional treatment plan of a meningioma patient. The CTV has been extended with a ring of 4 mm. This wish-list has been run with the functionality beamspot resampling enabled. The beamspots which resulted from this optimization have been used for the probabilistic treatment plans. Here, $\bar{d}(\boldsymbol{x}, \boldsymbol{0})$ denotes the mean dose.

Constraints		
Volume	Type	Limit
GTV	Maximize $\min_i d_i(\boldsymbol{x}, \boldsymbol{0})$	$0.97 \cdot 50.4$ Gy
GTV-ring (0-4 mm)	Maximize $\min_i d_i(\boldsymbol{x}, \boldsymbol{0})$	$0.97 \cdot 50.4$ Gy
Optic Nerve (L)	Minimize $\max_i d_i(\boldsymbol{x}, \boldsymbol{0})$	50 Gy
Optic Nerve (R)	Minimize $\max_i d_i(\boldsymbol{x}, \boldsymbol{0})$	50 Gy
Optic Chiasm	Minimize $\max_i d_i(\boldsymbol{x}, \boldsymbol{0})$	50 Gy

Objectives			
Priority	Volume	Type	Goal
2	GTV	Minimize $\max_i d_i(\boldsymbol{x}, \boldsymbol{0})$	$1.05 \cdot 50.4$ Gy
2	GTV-ring (0-4 mm)	Minimize $\max_i d_i(\boldsymbol{x}, \boldsymbol{0})$	$1.05 \cdot 50.4$ Gy
2	GTV-ring (0-5 mm)	Minimize $\max_i d_i(\boldsymbol{x}, \boldsymbol{0})$	$0.9 \cdot 50.4$ Gy
2	GTV-ring (5-10 mm)	Minimize $\max_i d_i(\boldsymbol{x}, \boldsymbol{0})$	$0.6 \cdot 50.4$ Gy
2	GTV-ring (10-15 mm)	Minimize $\max_i d_i(\boldsymbol{x}, \boldsymbol{0})$	$0.3 \cdot 50.4$ Gy
2	External	Minimize $\max_i d_i(\boldsymbol{x}, \boldsymbol{0})$	$1.05 \cdot 50.4$ Gy
3	Hippocampus (Combined)	Minimize $\bar{d}(\boldsymbol{x}, \boldsymbol{0})$	10 Gy
4	Hippocampus (L)	Minimize $\bar{d}(\boldsymbol{x}, \boldsymbol{0})$	1 Gy
4	Hippocampus (R)	Minimize $\bar{d}(\boldsymbol{x}, \boldsymbol{0})$	1 Gy
5	Optic Chiasm	Minimize $\bar{d}(\boldsymbol{x}, \boldsymbol{0})$	1 Gy
5	Optic Nerve (L)	Minimize $\bar{d}(\boldsymbol{x}, \boldsymbol{0})$	1 Gy
5	Optic Nerve (R)	Minimize $\bar{d}(\boldsymbol{x}, \boldsymbol{0})$	1 Gy
6	Unspecified Tissue	Minimize $\bar{d}(\boldsymbol{x}, \boldsymbol{0})$	1 Gy
7	Brainstem	Minimize $\bar{d}(\boldsymbol{x}, \boldsymbol{0})$	1 Gy
8	Retina (L)	Minimize $\bar{d}(\boldsymbol{x}, \boldsymbol{0})$	1 Gy
8	Retina (R)	Minimize $\bar{d}(\boldsymbol{x}, \boldsymbol{0})$	1 Gy
9	Lens (L)	Minimize $\bar{d}(\boldsymbol{x}, \boldsymbol{0})$	1 Gy
9	Lens (R)	Minimize $\bar{d}(\boldsymbol{x}, \boldsymbol{0})$	1 Gy
10	Cochlea (L)	Minimize $\bar{d}(\boldsymbol{x}, \boldsymbol{0})$	1 Gy
10	Cochlea (R)	Minimize $\bar{d}(\boldsymbol{x}, \boldsymbol{0})$	1 Gy
11	Cerebellum	Minimize $\bar{d}(\boldsymbol{x}, \boldsymbol{0})$	1 Gy
11	MU	Minimize $\max_i d_i(\boldsymbol{x}, \boldsymbol{0})$	1

Appendix B

Fractionation with PCE

This chapter is a continuation of Section 5.9.1

In general, as discussed in Chapter 4, a grid $(a_1, ..., a_N)$ is part of a Smolyak sparse grid characterized by the grid level $L = \sum_{j=1}^{N} a_j - N + 1$.

If the considered uncertainties are N-dimensional and T fractions are considered, a Smolyak sparse grid can be denoted as follows:

$$\left\{ (\mathcal{L}^{(1)}, ..., \mathcal{L}^{(T)}) : \sum_{t=1}^{T} |\mathcal{L}^{(t)}| - T + 1 = L \right\}, \tag{B.1}$$

with $\mathcal{L}^{(t)} = (L_1^{(t)}, ..., L_N^{(t)})$ and $|\mathcal{L}^{(t)}| = \sum_{j=1}^{N} L_j^{(t)} - N + 1$.

In Section 5.9.1, it has been demonstrated that different permutations of the same combination of cubature points yield the same stochastic response. This reasoning can also be extended in terms of grids. For $T = 3$ and for an arbitrary number of dimensions N, the grids described by Equations (B.2) to (B.4) yield sets of cubature points with the same stochastic response.

$$\left\{ (\mathcal{L}^{(1)}, \mathcal{L}^{(2)}, \mathcal{L}^{(3)}) : |\mathcal{L}^{(1)}| = 3, |\mathcal{L}^{(2)}| = 1, |\mathcal{L}^{(3)}| = 1 \right\} \tag{B.2}$$

$$\left\{ (\mathcal{L}^{(1)}, \mathcal{L}^{(2)}, \mathcal{L}^{(3)}) : |\mathcal{L}^{(1)}| = 1, |\mathcal{L}^{(2)}| = 3, |\mathcal{L}^{(3)}| = 1 \right\} \tag{B.3}$$

$$\left\{ (\mathcal{L}^{(1)}, \mathcal{L}^{(2)}, \mathcal{L}^{(3)}) : |\mathcal{L}^{(1)}| = 1, |\mathcal{L}^{(2)}| = 1, |\mathcal{L}^{(3)}| = 3 \right\} \tag{B.4}$$

Therefore it suffices to consider only one of these permutations. The permutation which is considered and kept in the sparse grid is characterized by $|\mathcal{L}^{(1)}| \geq |\mathcal{L}^{(2)}| \geq ... \geq |\mathcal{L}^{(T-1)}| \geq |\mathcal{L}^{(T)}|$. By doing so, the *reduced Smolyak sparse grid* with grid level L is formed:

$$\left\{ (\mathcal{L}^{(1)}, ..., \mathcal{L}^{(T)}) : \sum_{t=1}^{T} |\mathcal{L}^{(t)}| - T + 1 = L, |\mathcal{L}^{(1)}| \geq |\mathcal{L}^{(2)}| \geq ... \geq |\mathcal{L}^{(T-1)}| \geq |\mathcal{L}^{(T)}| \right\}. \tag{B.5}$$

Consider a random error of N dimensions. For the PCEs a grid level L is used. Let the number of fractions be equal to the grid level minus two, i.e. $T = L - 2$. The reduced Smolyak sparse grid consists of the grids described in Table B.2a.

Now consider a treatment consisting of $T = L - 1$ fractions. The grids which are contained in the corresponding reduced Smolyak sparse grid are described in Table B.2b.

If grids 1 to N_{gr} of Table B.2a are compared with grids 1 to N_{gr} of Table B.2b, it can be observed that these grids are almost identical, except that the total grid level of the added

fraction equals $|\mathcal{L}^{(L-1)}| = 1$ for these grids. Since $n_{lev} = 2 \cdot lev - 1$, the number of different cubature points corresponding to grids 1 to N_{gr} are the same for Tables B.2a and B.2b. However, it can also be observed that an extra grid is contained in Table B.2b (i.e. grid $N_{\text{gr}} + 1$). This extra grid results in more different cubature points associated with a reduced Smolyak sparse grid corresponding to $T = L - 1$ than with one corresponding to $T = L - 2$.

Now, consider a treatment consisting of $T = L$ fractions. The grids which are contained in the corresponding reduced Smolyak sparse grid are described in Table B.2c.

If we compare the grids corresponding to $T = L - 1$ with those corresponding to $T = L$, it can be noticed that no extra grids are added and that the grid levels are the same for the first $L - 1$ fractions. The total grid level of the added fraction equals $|\mathcal{L}^{(L)}| = 1$ for all grids corresponding to $T = L$. Since $n_{lev} = 2 \cdot lev - 1$, the total number of different cubature points contained in the reduced Smolyak sparse grid corresponding to Table B.2b and B.2c are the same.

The same reasoning applies if the grid level is increased to $T = L + 1$: the number of different cubature points remains the same. In general, we can conclude that for a certain number dimensions N and a certain grid level L with Δ_{lev}, the number of different cubature points contained in the reduced Smolyak sparse grid is constant for any $T \geq L - 1$. It should be noted that this constant increases if the number of dimensions, the grid level or the extra level increases. If this finding is applied to PCE, this means that the number of function evaluations $N_{\text{fval}}(N, L, \Delta_{lev})$ only depends on the number of dimensions, the grid level and the extra level for $T \geq L - 1$.

Although this statement is not properly mathematically proven, it is supported by data. It has been tested for different number of dimensions and for different grid levels whether the number of different cubature points in a reduced Smolyak sparse grid remains constant for $T \geq L - 1$. It appears that this is indeed true. The number of function evaluations for $T \geq L - 1$ is shown in Table B.1 for different grid levels and number of dimensions.

Table B.1: The number of required function evaluations N_{fval} as a function of the number of dimensions N and the grid level L with extra level Δ_{lev}, for $T \geq L - 1$. Note that N_{fval} equals the number of different cubature points in the reduced Smolyak sparse grid.

(L, Δ_{lev})	N_{fval}			
	$N = 1$	$N = 2$	$N = 3$	$N = 4$
$(4, 1)$	30	91	192	341
$(5, 1)$	77	306	780	1607

Table B.2: Grids $(\mathcal{L}^{(1)}, ..., \mathcal{L}^{(T)})$ contained within the reduced Smolyak sparse grid for any N and for any L. The number of different cubature points is equal for the reduced Smolyak sparse grids corresponding to $T = L - 1$ and $T = L$.

(a) $T = L - 2$

#	$|\mathcal{L}^{(1)}|$	$|\mathcal{L}^{(2)}|$	$|\mathcal{L}^{(3)}|$	$|\mathcal{L}^{(4)}|$	$|\mathcal{L}^{(5)}|$	$\cdots$	$|\mathcal{L}^{(L-2)}|$
1	L	1	1	1	1	...	1
2	$L-1$	2	1	1	1	...	1
3	$L-2$	3	1	1	1	...	1
4	$L-2$	2	2	1	1	...	1
5	$L-3$	4	1	1	1	...	1
6	$L-3$	3	2	1	1	...	1
7	$L-3$	2	2	2	1	...	1
$\vdots$	$\vdots$						
N_{gr}	3	2	2	2	2	...	2

(b) $T = L - 1$

#	$|\mathcal{L}^{(1)}|$	$|\mathcal{L}^{(2)}|$	$|\mathcal{L}^{(3)}|$	$|\mathcal{L}^{(4)}|$	$|\mathcal{L}^{(5)}|$	$\cdots$	$|\mathcal{L}^{(L-2)}|$	$|\mathcal{L}^{(L-1)}|$
1	L	1	1	1	1	...	1	1
2	$L-1$	2	1	1	1	...	1	1
3	$L-2$	3	1	1	1	...	1	1
4	$L-2$	2	2	1	1	...	1	1
5	$L-3$	4	1	1	1	...	1	1
6	$L-3$	3	2	1	1	...	1	1
7	$L-3$	2	2	2	1	...	1	1
$\vdots$	$\vdots$							
N_{gr}	3	2	2	2	2	...	2	1
$N_{\mathrm{gr}}+1$	2	2	2	2	2	...	2	2

(c) $T = L$

#	$|\mathcal{L}^{(1)}|$	$|\mathcal{L}^{(2)}|$	$|\mathcal{L}^{(3)}|$	$|\mathcal{L}^{(4)}|$	$|\mathcal{L}^{(5)}|$	$\cdots$	$|\mathcal{L}^{(L-2)}|$	$|\mathcal{L}^{(L-1)}|$	$|\mathcal{L}^{(L)}|$
1	L	1	1	1	1	...	1	1	1
2	$L-1$	2	1	1	1	...	1	1	1
3	$L-2$	3	1	1	1	...	1	1	1
4	$L-2$	2	2	1	1	...	1	1	1
5	$L-3$	4	1	1	1	...	1	1	1
6	$L-3$	3	2	1	1	...	1	1	1
7	$L-3$	2	2	2	1	...	1	1	1
$\vdots$	$\vdots$								
N_{gr}	3	2	2	2	2	...	2	1	1
$N_{\mathrm{gr}}+1$	2	2	2	2	2	...	2	2	1