Beamlet selection and energy layer reduction in IMPT by sparsity induced optimisation

Bachelor Thesis

by

A.F.J.G. Janssen

to obtain the degree of Bachelor of Science
at the Delft University of Technology,
to be defended publicly on Wednesday June 26, 2019 at 1:00 PM.

Student number:	4584171	
Project duration:	November 12, 2018 – June 26, 2019	
Thesis committee:	Dr. ir. S. Breedveld,	Erasmus MC, supervisor
	Dr. ir. M. Keijzer,	TU Delft, supervisor
	Dr. ir. D. Lathouwers	TU Delft, supervisor
	Dr. D. C. Gijswijt	TU Delft
	Dr. Z. Perkò	TU Delft

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

TU Delft

Erasmus MC
Universitair Medisch Centrum Rotterdam

Abstract

In Intensity Modified Proton Therapy, the number of energy layers and the number of beamlets determine the radiation time and the plan calculation time. The purpose of this research project is to test the ℓ_1-norm and the ℓ_1/ℓ_2-norm on the reduction of beamlets and energy layers.

A data set for a prostate cancer patient, that contains 18777 beamlets in 40 energy layers, and a reference plan, generated by 1705 beamlets of a regular grid in 37 energy layers, were included in this project. Both norms were added to the weighted sum. The weights of the objectives were extracted from a sequential ϵ-constraint optimisation. After deletion of beamlets and energy layers, sequential ϵ-constraint optimisation was performed to calculate the output plan. The *Dose Volume Histogram* (DVH) of the output plan and reference plan are compared to evaluate the plan quality.

The addition of an ℓ_1-norm is a useful tool for selecting beamlets. Using a weight for the norm, that is smaller than the lowest weight in the weighted sum, an output plan was generated by 1260 beamlets in 37 energy layers. If the weight of the norm was too large, it was not possible to generate an acceptable output plan with the beamlets that were selected.

The selection of 28 up to 40 energy layers using the ℓ_1/ℓ_2-norm did not differ much from the selection without the norm. The selection of fewer energy layers was different which resulted in a difference in the DVH for the selection of 21 energy layers. In all these selections on energy layers, the plans were generated by more than 14900 beamlets. The ℓ_1/ℓ_2-norm gave a good result for deleting both beamlets and energy layers, although more than double the number of the beamlets in the reference plan needed to be selected.

Using both the ℓ_1-norm and ℓ_1/ℓ_2-norm an almost equivalent plan to the output plan of the ℓ_1/ℓ_2-norm could be achieved with a reduction of another thousand beamlets.

The results encourage additional testing of the sparsity inducing norms on a data set with more energy layers and implementation of the ℓ_1-norm for the reduction of beamlets.

Preface

In autumn last year I started this research project within the Erasmus University Medical Centre in Rotterdam. At the start I was a bit overwhelmed, "What could I do, I am just a student?". Sometimes it felt like searching for a needle in a haystack, when a minor mistake in one of the so many scripts caused my scripts not to work. In the end, I absolutely loved working on a project and seeing all the results. This project has shown me how much you can do with mathematics and physics in an area that I would not have expected it.

Writing this preface I have mixed feelings, I am proud of the result but it also feels strange. For almost 8 months I have been working hard on this project, but it is also the end of three amazing years in Delft.

Now I would like to thank all three of my supervisors for reading this report and correcting me when I was wrong. Sebastiaan, for the meetings whenever I was in Rotterdam and helping me right away when I got stuck. Marleen besides putting me in contact with Sebastiaan and kindly asking me for chapters, for writing all the recommendation letters for my masters degree.

I am ready for the next chapter.

Anne-Fleur Janssen
Delft, June 2019

Contents

1 Proton therapy

Radiotherapy is used in the treatment of cancer, sometimes in addition to surgery, chemotherapy or hyperthermia therapy. The goal of radiotherapy is to irradiate the tumour cells, whilst sparing the healthy cells as much as possible. The introduction of proton therapy and the problem definition will be given in this chapter.

1.1. Proton radiation

When protons travel through matter, they ionise the material and leave a dose behind. The radiation damages and eventually kills cells in the body. Hence, radiation is an effective way of treating cancer. A typical dose deposition as a function of depth is visualised in the Bragg curve in Figure 1.1.
The dose a proton deposits during its travel is relatively low in comparison to the dose it deposits at its maximum. After reaching its peak, the dose deposition goes to zero. This limits the negative effects on the healthy tissue and organs. The range depends on the energy of the proton and the electron density of the tissue [1, 2]. This phenomenon is visualised in the Bragg curve. A detailed explanation of the Bragg curve and different interaction methods of the proton will be given in Chapter 2

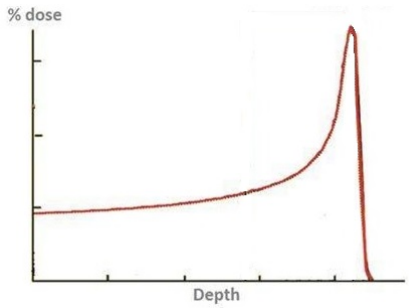


Figure 1.1: Schematic representation of a single Bragg curve. Adaptation from [1]

1.2. Treatment

As healthy cells repair much quicker than tumour cells and to avert side-effects, the treatment is spread out over multiple treatment fractions.

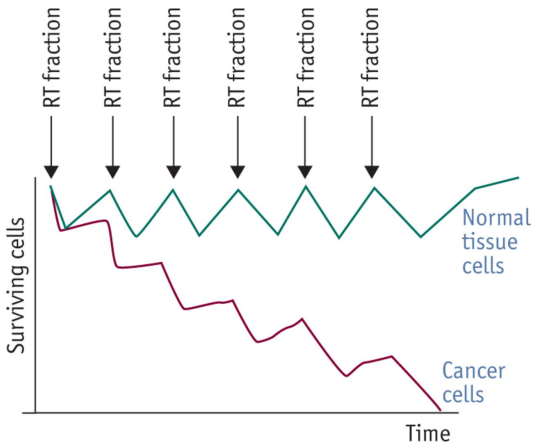


Figure 1.2: Normal tissue cells repair much quicker than cancerous cells. The cancerous cells will be killed whilst sparing the healthy cells. [3]

Each fraction usually takes twenty minutes to one hour of which most of the time is used for positioning the patient. The patient is irradiated for less than ten minutes. [4]
During the treatment, the patient should lie as still as possible. When the patient moves, the actual dose that is delivered might not correspond to the dose that was planned.

1.2.1. Treatment plan

Before the treatment can be performed, a personal treatment plan is created, since each patient is anatomically unique. The plan is based on various different constraints, such as a minimum dose for the volume of the tumour, the *Planning Target Volume* (PTV), and objectives, such as minimising the maximum doses for different organs, the *organs-at-risk* (OARs). The PTV of a prostate cancer patient is depicted in Figure 1.3. This makes it a multi-criteria problem. The objectives and constraints are contained in a wish list with their priorities.

Table 1.1: Wish list with all constraints and objectives. The priority of the objective indicates in which order the sequential ϵ-constraint optimisation optimises. A low number will be optimised first. PTV is planning targeting volume and Gp is gigaprotons

Constraints	Name	Type	Limit (Gy)
	PTV-high	Minimum	71.78
	PTV-intermediate	Minimum	54.45
	PTV-low	Minimum	54.45

Objectives	Name	Type	Goal (Gy)	Priority
f_1	PTV-high	Maximum	79.18	1
f_2	PTV-intermediate	Maximum	79.18	1
f_3	PTV-low	Maximum	58.85	1
f_4	Conformity ring PTV-high	Maximum	79.18	2
f_5	Conformity ring PTV 0-10mm	Maximum	58.85	2
f_6	Conformity ring PTV 10-15mm	Maximum	49.5	2
f_7	Femural heads	Maximum	50	3
f_8	Rectum	Mean	1	4
f_9	Large and small intestines	Mean	1	5
f_{10}	Bladder	Mean	1	6
f_{11}	Femural heads	Mean	1	7
f_{12}	All conformity rings	Mean	1	8
f_{13}	Rest of the conformity rings	Maximum	1	8
f_{14}	Total spot weight	Mean	1 Gp	9

A sequential ϵ-constraint optimisation as described by Breedveld [5] is used in the computation of a dose distribution that satisfies the requirements in the wish list as good as possible. The optimisation method is explained briefly in Section 3.2.1. The goal of the optimisation is to keep the dose in the OARs as low as possible whilst satisfying the minimum dose in the tumour to eradicate all malignant cells. If the plan is clinically acceptable, a treatment planning software computes the plan that is used within the treatment device.

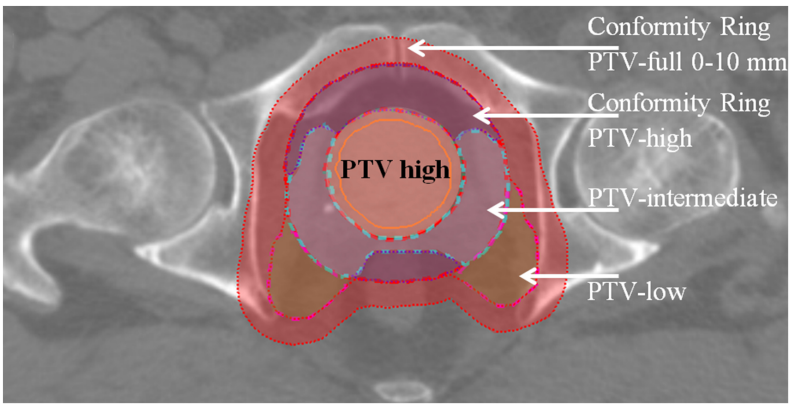


Figure 1.3: The PTV-high is an expansion of the prostate. The PTV-low consists of the expanded seminal vesicles and lymph nodes, that is not contained in PTV-intermediate, the 15 mm transition region between PTV-low and PTV-high. The conformity ring of PTV-high is the PTV-high with a 15 mm expansion. The conformity ring of PTV-full is the area within 10 mm outside the PTV-full, the PTV-high conformity ring and PTV-low. [6]

1.2.2. Treatment device

In *External Beam Radiotherapy* (EBRT) protons exit the treatment device through the head. The head contains magnets to steer a bundle of protons to different positions in the patient. The energy of the proton determines the range of the Bragg peak. The positioning of the proton using magnets will be explained in more detail in Chapter 2.

The treatment device most commonly used in EBRT can be seen in Figure 1.4.

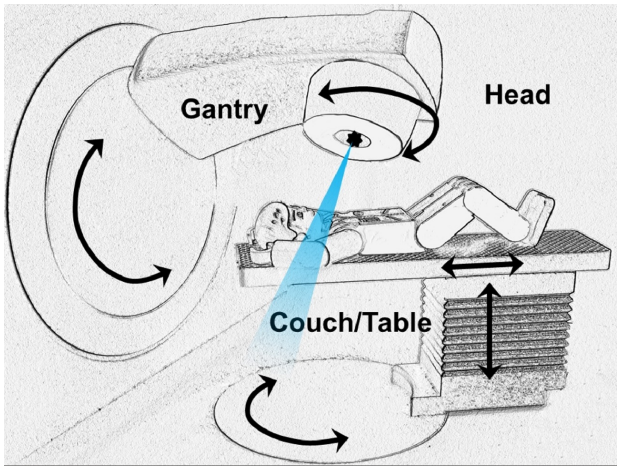


Figure 1.4: The machine most commonly used for radiotherapy. The gantry can rotate 360° to irradiate the patient from different angles. The table can move and rotate. Adapted from [7]

The machine can irradiate the patient from different directions, as the gantry can rotate around the patient and the table can turn. In this report we will only consider two opposite beam directions.

1.3. Movement of the patient and the PTV

Treatment with protons is very precise and the majority of the dose ends up in the tumour, which spares the surrounding tissues quite well. However, this precision has a downside when the tumour or OAR are displaced from their positions on the scan. As a consequence, the dose at the edge of the tumour might not be sufficient to prevent the tumour from growing back. Another consequence is that the dose in the surrounding organs might be too high.
The shift can be caused by several reasons. One of them is a natural movement as a cause of expanding organs such as the stomach or bladder in between the different treatment fractions. Another cause is the slight movement of the patient during the procedure even though the patient is held as still as possible. The longer the procedure, the higher risk of movement.

1.3.1. Movement of organs in between fractions

If the movements of organs is to be taken into account, a CT scan should be made before each treatment session. Using this CT scan a modified plan can be made. However, making a new plan for each session is time consuming, as the optimisation contains a large number of variables.
Reducing the number of variables enables us to make personalised plans for each treatment.

1.3.2. Movement of the patient during treatment

In order to reduce the risk of body movement during the treatment, the radiation time should be reduced. This does not include repetitive movements such as breathing. The length of the session depends among others on the number of different energies that are used. These energies can be ordered into layers; all beamlets with the same energy are in the same layer. Currelty switching between energy layers takes two to five seconds. Most of this time is used for calibrating the magnets in the head of the collimator. As the time of radiation is usually less than ten minutes, having a lot of energy layers is a major contribution on the total radiation time of the treatment. Besides the lower risk on movement, more patients can be treated due to the decrease in radiation time.

1.3.3. Reduction of energy layers and beamlets

The aim of this research project is to reduce the number of beamlets and energy layers used in the optimisation using sparsity inducing norms. The sparsity inducing norms will be introduced in Chapter 4 and in Chapter 5 it is described how the sparsity inducing terms are added.

2

Tumour coverage

The proton pencil beam can be modified to fully irradiate the tumour. This can be achieved by changing the energy of the proton, to choose the range of the proton, the intensity, to modify the size of the dose and the magnetic deflection, to steer the proton. [2]
Protons of different energies will create a *spread-out Bragg peak* (SOBP). The spread-out Bragg peak, range calculation and magnetic deflection will be covered in this chapter.

2.1. Spread-out Bragg peak

Intensity Modulated Proton Therapy, (IMPT), allows the alteration of the intensity of the proton beam. In order to get a sufficient dose in the tumour, the tumour is irradiated with protons of different energies and intensities. This results in the spread-out Bragg peak pictured in Figure 2.1.

2.1.1. Proton interaction mechanisms

The proton loses energy due to three main interactions, which are depicted in Figure 2.2.
Inelastic Coulomb scattering between the proton and atomic electrons causes energy loss and determines the range of the proton. The deflection of the proton is negligible as the mass of the proton is much greater than the mass of electrons.

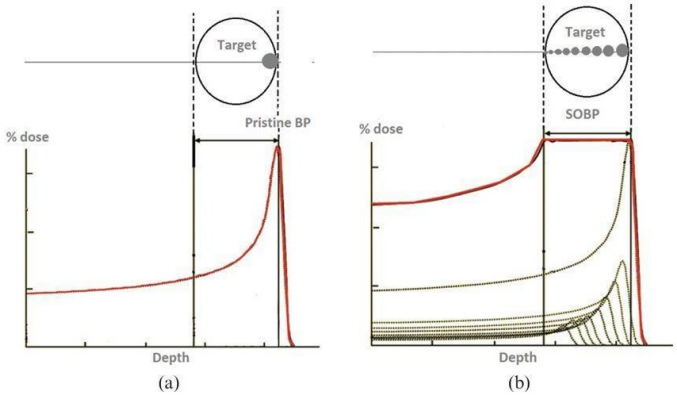


Figure 2.1: Schematic representation of a Bragg peak with the relative dose to the depth in tissue. a) A single Bragg peak. b) The spread-out Bragg Peak, obtained by irradiating with protons of different energies and intensities. Adapted from [1]

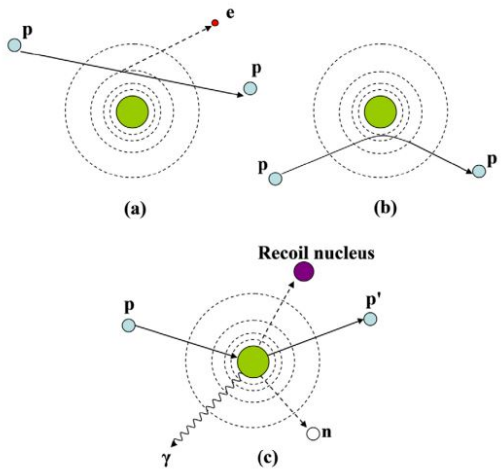


Figure 2.2: Three main interactions mechanisms. (a) Inelastic Coulomb scattering, (b) elastic Coulomb scattering and (c) non-elastic Coulomb scattering. p depicts a proton, e an electron, n a neutron and γ a gamma-ray. [8]

As the loss per interaction is small, the proton is continuously slowing down.
Elastic Coulomb scattering changes the trajectory of the proton. The proton experiences a repulsive force from the nucleus of the atom, which also has a positive charge, and is deflected from its original path.
A non-elastic nuclear interaction is less predictable. When the proton enters the nucleus, a deuteron, secondary proton, triton or heavier ion or one or more neutrons can be emitted. [8, 9]

2.1.2. Stopping power

The rate of energy loss of the projectile is defined as the quotient of the change in energy dE and the travelled path dx. This is the stopping power, S.

$$S(E) = -\frac{\mathrm{d}E}{\mathrm{d}x} \tag{2.1}$$

The Bethe formula appears to be an accurate formula to approximate the stopping power for radiotherapy.

$$S = \frac{4\pi n z^2}{m_e c^2 \beta^2}\left(\frac{e^2}{4\pi\epsilon}\right)^2\left[\ln\frac{2m_e c^2\gamma^2\beta^2}{I} - \beta^2\right] \tag{2.2}$$

n is the electron density of the material, z the atomic number of the projectile, m_e the mass of an electron, c the speed of light, e the electric charge and ϵ_0 the permittivity of vacuum. $\beta = v/c$, where v the velocity of the projectile. $\gamma = (1-\beta^2)^{-\frac{1}{2}}$ is the Lorentz factor and I the mean excitation potential of the absorbing material. [8, 10]

2.1.3. Bragg curve

Using Einsteins famous relationship for kinetic energy, it becomes clear that the velocity is high for high energy particles.

$$E_k = (\gamma - 1)m_p c^2 \tag{2.3}$$

Where m_p is the mass of the proton and γ the Lorentz factor. The Lorentz factor, $\gamma = (1-\beta^2)^{-\frac{1}{2}}$, is larger for large v.

As the particle has a certain velocity, the stopping power is not zero and energy is deposited to the tissue. With the decrease of energy, the velocity and thus β decreases as well. This results in a larger stopping power. At low energy, as $\beta \to 0$, S increases causing a peak in the deposition of energy to occur. This results in the Bragg-Peak that is depicted in Figure 2.1. Under the assumption that the path of the proton is straight and that protons lose energy continuously the range can be calculated.

$$R(E) = \int_0^E \frac{1}{S(E')}\,dE' \tag{2.4}$$

R is the range in m. [8]

2.2. Active scanning

To get a full tumour coverage, active scanning is used. Magnets in a magnetic scanner deflect the path of the pencil beam to specific positions in a vertical plane. A sketch of different paths of the pencil beam is given in Figure 2.3. [1, 11]

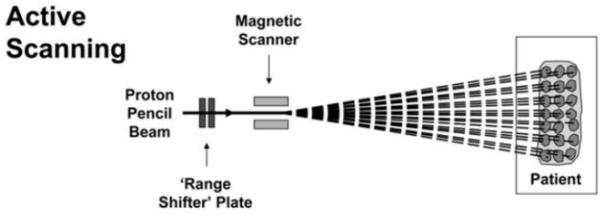


Figure 2.3: Pencil beam active scanning. The pencil beam is deflected to points on a vertical plane perpendicular to the initial pencil beam. [11]

The deflection of the particle will be explained with the Lorentz-force.

$$\mathbf{F}_L = q\mathbf{E} + q\mathbf{v} \times \mathbf{B} \tag{2.5}$$

All bold variables are vectors. $\mathbf{F}_L$ is the Lorentz force on the particle with charge q and velocity $\mathbf{v}$. $\mathbf{E}$ and $\mathbf{B}$ are the electric field and magnetic field respectively. The force on the proton is summation of a force perpendicular on both the velocity of the particle and direction of the magnetic field and a force in the direction of the electric field.

Consider the situation where $\mathbf{E} = 0$ and $\mathbf{B} = B\hat{y}$ is constant in the y-direction. The direction of the velocity is in the z-direction. This is sketched in Figure 2.4.
In this situation, the Lorentz force simplifies to a simple equation.

$$\mathbf{F}_L = -qB(v_z\hat{x} - v_x\hat{z}) \tag{2.6}$$

$\hat{x}$ and $\hat{z}$ are the unit vectors in the x-direction and z-direction respectively. Using Newton's law of motion the centripetal force is given as

$$\mathbf{F}_c = \frac{m_p v^2}{r} \tag{2.7}$$

With m_p the mass of the proton and r the radius of the circle. Solving

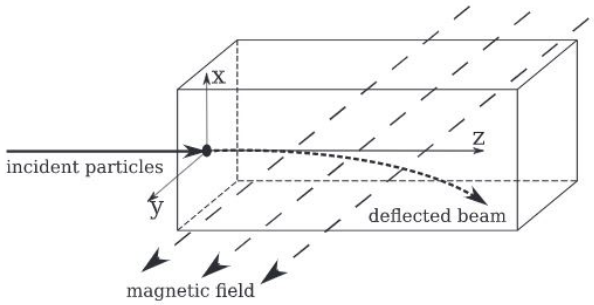


Figure 2.4: The incident particle is deflected in a circular path towards the negative x-direction under the influence of a uniform magnetic field in the y-direction. [12]

these equations for the initial velocity of $v_x = 0$, r can be calculated. The proton will travel in a circular path with radius r.

Changing the strength and direction of the magnetic field the path of the proton can be modified.

3

Treatment plan optimisation

In the optimisation of treatment plans, discretised values are used. In this chapter the discretisation of the patient's volume and the beam positioning is covered. Next, the sequential ϵ-constraint optimisation method will be compared with the weighted sum optimisation method for the multi-criteria problem.

3.1. Discretisation

A beam source of high energy protons exits the head of the treatment device as shown in Figure 1.4. From the beam source, the protons travel through the head to the patient. The modulation device is divided into *beamlets*, x_i, and the volume of the patient into *voxels*, d_i. A simple figure of the discretisation of the patient and the modulation device is pictured in the following figure.

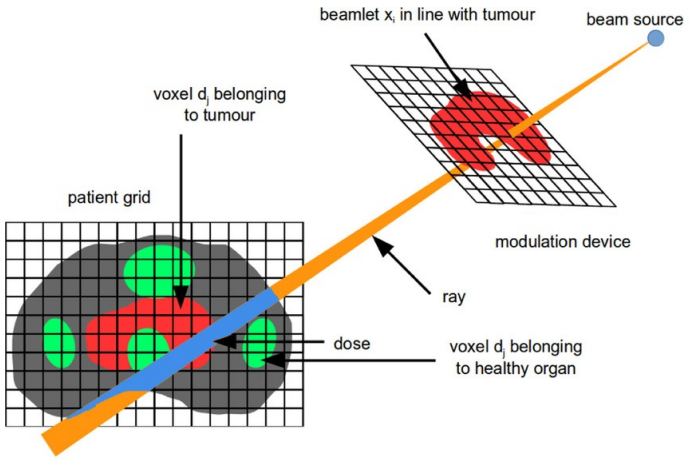


Figure 3.1: The modulation device is discretised into beamlets, x_i. A beam source sends out ionising radiation, that is deflected through the grid of the modulation device. The patient's volume is discretised into voxels, d_j. The higher the intensity of the beam, the higher the resulting dose in the patient. Adapted from [7]

The numerical value of x_i is the intensity of the pencil beam that passes through the grid element. Using the *dose influence matrix* A, the discretised dose in the patient, d_j, can be calculated. There is a different dose influence matrix for each organ in order to optimise different organs separately. $\mathbf{d}$ is a vector with all doses d_j and $\mathbf{x}$ of all beamlets x_i.

$$\mathbf{d}(\mathbf{x}) = A\mathbf{x} \tag{3.1}$$

The total dose distribution is pictured in a 3D dose distribution, overlayed on the CT. The 2D representation can be given in a *Dose Volume Histogram* (DVH). To keep the problem as small as possible, the most important volumes are sampled with a higher sampling resolution than less important volumes. [7]

3.2. Optimisation

In radiotherapy various multi-criteria optimisation methods can be used, such as weighted sum optimisation and sequential ϵ-constraint optimisation. The two methods are interesting, because it is possible to switch between these. The switch between the two methods is useful, as it is easier to add, modify or delete objectives and constraints in a weighted sum than in a sequential ϵ-constraint optimisation. This way all treatment objectives are optimised simultaneously while preserving patient specific trade-offs.

In this section the sequential ϵ-constraint optimisation and the weighted sum

optimisation are introduced. The objectives are denoted as f_i where $i \in \{1,\ldots,s\}$ and s the number of objectives and the constraints on the tumour as g_j where $j \in \{1,\ldots,r\}$ and r the number of constraints. The constraints g_j are combined in $\mathbf{g}(x)$, such that $\mathbf{g(x)} \leq \mathbf{0}$. [5]
An example of a wish list can be found in Table 1.1.

3.2.1. Sequential ϵ-constraint optimisation

The method of sequential ϵ-constraint optimisation, starts by minimising the dose in the OAR with the highest priority, f_1, whilst satisfying all constraints $\mathbf{g}(\mathbf{x})$. This will generate a plan that has a minimal dose everywhere in the tumour and the lowest dose possible in the OAR with the highest priority. When this optimisation is finished, the result will be relaxed by a value ϵ_1 and added as a constraint in the next steps in the optimisation. Relaxation is necessary to allow clinically interesting trade-offs.
In the next step the OAR with the second to highest priority will be minimised satisfying all the new and old constraints. This will generate a plan that has a minimal dose everywhere in the tumour and a dose in the OAR with the highest priority that is smaller than ϵ_1.
These steps are repeated until all OARs have been processed. The optimisation problem of the i'th step is given below.

$$\begin{aligned} \text{minimise} \quad & f_i(x) \\ \text{subject to} \quad & \mathbf{g}(x) \leq \mathbf{0} \\ & f_k(x) \leq \epsilon_k \qquad\qquad k \in \{1,...,i-1\} \end{aligned} \tag{3.2}$$

All the objectives f_k, with a higher priority than f_i, are added as constraints with value ϵ_k:

$$\epsilon_k = \begin{cases} b_k & f_k(\mathbf{x}^*)\delta \leq b_k \\ f_k(\mathbf{x}^*)\delta & f_k(\mathbf{x}^*)\delta > b_k \end{cases} \tag{3.3}$$

b_k is the goal set for the minimisation of the objectives and δ the relaxation value. In most cases the relaxation value is 1.03 (3%). Without relaxation of the result, it might occur that further calculations do not give a solution. A detailed explanation can be found in Breedveld [5].

3.2.2. Weighted sum optimisation

In the weighted sum method, the objective function is obtained by summing the objectives f_i with a specific weight w_i for each objective.

$$\begin{aligned} &\text{minimise} && \sum w_i f_i(x) \\ &\text{subject to} && \mathbf{g}(x) \leq \mathbf{0} \end{aligned} \tag{3.4}$$

3.2.3. Comparison weighted sum and ϵ-constraint optimisation

To switch between the two optimisation methods, it needs to be confirmed that the methods give similar results. The optimal solutions of both methods will be calculated and compared.

First the optimal solution is calculated using the ϵ-constraint method. In Breedveld it is proven that if particular weights are chosen, the weighted sum optimisation results in an optimal solution identical with the optimal solution of the ϵ-constraint optimisation. [5]

Therefore, these particular weights will be extracted from the plan generated with the sequential ϵ-constraint method and used in the weighted sum method.

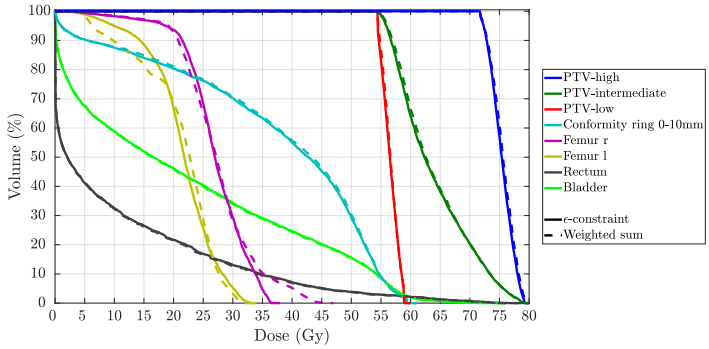


Figure 3.2: DVH of a treatment plan generated by ϵ-constraint optimisation and by the weighted sum optimisation.

It appears that the solutions are quite similar. Numerical and rounding errors can be a cause for the differences in the solutions.

4

Sparsity inducement using norms

In this chapter the cost function $\Omega(x)$ for sparsity-induced optimisation is described. The goal of a sparsity-inducing cost function is to make the solution more sparse. A vector or matrix is sparse when more elements are zero than non-zero.
Such a criterion should be active during optimisation of each objective, making inclusion in the sequential ϵ-constraint method impractical. As an alternative, the weighted-sum formulation is used, that is described in Section 3.2.2. The sparsity-induced cost function is added by some small weight λ.

$$\begin{aligned} &\text{minimise} && \sum w_i f_i(x) + \lambda\Omega(x) \\ &\text{subject to} && \mathbf{g}(x) \leq \mathbf{0} \end{aligned} \tag{4.1}$$

It is suggested that the sparsity is induced by an ℓ_1-norm or a mixed ℓ_1/ℓ_p-norm. [13, 14]
In Chapter 5 the inclusion of the sparsity-induced cost functions on the optimisation of a treatment plan in radiotherapy is explained.

4.1. ℓ_1-norm sparsity term

At first glance, it is not immediately clear why an addition of an ℓ_1-norm induces sparsity. In order to understand this, the ℓ_1-norm is sketched and compared to the general ℓ_q-norms. The ℓ_q-norm is defined as $\|\mathbf{x}\|_q = (\sum_{i=1}^{n} |x_i|^q)^{\frac{1}{q}}$.

4.1.1. ℓ_q in a simple 2D problem

In Figure 4.1a $\|\mathbf{x}\|_q = x_1^q + x_2^q = 1$ is sketched for $q = 1, 2$ and 10. In this figure an important result can be seen. The curve of $\|\mathbf{x}\|_q = 1$ is a diamond if $q = 1$, a circle if $q = 2$ and tends to look like a square if $q = 10$. For $q \to \infty$ the curve approaches a square. For all other q, where $a \leq q \leq b$, the curve of $\|\mathbf{x}\|_q = 1$ is positioned outside the curve of $\|\mathbf{x}\|_a = 1$ and inside the curve of $\|\mathbf{x}\|_b = 1$.
In Figure 4.1b only the positive quadrant is taken into account. A simple optimisation problem with only one constraint, $x_2 + \frac{3}{2}x_1 \geq 2$, is considered. Then all points on and above the red line satisfy the constraint. The objective is the minimisation of $x_2 + \frac{3}{2}x_1$. Then all points on the red line are possible solutions. Now, $\Omega(\mathbf{x}) = \|\mathbf{x}\|_q$ is added with weight $\lambda = 1$ to the problem with the single constraint. Thus the goal of the optimisation is to choose the smallest c, such that $\|\mathbf{x}\|_q = c$ and the red line intersect. This can be visualised as an expanding balloon, where the balloon is inflated until it touches the line of the constraint.
As the curve of $\|\mathbf{x}\|_1 = c$ is a diamond, the point of intersection is very likely to be on one of the axes, which results to a sparse solution. The curve of $\|\mathbf{x}\|_{10} = c$ resembles a square. Therefore, the intersection is very likely to be in the middle of the line and the solution is not sparse at all. The ℓ_2-norm is not as straightforward. In that case, the probability of getting a sparse or almost sparse solution is quite low. We say that the solution is almost sparse if x_1 or x_2 is small in comparison to the other.
Visually, it is very clear that the addition of an ℓ_1-norm increases sparsity and the addition of an ℓ_q-norm where $q \geq 2$ does not.

4.1.2. Extension of ℓ_q to a larger problem

In the previous section a problem with two variables and one constraint is considered. This can be extended to a problem with more constraints and variables. The extension to a problem with multiple variables is rather straightforward. Then, the surface can be seen as a hypervolume and the minimisation of the norm corresponds with expanding the volume until it intersects with the constraint.

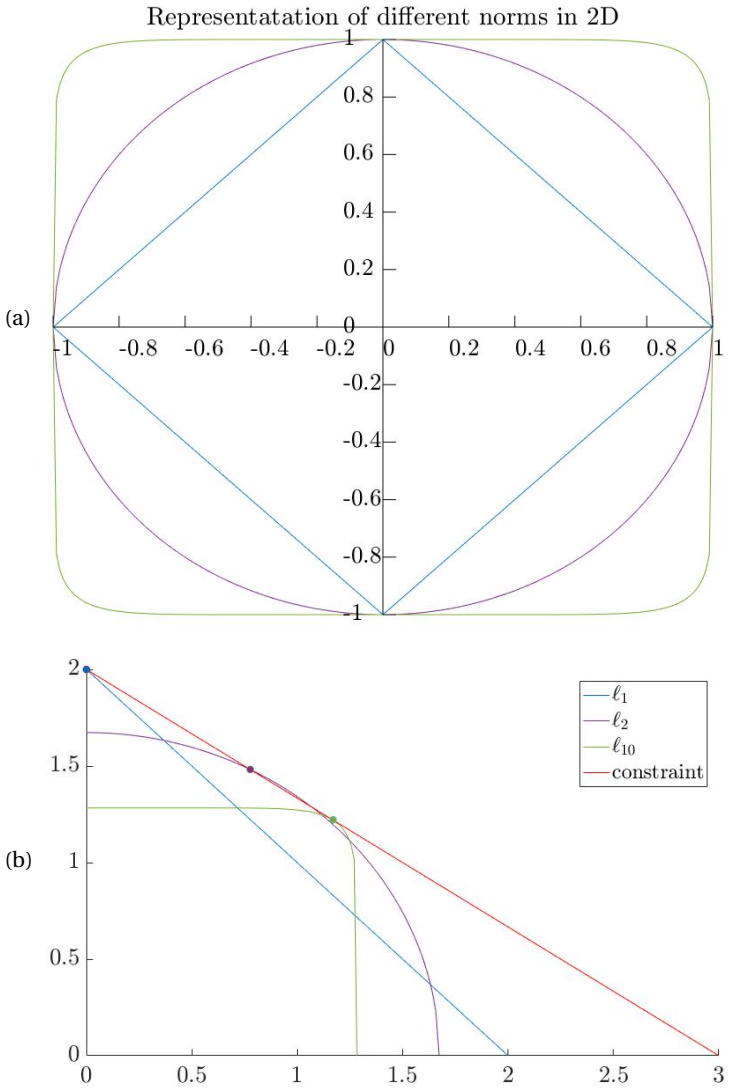


Figure 4.1: a) $\|\mathbf{x}\|_q = 1$ for $q = 1, 2$ and 10. b) $\|\mathbf{x}\|_q = c$ and constraint $x_2 + \frac{2}{3}x_1 \geq 2$ for positive values x_1 and x_2. The point of intersection is indicated with dots.

If there is more than one constraint in the problem, the same thought process can be used. Even though the intersection of $\|\mathbf{x}\|_1$ with the constraints might not be exactly at one of the axis, the probability of getting a sparse or almost sparse solution is higher.

4.2. ℓ_1-grouped sparsity term

A grouped sparsity term, ℓ_q/ℓ_p-norm, is used if the goal is to induce sparsity in a problem, where all variables of a group in a partition should either be selected or ignored. The ℓ_q/ℓ_p is defined as

$$\Omega(\mathbf{x}) = (\sum_{g \in G} \|\mathbf{x}_g\|_p^q)^{\frac{1}{q}} \tag{4.2}$$

G is a partition of the set $S = \bigcup x_i$, the set with all x_i. g is an element of partition G and therefore a subset of S. $\|\mathbf{x}_g\|_p$ denotes the ℓ_p-norm performed on all x_i that are in set g.
The grouped sparsity term will be clarified using a simple example.

Consider Figure 4.2 with set $S = \{x_i : i \in [1,6]\}$ and assume that the variables of set S are either red, blue or green. Each subset S_1, S_2 and S_3 corresponds to one of these colours. $G = \{\{x_1\},\{x_2,x_5\},\{x_3,x_4,x_6\}\} = \{S_1,S_2,S_3\}$ is a partition of the set with elements S_1, S_2 and S_3 based on the different colours.

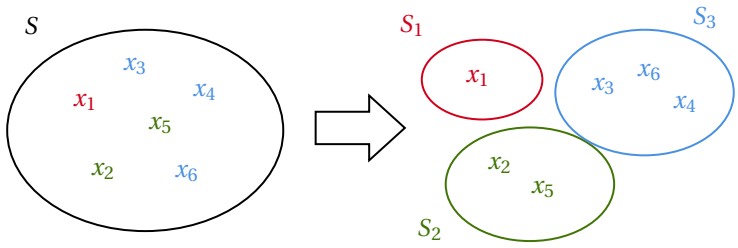


Figure 4.2: Set S can be divided in three disoint sets S_1, S_2 and S_3 based on the colour of x_i.

The ℓ_p-norm of each group is

$$\begin{aligned} S_1: &\quad \|\mathbf{x}_{\mathbf{S_1}}\|_p = \sqrt[p]{x_1^p} \\ S_2: &\quad \|\mathbf{x}_{\mathbf{S_2}}\|_p = \sqrt[p]{x_2^p + x_5^p} \\ S_3: &\quad \|\mathbf{x}_{\mathbf{S_3}}\|_p = \sqrt[p]{x_3^p + x_4^p + x_6^p} \end{aligned} \tag{4.3}$$

Calculating the ℓ_q/ℓ_p-norm results in

$$\Omega(\mathbf{x}) = \sqrt[q]{\sqrt[p]{x_1^p}^q + \sqrt[p]{x_2^p + x_5^p}^q + \sqrt[p]{x_3^p + x_4^p + x_6^p}^q} \tag{4.4}$$

Logically, the ℓ_1-norm is used on the disjoint groups, as the goal is to select or ignore whole groups. Another norm will be used on the variables of the groups. The norm on the variables within the group depends on the preferred result. If the variables need to be equal, it is advised to use an ℓ_p-norm with high p, whereas a low p is advised if the sparsity of the variables within the group need to be induced even more. These norms are combined in a mixed ℓ_1/ℓ_p-norm.

$$\Omega(\mathbf{x}) = \sum_{g \in G} \left\| \mathbf{x}_g \right\|_p \tag{4.5}$$

Now the ℓ_1-norm is calculated over the ℓ_p-norm of each group, $\left\| \mathbf{x}_g \right\|_p$.

The ℓ_1/ℓ_2-norm is defined on the coloured groups of set S as,

$$\Omega(\mathbf{x}) = \sqrt{x_1^2} + \sqrt{x_2^2 + x_5^2} + \sqrt{x_3^2 + x_4^2 + x_6^2} \tag{4.6}$$

5

Method

The goal of this research project is to reduce the number of energy layers and the number of beamlets, that are used in a radiotherapy treatment plan. In this chapter it is described how to add the sparsity inducing terms to the weighted sum and to find a solution with fewer beamlets and energy layers.
First the data set and scripts that are used are described. Second, the overall process is explained using a workflow. Then the ℓ_1-norm and ℓ_1/ℓ_2-norm are written in forms such that these can be included in the optimisation problem. Finally the method that is used for the selection of beamlets is explained.

5.1. Data set and scripts

The data set includes the details of two laterally opposing beams, the wishlist (see Table 1.1) of a prostate cancer patient, a CT scan of the patient and other information that is required to visualise the data. In total there were 18777 beamlets in 40 energy layers, 18 in the first beam and 22 in the second beam.
All plans and calculations are made in the house-build program iCycle using already existing routines such as "*cons2weights*", to extract the weights for the weighted sum optimisation, "*primaldual*" to solve an optimisation problem by a weighted sum and "*mcopt*" to calculate a plan using sequential ϵ-constraint optimisation. The script "*cons2weights*" is based on the theory in Breedveld. [5]. The scripts "*primaldual*" and "*mcopt*" are able to optimise for objectives and

constraints that have a specific cost function. For each function the function value, gradient and Hessian need to be defined given a matrix or vector that is specific for this function.

Take for example the minimisation of the dose in an OAR. The function is defined as $\mathbf{d} = A\mathbf{x}$, with A dose influence matrix. If the matrix A is known, the function value, gradient and Hessian can be calculated.

5.2. Process of making a sparse treatment plan

The workflow sketched below will be used to reduce the number of beamlets and energy layers.

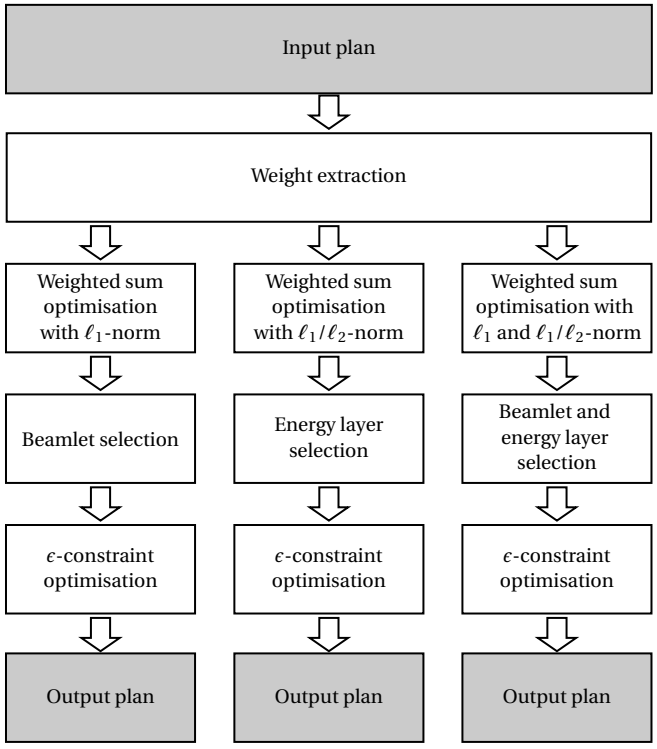


Figure 5.1: Workflow of the reduction of spots, energy layers or reduction of spots and energy layers.

The input plan is a plan created with sequential ϵ-constraint optimisation on all 18777 beamlets in the data set. In all cases the weighted sum method is used with the addition of one or two sparsity inducing terms as described in Chapter 4. The plans with the added norm(s) will be called "*intermediate plans*".

These intermediate plans will be compared with the plan calculated with the weighted sum method without the addition of a norm. This comparison is used to analyse the influence of the norms on the weighted sum.
The selection of beamlets and energy layers is performed on the intermediate plans. Using the remaining beamlets and energy layers a new (output) plan is created that satisfies the minimum doses in the PTVs, using the sequential ϵ-constraint optimisation.
The output plan is then compared to the reference plan in Figure 5.2. The reference plan is created with beamlets on a "regular grid" with a lateral spacing of 6 mm and a relative energy spacing of 2. All beamlets of this reference plan were included in the data set.

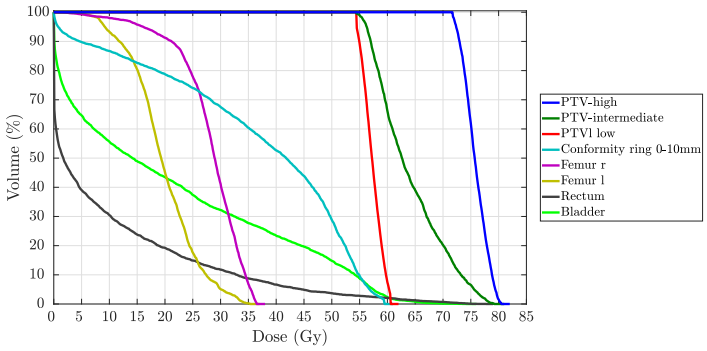


Figure 5.2: The reference plan of a prostate cancer patient. The plan is generated by 1705 beamlets in 37 different energy layers.

5.3. ℓ_1-norm

As all $x \geq 0$, the ℓ_1-norm simplifies to $\|\mathbf{x}\|_{i=1} = \sum_{i=1}^{n} x_i$, where n is the number of beamlets. The norm results in a linear cost function and is denoted in matrix form as

$$\|\mathbf{x}\|_1 = \mathbf{c}\mathbf{x} \tag{5.1}$$
$$\mathbf{c} = \begin{bmatrix} 1 & 1 & \dots & 1 \end{bmatrix}$$

where $\mathbf{c}$ is a $1 \times n$ matrix.
The function value f, the i'th element of the gradient $\nabla\mathbf{f}$, $\frac{\partial f}{\partial x_i}$, and element $h_{i,k}$

of the Hessian, $H = \nabla^2 f$, of a linear function are defined as,

$$f = \sum_{i=1}^{n} c_i x_i \tag{5.2}$$

$$\frac{\partial f}{\partial x_i} = c_i \tag{5.3}$$

$$h_{i,k} = 0 \tag{5.4}$$

Where c_i is the i'th element of the cost function vector $\mathbf{c}$.

5.4. Energy layer reduction

In the previous chapter it is stated that an ℓ_1-grouped sparsity term is a useful tool to induce sparsity in a problem where there is a partition of all variables. In this part the ℓ_1-grouped sparsity term is applied to energy layers and a mixed ℓ_1/ℓ_2-norm is written in a form that can be added to the optimisation problem.

5.4.1. Energy layers

As mentioned in Chapter 1, protons of different energies are used during treatment. Additionally to the spatial information, each beamlet also contains the information of the energy of the protons passing through the head. So each beamlet comes from a specific position and is in a certain energy level.
Therefore, a partition, E, of beamlets based on the energy can be constructed. E is a set with elements E_j, where $j \in \{1, \ldots, m\}$ and m the number of energy layers. E_j is the set containing all beamlets that are in energy layer j such that the union of all E_j is an exact cover of the set with all beamlets, S. Using this exact cover, a grouped sparsity term can be added.

$$S = \bigcup_{j=1}^{m} E_j \tag{5.5}$$

The mixed ℓ_1/ℓ_p-norm is an ℓ_1-norm on all $\left\|\mathbf{x}_{E_j}\right\|_p$ for $E_j \in E$. The grouped sparsity term will result in the fact that whole energy layers can be neglected.

5.4.2. ℓ_1/ℓ_2-norm

The ℓ_1/ℓ_2-norm does not have a simple function value that can be written in a linear or other simple matrix form. First, the function value f will be calculated.

$$f = \sum_{j=1}^{m} z_j \tag{5.6}$$

z_j is the ℓ_2-norm of layer j.

$$z_j = \sqrt{\sum_{i=1}^{n} x_{E_j}^2} \tag{5.7}$$

f can be divided into a part that is dependent on a specific x_i that is in energy layer E_j and a part that is not.

$$f = z_j(x_i) + g(x_k) \tag{5.8}$$

Where $g(x_k)$ is a function on all x_k's that are not in the same energy layer as x_i, $x_k \notin E_j$. The gradient of the i'th element that is in the j'th energy layer is the derivative of $z_j(x_i)$.

$$\frac{\partial f}{\partial x_i} = \frac{x_i}{z_j} \tag{5.9}$$

The Hessian has the following elements $h_{i,k}$ where $i,k \in \{1,\dots n\}$ with n the number of beamlets and $j \in \{1,2,\dots m\}$ with m the number of energy layers.

$$h_{i,k} = \begin{cases} -\frac{x_i^2}{z_j^3} + \frac{1}{z_j} & i = k \\ -\frac{x_i x_k}{z_j^3} & x_i, x_k \in E_j \\ 0 & \text{otherwise} \end{cases} \tag{5.10}$$

Using the script that is given in Appendix A, the function value, gradient an Hessian matrix can be calculated with matrix B. $b_{j,i}$, where $i \in \{1,2,\dots,n\}$ with n the number of beamlets and $j \in \{1,2,\dots,m\}$ with m the number of energy layers, is an element of B.

$$b_{j,i} = \begin{cases} 1 & x_i \in E_j \\ 0 & \text{otherwise} \end{cases} \tag{5.11}$$

B is $m \times n$ matrix, where m indicates the number of energy layers and n the total number of beamlets. Each $b_{j,i}$ indicates whether beamlet x_i with $i \in \{1,2,\dots,n\}$ is in energy layer E_j with $j \in \{1,2,\dots,m\}$.

5.5. Selection of beamlets and energy layers

The plan that is generated by the weighted sum with either the ℓ_1-norm and/or ℓ_1/ℓ_2-norm contains all of the beamlets and energy layers that were used in the input plan. Depending on the weight of the norm, a number of the numerical values of the beamlets or energy layers will be small or zero. Selecting all beamlets and energy layers that have a large numerical value (or deleting the lower ones), hopefully a new clinically acceptable plan can be made with fewer beamlets and energy layers.

5.5.1. Selection of beamlets

All beamlets $x_i \geq b$ will be selected, where b is the lower bound. In the results there will be looked at different values of b for different weights for the sparsity terms. Deleting $x_i < b$ will delete two kinds of beamlets; beamlets with a numeric value that is zero or that is non-zero.

The deletion of the beamlets with zero contribution can be done without any problem if the plan of the weighted sum is clinically acceptable. In that case the resulting x_i's still satisfy the constraints and a plan without the norm will result in an equivalent or better plan. In the final optimisation, using the ϵ-constraint optimisation, the optimisation is performed on all the beamlets and energy layers that were selected without the extra term. Thus the DVH might result in a better plan, as the minimisation of the norm(s) is not one of the objectives anymore. The minimisation of the norms caused that different trade-offs were made.

However, after deleting the beamlets with a non-zero numeric value the intermediate plan is insufficient to satisfy the minimal dose in the tumour and the numeric value of other beamlets need to change. To explain this consider the following case.

> Suppose that the dose in the tumour is sufficient after the deletion of beamlets and the numerical value of the beamlets is non-zero. Then, the deletion will only influence the maximum doses in organs and in the tumour. However the dose in all organs and in the tumour is minimised. Thus these beamlets would have been zero or the dose in the tumour is not sufficient anymore.

As a result of an insufficient dose, the numerical values of other beamlets need to increase. This will cause the DVH to change negatively for the healthy organs.

The lower the numerical value of the beamlets, the less the numerical values of other beamlets need to change. Thus ideally the beamlets with a low numerical value are deleted given that the DVH is clinically acceptable.

5.5.2. Selection of energy layers

The same thought process and selection process as in the previous subsection can be used with energy layers instead of beamlets.
Vector $\mathbf{y} = B\mathbf{x}$ will be introduced, the vector of y_j. The numerical value of $y_j = \sum_{x_i \in E_j} x_i$ is the sum of all x_i in the j'th energy layer. B is the matrix that is defined in Section 5.4.2. Then, the energy layer can be deleted in full if y_j is small.
If the goal is to have a plan with m' energy layers, the m' energy layers with the largest value for y_j will be selected en the other layers deleted.

6

Results

In this chapter the intermediate results and output plans after spot and/or energy layer reduction using different sparsity-inducing terms will be discussed. In all cases the goal of the output plans is to approximate the reference plan in number of beamlets, energy layers and in the 2D dose distribution.
First the results of the ℓ_1-norm will be analysed, followed by the mixed ℓ_1/ℓ_2-norm. Finally, the results of the addition of both of these norms will be shown. The weighted sum optimisation was performed with the weights given in Table 6.1.

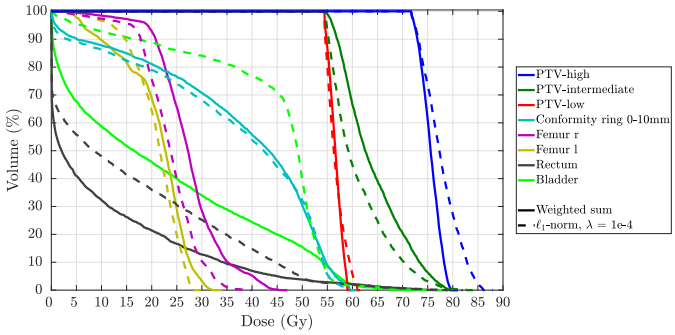
Table 6.1: Weights for all the objectives in the weighted sum. PTV is planning targeting volume and Gp is gigaprotons

Objectives	Name	Type	Weight
f_1	PTV-high	Maximum	3.30e-02
f_2	PTV-intermediate	Maximum	3.60e-03
f_3	PTV-low	Maximum	3.75e-01
f_4	Conformity ring PTV-high	Maximum	1.58e-02
f_5	Conformity ring PTV 0-10mm	Maximum	6.18e-02
f_6	Conformity ring PTV 10-15mm	Maximum	1.82e-03
f_7	Femural heads	Maximum	4.08e-06-4.73e-03
f_8	Rectum	Mean	1.42e-01
f_9	Large and small intestines	Mean	3.45e-08-2.13e-01
f_{10}	Bladder	Mean	9.89e-02
f_{11}	Femural heads	Mean	1.93e-02-2.09e-02
f_{12}	All conformity rings	Mean	3.96e-08-2.03e-03
f_{13}	Rest of the conformity rings	Maximum	1.75e-03-2.81e-03
f_{14}	Total spot weight	Mean	2.26e-07

6.1. ℓ_1-norm

To examine the effect of the ℓ_1-norm, the intermediate solutions using different weights for the sparsity inducing term will be compared to the result of the weighted sum without an added sparsity term. In this section addition of the ℓ_1-norm with weights λ = 1e-4, λ = 1e-6, λ = 1e-8 and λ = 1e-10 will be compared. The DVHs are given in Figure 6.1.

(a)



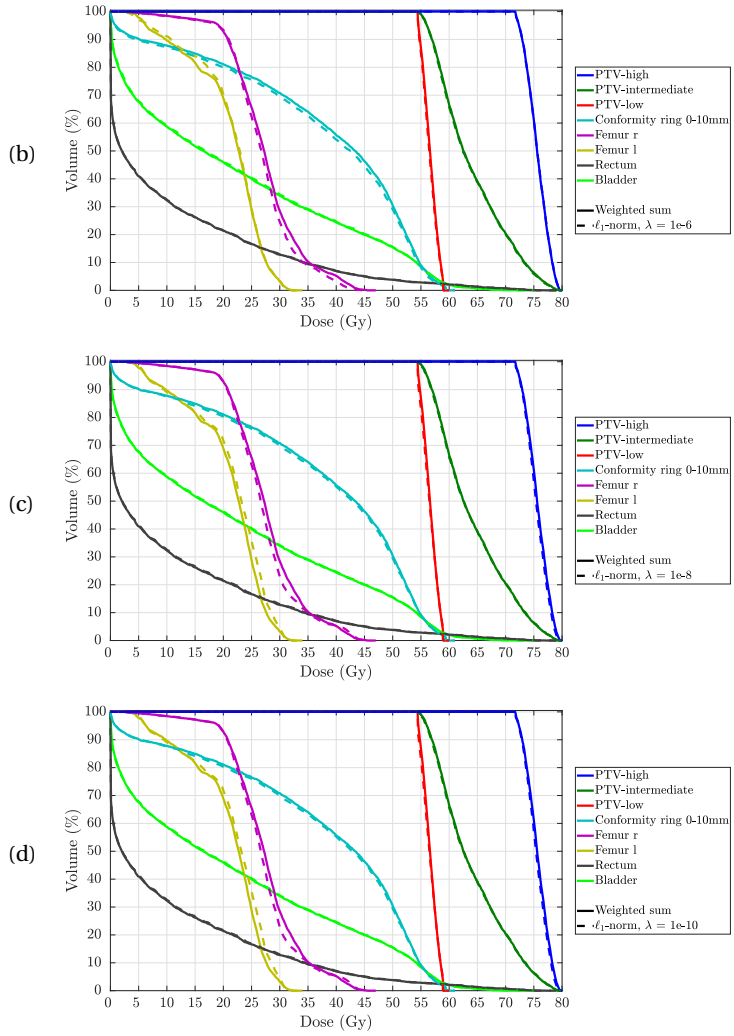


Figure 6.1: DVHs of the solution of the weighted sum without a sparsity term and of the intermediate result of the weighted sum method with addition of an ℓ_1-norm with weight λ. (a) λ = 1e-4, (b) λ = 1e-6, (c) λ = 1e-8 and (d) λ = 1e-10.

In the intermediate result with weight λ = 1e-4 large differences in the DVH can be seen. Adding the norm with this weight a different trade-off is made, as the minimisation of the ℓ_1-norm weighed against minimising some maximum doses, which were not hard constraints in this formulation. This happens if the weight of the norm is higher than the weights of the objectives of the weighted sum. This, however, does not imply that the solution after the reduction of the beamlets and re-optimisation is not acceptable. But, it might also be possible

that the norm with a higher weight λ caused the wrong beamlets to be selected. The intermediate plans with weights λ = 1e-6, λ = 1e-8 and λ = 1e-10 have very similar DVHs as the weighted sum. This is not surprising, as the weights λ = 1e-8 and λ = 1e-10 are lower than the lowest weight of the objectives and λ = 1e-6 is smaller than most weights in the weighted sum. As weight λ = 1e-6 is larger than some of the objectives it was expected that the dose distribution changed a bit more. The only noticeable difference is the change of dose in the femurs.
Adding the sparsity term with one of these weights, the optimisation can be performed with enough freedom to minimise the maximum doses whilst trying to make the ℓ_1-norm as low as possible.

6.1.1. Sparsity of the intermediate solutions

Now remains the question if the intermediate solutions are sparser than the weighted sum. In the table on the next page the number of beamlets that have a higher numeric value than b = 0.01, 0.1, 1 and 10, is shown, that is $|\{x_i : x_i \geq b\}|$. The lower the numbers, the sparser the solution. Ideally $|\{x_i : x_i \geq b\}|$ is small when b is small and the DVH is almost identical. Then, all these beamlets can be deleted and since the beamlets already have a low value, the DVH will not change a lot.

Table 6.2: Indicates the number of beamlets that have a numerical value higher than b = 0.01, 0.1, 1 and 10, $|x_i \geq b|$. The total number of beamlets is 18777. λ indicates the weight, that is used to add the ℓ_1-norm to the weighted sum. "Weighted sum" and "ϵ-constraint" are plans without a sparsity term calculated using a weighted sum and the sequential ϵ-constraint optimisation.

	$b = 0.01$	$b = 0.1$	$b = 1$	$b = 10$
λ = 1e-4	18757	5512	2363	1271
λ = 1e-6	18777	16973	11193	2874
λ = 1e-8	11868	5375	2149	1771
λ = 1e-10	12137	5915	2228	1775
weighted sum	18777	17328	11850	4178
ϵ-constraint	10698	3931	2042	1758

As can be seen in Table 6.2, the ℓ_1-norm does indeed increase the sparsity in the solution as all values of the plans with a sparsity term are lower than or close to the values of the weighted sum without an added ℓ_1-norm. It appears that the solution of the ϵ-constraint optimisation is already quite sparse in comparison to the other plans, especially in comparison to the weighted sum method. By optimising over a different set of beamlets it appeared that this does not always

need to be the case.
As expected, the sparsity increases for weights λ = 1e-10 and λ = 1e-8 whilst keeping the DVH almost similar. Then for λ = 1e-6 the solution is less sparse in comparison to the plans with a lower weight, which is not what was expected. Then the sparsity is again increased for λ = 1e-4. In this case, the number of beamlets that is larger than $b = 0.01$, thus $|x_i \geq 0.01|$, is high, whereas the number of beamlets that is larger than $b = 10$ is lower than in every other method. Thus a lot more beamlets have a value in the interval $0.01 \leq x_i < 10$. It can be concluded that in order to reduce some beamlets, other beamlets cannot be reduced to almost zero. This is different from with weights λ = 1e-8 and λ = 1e-10, where a lot more beamlets could be reduced to a value of $x_i < 0.01$.
As a large number of beamlets have a low value in the intermediate results of λ = 1e-8, λ = 1e-10 and the ϵ-constraint optimisation and the DVHs are similar, these plans will very likely result in an acceptable output plan. λ = 1e-4 does have a large number of beamlets that have a value $x_i < 1$ as well, however the corresponding DVH is not acceptable and it is not certain that re-optimisation will return an acceptable plan.

6.1.2. Spot reduction to 1705 beamlets

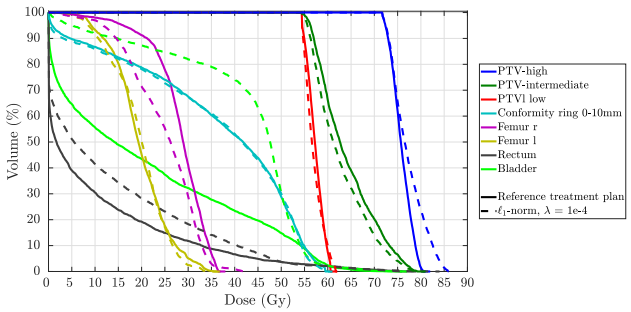
Now the results of the second sequential ϵ-constraint optimisation with 1705 beamlets is shown. The beamlets are selected using the method described in Section 5.5 using different values of b for each weight λ. By comparing the output plan and the reference plan both with the same number of beamlets, the quality of the norms with different weights can be discussed. Besides that the different values for b, that are chosen for the same reduction, can be compared.

Table 6.3: b, the lower bound, for each intermediate plan such that $|\{x_i : x_i \geq b\}| = 1705$. b is rounded off to one decimal digit. λ indicates the weight, that is used to add the ℓ_1-norm to the weighted sum. "Weighted sum" and "ϵ-constraint" are plans without a sparsity term calculated using a weighted sum and the sequential ϵ-constraint optimisation.

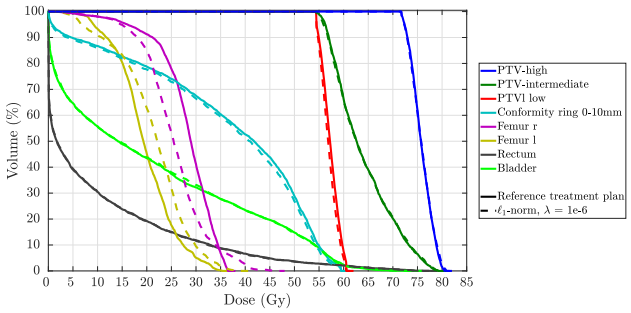
	b
λ = 1e-4	2.3
λ = 1e-6	47.7
λ = 1e-8	22.0
λ = 1e-10	21.6
weighted sum	60.6
ϵ-constraint	17.0

b for λ = 1e-4 is very low. This means that all beamlets that are deleted already had a small numeric value. Due to the high weight of the norm present in the weighted sum, the quality of the plan calculated with ϵ-constraint optimisation might increase. This is different from $b = 60.6$ in the weighted sum and $b = 47.7$ in the addition with weight λ = 1e-10, where b is quite large and the beamlets are selected more randomly. Then, the numeric values of the other beamlets need to change more and it is not certain this will result in an acceptable output plan. The DVHs of the plans with 1705 beamlets in comparison to the reference plan are given in the figures below.

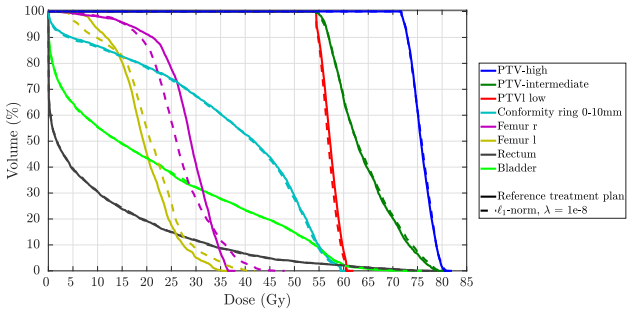
(a)



(b)



(c)



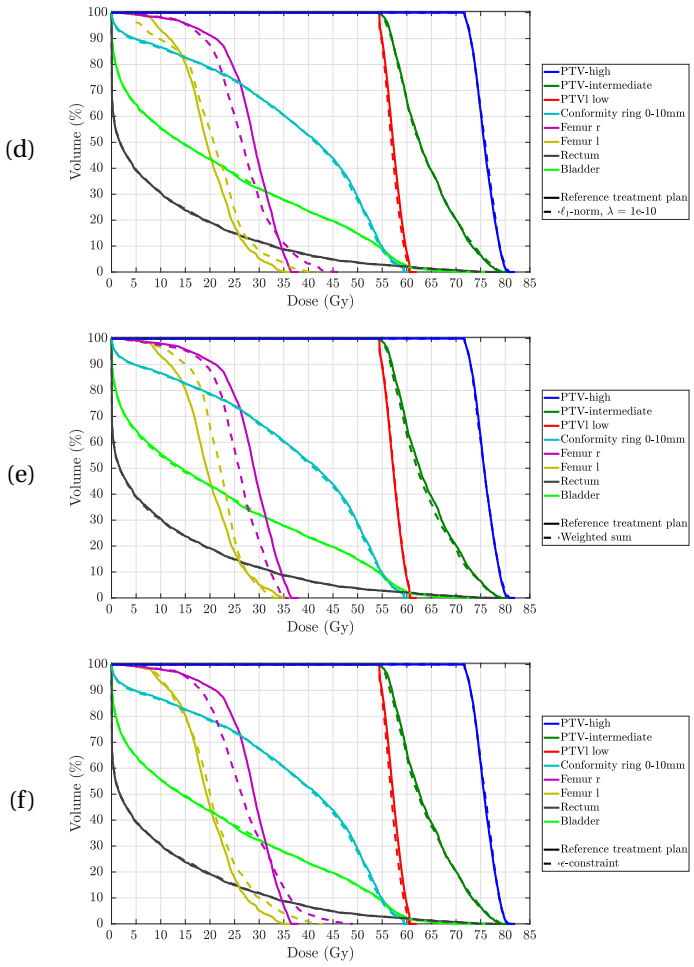


Figure 6.2: Solution of sequential ϵ-constraint optimisation using 1705 beamlets in comparison with the reference plan. The selection of beamlets was performed on the solution of the weighted sum with addition of an ℓ_1-norm with weight λ. (a) λ = 1e-4, (b) λ = 1e-6, (c) λ = 1e-8 and (d) λ = 1e-10, (e) the weighted sum without an extra term and (f) the sequential ϵ-constraint optimisation of all 18777 beamlets.

The plan which was calculated after the selection based on the ℓ_1-norm with weight λ = 1e-4 differs from the reference plan quite a lot. The plan did not get better after the re-optimisation without the sparsity-inducing term. Looking at the DVH, the selection from the intermediate plan with λ = 1e-4 did not result in an admirable selection as the maximum doses increased. The ℓ_1-norm with a higher weight caused the wrong choices in the selection of beamlets to be made as suggested in the previous section.

The rest of the DVHs are nearly identical to the reference plan. In almost all plans the maximum dose of the left thighbone (Femur l) is increased. Besides that, there is a trade-off between the two thighbones (Femur l and Femur r).
As the DVHs are nearly identical, it can be concluded that is possible to achieve a similar plan only using the result of the sequential ϵ-constraint optimisation or the weighted-sum with or without the addition of an ℓ_1-norm. But, it might be a lucky case that the solution of the ϵ-constraint optimisation is sparse.
In the table below the number of similar selected beamlets of two intermediate plans is shown.

Table 6.4: Number of similar beamlets, that were selected from the intermediate plans. λ indicates the weight, that is used to add the ℓ_1-norm to the weighted sum. "WS" and "ϵ-constr" are plans without a sparsity term calculated using a weighted sum and the sequential ϵ-constraint optimisation.

	λ = 1e-4	λ = 1e-6	λ = 1e-8	λ = 1e-10	WS	ϵ-constr
λ = 1e-4		388	316	312	342	308
λ = 1e-6	388		1313	1307	1376	1282
λ = 1e-8	316	1313		1683	1468	1542
λ = 1e-10	312	1307	1683		1469	1546
WS	342	1376	1468	1469		1426
ϵ-const	308	1282	1542	1546	1426	

As suggested before, the plan with λ = 1e-4 selected very different beamlets than the rest of the plans. Only 300-400 beamlets out of the 1705 beamlets were similar as the selected beamlets of other plans. On the other hand, the beamlets that were selected from the plan with λ = 1e-8 and λ = 1e-10 are very similar and only 32 beamlets were not the same.
As the output plan with λ = 1e-4 is significantly worse than the other plans and is not clinically acceptable, this plan will not be considered in further reduction of the beamlets.

6.1.3. Further reduction of beamlets

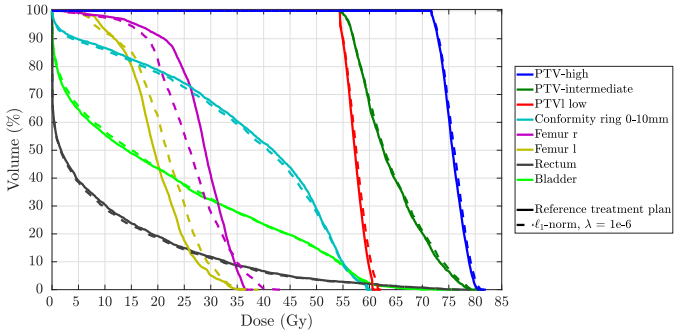
It appears that a good plan with 1705 beamlets can be made using different intermediate plans even though according to Table 6.4, the beamlets that are selected are not the same. Thus it is interesting to reduce the beamlets even more and see if any differences between the plans arise. In this subsection the number of beamlets is reduced to 1260 beamlets. All beamlets larger than b are selected with b in Table 6.5.

Table 6.5: b for each intermediate plan such that $|\{x_i : x_i \geq b\}| = 1260$ rounded off to one decimal.

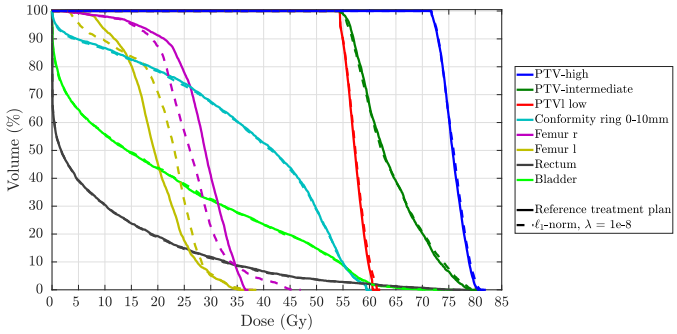
	b
ℓ_1, λ = 1e-6	121.0
ℓ_1, λ = 1e-8	149.9
ℓ_1, λ = 1e-10	145.5
weighted sum	126.0
ϵ-constraint	149.0

As all b's are quite large, deviations from the reference plan are expected.
In the following figures, the DVHs are shown of the plans generated by 1260 beamlets.

(a)



(b)



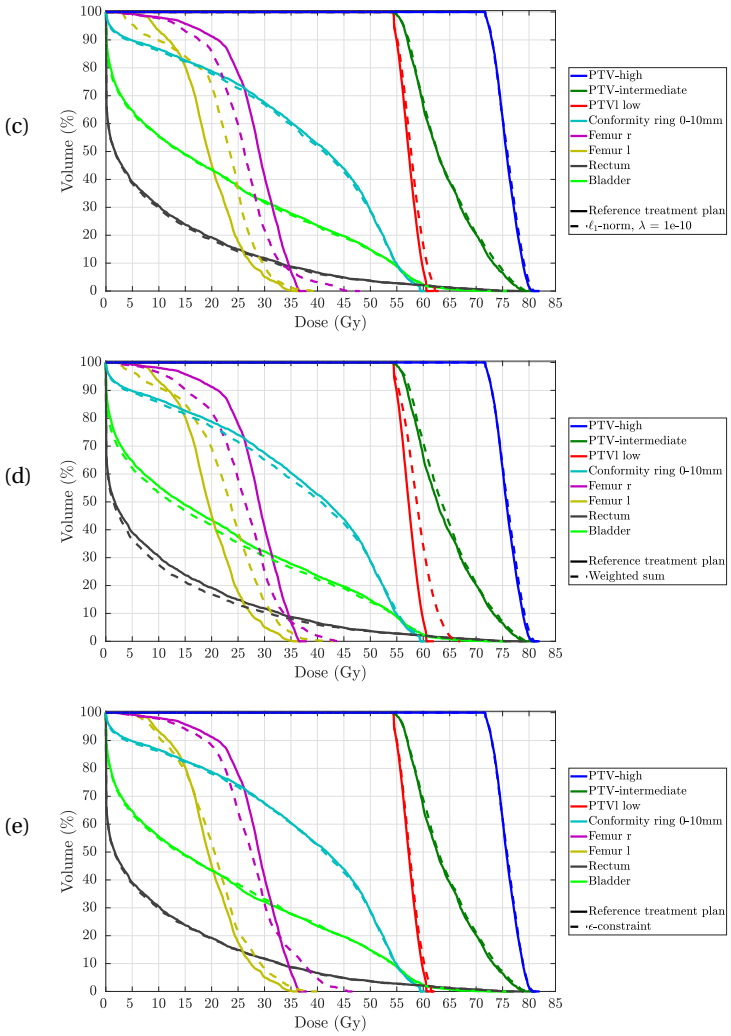


Figure 6.3: Solution of sequential ϵ-constraint optimisation using 1260 beamlets in comparison with the reference plan with 1705 beamlets. The selection of beamlets was performed on the solution of the weighted sum with addition of an ℓ_1-norm with weight λ. (a) λ = 1e-6, (b) λ = 1e-8 and (c) λ = 1e-10, (d) the weighted sum without an extra term and (e) the sequential ϵ-constraint optimisation of all 18777 beamlets.

Even with 1260 beamlets, the DVHs are quite similar. A difference in maximum dose appears in the DVH of the weighted sum and a small difference in the other DVHs. The weighted sum does not have enough freedom to keep the dose in PTV-low at 60 Gy whilst satisfying the dose constraints in the tumour. In all cases further reduction of the beamlets increases the maximum doses of PTV-low and

eventually all other volumes. This is shown in the next figure.

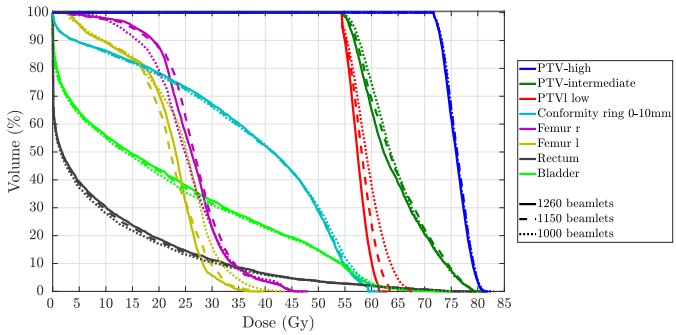


Figure 6.4: DVHs of the ϵ-constraint optimisation with 1260, 1150 and 1000 beamlets. The beamlets are selected from the result of the weighted sum with an ℓ_1-norm with weight λ =1e-8.

Even though b is larger for λ = 1e-8 than λ = 1e-10 and the weighted sum, this does not result in a worse plan. As stated before, the numerical value of some beamlets need to increase in order to decrease other beamlets. In case of weight λ = 1e-8, more beamlets have a smaller numerical value and therefore the numerical value of other beamlets have already changed. This way, the re-optimisation with the selected beamlets will result in a better plan. The next example will clarify this.

> Consider beamlets x_1, x_2 and x_3 that have numerical values $x_1 = 4$, $x_2 = 6$ and $x_3 = 12$. Due to the sparsity term with a higher weight the new values are $x_1 = 0$, $x_2 = 13$, $x_3 = 9$ and with the lower weight $x_1 = 3$, $x_2 = 7$ and $x_3 = 10$. Then assume $b = 10$ for the higher weight and $b = 8$ for the lower weight. In the first case, x_1 and x_3 are deleted, whereas in the second case x_1 and x_2 are deleted. In both cases x_2 needed to increase to allow the other beamlets to become smaller. In the second case however the weight was not high enough to make $x_3 < x_2$. As a result when two x_i's are deleted, x_2 is deleted instead of x_3 in the case of the lower weight.

6.2. ℓ_1/ℓ_2-norm

Another goal is to reduce the number of energy layers that are used in the treatment of the patient. To reduce the number of energy layers it was suggested to add an ℓ_1/ℓ_2-norm to the weighted sum. The norm is added with weights λ = 1e-4, λ = 1e-6, λ = 1e-8. The solutions of these optimisations are compared to the

solution of the weighted sum without the added norm. In the previous section it was concluded that the solution of the sequential ϵ-constraint optimisation was a good plan to reduce the spots if the plan is already sparse. This might also be the case with energy layers. Therefore this plan will also be used to reduce the energy layers.

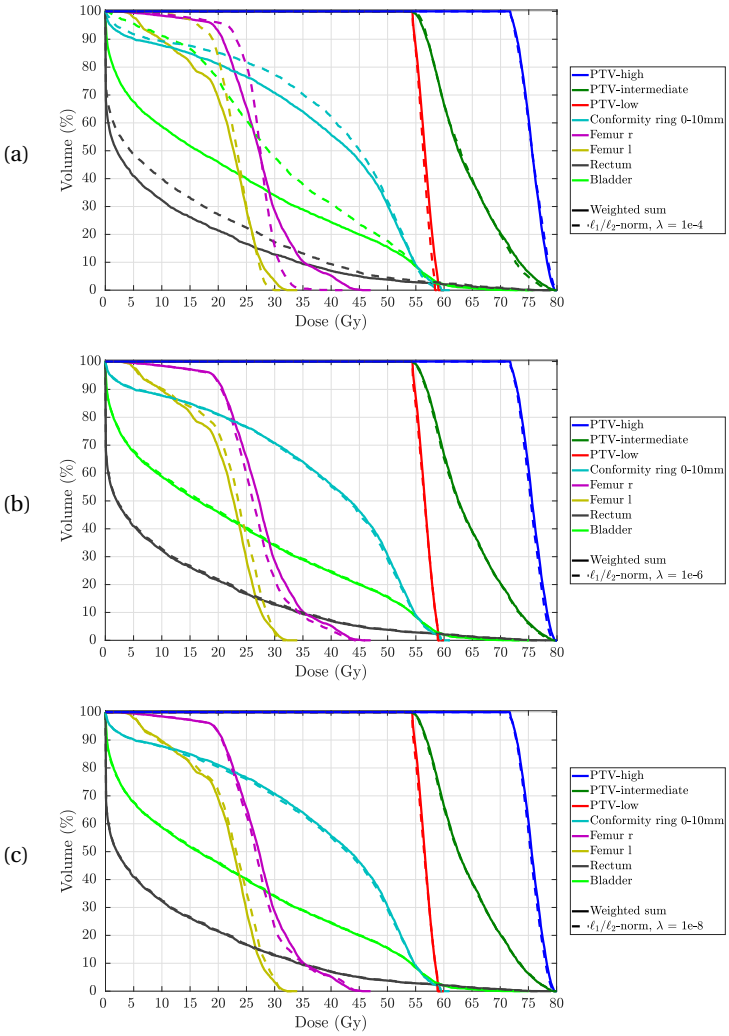


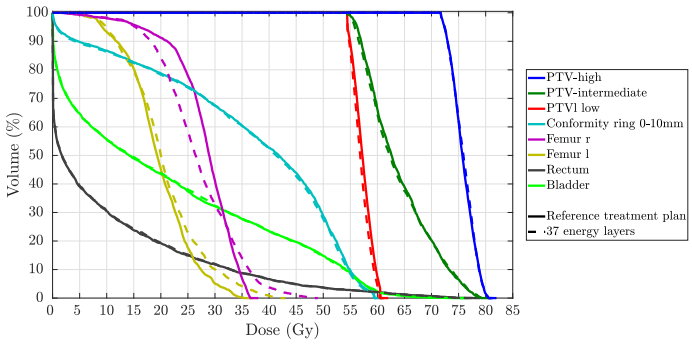
Figure 6.5: DVHs of solution of the weighted sum without a sparsity term and of the intermediate result of the weighted sum method with addition of an ℓ_1/ℓ_2-norm with weight λ. (a) λ = 1e-4, (b) λ = 1e-6 and (c) λ = 1e-8.

Just like the ℓ_1-norm, the plan with the addition with weight λ = 1e-4 has the largest difference in the DVH with the weighted sum without the sparsity term. The other plans look quite similar to the weighted sum without the added ℓ_1/ℓ_2-norm

6.2.1. Selection of energy layers

Using matrix B, introduced in Section 5.4.2, vector $\mathbf{y}$ is calculated. Recall that $\mathbf{y}$ is the vector of all $y_j = \sum_{x_i \in E_j} x_i$, the sum of all x_i in every energy layer. The values for y_j of the solutions of the intermediate results are given in Appendix B. These values are sorted from lowest to highest for each intermediate plan. In the table an interesting result can be seen. The order of the first eight layers is the same for all plans and the values of the solution with weight λ = 1e-8 is quite similar to the solution of the ϵ-constraint optimisation. In the layers 9 to 12 a minor change in order can be observed. The only large difference can be seen in layer 39. That layer became very large in the weighted sum with λ = 1e-4.
As the order of the energy layers is the same for most layers, deleting an arbitrary number of layers up to the eighth layer or all the lowest twelve layers without deleting spots will result in the same DVH. In the figures below the DVHs with the highest 37, 32 and 28 energy layers will be compared with the reference plan.

(a)



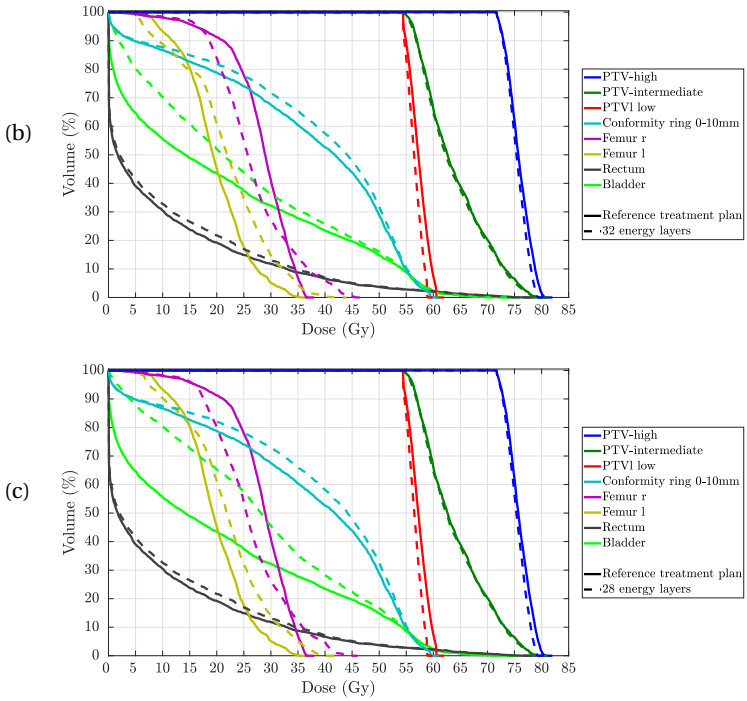


Figure 6.6: DVHs with the reference plan and plans with (a) 37, (b) 32 and (c) 28 energy layers generated by sequential ϵ-constraint optimisation.

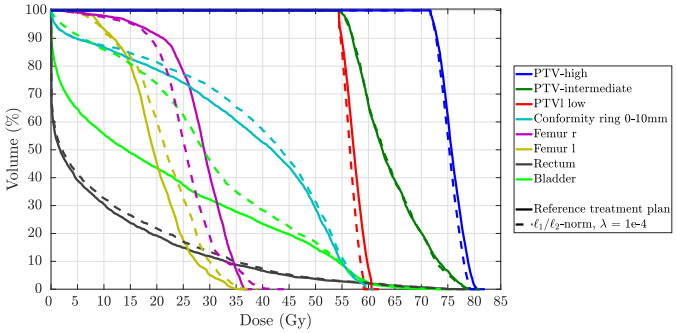
The plan with 37 energy layers is very similar with the reference plan. Deleting more than these three layers does not change the DVH as much, only the dose in both the femurs is changed and the dose in the bladder increases as more energy layers are deleted. In all plans, more than 15000 beamlets are used in the optimisation, thus having more than enough freedom to make a plan. 19 or even more more layers can be deleted, when there is no selection on beamlets.

6.2.2. Beamlet and energy layer reduction

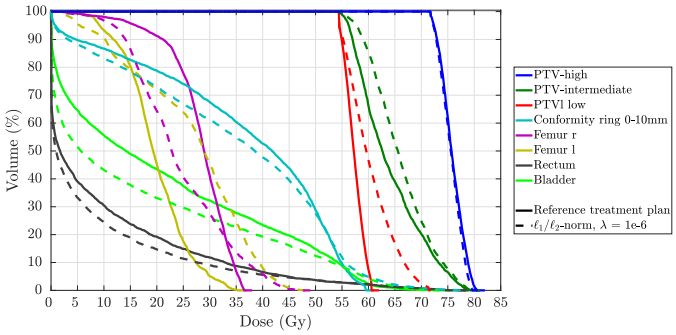
As the plans in the previous section have a large number of beamlets, the number of beamlets will be reduced as well. In Section 6.1 it was shown that the ℓ_1-norm is not always necessary to reduce the number of beamlets. It was already covered that the deletion of the first 8 layers will be similar if the deletion was performed on the input result or the intermediate result with any ℓ_1/ℓ_2-norm. This is useful if the outcome plans after deletion of spots is compared, as the deletion of different energy layers cannot be the cause of different DVHs.

In Figure 6.7 the DVHs of plans with 5000 beamlets and 32 energy layers will be compared with the reference plan.
In contrast to the ℓ_1-norm, the norm with weight λ = 1e-4 results in the best plan and all other plans have large deviations. This might be the result of the shift in numerical values of the beamlets in this plan. As the weight is higher, a larger number of numerical values of the beamlets became smaller. As a result the beamlets that should take over the contribution of the beamlets that are in the layers that will be deleted, are larger and thus not deleted.

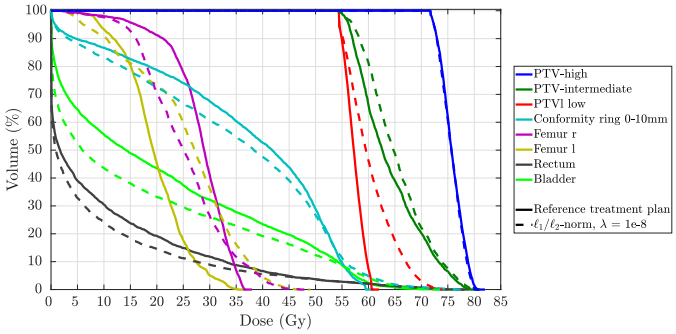
(a)



(b)



(c)



(d)

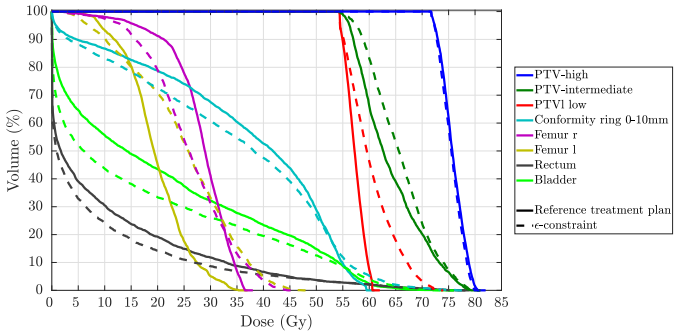


Figure 6.7: Solution of sequential ϵ-constraint optimisation with 5000 beamlets and 32 energy layers in comparison with the reference plan with 1705 beamlets and 37 energy layers. The selection of beamlets was performed on the solution of the weighted sum with addition of an ℓ_1/ℓ_2-norm with weight λ. (a) λ = 1e-4 , (b) λ = 1e-6 and (c) λ = 1e-8 and (d) the sequential ϵ-constraint optimisation of all 18777 beamlets.

6.3. ℓ_1-norm and ℓ_1/ℓ_2-norm

The ℓ_1-norm is a useful tool to reduce beamlets and the ℓ_1/ℓ_2-norm to select energy layers. However if both reductions are wanted, these norms individually are not satisfactory. In this section an improvement will be made on the plan with the selection on the intermediate plan of the ℓ_1/ℓ_2-norm with weight λ = 1e-4.

6.3.1. Improvement on Figure 6.4

In Section 6.2.2 a plan was shown with 5000 beamlets in 32 energy layers. In this section the ℓ_1-norm and ℓ_1/ℓ_2-norm will be used to find an equivalent plan with less beamlets.

For the removal of beamlets, an ℓ_1-norm with weights λ = 1e-6 and λ = 1e-8 seem to give a good result. For the selection of energy layers it is better to use a higher weight on the ℓ_1/ℓ_2-norm, for example λ = 1e-4. Therefore a selection was made from the intermediate result with λ = 1e-8 for the ℓ_1-norm and λ = 1e-4 for the grouped norm.

The addition of the ℓ_1-norm with λ = 1e-8 to the ℓ_1/ℓ_2-norm with weight λ = 1e-4 did however not change the number of beamlets that were needed for the optimisation. Looking at the beamlets that were selected of the intermediate plans it could be concluded that almost 100 percent of the beamlets that were selected were similar. Then λ = 1e-6 was used for the ℓ_1-norm. This did not improve the DVH as well. In this case 97 percent of the beamlets were similar to

the selection from the intermediate plan of the ℓ_1/ℓ_2-norm alone.
A smaller weight than λ = 1e-4 for the ℓ_1/ℓ_2-norm in combination with the ℓ_1-norm with weight λ = 1e-8 or λ = 1e-6 gave worse results. Just like in the addition of the ℓ_1/ℓ_2-norm alone, the numeric values of the beamlets in a specific energy layer did not change enough that a good selection was made (see Section 6.2.2). Using a weight for the ℓ_1/ℓ_2-norm that is higher than λ = 1e-4 was to aggressive on the energy layers. For example with λ = 1e-3, if 5000 beamlets are selected, these beamlets were only in 17 different layers.
It seems as if the addition of a higher weight for the ℓ_1-norm is better to influence the selection of beamlets when the ℓ_1/ℓ_2-norm is also present. Using a weight for the ℓ_1-norm that was bigger than λ = 1e-6 did give some interesting results, for example in the case of λ = 1e-5. In that case, the number of beamlets that were small increased a lot and due to the ℓ_1/ℓ_2-norm the numbers of layers that had beamlets with a small numeric value increased even more. This resulted in the fact that, for example, the selection 4000 beamlets immediately included that only 30 energy layers were used instead of 37 energy layers with the ℓ_1-norm alone or 34 with the ℓ_1/ℓ_2-norm alone. The DVH of the output plan is given below in comparison to both the reference plan and the output plan of the selection on the ℓ_1/ℓ_2-norm with weight λ = 1e-4. In order to keep the comparison fair, the plan is compared with a plan that contains 4000 beamlets in 30 energy layers.

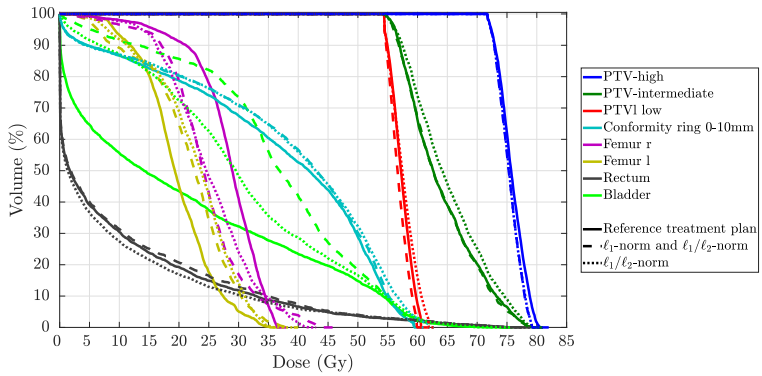


Figure 6.8: DVH of the reference plan, selection of the intermediate result with an ℓ_1/ℓ_2-norm, weight λ = 1e-4, and an ℓ_1-norm, weight λ = 1e-5 and selection from from the intermediate result with an ℓ_1/ℓ_2-norm, weight λ = 1e-4. The plans from the intermediate results have 4000 beamlets in 30 energy layers.

The result is better than the result of only the ℓ_1/ℓ_2-norm in the doses of the

PTVs. However, the dose in the bladder increased in both plans a lot in comparison with the reference plan. The addition of both norms made a trade-off between the dose in the PTVs and the bladder in order to reduce the numeric values in the beamlets.
Thus the weights of the norms should be closer to each other to have an effect and not too high such that too many energy layers are deleted.

6.3.2. Difference in number of beamlets

λ = 1e-5 for the ℓ_1-norm and λ = 5e-5 for the ℓ_1/ℓ_2-norm are chosen. The DVHs of the output plans will be compared to both the reference plan and the output plan of the selection on the ℓ_1/ℓ_2-norm with weight λ = 1e-4.

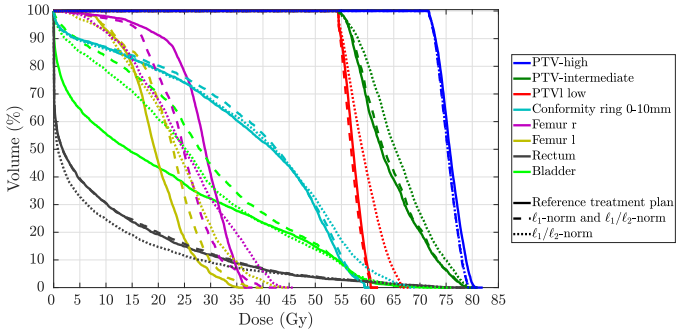


Figure 6.9: DVH of the reference plan, selection of the intermediate result with an ℓ_1/ℓ_2-norm, weight λ = 5e-5, and an ℓ_1-norm, weight λ = 1e-5 and selection from from the intermediate result with an ℓ_1/ℓ_2-norm, weight λ = 1e-4. The plans from the intermediate results have 3500 beamlets in 32 energy layers.

The result is better in almost organs and PTVs. The dose in the bladder is slightly increased, but the maximum dose is the same. Further reduction of beamlets while keeping the number of energy layers the same was not possible. In the following figure the DVH of the output plan with the ℓ_1/ℓ_2-norm and both norms are shown containing 32 energy layers and a different number of beamlets, so that the DVHs are quite similar.
The DVHs in the figure are quite similar even though the plan with both norms contains 3000 beamlets and the plan with only the ℓ_1/ℓ_2-norm 4000. Thus the combination of both norms indeed increases the ability to reduce the beamlets even further given a number of energy layers.
A similar result can be seen with 21 energy layers.
Now, the plan with the selection on the intermediate plan with the ℓ_1/ℓ_2-norm

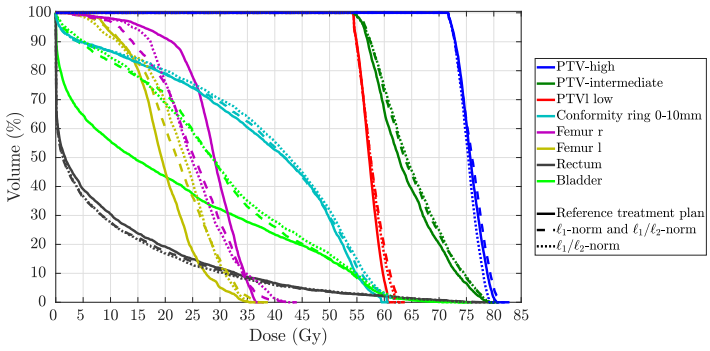


Figure 6.10: ϵ-constraint optimisation of different selections of beamlets and the reference plan with 1705 beamlets in 37 energy layers. Selection of 3000 beamlets in 32 energy layers from the intermediate result with an ℓ_1/ℓ_2-norm, weight λ = 5e-5, and an ℓ_1-norm, weight λ = 1e-5. Selection of 4000 beamlets in 32 energy layers from the intermediate result with an ℓ_1/ℓ_2-norm, weight λ = 1e-4.

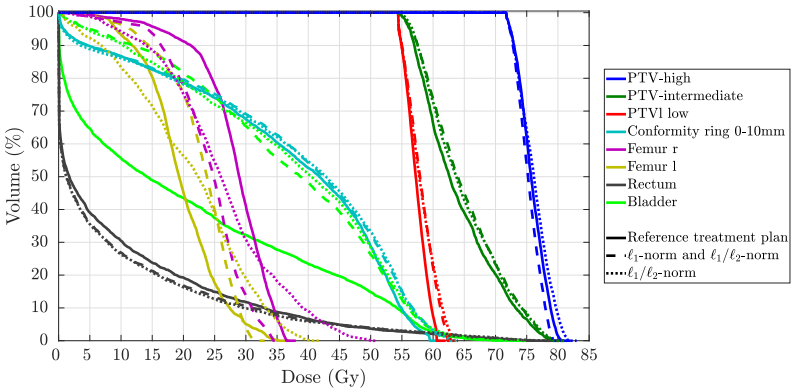


Figure 6.11: ϵ-constraint optimisation of different selections of beamlets and the reference plan with 1705 beamlets in 37 energy layers. Selection of 3500 beamlets in 21 energy layers from the intermediate result with an ℓ_1/ℓ_2-norm, weight λ = 5e-5, and an ℓ_1-norm, weight λ = 1e-5. Selection of 4500 beamlets in 21 energy layers from the intermediate result with an ℓ_1/ℓ_2-norm, weight λ = 1e-4.

contains 4500 beamlets and the plan with the selection on both norms 3500 beamlets. In both cases the addition of both norms reduced the number of beamlets with 1000 beamlets.

From these figures it can be concluded, that there is a trade-off between the reduction of energy layers and beamlets. The more energy layers that are deleted, the more beamlets should be selected. It would be interesting to see, what combination of number of energy layers and beamlets is the most time efficient.

In all cases, the weights that are chosen might not give the best results, thus it might be possible that better output plans can be achieved by choosing differ-

ent weights. As the combination of the two norms has two weights that can be modified, these results can most likely be improved.

7

Conclusion

The numeric values of the beamlets could be influenced with the ℓ_1-norm, so that after selecting beamlets with the highest values, the DVH of the output plan was similar to that of the reference plan. In this data set the input plan was already quite sparse. Thus the same result could be achieved by selecting beamlets on the solution of the initial ϵ-constraint optimisation. Using the ℓ_1-norm with a weight that is slightly lower than the weights of the objectives, the best result can be achieved with an output plan of 1260 beamlets. If a weight is chosen that is too high, there might be a selection of the wrong beamlets. When the weight is too low, the numeric values are not changed enough, resulting in a more random selection of beamlets.

The ℓ_1/ℓ_2-norm did not change the order of the energy layers in most cases. However it did influence the optimisation in such a way that the numeric values of the beamlets were shifted and, besides energy layers, beamlets could be deleted because of this shift. The addition of both norms resulted in a plan with less energy layers but still a larger number of beamlets than the reference plan, namely 5000 instead of 1705.

With the addition of both norms, the numeric values of the beamlets and energy layers can be reduced. However there is a trade-off between energy layers and beamlets. This resulted in an output plan with less beamlets than with the addition of the ℓ_1/ℓ_2-norm alone.

It might be possible to refine the exact weights to improve the output plans even

more. Furthermore, it can be interesting to investigate the radiation time of all plans so that the trade-off between the number of energy layers and beamlets can be made.

A

Scripts

```
%calculate L2-norm of each group
L2_group     = sqrt(B*(x.*x));                %vector m(groups)x1

%calculate function value, f and gradient, grad
f            = sum(L2_group);
grad         = x./(B'*L2_group);              %vector n(beamlets)x1

% calculate Hessian
hess         = diag(-x.*x);                   %matrix n(beamlets)xn

for i = 1:size(B,1)
      group  = (~any(B(:,1:(end-i))~= B(:, (i+1):end)))';

      if sum(group) == 0                      %then x_i and x_k are
            break                             %not in the same
      end                                     %group anymore

      D      = group.*(-x(1:(end-i)).*x((i+1):end));
      hess   = hess  + diag(D, i) + diag(D, -i);

end
```

```
23 hess  = hess ./(B'*(L2_group.^3)) + diag(grad./x);
```

B

Table of energy layers

Table B.1: λ indicates the weight, that is used to add the ℓ_1/ℓ_2-norm to the weighted sum. "Weighted sum" and "ϵ-constraint" are plans without a sparsity term calculated using a weighted sum and the sequential ϵ-constraint optimisation. The number of "Layer" is the index of the beam. All numbers 1 to 18 are in the first beam and numbers 19 to 40 in the second beam. The lowest value of each beam corresponds to the layer with the lowest energy.

	λ = 1e-4		λ = 1e-6		λ = 1e-8	
#	y_j	Layer	y_j	Layer	y_j	Layer
1	0,0	40	0,1	40	0,1	40
2	2,9	19	4,5	19	1,4	19
3	12,3	1	7,6	1	2,0	1
4	20,1	20	299,0	20	318,6	20
5	167,9	2	707,6	2	669,6	2
6	1315,4	21	2322,9	21	2087,2	21
7	1661,3	22	3226,2	22	3078,8	22
8	2544,5	3	3754,1	3	3385,6	3
9	3394,8	23	5553,6	23	5625,2	24
10	3895,3	4	5848,5	4	5993,6	23
11	4035,1	5	6792,4	24	6254,5	4
12	4531,5	24	7675,0	5	8272,4	5
13	6305,6	25	8794,1	39	10468,1	25
14	7548,8	6	10712,6	25	10651,6	39
15	8152,7	27	10820,8	6	10877,0	6
16	8421,8	26	12464,7	27	12203,7	27
17	10238,3	28	13500,9	26	13015,1	14
18	11518,9	7	13562,5	14	13258,0	26
19	12381,2	29	14676,7	35	14773,5	35
20	13361,6	14	15145,2	28	15080,4	28
21	13580,7	8	17248,0	8	16762,5	18
22	14729,8	39	17395,1	18	16778,3	8
23	14948,8	18	18083,5	7	16860,0	7
24	16895,4	35	18660,6	29	18990,6	29
25	19610,6	30	19696,4	15	19020,7	15
26	22617,7	9	22812,0	30	23111,6	30
27	22866,9	31	24442,6	34	25804,8	34
28	27088,2	34	26780,5	9	27757,4	9
29	28650,4	15	27657,6	31	28188,5	31
30	28999,1	10	31398,2	36	32072,6	10
31	33114,0	13	33013,1	10	32715,4	36
32	35797,1	36	37562,7	13	37840,3	13
33	36916,5	32	39600,5	32	39311,2	32
34	40028,8	11	42813,7	16	40556,2	16
35	42152,8	33	43153,7	33	43921,4	11
36	49224,5	12	43555,0	11	43979,6	33
37	54991,0	38	46865,9	17	45252,7	38
38	57214,0	16	46973,6	38	46615,7	17
39	61394,5	17	54980,2	37	54310,7	37
40	73182,4	37	56759,8	12	57407,3	12

#	λ = 1e-10		Weighted sum		ϵ-constraint	
	y_j	Layer	y_j	Layer	y_j	Layer
1	0,2	40	11,0	40	0,1	40
2	3,0	19	178,0	19	1,4	19
3	4,3	1	285,1	1	1,8	1
4	322,0	20	708,8	20	309,3	20
5	672,8	2	991,7	2	668,2	2
6	2105,1	21	2446,9	21	2106,4	21
7	3036,7	22	3616,5	22	3022,9	22
8	3371,0	3	3790,3	3	3378,8	3
9	5638,3	24	6220,9	23	5666,8	24
10	5996,4	23	6434,3	4	5942,2	23
11	6249,7	4	6485,7	24	6246,5	4
12	8281,8	5	8662,2	5	8203,9	5
13	10224,8	25	10119,4	39	9722,0	39
14	10682,1	39	11182,7	25	10275,7	25
15	10843,2	6	11369,7	6	10796,8	6
16	12053,2	27	12900,4	27	12114,2	27
17	12979,4	14	13430,4	26	13231,9	14
18	13714,5	26	13814,0	14	13579,2	26
19	14833,6	35	15442,0	28	15108,9	28
20	15084,6	28	15571,4	35	16086,2	8
21	16452,4	8	16276,8	18	16305,9	35
22	16738,2	18	17684,5	7	16656,1	7
23	16837,0	7	18329,5	8	17456,0	18
24	19021,8	15	19763,9	29	18926,0	29
25	19025,6	29	19880,1	15	18963,4	15
26	23136,5	30	24446,5	30	23098,3	30
27	25834,2	34	27322,8	34	25265,5	34
28	28168,6	31	28609,8	9	28408,6	31
29	28196,9	9	29151,1	31	28646,6	9
30	32117,6	10	33099,8	36	32370,5	10
31	32827,2	36	34044,3	10	32589,3	36
32	37763,4	13	39194,1	13	37949,9	13
33	39281,2	32	40333,3	16	38962,5	32
34	40580,3	16	41197,1	32	41946,2	16
35	43918,1	11	44027,5	17	43660,6	11
36	43978,0	33	45198,3	38	43711,6	33
37	45257,9	38	45233,3	33	44777,6	38
38	46636,0	17	45402,3	11	46512,8	17
39	54220,7	37	52043,5	37	53525,0	37
40	57458,9	12	58704,7	12	57273,5	12

Bibliography

[1] S.E. McGowan, N.G. Burnet, and A.J. Lomax. Treatment planning optimisation in proton therapy. *The British journal of radiology*, 86(1021):20120288–20120288, 2013.

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

[3] P.M. Putora, M. Früh, and L. Kern. The place of radiotherapy in the palliative management of nsclc. *Breathe*, 8(2):134–143, 2011. ISSN 1810-6838. doi: 10.1183/20734735.012711. URL `https://breathe.ersjournals.com/content/8/2/134`.

[4] External beam radiation, Aug 2017. URL `https://www.cancer.gov/about-cancer/treatment/types/radiation-therapy/external-beam`.

[5] S. Breedveld. Towards Automated Treatment Planning in Radiotherapy: A Mathematical Approach to Automated and Integrated Multi-Criterial Optimization of Beam Angles and IMRT Fluence Profiles. 2013.

[6] T. Jagt, S. Breedveld, R. Van Haveren, B. Heijmen, and M. Hoogeman. An automated planning strategy for near real-time adaptive proton therapy in prostate cancer. *Physics in Medicine & Biology*, 63(13):135017, 2018.

[7] S. Breedveld, D. Craft, R. Van Haveren, and B. Heijmen. Multi-criteria optimisation and decision-making in radiotherapy. *European Journal of Operational Research*, 2018.

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

[9] L. Verhey, H. Blattman, P. M. Deluca, and D. Miller. 4. Proton Interactions with Matter. *Journal of the International Commission on Radiation Units and Measurements*, os30(2):13–14, 04 2016. ISSN 1473-6691. doi: 10.1093/jicru/os30.2.13. URL `https://doi.org/10.1093/jicru/os30.2.13`.

[10] D. Grimes, D.R. Warren, and M. Partridge. An approximate analytical solution of the bethe equation for charged particles in the radiotherapeutic energy range. *Scientific Reports*, 7, 12 2017. doi: 10.1038/s41598-017-10554-0.

[11] R. Leroy, N. Benahmed, F. Hulstaert, N. Van Damme, and D. De Ruysscher. Proton therapy in children: a systematic review of clinical effectiveness in 15 pediatric cancers. *International Journal of Radiation Oncology* Biology* Physics*, 95(1):267–278, 2016.

[12] H. Fuchs, P. Moser, M. Gröschl, and D. Georg. Magnetic field effects on particle beams and their implications for dose calculation in mr-guided particle therapy. *Medical physics*, 44(3):1149–1156, 2017.

[13] W. Gu, D. O'Connor, D. Nguyen, V. Y. Yu, D. Ruan, L. Dong, and K. Sheng. Integrated beam orientation and scanning-spot optimization in intensity-modulated proton therapy for brain and unilateral head and neck tumors. *Medical Physics*, 45(4):1338–1350.

[14] F. Bach, R. Jenatton, and J. Mairal. Optimization with Sparsity-Inducing Penalties. Foundations and Trends in Machine Learning Series. Now Publishers, 2011. ISBN 9781601985101.