Dissertation submitted to the Combined Faculties for the Natural Sciences and for Mathematics of the Ruperto-Carola University of Heidelberg, Germany, for the degree of Doctor of Natural Sciences

Put forward by Mark Bangert, born in Weinheim.
Oral examination: July 20, 2011.

New concepts for beam angle selection in IMRT treatment planning:
From heuristics to combinatorial optimization

Referees

Prof. Dr. Uwe Oelfke
Prof. Dr. Wolfgang Schlegel

Abstract

This thesis investigates beam ensemble selection strategies in intensity-modulated radiation therapy treatment planning. Beam ensemble selection strategies are applied to find the very beam ensembles that meet the treatments' objectives at the best possible rate.

(1) A formal description of the beam ensemble selection problem is presented and the characteristics of the search space is discussed with a focus on its non-convexity and exponential complexity. (2) We review existing approaches to beam ensemble selection and provide a comprehensive overview of the field. (3) Conceptual advancements of beam ensemble selection strategies relying on score functions and geometric considerations are introduced. For photons, we demonstrate a clear benefit regarding organ at risk sparing for asymmetric patient geometries as regularly observed within the abdomen or skull. For protons, phantom studies yield plausible beam configurations. The measures taken to guarantee robustness regarding potential uncertainties are promising but require refinements. (4) The simultaneous optimization of beamlet weights and beam orientations is investigated at a very high precision. We apply different metaheuristics for the combinatorial optimization of beam ensembles and confirm the beneficial performance of genetic algorithms in this context.

Both heuristic selection and combinatorial optimization of beam ensembles may yield extensive benefits for complicated planning cases. In the future it will be critical to transfer automated beam ensemble selection to the clinic for the benefit of the patient.

Zusammenfassung

Die vorliegende Arbeit beschäftigt sich mit Verfahren zur Auswahl von geeigneten Einstrahlrichtungen in der intensitätsmodulierten Strahlentherapie. Diese Verfahren werden eingesetzt, um ein Ensemble von Einstrahlrichtungen zu finden, welches die bestmögliche Realisierung der klinischen Bestrahlungsziele ermöglicht.

(1) Wir geben eine formale, mathematische Beschreibung des Auswahlproblems von optimalen Einstrahlrichtungen und diskutieren die Eigenschaften des Lösungsraums, insbesondere im Hinblick auf Nicht-Konvexität und exponentielle Komplexität. (2) Wir diskutieren bestehende Ansätze im Rahmen einer umfassenden Literaturauswertung. (3) Konzeptionelle Weiterentwicklungen von Strategien zur Auswahl von Einstrahlrichtungen, die auf skalaren Gütefunktionen oder geometrischen Überlegungen basieren, werden vorgestellt. Für Photonen demonstrieren wir deutliche Verbesserungen für asymmetrische Patientengeometrien, wie sie häufig im Abdomen oder Schädel auftreten. Erste Experimente für Protonen liefern plausible Einstrahlrichtungen. Die Ansätze zur Gewährleistung von Robustheit sind vielversprechend, bedürfen aber weiterer Verbesserungen. (4) Wir untersuchen die gleichzeitige Optimierung von Einstrahlrichtungen und zugehörigen Fluenzprofilen mit sehr hoher Genauigkeit. In einem Vergleich verschiedener Metaheuristiken zur kombinatorischen Optimierung zeichnet sich ein genetischer Algorithmus durch die besten Konvergenzeigenschaften und die besten resultierenden Bestrahlungspläne aus.

Sowohl der Einsatz von Heuristiken als auch der Einsatz von kombinatorischen Optimierungsverfahren zur Auswahl von Einstrahlrichtungen kann eine erhebliche Verbesserung komplizierter Bestrahlungspläne ermöglichen. Wir erachten die klinische Implementierung und Evaluierung der entwickelten Konzepte als Hauptaufgabe für die Zukunft.

Contents

About this manuscript

This thesis does not include a general introduction to radiation therapy. A good and free reference may be the Radiation Oncology Physics Handbook published by the International Atomic Energy Agency[1]. We assume that the reader is familiar with external beam radiotherapy, particularly with the concept of intensity-modulation.

Parts of this work have already been published in accordance with the regulations of the Faculty of Physics and Astronomy of Heidelberg University[2]. Section 4.2 is based on a research paper published in Physics in Medicine and Biology (Bangert and Oelfke 2010c), sections 4.3 and 4.5 are based on conference papers for the Ninth International Conference on Machine Learning and Applications (Bangert et al. 2010) and the XVIth International Conference of the Use of Computers in Radiation Therapy (Bangert and Oelfke 2010a), respectively. The results of section 4.4 have been presented during the World Congress on Medical Physics and Biomedical Engineering 2009 (Bangert and Oelfke 2009b) and during the Annual Meetings of the Particle Therapy Co-Operative Group in 2009 and 2010 (Bangert and Oelfke 2009a; 2010b). Parts of chapter 5 will be presented at the Meeting of the American Association of Physicists in Medicine 2011; section 5.1.1 describes joint work with Peter Ziegenhein.

[1] Free download available at http://www-naweb.iaea.org/nahu/dmrp/syllabus.shtm.

[2] Promotionsordnung der Universität Heidelberg für die Naturwissenschaftlich-Mathematische Gesamtfakultät §7
http://www.physik.uni-heidelberg.de/md/physik/studium/Promotion/promotionsordnung_nawi_ma.pdf

Preface

> "If we didn't understand something [on Melmac], we broke it."
>
> - ALF, December 7, 1987 on NBC (episode 11, season 2).

Cancer is a class of disease in which a group of cells displays uncontrolled growth through division beyond normal limits, invasion of adjacent tissues, and sometimes metastasis in other locations of the body. It may emerge if the delicate equilibrium of information processing within a cell, which involves constant signaling of almost all cellular components, gets out of control while intrinsic repair mechanisms fail. Both environmental factors such as smoking, obesity, radiation exposure or infections and genetic predispositions may favor the development of cancer.

The adequate treatment of cancer would be a direct re-establishment of normal cell signaling on a molecular level by highly targeted agents. In the foreseeable future, however, the development of such agents for all cancers is beyond our capabilities. Treatment modalities broadly deploying toxic agents to cancerous tissues, such as radiation therapy, will continue to play an important role in cancer therapy and it is worthwhile to further improve these techniques.

1. Introduction

Radiation therapy is used to treat localized cancers. It may be applied on its own or in combination with surgery and/or chemotherapy. For every patient, the actual treatment is preceded by a complex decision-making process called radiation therapy treatment planning.

1.1. Overview of radiation therapy treatment planning

Radiation therapy treatment planning for external beam radiotherapy is not a sequential and linear process. Some decisions may have implications on subsequent treatment options. The most important steps are:

- Definition of clinical objectives
- Translation of clinical objectives to radiation prescription and fractionation scheme
- Selection of radiation modality and dose delivery mode
- Selection of irradiation directions and optimization of radiation fluences

The clinical objectives are defined by the medical condition of the patient. For a prostate cancer patient, for example, curative radiation therapy is applied to control tumor growth while minimizing adverse side effects in normal tissues such as bowel or urinary dysfunction (Andreyev 2007).

Therapeutic decisions regarding the dose prescription and fractionation scheme are based on results from clinical studies (Bentzen et al. 2010). However, every patient may exhibit a different trade-off between conflicting clinical objectives and consequently the radiation oncologist may have to reconsider on an individual basis.

The same holds - to some extent - for the selection of a suited radiation modality and delivery mode. While very complex tumor geometries in proximity to organs at risk (OAR) might benefit from a full-fledged intensity-modulated particle therapy treatment, convex tumors may be treated effectively with 3D conformal photon radiation therapy (Cozzi et al. 2001). Of course, this decision about radiation modality and delivery mode also underlies ordinary restrictions of the infrastructure: not all radiation oncology centers have access to all technologies.

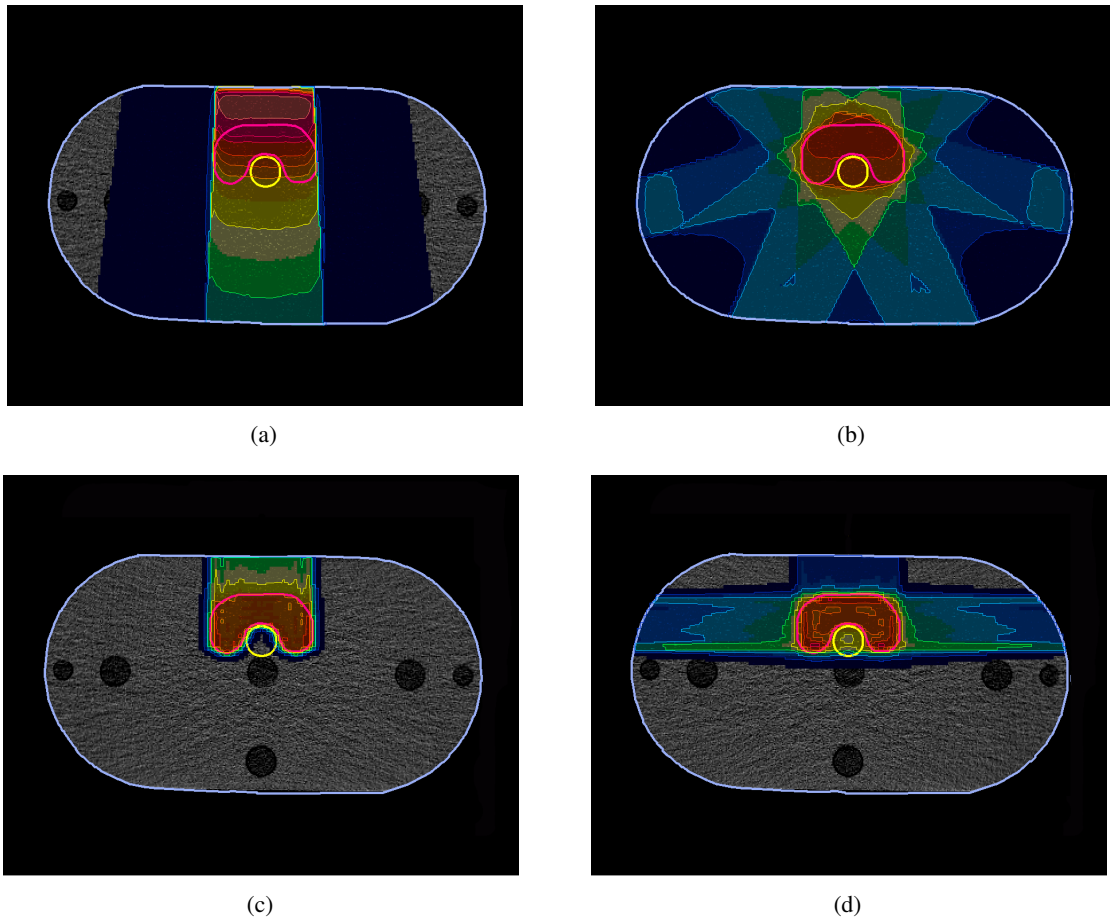
The selection of a beneficial beam ensemble for irradiation is the very subject of this work. It will be introduced in due detail in the following section 1.2.

For treatment strategies applying intensity modulation techniques, the lateral radiation fluence (and, for particle therapy, the fluence in depth) of every irradiation direction is routinely optimized (Bortfeld 2006). This calculation of ideal non-uniform radiation intensities according to a predefined dose prescription is called inverse planning. Finally, the fluence profiles are translated to deliverable segments for the actual treatment.

Eventually, all treatment parameters defining the radiation therapy treatment plan have been determined and the treatment of the patient can begin. Of course, all treatment parameters have to be monitored and re-evaluated during the course of the treatment. Any unexpected response of the patient to the treatment requires an adequate adjustment of the treatment plan.

1.2. Beam ensemble selection

Ionizing radiation effects all tissues along its penetration path within the patient. Radiation damage is not restricted to cancerous tissues. For 6 – 20 MeV photons, the maximum energy deposition is observed after a build up effect in a depth of 1.0 – 4.0 cm. Behind this maximum, the dose deposition decreases exponentially with increasing depth as depicted in figure 1.1(a). Charged particles, in contrast, continuously lose energy in electromagnetic interactions yielding a maximum dose deposition at their end of range called Bragg peak. Figure 1.1(c) shows that it is possible to facilitate this physical property in the clinic to achieve a homogeneous dose deposition in the entire target volume with intensity-modulated proton therapy (IMPT) using one beam direction only.



(a) (b)

(c) (d)

Figure 1.1.: Dose distributions of treatment plans for a phantom applying (a) one conformal photon beam, (b) seven conformal photon beams, (c) one 3D intensity-modulated proton beam, and (d) three 3D intensity-modulated proton beams. Red contour: target, yellow contour: OAR.

If we are to achieve a homogeneous dose deposition in the entire target volume with photons, we have to superimpose multiple beams from different directions within the target volume, as shown in figure 1.1(b). This has a positive side effect, which is also observed for particles, as visualized in figure 1.1(d): With multiple irradiation angles, we can spread the dose deposition over the normal tissues. A higher dose deposition in a smaller volume is traded in for a lower dose deposition in a larger volume.

Furthermore, the dose gradient around the entire target volume can be improved using multiple superimposed beams. For photons, this emerges from the linear superposition of the steep lateral dose gradients of the in-

dividual photon beams (see figure 1.1(a)); for protons, this emerges from the linear superposition of both the steep lateral dose gradients and the distal dose fall-off behind the Bragg peak (see figure 1.1(c)).

It is evident that there are better and worse beam orientations regarding the clinical objectives of radiation therapy treatment planning. Of course, we want to avoid irradiating directly through an OAR. The beam directions should be well separated to spread high doses over the normal tissues. For photons, it makes sense to include beam orientations tangentially to OAR-target interfaces in order to model a steep dose gradient between the OARs and the target. For particles, in contrast, we rather facilitate the distal fall-off to establish a conformal treatment plan.

The mentioned basic principles, however, do not suffice to find the ideal beam configuration for every individual patient. The interdependence of multiple intensity-modulated fields is too complex to be overseen a priori by a human planner. Finding the true optimum beam ensemble for every patient is a very difficult problem.

Numerous authors address the beam angle selection (BAS) problem and demonstrate the benefit of an automated BAS strategy for the quality of 3D conformal and/or intensity-modulated radiation therapy (IMRT). In the current clinical practice, however, the majority of patients are irradiated with evenly spaced coplanar beams or the orientations of treatment beams are adjusted based on the experience of a human expert in a tedious trial and error process. Unlike the optimization of fluence profiles, BAS does not rely on the minimization of a mathematical objective (Orton et al. 2008); it is typically excluded from the inverse planning process.

We see four reasons for the reluctance to apply published BAS strategies in the clinic. First, equi-spaced coplanar beams or a set of manually adjusted beams usually yield clinically acceptable treatment plans (Bortfeld and Schlegel 1993). Second, the incorporation of BAS, a highly non-convex optimization problem with exponential complexity (Bortfeld and Schlegel 1993, Craft 2007, Ehrgott et al. 2008), into conventional inverse treatment planning results either in non-intuitive heuristics or very complex and time-consuming optimization processes. Third, the main vendors of treatment planning software have not yet implemented BAS in a commercially available product. Fourth, an irradiation with customized non-coplanar beams may prolong the treatment and reduce the patient throughput in the clinic. Though additional couch rotations can be performed automatically by all major vendors of linear accelerators by now, patient positioning might require additional attention. Unfortunately, this issue has not yet been studied by the radiation oncology community.

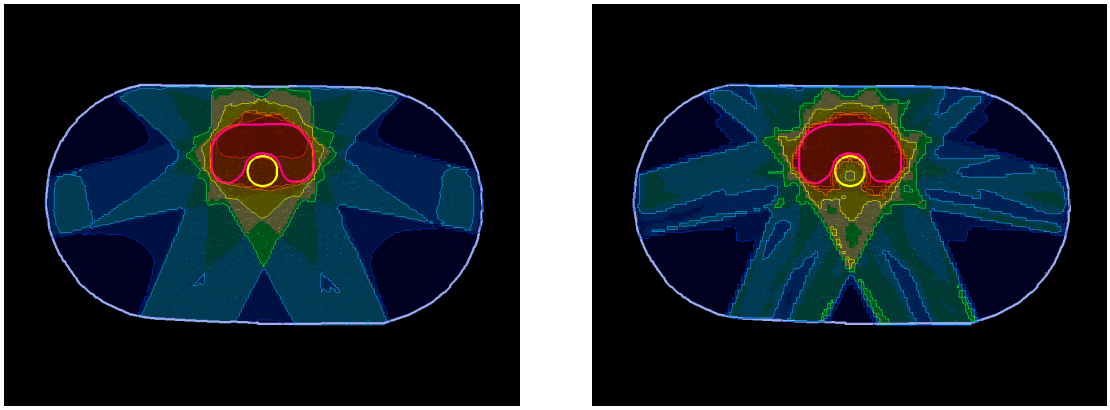
It is the objective of this work to investigate the role of BAS in radiation therapy treatment planning. We present a mathematical description of the BAS problem in radiotherapy in chapter 2 and we review approaches to BAS which have already been published in chapter 3. Furthermore, we introduce novel strategies for BAS applying machine learning concepts in order to overcome shortcomings of the existing approaches in chapters 4 and 5. Advantages and disadvantages of the underlying methodology are discussed and the clinical impact of the developed strategies are evaluated in treatment planning studies. The summary and outlook in chapter 6 conclude this thesis. Our investigations focus on applications for photon radiotherapy, but wherever applicable we include information regarding particle therapy.

2. Formal definition of the beam ensemble selection problem

This chapter gives a detailed overview of conventional inverse planning, which was introduced as the final step of the overall radiation therapy treatment planning process in section 1.1, and it discusses a formal extension of conventional inverse planning that incorporates automated BAS.

2.1. Conventional inverse planning

Figure 2.1(a) visualizes that it is impossible to spare an OAR which is surrounded by a C-shaped target using conformal radiation therapy. During an IMRT treatment, however, non-uniform radiation fluences may be delivered from different beam angles. This yields a higher conformity of the dose distribution, as visualized in figure 2.1(b). The process of finding the ideal fluence modulation according to the clinical objectives for every patient is called inverse planning. It corresponds to the mathematical optimization of $500 - 60,000$ beamlet weights, which constitute a discrete representation of the modulated fluence profiles.



(a) (b)

Figure 2.1.: Dose distributions of (a) a conformal photon therapy treatment plan and (b) an IMRT treatment plan. Both treatment plans apply seven fields. Red contour: target, yellow contour: OAR, and blue contour: normal tissue.

2.1.1. Mathematical formulation

The clinical objectives of radiation therapy are usually translated to a convex - and therewith unimodal (Nocedal and Wright 1999) - objective function which depends on the treatment parameters. Hence, inverse planning is reduced to a mathematically tractable optimization problem. At the German Cancer Research Center, we usually apply a quadratic objective function F which is formulated on a discrete representation of the patient anatomy (Oelfke and Bortfeld 2001).

$$F = \sum_{i \in \text{Target}} \left\{ p_i^{\max} \cdot \left[D_i - D_i^{\max} \right]_+^2 + p_i^{\min} \cdot \left[D_i^{\min} - D_i \right]_+^2 \right\} + \sum_{i \in \text{OAR}} \left\{ p_i^{\max} \cdot \left[D_i - D_i^{\max} \right]_+^2 \right\}$$
$$\text{where } [x]_+ = \begin{cases} x \text{ for } x > 0 \\ 0 \text{ else.} \end{cases} \tag{2.1}$$

D_i denotes the actual dose in voxel i. $D_i^{\max}$ and $D_i^{\min}$ correspond to the desired maximum and minimum doses in voxel i. For target volumes, $D_i^{\max}$ and $D_i^{\min}$ are often the same. For normal tissues and potential OARs, the desired minimum dose $D_i^{\min} = 0$ Gy. This constraint is never violated and consequently not included in equation 2.1. The positivity operator $[\cdot]_+$ ensures that only violated constraints contribute to the objective function F. $p_i^{\min}$ and $p_i^{\max}$ are custom penalties to assign different weights to the individual contributions. In clinical practice, the penalties and tolerance doses are set individually for every region of interest - not for every voxel.

By defining effective penalties p_i and tolerance doses D_i^{pres} as

$$p_i^{\min} = 0 \; \forall \; i \in \text{OAR} \qquad D_i^{\min} = D_i^{\max} \; \forall \; i \in \text{OAR}$$
$$p_i = \begin{cases} p_i^{\max}, \; D_i > D_i^{\max} \\ p_i^{\min}, \; D_i \leq D_i^{\min} \end{cases} \qquad D_i^{\text{pres}} = \begin{cases} D_i^{\max}, \; D_i > D_i^{\max} \\ D_i^{\min}, \; D_i \leq D_i^{\min} \end{cases} \tag{2.2}$$

we may rewrite equation 2.1 in a more compact way:

$$F = \sum_i p_i \left\{ D_i - D_i^{\text{pres}} \right\}^2 . \tag{2.3}$$

This representation suggests to interpret the minimization of the objective function F as a weighted least squares fit of the actual dose distribution to a desired dose distribution. However, this analogy has to be taken with care because F is a continuous but not continuously differentiable function due to the piecewise definition of the effective penalties p_i and tolerance doses D_i^{pres}.

Equation 2.3 hides that the dose D_i is given by a weighted linear superposition of multiple beamlets j. Using a dose influence matrix D_{ij} which specifies the dose contribution from beamlet j to voxel i, this dependence becomes explicit:

$$F(\boldsymbol{w}) = \sum_i p_i \left\{ \Sigma_j (w_j D_{ij}) - D_i^{\text{pres}} \right\}^2 \tag{2.4}$$

With the number of beamlets n_b and number of voxels n_v we can define the vectors and matrices

$$\boldsymbol{w} = \begin{pmatrix} w_1 \\ w_2 \\ \vdots \\ w_{n_b} \end{pmatrix}, \quad \boldsymbol{d}^{\text{p}} = \begin{pmatrix} D_1^{\text{pres}} \\ D_2^{\text{pres}} \\ \vdots \\ D_{n_v}^{\text{pres}} \end{pmatrix}, \quad I = \begin{pmatrix} D_{1,1} & D_{1,2} & \cdots & D_{1,n_b} \\ D_{2,1} & D_{2,2} & \cdots & D_{2,n_b} \\ \vdots & \vdots & \ddots & \vdots \\ D_{n_v,1} & D_{n_V,2} & \cdots & D_{n_v,n_b} \end{pmatrix}, \quad P = \begin{pmatrix} p_1 & & \cdots & 0 \\ & p_2 & & \vdots \\ \vdots & & \ddots & \\ 0 & \cdots & & p_{n_v} \end{pmatrix} \tag{2.5}$$

to rewrite equation 2.4 in matrix form:

$$F(\boldsymbol{w}) \; = \; (I\boldsymbol{w} - \boldsymbol{d}^{\text{p}})^{\intercal} \; P \; (I\boldsymbol{w} - \boldsymbol{d}^{\text{p}}). \tag{2.6}$$

Equation 2.6 is a clear and concise representation of the objective Function F. The first and second derivatives of F with respect to the beamlet weight vector $\boldsymbol{w}$ are

$$\nabla_{\boldsymbol{w}} \, F(\boldsymbol{w}) \; = \; 2 \cdot I^{\intercal} \; P \; (I\boldsymbol{w} - \boldsymbol{d}^{\text{p}}) \tag{2.7}$$
$$\nabla_{\boldsymbol{w}}^2 \, F(\boldsymbol{w}) \; = \; 2 \cdot I^{\intercal} \; P \; I. \tag{2.8}$$

This yields a formal definition of the optimization problem of inverse planning:

$$\boldsymbol{w}^* = \underset{\boldsymbol{w} \in \mathbb{R}_+^{n_b}}{\arg\min} \{(I\boldsymbol{w} - \boldsymbol{d}^{\mathrm{p}})^\top \, P \, (I\boldsymbol{w} - \boldsymbol{d}^{\mathrm{p}})\} \tag{2.9}$$

$\mathbb{R}_+^{n_b}$ is the n_b-dimensional positive orthant. Note that it is physically impossible to deliver negative radiation fluences. The beamlet weights are optimized subject to a positivity constraint.

2.1.2. Convexity of the beamlet weight optimization problem

Convex optimization problems may be handled efficiently with algorithmic optimization techniques because they exhibit only one minimum (Nocedal and Wright 1999).

An optimization problem is convex, if the Hessian matrix is positive semidefinite on its domain of definition and its domain of definition is a convex set (Nocedal and Wright 1999). It is evident that the domain of definition of the inverse planning problem, i.e. the n_b-dimensional positive orthant $\mathbb{R}_+^{n_b}$, is a convex set. To show that the Hessian of the inverse planning problem is positive semidefinite we use that a matrix M is positive semidefinite on the domain of definition $\mathbb{D}$ if

$$\boldsymbol{r}^\top \, M \, \boldsymbol{r} \geq 0 \;\; \forall \, r \in \mathbb{D}. \tag{2.10}$$

Based on this definition, it is straightforward to show that the inverse planning problem is convex:

$$\begin{aligned} & \boldsymbol{r}^\top \, \nabla_{\boldsymbol{w}}^2 F(\boldsymbol{w}) \, \boldsymbol{r} \;\; \forall \, r \in \mathbb{R}_+^{n_b} \\ &= 2 \cdot \boldsymbol{r}^\top \, I^\top \, P \, I \, \boldsymbol{r} \\ &= 2 \cdot (I \, \boldsymbol{r})^\top \, P \, I \, \boldsymbol{r} \\ &= 2 \cdot \boldsymbol{s}^\top \, P \, \boldsymbol{s} \quad \text{with } \boldsymbol{s} = I \, \boldsymbol{r} \\ &= 2 \cdot \sum_i s_i^2 p_i \geq 0 \;\; \forall \, \boldsymbol{s} \;\; \in \mathbb{R}^{n_b} \qquad \square \end{aligned} \tag{2.11}$$

Note that P is a diagonal matrix and the effective penalties p_i are always greater than or equal to 0. For any realistic inverse planning process, a few effective penalties p_i will always be greater than 0. Hence, a realistic Hessian matrix will even be strictly positive definite.

2.1.3. Degeneracy of the beamlet weight optimization problem

Even though the beamlet weight optimization problem is convex, there are many solutions to the inverse planning problem which result in clinically equivalent treatment plans. In this context, Alber et al. (2002) coined the term of the degeneracy of the IMRT optimization problem.

In their 2002 and 2006 publications, Alber et al. investigated the curvature of the solution space, i.e. the space of beamlet weights, around the minimum of the objective function. In analyses of the eigenvalues of the Hessian matrix Alber et al. found that only a small subset of the solution space has significant curvature. As the first derivative also vanishes around the minimum, the objective function is almost flat in a neighborhood of the minimizer in most directions. A descriptive analogy in two dimensions is a deep, long, and narrow valley which is almost flat along its bottom. For the inverse planning problem, the bottom of this valley corresponds to a set of treatment plans with different beamlet weights $\boldsymbol{w}$ but almost the same objective function value $F(\boldsymbol{w})$.

Surprisingly, Alber et al. (2002) did not elaborate on the condition number of the Hessian matrix κ, i.e. the ratio between its largest and smallest eigenvalue. The condition number κ is an important characteristic of an optimization problem because the convergence rate is proportional to $1/\kappa$. The data of Alber et al. (2002) clearly shows that the condition number κ is larger than 10^5. Hence the inverse planning problem is rather ill-conditioned and comes with slow convergence rates.

2.1.4. Solving the beamlet weight optimization problem

Newton methods

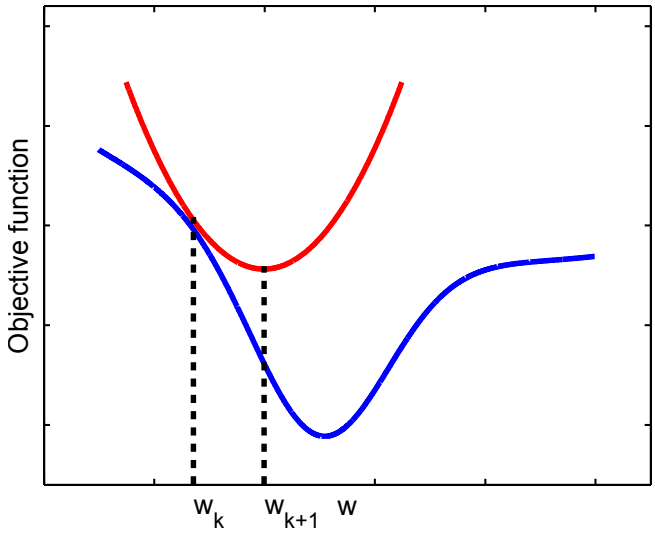


Figure 2.2.: Schematic workflow of a Newton algorithm: A second order approximation (red) of the objective function (blue) is considered during iteration k at the current iterate w_k. Here, the next iterate w_{k+1} is given by the minimum of the second order approximation of the objective function, which corresponds to a full Newton step.

Convex continuous optimization problems may be efficiently solved with iterative Newton methods. Therefore, the optimization is initialized with a beamlet weight vector $\boldsymbol{w}_0 \in \mathbb{R}_+^{n_b}$. To find the search direction $\boldsymbol{\delta}_k$ for iteration k a second order approximation of the objective function is considered.

$$F(\boldsymbol{w}_k + \Delta\boldsymbol{w}) = F(\boldsymbol{w}_k) + \nabla_{\boldsymbol{w}} F(\boldsymbol{w}_k)\, \Delta\boldsymbol{w} + \frac{1}{2}\Delta\boldsymbol{w}^{\mathsf{T}}\, \nabla_{\boldsymbol{w}}^2 F(\boldsymbol{w}_k)\, \Delta\boldsymbol{w} \tag{2.12}$$

As visualized in figure 2.2, it is a valid assumption that the minimizer of the second order approximation of the objective function is a better approximation of the minimum of the objective function.

$$\begin{aligned} \frac{\partial}{\partial \Delta\boldsymbol{w}}\, F(\boldsymbol{w}_k + \Delta\boldsymbol{w}) = \nabla_{\boldsymbol{w}} F(\boldsymbol{w}_k) + \nabla_{\boldsymbol{w}}^2 F(\boldsymbol{w}_k)\, \Delta\boldsymbol{w}_k \overset{!}{=} 0 \\ \Delta\boldsymbol{w}_k = -\left\{\nabla_{\boldsymbol{w}}^2 F(\boldsymbol{w}_k)\right\}^{-1}\, \nabla_{\boldsymbol{w}} F(\boldsymbol{w}_k) =: \boldsymbol{\delta}_k \\ \Rightarrow \boldsymbol{w}_{k+1} = \boldsymbol{w}_k - \left\{\nabla_{\boldsymbol{w}}^2 F(\boldsymbol{w}_k)\right\}^{-1}\, \nabla_{\boldsymbol{w}} F(\boldsymbol{w}_k) \end{aligned} \tag{2.13}$$

Line search methods

The second order approximation of the objective function, however, is only valid in a local neighborhood around the current iterate $\boldsymbol{w}_k$. Taking the minimum of the second order approximation of the objective function as the next iterate, i.e. performing a full Newton step, may sometimes fail and even yield a higher objective function value. The update instruction derived in equations 2.13 has to be combined with a line search method that evaluates the objective function F along the search direction to guarantee convergence. We advocate the use of a backtracking line search that applies an Armijo rule (Nocedal and Wright 1999). A backtracking line search systematically decreases the step length α_k by a constant factor f along the search direction $\boldsymbol{\delta}_k$ to find a new iterate. The Armijo rule yields sufficient descent by enforcing that the objective function value of the new iterate $F(\boldsymbol{w}_{k+1})$ is smaller than an estimated objective function value defined as $F(\boldsymbol{w}_k) - c\alpha_k \cdot \boldsymbol{\delta}^{\mathsf{T}} \nabla_w F(\boldsymbol{w}_k)$ (with

typically $0.1 > c > 0.0001$). A small enough α will always fulfill this condition because the Hessian of the beamlet weight optimization problem is positive semidefinite.

We also investigated a more sophisticated line search algorithm. During every iteration, it establishes a one-dimensional quadratic approximation of the objective function along the search direction $\boldsymbol{\delta}_k = -[\nabla_{\boldsymbol{w}}^2 F(\boldsymbol{w}_k)]^{-1} \cdot \nabla_{\boldsymbol{w}} F(\boldsymbol{w}_k)$. It is based on the current objective function value $F(\boldsymbol{w}_k)$, the current gradient projected onto the current search direction $\boldsymbol{\delta}_k^\mathrm{T} \nabla_w F(\boldsymbol{w}_k)$, and a third objective function value $F(\boldsymbol{w}_k - \alpha_3 \boldsymbol{\delta}_k)$. Due to the positivity constraint of the beamlet weight optimization problem, we have to take care that the actually "allowed" gradient is not over estimated. As we may not enter the negative domain, a component of $\nabla_w F(\boldsymbol{w}_k)$ is set to zero for the calculation of $\boldsymbol{\delta}_k^\mathrm{T} \nabla_w F(\boldsymbol{w}_k)$ if the corresponding beamlet weight is already zero. The step length α_k, given as the minimum of the one-dimensional quadratic approximation, may be calculated analytically:

$$\alpha_k = \frac{\alpha_3^2 \boldsymbol{\delta}_k^\mathrm{T} \nabla_w F(\boldsymbol{w}_k)}{2 \cdot \left(F(\boldsymbol{w}_k - \alpha_3 \boldsymbol{\delta}_k) - F(\boldsymbol{w}_k) - \alpha_3 \boldsymbol{\delta}_k^\mathrm{T} \nabla_w F(\boldsymbol{w}_k) \right)}. \tag{2.14}$$

Figure 2.3 shows the one-dimensional quadratic approximation for a representative iteration. The objective function F is almost a pure quadratic function. Hence, the computed step length corresponds almost to the exact location of the minimum along search direction $\boldsymbol{\delta}_k$. Even though the quadratic line search yields superior algorithmic convergence properties (i.e. we need less iterations) we observed a 10% increase in computation time because an additional function evaluation is needed to establish the quadratic approximation. The simple back tracking Armijo line search is faster.

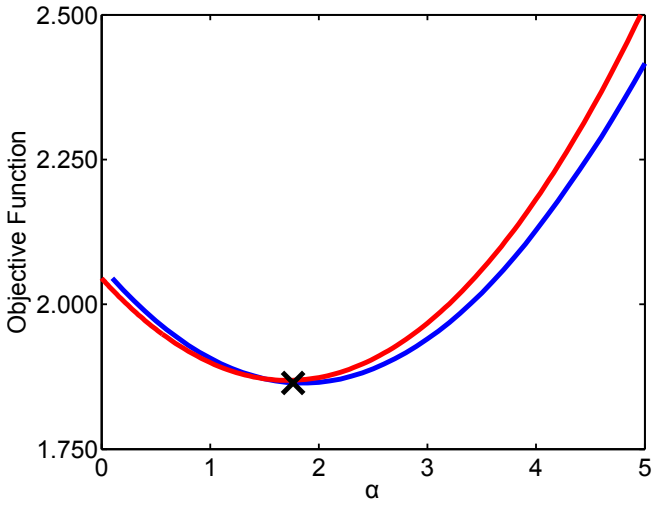


Figure 2.3.: One-dimensional quadratic approximation (red) of the objective function F (blue) along the search direction $\boldsymbol{\delta}_k$. The computed step length α_k (black cross) corresponds almost exactly to the minimum of the objective function along the search direction $\boldsymbol{\delta}_k$.

The algorithm panel 2.1 gives an overview of a Newton method with a backtracking Armijo line search as applied for beamlet weight optimization. Note that it is necessary to project the beamlet weights $\boldsymbol{w}_k$ back onto the feasible set $\mathbb{R}_+^{n_b}$ in every iteration to meet the positivity constraint. For clinical application, the algorithm stops if three subsequent iterations improved the objective function value by less than 0.1% each.

Algorithm 2.1 Algorithm of a Newton method applied for beamlet weight optimization

```
init w_0;  k = 0;  c = 0.01;  f = 0.5;
while (Stopping criterion == false) do
    δ_k = {∇²_w F(w_k)}^{-1}  ∇_w F(w_k);
    α_k = 1;
    while F(w_k) − cα_k · δ_k^T ∇_w F(w_k) < F(w_k − α_k · δ_k) do
        α_k = f · α_k;
    end while
    w_{k+1} = w_k − α_k · δ_k
    project w_{k+1} to ℝ_+^{n_b}
    k + +;
end while
```

Quasi Newton methods

Table 2.1.4 shows that an evaluation of the Hessian matrix requires n_b times more floating point operations than a gradient evaluation. The number of beamlets n_b typically ranges between 1,000 and 60,000. For a large number of beamlets the computation of the Hessian matrix is also associated with memory issues. Storing a full $n_b \times n_b$ Hessian matrix for 20,000 beamlets requires 3 GB in double precision. Furthermore the Hessian matrix has to be inverted according to equations 2.13 in order to determine the search direction $\boldsymbol{\delta}_k$. Hence, it is computationally more efficient to apply a quasi Newton method that approximates the Hessian matrix.

		# additions	# multiplications
Objective function	$(I\boldsymbol{w} - \boldsymbol{d}^{\mathrm{p}})^{\mathrm{T}}\ P\ (I\boldsymbol{w} - \boldsymbol{d}^{\mathrm{p}})$	$n_v(n_b + 1)$	$n_v(n_b + 2)$
Gradient	$2 \cdot I^{\mathrm{T}}\ P\ (I\boldsymbol{w} - \boldsymbol{d}^{\mathrm{p}})$	$n_v(n_v + n_b + 1)$	$n_v(n_v + n_b + 2)$
Hessian	$2 \cdot I^{\mathrm{T}}\ P\ I$	$n_v n_b(n_v + n_b)$	$n_v n_b(n_v + n_b)$

Table 2.1.: Number of floating point operations required for the evaluation of the objective function, the gradient, and the Hessian matrix. The numbers given are an upper bound because the dose influence matrix I is usually very sparse. Note that $n_v \gg n_b$.

At the German Cancer Research Center, we use an L-BFGS two loop recursion algorithm, as shown in algorithm panel 2.2, that directly approximates the search direction $\boldsymbol{\delta}_k$ (Nocedal and Wright 1999). The inverse Hessian is never explicitly constructed. The influence of the curvature of the solution space is modeled based on the changes of the gradient $\boldsymbol{y_k}$ relative to the change of the beamlet weights $\boldsymbol{s_k}$. Algorithm 2.2 replaces the computation of $-[\nabla^2_{\boldsymbol{w}} F(\boldsymbol{w}_k)]^{-1}\ \nabla_{\boldsymbol{w}} F(\boldsymbol{w}_k)$ in algorithm 2.1 to obtain the search direction $\boldsymbol{\delta}_k$ in algorithm 2.1.

The only free parameter of an L-BFGS update is the length of the history m which is used for the update. We found that the algorithm is more stable, especially for particle therapy treatment plans featuring many beamlets, with a rather short history $m = 4$. In experiments with an exact line search we found that the L-BFGS algorithm yields a very good approximation of the curvature of the objective function F because the ideal step length α^* is always around 1. Due to the good curvature approximation, the first evaluated Armijo step length $\alpha = 1$ is often a near optimal step length and consequently, the Armijo line search may outperform a more sophisticated line search applying a quadratic approximation as discussed earlier.

2.1.5. Critical remarks regarding state-of-the-art inverse planning

With the methods described in the last section it is possible to determine suited fluence profiles according to the clinical objectives defined by the penalties P and tolerance doses $\boldsymbol{d}^{\mathrm{p}}$ for every patient in an efficient manner.

However, this approach also has inherent disadvantages.

Shortcomings of a global objective function

The success of the optimization process stands or falls with the formulation of the penalties P and tolerance doses $\boldsymbol{d}^{\mathrm{p}}$. It is very difficult to establish constraints so that they capture the clinical objectives to 100%. The optimization process often yields unexpected and unacceptable solutions that exploit some weakness or incompleteness in the prescription (Bortfeld 2006). Sometimes it may be very difficult to handle local deviations from the desired dose distribution, like hot and cold spots, with the global formulation of the objective function F. Consequently, the prescription has to be constantly readjusted during a tedious trial and error process.

Sequencing issues

For the actual treatment, the modulated fluence profiles have to be translated into deliverable multileaf collimator segments. The number of segments should be reduced to a minimum with regard to treatment times and applied monitor units. Hence, this segmentation may compromise the quality of the resulting dose distribution, especially when the fluence profiles are highly modulated. It is possible to address this issue by incorporating smoothness constraints on the modulation of the fluence profiles (Webb et al. 1998, Alber and Nüsslin 2001, Matuszak et al. 2007; 2008) or by directly optimizing the multileaf collimator segments (De Gersem et al. 2001, Shepard et al. 2002).

Algorithm 2.2 L-BFGS two loop recursion algorithm (Nocedal and Wright 1999).

$\boldsymbol{s_k} = \boldsymbol{w_{k+1}} - \boldsymbol{w_k}, \quad \boldsymbol{y_k} = \nabla_w F(\boldsymbol{w_{k+1}}) - \nabla_w F(\boldsymbol{w_k}), \quad \rho_k = \frac{1}{\boldsymbol{y_k}^{\mathsf{T}} \boldsymbol{s_k}}, \quad H_k^0 = \frac{\boldsymbol{s}_{k-1}^{\mathsf{T}} \boldsymbol{y}_{k-1}}{\boldsymbol{y}_{k-1}^{\mathsf{T}} \boldsymbol{y}_{k-1}} \mathbb{1}$
$\boldsymbol{q} = \nabla_w F(\boldsymbol{w_k})$;
for $i = k-1, k-1, ..., k-m$ **do**
  $a_i = \rho_i \boldsymbol{s_i}^{\mathsf{T}} \boldsymbol{q}$;
  $\boldsymbol{q} = \boldsymbol{q} - a_i \boldsymbol{y_i}$;
end for
$\boldsymbol{\delta}_k = H_k^0 \boldsymbol{q}$;
for $i = k-m, k-m+1, ..., k-1$ **do**
  $\beta = \rho_i \boldsymbol{y_i}^{\mathsf{T}} \boldsymbol{\delta}_k$;
  $\boldsymbol{\delta}_k = \boldsymbol{\delta}_k + \boldsymbol{s_i}(a_i - \beta)$;
end for

2.2. Beam angle selection

Before the optimization of beamlet weights, which is the last step of the overall treatment planning process, a suited beam configuration is established. Just like all other treatment parameters that are fixed before beamlet weight optimization, the beam configuration is usually selected based on clinical experience rather than based on a mathematical objective. As we still lack adequate quantitative models for decisions regarding the fractionation scheme, radiation modality, or delivery mode, it is unavoidable to consult clinical experience. For the decision about a suited beam configuration, however, any objective function that may be used for beamlet weight optimization may also be used to quantify the benefit of a certain beam configuration. Nevertheless, BAS has not been incorporated into standard inverse planning yet.

In the remainder of this chapter we elaborate on the reasons why the beam angle optimization problem is so "difficult to crack" - as Craft (2007) put it. Without loss of generality, the characteristics of the BAS problem are therefore discussed by means of an extended formulation of the inverse problem that uses the quadratic objective function F, which was already introduced for beamlet weight optimization in section 2.1.

2.2.1. Formal definition of the beam angle selection problem

BAS denotes the process of identifying the very beam ensemble B^* within a set of candidate beam directions $\mathcal{B}$ that meets the underlying objectives of the radiation treatment at the best possible rate.

The set of candidate beam directions $\mathcal{B}$ is limited to beam orientations where the treatment head does not collide with the patient or the treatment couch. Figure 2.4(a) shows that $\mathcal{B}$ is a continuous infinite set but it is usually approximated as a discrete set for practical reasons, as shown in figure 2.4(b).

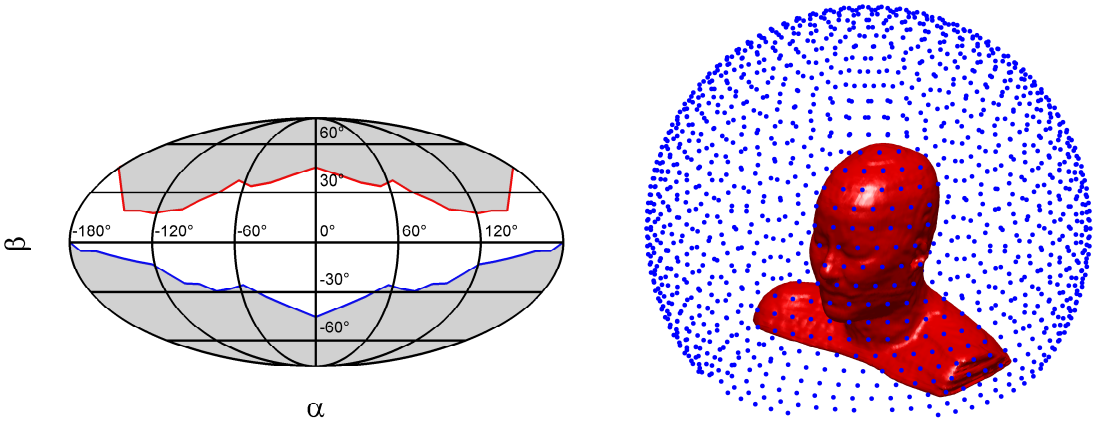


Figure 2.4.: (a) Feasible beam orientations for the Siemens ARTISTE linear accelerator as commissioned at the German Cancer Research Center. α denotes the angle around the patient axis, β denotes the angle towards the transversal plane. Please see appendix A for a sketch of the underlying coordinate system. The shaded area outlined in blue yields couch gantry collisions for all treatment sites. The shaded area outlined in red is only accessible for head and neck cases where superior beams do not interfere with the patient. (b) A head and neck patient relative to a discrete set of 1376 candidate beam directions.

The set of possible solutions to the BAS problem $\widetilde{\mathcal{B}}$ is given by the powerset of of $\mathcal{B}$, i.e. the set of all subsets of $\mathcal{B}$ (Ehrgott et al. 2008).

$$\widetilde{\mathcal{B}} = \mathcal{P}(\mathcal{B}) := \{U \mid U \subseteq \mathcal{B}\} \tag{2.15}$$

Normally we are interested in a solution B^* applying a fixed number of beams, i.e. B^* has a fixed and finite cardinality. Note that this restriction is not valid in general. If we are to find the best trajectory around the patient for arc therapy, we are in fact looking for an infinite subset $B^* \subseteq \mathcal{B}$.

In order to make a clear decision which beam ensemble should be chosen for the irradiation of a patient, every treatment planning agent - that includes both automated BAS frameworks and human experts - requires a transformation $\mathcal{T}$ that projects a dose distribution $\mathcal{D}$, which is associated with a beam ensemble B, onto the real axis $\mathbb{R}$. Otherwise it is impossible to compare different beam configurations and make a clear choice. For an automated BAS framework, $\mathcal{T}$ may be given in precise mathematical terms but for a human expert, $\mathcal{T}$ may be a vague formalism based on clinical experience that cannot be condensed to one formula. In both cases, however, we can find a joint and general definition of the BAS problem:

$$B^* = \underset{B \in \widetilde{\mathcal{B}}}{\arg\min}\, \mathcal{T}\left(\mathcal{D}(B)\right). \tag{2.16}$$

Without loss of generality, we apply the quadratic objective function F, which was introduced earlier for beamlet weight optimization, as transformation $\mathcal{T}$. F depends not only on the beamlet weights $\boldsymbol{w}$ but also on the beam orientations B because the dose influence matrix I implicitly depends on the beam orientations B.

$$I = I(B) \tag{2.17}$$

$$\Rightarrow F = F(B, \boldsymbol{w}) = (I(B)\,\boldsymbol{w} - \boldsymbol{d}^{\mathrm{p}})^{\mathsf{T}}\; P\;(I(B)\,\boldsymbol{w} - \boldsymbol{d}^{\mathrm{p}}) \tag{2.18}$$

This allows for a mathematically more concrete formulation of the BAS problem.

$$\{B^*, \boldsymbol{w}^*\} = \underset{B \in \widetilde{\mathcal{B}},\, \boldsymbol{w} \in \mathbb{R}_+^{n_b}}{\arg\min} \quad (I(B)\,\boldsymbol{w} - \boldsymbol{d}^{\mathrm{p}})^{\mathsf{T}}\; P\;(I(B)\,\boldsymbol{w} - \boldsymbol{d}^{\mathrm{p}}) \tag{2.19}$$

Equation 2.19 is an extended version of the conventional inverse planning problem defined in equation 2.9 that includes BAS.

2.2.2. Derivatives of the objective function with respect to beam angles

In order to solve the optimization problem defined in equation 2.19, it would be desirable to obtain information about the gradient of the objective function F with respect to the beam orientations B.

Unfortunately, it is impossible to differentiate the objective function F directly with respect to the beam orientations B because the dependence of the dose influence matrix I from the beam ensemble B is not described by a mathematical function. Consequently, it is also impossible to implement an automatic differentiation strategy (Griewank and Walther 2008) for the BAS problem.

The limitations of the BAS problem regarding differentiation are not a special case for the objective function F. This restriction holds for all transformations $\mathcal{T}$ that operate on a discrete and discontinuous representation of the patient anatomy.

A numerical differentiation (Richard 1988) of F using central differences would be computationally very expensive. For a beam ensemble B featuring η beams, we would need 2η additional dose calculations and beamlet weight optimizations to compute $\nabla_B F$.

Craft (2007) showed that it is possible to compute $\nabla_B F$ efficiently by means of linear programming duality theory. His optimization process, however, became quickly stuck in local minima. A direct optimization of the beam configuration in the continuous space of beam orientations does not yield the global optimum solution. It is necessary to apply techniques for global optimization. These techniques, however, have to operate on a discrete set of candidate beam directions because the ray tracing and dose calculation for any given beam orientation is time consuming and should be reduced to a minimum.

2.2.3. Non-convexity of the beam angle selection problem

Unlike the beamlet weight optimization problem, the BAS problem is non-convex and exhibits multiple local minima. Note that non-convexity of a mathematical function does not imply multiple local minima, but multiple local minima of a continuous function (defined on a convex domain) imply non-convexity.

Unfortunately it is impossible to show the non-convexity of the BAS problem with a mathematical analysis of the Hessian matrix because we cannot formulate the derivatives of F with respect to the beam orientations. However, the non-convex nature of BAS can be understood by simple inspection of the problem. In a 1993 paper Bortfeld and Schlegel (1993) suggest the following Gedankenexperiment:

> "[...] let us consider a single-beam irradiation of a target situated in the center of a circular phantom. Let the target be a rectangle, of which one dimension is only slightly longer than the other. Without using any mathematics it can be seen that a beam orientation parallel to the rectangle's sides results in a dose distribution that fits the target better than an obliquely incidenting beam. Thus, each orientation that is parallel to the target's sides corresponds to a local minimum of F. However, only those orientations that are parallel to the longer sides of the rectangle correspond to the global minimum of F, which means that these orientations fit the target best."

Bortfeld and Schlegel (1993) conclude:

> "The objective function [...], if regarded as a function of the beam orientations, is in general not convex and has local minima."

This conclusion is supported by multiple publications that introduce score functions to rank different beam orientations, which were reviewed in section 3.3.3 (Soderstrom and Brahme 1992, Pugachev and Xing 2001b, Vaitheeswaran et al. 2010).

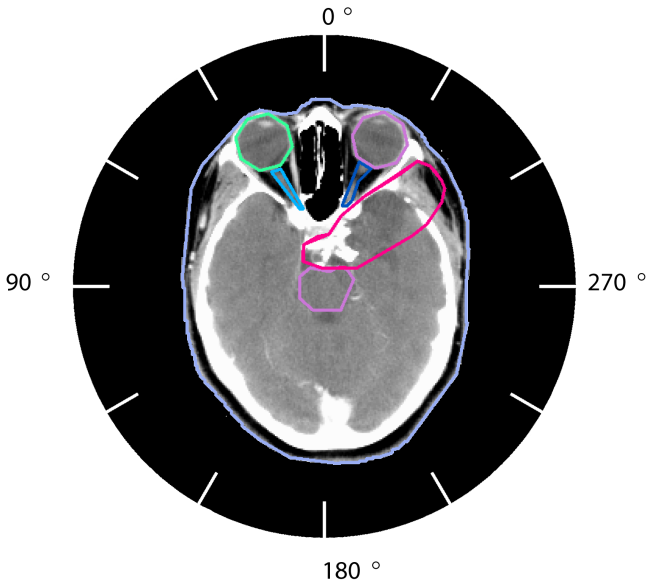


Figure 2.5.: Sketch of the anatomy of the intracranial patient investigated relative to the gantry angle α. The target volume is shown in red.

We calculated explicit visualizations of the solution space of the BAS problem to investigate its non-convexity. First, we studied an intracranial lesion, shown in figure 2.5. 2,628 treatment plans were calculated for beam configurations with three (2.6(a)), five (2.6(b)), seven (2.6(c)), and nine (2.6(d)) coplanar photon beams. Each beam configuration had all but two beams fixed: In figure 2.6(a) one beam was fixed at gantry angle 0°; in figure 2.6(b) three beams were fixed at gantry angles 0°, 120°, and 240°; in figure 2.6(c) five beams were fixed at gantry angles 0°, 72°, 144°, 216°, and 288°; in figure 2.6(d) seven beams were fixed at gantry angles 51°, 103°, 154°, 206°, 257°, and 309°. The gantry angles α_1 and α_2 of the remaining two beams were varied in 5° increments. The corresponding objective function values are shown in figure 2.6.

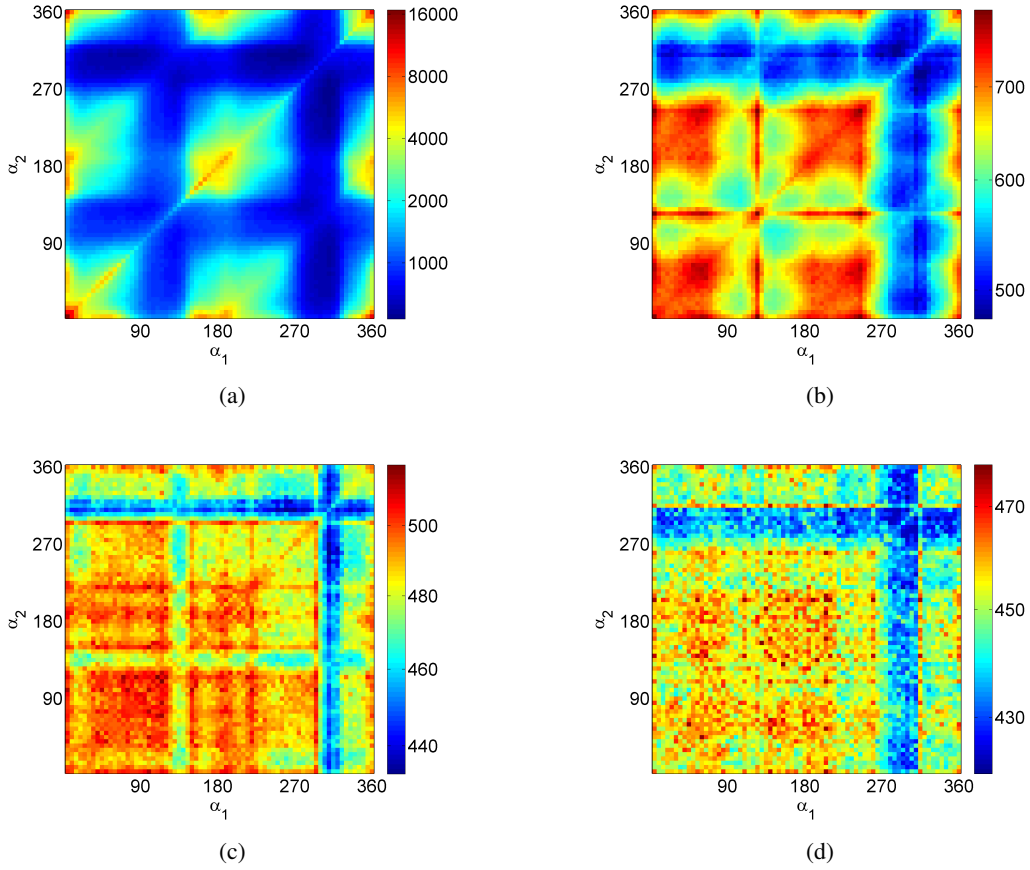


Figure 2.6.: Objective function landscapes of 2,628 treatment plans each for (a) $n = 3$, (b) $n = 5$, (c) $n = 7$, and (d) $n = 9$ coplanar beams for an intra-cranial patient. $n - 2$ beams were always fixed evenly distributed around the patient: for (a) one beam was fixed at gantry angle 0°; for (b) three beams were fixed at gantry angles 0°, 120°, and 240°; for (c) five beams were fixed at gantry angles 0°, 72°, 144°, 216°, and 288°; for (d) seven beams were fixed at gantry angles 51°, 103°, 154°, 206°, 257°, and 309°. The gantry angles α_1 and α_2 of the remaining two beams were varied in 5° increments. Note that the objective function landscapes are symmetric because a permutation of α_1 and α_2 results in the same beam configuration. A logarithmic color scale is applied.

For all four scenarios, the objective function values increase where the variable beams α_1 and α_2 approach the fixed beams. Here, an n-beam plan becomes effectively an $(n-1)$-beam plan. The same holds for the diagonal elements because there the two variable beams α_1 and α_2 approach one another. As expected, the objective function value decreases with increasing number of beams and the relative influence of the beam ensemble on the objective function value decreases with increasing number of beams. For nine beams (figure 2.6(d)), the influence of the beam ensemble almost vanishes for $\alpha_{1/2} < 200°$ within the numerical noise of the beamlet weight optimization. Nevertheless there is a distinguished global optimum solution around $\alpha_{1/2} \approx 315°$. This is a beneficial beam orientation for smaller number of beams, too. The target volume "points" directly in this direction (compare figure 2.5). All objective function landscapes exhibit multiple local minima. With a poor starting value, every greedy neighborhood search algorithm applied to find the best beam ensemble will get stuck in an inferior local minimum. Figure 2.6 indicates that local minima might become more frequent with an increasing number of beams.

Figure 2.7 compares objective function landscapes for 2-beam proton plans and 5-beam photon plans for dif-

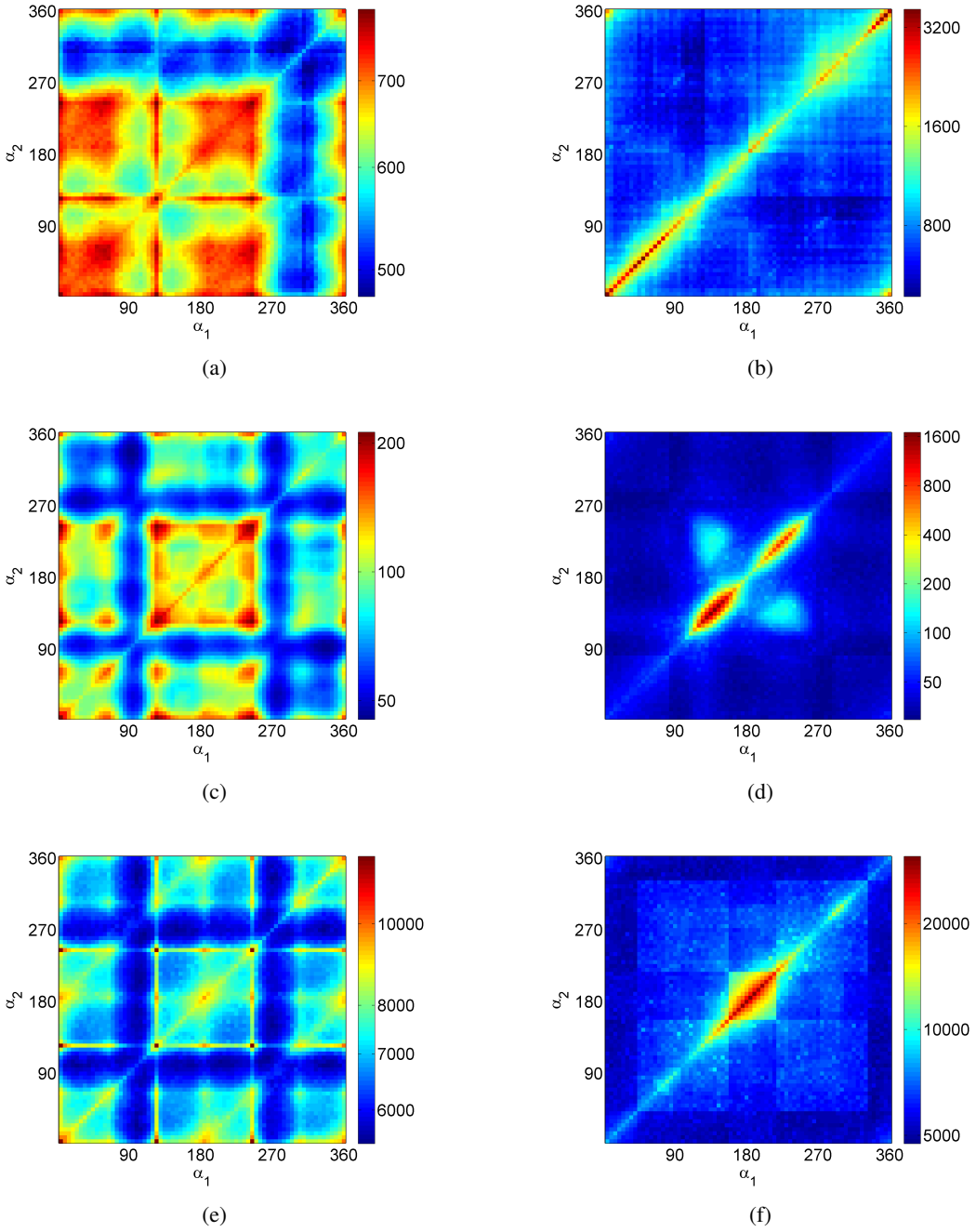


Figure 2.7.: Objective function landscapes of 2,628 treatment plans for (a), (b), (c) photon treatment plans applying five coplanar beams and (b), (d), (f) proton treatment plans applying two coplanar beams. (a) and (b) show objective function landscapes for the intracranial patient shown in figure 2.5, (c) and (d) shown objective function landscapes for a pancreas lesion, and (e) and (f) show objective function landscapes for a prostate lesion. For the photon plans, three beams were fixed at gantry angles 0°, 120°, and 240°. The gantry angles α_1 and α_2 of the remaining two photon beams, and the two proton beams respectively, were varied in 5° increments. Note that a logarithmic color scale is applied.

ferent sites. Apparently it is possible to achieve comparable objective function values with protons while significantly reducing the number of beams. For all sites and radiation modalities we observe local minima. In comparison to photon treatment plans applying three beams (see figure 2.6(a)), the experiments with protons indicate that the problem with local minima might be less severe for particles, even for a similar number of beams. For protons, bad beam configurations clearly stand out. Often they can be explained with a simple anatomical interpretation. For the prostate case (figure 2.7(f)) the peak at (180°|180°) corresponds to an irradiation from the posterior through the rectum. For the pancreas case case (figure 2.7(f)) the four peaks around (180°|180°) correspond to an irradiation from the posterior through one or both kidneys. The fraction of good beam configurations is larger for protons compared to photons. This effect does not result from a smaller number of beams, as can be seen in comparison to the photon plan for the intracranial lesion that applies only three beams (see figure 2.6(a)). The objective function landscapes of the proton treatment plans are not as smooth as the objective function landscapes of the photon treatment plans. They exhibit "edges", e.g. for the intracranial lesion around 120° and for the prostate lesion around 40° and 330°. It is unclear if these edges stem from different lateral scattering properties of a proton beam, the distal fall-off behind the Bragg peak, or another effect.

2.2.4. Complexity of the beam angle selection problem

The selection of of a beam ensemble for radiation therapy within a set of candidate beam directions is a combinatorial optimal decision problem. It can be considered equivalent to a standard problem of operations research: the facility location problem. In a basic formulation, the facility location problem consists of a set of potential facility sites where a facility can be opened, and a set of demand points that must be serviced. The goal is to pick a subset of facilities to open in order to minimize the sum of distances from each demand point to its nearest facility (Wikipedia 2010). For BAS, a facility corresponds to a beam orientation. We want to open up or pick those beam orientations within a set of candidate beams that minimize the objective function F. As there is no direct geometric link between beam configurations and the objective function values, BAS corresponds to a non-metric facility location problem.

The facility location problem, and consequently the BAS problem, is NP-hard. A formal proof of NP-hardness requires a thorough introduction to complexity theory, which is beyond the scope of this manuscript. Interested readers are referred to the PhD thesis of Sultan who shows that the BAS problem is NP-hard by reduction of a Knapsack problem (Sultan 2006, chapter 3.2). At this point we just want to elaborate on the implications of NP-hardness.

To this day there is no algorithm known that solves an NP-hard problem in polynomial time $O(n^k)$[1]. This means there is no constant k so that a polynomial n^k describes the asymptotic runtime of the algorithm in dependence of the size of the input n (Cormen 2001). For BAS, NP-hardness implies drastically increasing runtimes with an increasing number of beams. More precisely, the runtime of an exact BAS algorithm increases at least superpolynomially with the number of beams.

An exact BAS algorithm corresponds to a brute force evaluation of all possible solutions. We already know that, given a finite set of candidate beams $\mathcal{B}$, the set of possible solutions $\widetilde{\mathcal{B}}$ is given by the powerset of $\mathcal{B}$. The number of possible solutions, i.e. the cardinality of $\widetilde{\mathcal{B}}$, is given by

$$|\widetilde{\mathcal{B}}| = 2^{|\mathcal{B}|}. \tag{2.20}$$

For 36 candidate beam directions, which corresponds to a 10° spacing in the coplanar plane, this results in $7 \cdot 10^{10}$ possible beam ensembles. It is beyond prohibitive to evaluate that many solutions. When we are

[1] It is still unclear whether such an algorithm cannot or does exist. The so called P versus NP problem is one of the hottest topics in theoretical computer science. It is one of the seven Millennium Prize Problems selected by the Clay Mathematics Institute.

looking for the best beam ensemble featuring exactly η beams, we only have to evaluate the subsets of $\mathcal{B}$ with η elements.

$$\widetilde{\mathcal{B}}_\eta = \mathcal{P}_n(\mathcal{B}) := \{U \mid U \subseteq \mathcal{B} \wedge |U| = \eta\} \tag{2.21}$$

The cardinality of $\widetilde{\mathcal{B}}_\eta$, i.e. the number of possible beam configurations with η beams, is given by (Hou et al. 2003)

$$|\widetilde{\mathcal{B}}_\eta| = \binom{|B|}{\eta} = \frac{|B|!}{\eta!\,(|B|! - \eta!)}. \tag{2.22}$$

Figure 2.8 shows the number of possible beam configurations $|\widetilde{\mathcal{B}}_\eta|$ versus the number of candidate beams $|B|$ for beam ensembles applying a varying number beams η. The diagram also includes the number possible solutions $|\widetilde{\mathcal{B}}|$ without a fixed number of beams. It is obvious that, if we are to evaluate all possible beam combinations, we have to reduce the number of candidate beams or the size of the beam ensemble. Even selecting only the best five beams in a set of 36 candidate beam directions corresponds to the evaluation of 376, 922 treatment plans. If we could evaluate a treatment plan in 1 sec, we would still require 105 h of computation time for this task. In simple terms, the solution space of the BAS problem is not only highly non-convex with multiple local minima but also extremely large.

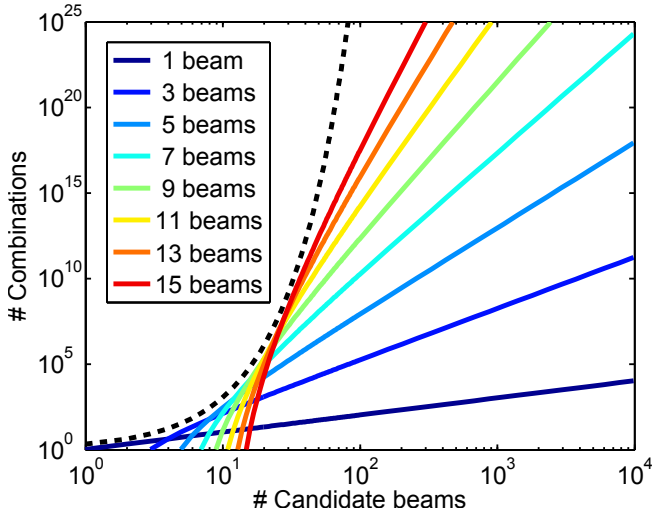


Figure 2.8.: Number of beam combinations $|\widetilde{\mathcal{B}}_\eta|$ vs. number of candidate beams $|B|$ for different number of beams η. The dashed black curve depicts the number of possible solutions $|\widetilde{\mathcal{B}}| = 2^{|\mathcal{B}|}$ without a fixed number of beams.

2.2.5. Degeneracy of the beam angle selection problem

Meedt et al. (2003) found strong indications that the BAS problem exhibits a degeneracy just like the beamlet weight optimization problem, which was discussed in section 2.1.3. Meedt et al. claim that within the vast space of possible solutions, there is a large number of beam configurations yielding near optimal treatment plans. This was later confirmed by Llacer et al. (2009), who investigated the degeneracy of the BAS problem by analyzing the eigenvalues of a target volume coverage matrix.

The findings of Meedt et al. and Llacer et al. are in agreement with the results of our studies of the solution space of the BAS problem regarding non-convexity, which were already presented in section 2.2.3. Figures 2.6

and 2.7 show that there are many treatment plans that result in almost the same objective function value like the true optimum. These beam configurations may represent dose distributions that are equivalent for the clinical application.

2.2.6. Optimizing the number of beams

In clinical routine, it is desirable to keep the number of beams at a minimum in order to reduce the overall treatment time. A reduction of the number of beams, however, compromises the quality of the resulting treatment plans. In order to show that the quality of a dose distribution, as measured by any transformation used to choose between different beam configurations $\mathcal{T}$, monotonically increases (i.e. the absolute value of $\mathcal{T}$ monotonically decreases) with an increasing number of beams

$$\mathcal{T}_1^* \geq \ldots \geq \mathcal{T}_{\eta-1}^* \geq \mathcal{T}_\eta^* \geq \mathcal{T}_{\eta+1}^* \geq \ldots \geq \mathcal{T}_\infty^* \tag{2.23}$$

we assume that B_η^* is the ideal beam ensemble with η beams (Ehrgott et al. 2008). B_η^* is associated with the dose distribution $\mathcal{D}_\eta^*$ and an absolute value $\mathcal{T}_\eta^*$ of a transformation used to choose between different beam configurations. The dose distribution $\mathcal{D}_\eta^*$ is determined by the corresponding fluence profile $\mathcal{F}_\eta^*$ which may be decomposed in a set of individual fluence profiles per beam

$$\mathcal{F}_\eta^* = \left\{\mathcal{F}_\eta^{*1}, \ldots, \mathcal{F}_\eta^{*\eta}\right\}. \tag{2.24}$$

We see by inspection that a beam ensemble $B_{\eta+1}$ defined as

$$B_{\eta+1} = B_\eta^* \cup \beta \in \mathcal{B} \setminus B_\eta^* \tag{2.25}$$

with the fluence profile

$$\mathcal{F}_{\eta+1} = \left\{\mathcal{F}_{\eta+1}^1 = \mathcal{F}_\eta^{*1}, \ldots, \mathcal{F}_{\eta+1}^\eta = \mathcal{F}_\eta^{*\eta}, \mathcal{F}_{\eta+1}^{\eta+1} = 0\right\} \tag{2.26}$$

results in the same dose distribution and consequently in the same absolute value of $\mathcal{T}$:

$$\mathcal{D}(\mathcal{F}_{\eta+1}) = \mathcal{D}(\mathcal{F}_\eta^*) \tag{2.27}$$

$$\Rightarrow \mathcal{T}\left(\mathcal{D}(\mathcal{F}_{\eta+1})\right) = \mathcal{T}\left(\mathcal{D}(\mathcal{F}_\eta^*)\right) = \mathcal{T}_\eta^*. \tag{2.28}$$

Thus $\mathcal{T}_\eta^*$ is an upper bound for $\mathcal{T}_{\eta+1}^*$.

$$\mathcal{T}_\eta^* \geq \mathcal{T}_{\eta+1}^* = \mathcal{T}\left(\mathcal{D}\left(B_{\eta+1}^*\right)\right) \text{ with } B_{\eta+1}^* = \underset{B \in \widetilde{\mathcal{B}}_{\eta+1}}{\arg\min}\, \mathcal{T}\left(\mathcal{D}(B)\right) \tag{2.29}$$

The ideal beam configuration applying $\eta + 1$ beams has to better or at least as good as the best beam configuration applying η beams.

At the same time, it is obvious that there must be a point of diminishing return, where adding another beam only yields clinically negligible benefits. Valuable insight regarding this issue comes from a paper by Bortfeld (2010). For his derivations, Bortfeld assumed a flat photon depth dose and parametrized the lateral photon fluence by Chebyshev polynomials of the first kind. Thereby he was able to show that, contrary to intuition, there is no benefit whatsoever in increasing the number of beams beyond a certain limit. This limit depends on the achievable amount of intensity modulation per beam. He found that the required number of beams is of the order $10 - 20$ for realistic cases. Of course, Bortfeld's derivations are based on wide approximations. Yet still his results are in agreement with current clinical practice, where sophisticated IMRT treatments apply nine or more beams.

If we do not make any approximations, however, an answer to the question regarding the ideal number of beams will necessarily depend on a subjective definition of what benefit of adding another beam is or is not clinically negligible. For routine treatments an answer will also be biased by the elongated treatment times by irradiation with more beams.

Unfortunately, there is only a single publication that investigates the ideal number of beams in a treatment planning study: based on target homogeneity and normal tissue control probability, Liu et al. (2006) show that nine beams are ideal for the irradiation of standard lung and liver lesions. Lesions with a diameter smaller than 2 cm, however, could benefit from an irradiation using up to 13 beams.

At this point we want to mention rotational therapy modalities, where the beam is turned on during rotation around the patient. In principle, these techniques correspond to an irradiation from infinitely many beam orientations. However, you have to consider that the beam orientations are restricted to a single trajectory around the patient. Beams do not impinge from all over 4π. Conventional rotational therapy is performed by a rotation of the treatment head in the coplanar plane. Only recently, alternative rotation parts have been investigated (Yang et al. 2011). Furthermore, rotational therapy techniques come with very restricted capabilities to modulate the radiation fluence from different beam orientations. It is impossible to deliver multiple segments per beam orientation using a single arc and the field shapes of neighboring beam orientations underly smoothness constraints due to the limited leaf speeds of multileaf collimators (Ulrich et al. 2007).

2.2.7. Beam angle selection for particle therapy

So far we have assumed that we have (1) perfect information about the patient anatomy, (2) 100% accurate models to quantify the radiation transport within the patient, (3) complete knowledge about the effects of ionizing radiation on biological tissues, and (4) capabilities to deliver every radiation treatment exactly as it was planned. Unfortunately, this is not true in clinical routine.

Both radiation therapy treatment planning and delivery are subject to uncertainties. Especially for particle therapy, deficiencies of the mathematical models applied for treatment planning and errors during patient setup may compromise the treatment (Lomax 2008a;b).

For beamlet weight optimization, there are powerful concepts to account for these uncertainties during inverse planning. Based on probabilistic (Unkelbach and Oelfke 2004) and worst case optimization techniques (Pflugfelder et al. 2008b) it is possible to establish treatment plans that are robust regarding potential errors. For BAS for particle therapy, it is essential to incorporate these ideas. Beam orientations have to be selected according to both dosimetric aspects and the criterion of treatment plan robustness.

2.2.8. Clarification of nomenclature

In the literature, the terms beam angle optimization, beam angle selection, and beam angle customization are often used as synonyms for strategies to establish beneficial beam ensembles for radiation therapy. We advocate a more precise language.

Mathematically, the term optimization refers to the minimization or maximization of a real-valued objective function by systematically choosing the values of input variables subject to potential constraints. Hence, the term beam angle optimization should be reserved for approaches that translate the BAS problem solely to the optimization of a real-valued objective function.

Approaches that identify a beneficial beam ensemble based on heuristic strategies should not be advertised as beam angle optimization because the resulting beam configurations do not necessarily fulfill a criterion of op-

timality. Consequently, the terms beam angle selection or beam angle customization are more appropriate as heuristic BAS strategies may yield beneficial - but not optimal - beam ensembles.

We are not suggesting this subtle distinction in order to imply that a “proper” beam angle optimization strategy should be preferred to a “simple” beam angle selection strategy. Both approaches are based on equally complex models and may be equally useful to improve the quality of radiation therapy treatment plans. We are suggesting this subtle distinction in order to establish a precise nomenclature.

3. Lessons learned from earlier approaches to beam ensemble selection

BAS attracts continuous scientific attention. By 2011, more than 50 publications with a clear focus on BAS have been published in peer-reviewed journals. The earliest manuscripts discussing BAS in the context of automated treatment planning date back more than 30 years (Hope et al. 1967, Hope and Cain 1972, der Laarse and Strackee 1976, Redpath et al. 1976).

Even though many competing ideas to select beneficial beam ensembles for radiation therapy have been published, none won broad clinical recognition (Orton et al. 2008). In this chapter, we want to review earlier BAS approaches with a focus on the benefits and the shortcomings of the underlying models. We want to highlight the reasons for their limited clinical impact and touch on the starting points for the novel BAS strategies presented in chapters 4 and 5.

3.1. Two classes of beam angle selection strategies

It is possible to identify two different classes of BAS strategies in the literature:

- **Joint beam angle and beamlet weight optimization strategies** intertwine BAS and the optimization of beamlet weights. Initial beam configurations are iteratively readjusted according to the results of a more or less full optimization of beamlet weights. These strategies often correspond to an optimization of beam angles as defined in section 2.2.8.

- **Heuristic beam ensemble selection strategies** exploit prior knowledge about the BAS problem in order to derive beneficial beam ensembles before the optimization of beamlet weights. These strategies do not correspond to an optimization of beam angles as defined in section 2.2.8.

The first class of BAS strategies intends to find an exact solution to an extended inverse planning problem (as defined in equation 2.19) albeit long computation times. The second class BAS strategies intends to quickly identify improved beam ensembles albeit some approximation error.

3.2. Joint beam angle and beamlet weight optimization strategies

3.2.1. Search strategies

Search strategies evaluate different beam configurations in a predefined order. They suffer from long computation times if the entire solution space is covered. Hence, computation times are sometimes reduced by considering only a small fraction of the entire solution space or the search strategies are merely applied to solve a reduced BAS problem, i.e. finding the best treatment plan with a limited number of beams at a sparse angular or dose resolution.

Das et al. (2003) iteratively added and replaced beams from a set of 72 candidate beam directions according to an equivalent uniform dose maximization strategy. For the prostate cases under investigation they reduced the

number of beams by smart beam placement without compromising clinical objectives.

Wang et al. (2004; 2005) developed an elegant divide and conquer strategy for BAS. They performed exhaustive searches on subspaces of the entire solution space of the BAS problem using parallel computing architectures. Within these subspaces, beneficial beam angles were identified which spanned the sub spaces for exhaustive search in the next iteration. For lung and prostate carcinomas (Wang et al. 2004) as well as paranasal sinus (Wang et al. 2005) carcinomas, they were able to reduce the number of beams without compromising clinical objectives, even though they only considered 46 candidate beam directions.

The mathematicians Engel and Tabbert (2005) developed a scheme to iteratively remove beams from a treatment plan initially applying all 360 candidate beam orientations. Operating with a rather sparse resolution of the underlying CT data of $7 \times 7 \times 10$ mm^3, Engel and Tabbert could supply the full Hessian matrix to the projected Newton algorithm which was used for the optimization of beamlet weights. Interestingly, they did not observe significantly improved convergence properties of the beamlet weight optimization process.

Potrebko et al. (2007) investigated the influence of a very basic search algorithm. By varying the starting gantry angle of equi-spaced coplanar beam configurations, they were able to show that prostate treatment plans could be improved by an appropriate placement of five equi-spaced coplanar beams.

3.2.2. Simulated annealing

Simulated annealing (Kirkpatrick et al. 1983) is a probabilistic heuristic for global optimization. With a finite probability, a simulated annealing algorithm will also accept steps increasing the objective function value. Hence, the algorithm may escape local minima. To ensure convergence of the algorithm, this probability decreases according to a predefined cooling schedule during the optimization process.

Bortfeld and Schlegel (1993) published the one of the first papers applying simulated annealing in the context of BAS. In order to arrive at reasonable computation times, the optimization problem was expressed in the spatial frequency domain and restricted to coplanar beam configurations. Bortfeld and Schlegel had to neglect lateral scatter and assume a purely exponential depth dose profile. For the phantom cases studied, they found that the best beam configuration tends to be an even distribution of beams around the patient. Four years later, Stein et al. (1997) published the most cited paper in the context of BAS[1]. They replaced the frequency domain formulation with a full beamlet weight optimization and enhanced the cooling schedule of the simulated annealing algorithm. The concluding treatment plan comparison showed that, for prostate cases, BAS had a significant impact only when applying five beams or less. Both the work of Bortfeld and Schlegel and Stein et al. considered 36 candidate beam directions in the coplanar plane; BAS took 5 minutes and 45 minutes, respectively.

Rowbottom et al. (2001a) used only 10% of all voxels for fluence optimization accelerating the simulated annealing strategy by one order of magnitude. The algorithm converged after evaluating only 500 beam configurations. In relation to a search space featuring 10^{18} combinations, this suggests that Rowbottom et al. might be trapped in local minima. Bortfeld and Schlegel evaluated $10,000$ beam combinations.

Pugachev et al. (2001) were the first to perform simulated annealing including a full beamlet weight optimization on a large-scale: They sampled $5,000 - 10,000$ beam configurations per patient. For this study Pugachev et al. accepted computation times of $210 - 250$ h to find a big impact of their BAS algorithm for nasopharynx and kidney cases. Non-coplanar beam ensembles provided additional advantages over coplanar beam ensembles for these complicated cases. For prostate cases, however, they did not observe a clear benefit of their BAS strategy. Pugachev et al. (2000) applied simulated annealing for beam angle selection in combination with a

[1] 194 citations according to http://www.scholar.google.com by February 26, 2011.

fast filtered back projection for beamlet weight optimization on several two dimensional phantom cases. They demonstrated a clear benefit of automated BAS, even when using up to nine beams. Rowbottom et al. (2001b) and Steadham et al. (1999) confirmed these findings for parotid gland tumors and pancreatic cancer, respectively. Djajaputra et al. (2003) applied fast simulated annealing on a set of 180 – 216 candidate beam directions in combination with an accelerated dose calculation.

3.2.3. Genetic algorithms

Genetic algorithms (Mitchell 1998) mimic an evolution process for optimization by applying operations such as selection, inheritance, reproduction, and mutation. Evolutionary algorithms are very powerful global optimization strategies requiring little information about the optimization problem. They proved useful for applications in fields like bioinformatics, economics, engineering, and manufacturing.

Hou et al. (2003) suggested a genetic algorithm which focused on searching in depth with a rather small population of 20 individuals over 500 generations. The algorithm was validated in a treatment planning study for head and neck cases, where they demonstrated that the genetic algorithm allowed for the reduction of the number of beams without compromising the quality of the treatment plans. Li et al. (2004) also investigated genetic algorithms for BAS. They adopted the in depth search strategy of Hou et al. but modified the selection process borrowing from simulated annealing concepts. Furthermore they introduced an immunity operation that excluded beam configurations featuring opposing and adjacent beams from the population. For prostate and paraspinal cases, they documented a benefit of BAS by genetic algorithms for resulting dose distributions. Schreibmann et al. (2004) suggested a genetic search with a larger population of 100 – 200 individuals for their multi objective optimization framework.

Li et al. (2005) tackle the BAS problem with a particle swarm algorithm. Swarm algorithms benefit from the social behavior within a swarm which evaluates the search space together and shares information to find a global optimum. In the experiments of Li et al., the particle swarm algorithm outperformed a conventional genetic algorithm yielding improved treatment plans for prostate and head and neck patients. Later, Lei and Li (2009) presented a modified genetic algorithm that incorporated aspects from swarm intelligence. For a phantom and nasopharynx case, they reported improved convergence properties as compared to standard genetic algorithms. However, they were operating on a very sparse beam angle grid with 10° spacing in the coplanar plane. The same authors also reported about improved convergence of DNA genetic algorithms for BAS as compared to conventional genetic algorithms (Li and Lei 2010). The approaching genetic algorithm and the DNA genetic algorithm, however, were not compared. Nazareth et al. (2009) presented a parallel implementation of a genetic algorithm that optimizes only 240 treatment plans in six generations. Nevertheless they were able to clearly improve treatment plans for prostate treatments.

3.2.4. Mixed integer programming

Mixed integer programming was introduced for BAS in conformal radiotherapy by Wang et al. (2003) and later extended to IMRT by Yang et al. (2006). The commercially available solver which was applied by Wang et al. and Yang et al. facilitated a branch and bound strategy to solve the mixed integer linear programming problem. This may be a mathematically inadequate approach, as we do not see a suited branch and bound strategy for the NP-hard combinatorial BAS problem (compare section 2.2.4). Unfortunately, the authors did not address this issue. Yet still, they were able to improve treatment plans for prostate and head and neck cases. Additionally, Wang et al. investigated the influence of the angular resolution on the quality of the resulting treatment plans. They stated that an angular resolution of 10° is probably already enough to obtain an almost ideal treatment plan. The data presented indicated only minor improvements if the angular resolution was increased to 5°.

3.2.5. Continuous optimization in the space of beam angles

It is remarkable that there is only one publication which investigated a full-fledged optimization of beam angles in the continuous beam angle space (Craft 2007) - all other authors focused on solving a combinatorial optimization problem, i.e. finding the best solution in a pool of feasible solutions. Craft implemented a local gradient search where the gradient was obtained by linear programming duality theory. The dose influence matrices for continuous angles were linearly interpolated. Hence, Craft was able to compute a descent direction in the beam angle space and perform small refinements. This optimization processes, however, became quickly stuck in local minima. For clinical purposes, Craft recommended to combine his strategy with a global search.

3.3. Heuristic beam ensemble selection strategies

3.3.1. Iterative strategies

Woudstra and Storchi (2000) introduced the iterative BAS strategy CYCLE for conformal radiation therapy based on precomputed dose distributions of single beams. It was later extended to include IMRT (Woudstra et al. 2005). By iterative construction of the best treatment plan with n beams from the best treatment plan with $n-1$ beams, the search space for the BAS problem increases only linearly - not exponentially - with the number of candidate beams. While utilizing this drastic simplification of the combinatorials of the BAS problem, CYCLE demonstrated a clear benefit for the treatment of pancreas tumors (Woudstra and Heijmen 2003), oesophagus tumors (Woudstra et al. 2005), and liver tumors (de Pooter et al. 2006; 2008). In agreement with earlier publications (Pugachev et al. 2001), de Pooter et al. (2006) observed that optimized non-coplanar beam configurations yield superior treatment plans in comparison to optimized coplanar beam configurations. Compared to a custom simulated annealing strategy, CYCLE yielded clinically equivalent dose distributions. Meedt et al. (2003) reported about an alternative iterative beam adding and replacement strategy without fluence optimization.

3.3.2. Geometric strategies

Geometric approaches for BAS are typically very fast because they do not require time consuming dose calculations to generate dosimetric information about potential beam orientations. A beam ensemble is established based on geometric information of the patient anatomy only.

Haas et al. (1998) developed a geometric approach to the BAS problem in two dimensions. A beam orientation is parametrized by four points forming a trapezoid. The final beam ensemble is identified by a multi objective genetic algorithm maximizing the overlap of beams within the target and minimizing the overlap of beams in normal tissues. This model only proved successful for the optimization of treatment plans with a limited number of beams. For IMRT treatment plans featuring five or more beams, Haas et al. could not report an improvement resulting from their geometric approach.

An exciting approach using neural networks for BAS was presented by Rowbottom et al. (1999). In order to train a neural network classifier with 45 training cases, the patient anatomy was translated to a twelve dimensional feature vector representing the corner points of the volumes of interest in three dimensional space. Despite these radical approximations, Rowbottom et al. obtained beam configurations which were comparable to alternative BAS strategies for prostate patients.

Llacer et al. (2009) presented a very interesting BAS algorithm that exploited geometric information with neural networks. They introduced a binary PTV coverage matrix which may be the first attempt to incorporate the concept of intensity-modulated fields that complement one another from different directions into a heuristic

BAS algorithm.

One of the fastest BAS selection strategies was introduced by Potrebko et al. (2008). For their anatomic beam orientation optimization, triangulated surface mesh data was analyzed to identify beam orientations that impinge tangentially to surface elements of the target volume. The influence of a target surface element on beam selection was weighted according to its area and its distance to the closest OAR. Anatomic beam orientation optimization provided a ranking of candidate directions which served as guidance for the radiation oncologist composing the final beam ensemble for an IMRT treatment plan. It helped to significantly improve the plan quality for gastric, prostate, and oropharynx cases.

3.3.3. Ranking strategies

A ranking strategy condenses information about the benefit of a beam direction to a single number. Based on this number it is possible to rank the candidate beams according to their expected merit. Ranking strategies are not contained automatic frameworks for BAS; they require significant user interaction as the final treatment plan is usually composed by a human expert based on the ranking. Furthermore, ranking strategies cannot represent potential synergetic effects of certain beam combinations.

Soderstrom and Brahme (1992) suggested two of the first ranking schemes. They were based on an entropy and Fourier measure of the intensity modulation of individual candidate beams. Chen et al. (1992) and Myrianthopoulos et al. (1992) introduced beam's-eye-view volumetrics, which quantify the overlap of treatment beams and critical structures, in the context of BAS. Rowbottom et al. (1998) picked up this idea and developed a ranking of candidate beam directions according to beam's-eye-view volumetrics for the optimization of prostate treatments. The reciprocal approach of target-eye-view maps was introduced by Cho et al. (1999).

Based on theoretical considerations borrowed from tomographic image reconstruction, Braunstein and Levine (2000) derived a ranking scheme of candidate beam orientations regarding their potential contribution to the prescribed dose. A clinical evaluation was not provided; plausibility was demonstrated by means of different phantom cases.

Pugachev and Xing (2001b) suggested to calculate a score according to the maximum target dose deliverable by a single candidate beam without exceeding the tolerance doses of the critical structures. Based on this information, beam ensembles were selected with a clear benefit regarding target homogeneity and OAR sparing for nasopharynx and paraspinal carcinomas (Pugachev and Xing 2001a). The approach of Pugachev and Xing is presumably the most popular BAS strategy relying on a score function. An alternative score containing information about the mean depth of the target volume and about the penetrated volume of the OARs was introduced and validated by Meyer et al. (2005).

Vaitheeswaran et al. (2010) introduced a ranking scheme for 36 coplanar candidate beam directions according to a beam intensity profile perturbation score which is proportional to the absolute changes of the beamlet intensities induced by variations of dose volume constraints. The resulting ranking served as guidance for a human planner which composed the final treatment plan while also considering minimum separation conditions, avoiding opposing beams, and bypassing critical structures. Apparently significant additional information was introduced by the human planner during the final composition of the beam ensemble. It is debatable whether the observed improvements of the dose distributions for the phantom, prostate, pancreas, and head and neck cases originate from the BAS strategy of Vaitheeswaran et al. or if they were introduced ex post by the human expert.

3.3.4. Clustering

Lim et al. (2009) facilitated a clustering algorithm in the context of BAS. They introduced characteristic vectors for 36 coplanar candidate directions and performed a K-means clustering according to an Euclidean metric in the space of characteristic vectors. Unfortunately, they only provided clinical data for one prostate case without comparison to a standard treatment plan.

3.3.5. Guided search

Pugachev and Xing (2002) presented a simulated annealing strategy which is guided by prior knowledge from beam's-eye-view dosimetrics. This hybrid approach, which combined a joint optimization of beamlet weights and beam orientations with a heuristic bias, accelerated conventional simulated annealing by one order of magnitude.

Using a nested partitions framework for BAS was suggested by (D'Souza et al. 2008). The nested partitions framework iteratively divided the entire search space into a promising region and a complementary region based on a heuristic promise index - not based on the result of a full beamlet weight optimization. Samples from these sub spaces were used to identify a more confined promising region and corresponding complementary region for the subsequent iteration. This approach is comparable to earlier work by Wang et al. (2004; 2005). D'Souza et al. evaluated only 25 treatment plans per iteration. It is debatable whether this was enough to make a reliable statement regarding new promising and complementary regions.

3.4. Beam ensemble selection for particle therapy

There are only two papers discussing BAS for particle therapy. Even though the focus of this manuscript is BAS for photon treatments we want to mention the publications of of Jäkel and Debus and Moravek et al. at this point for the sake of completeness.

Jäkel and Debus (2000) are the first to discuss the role of beam orientations in particle therapy. Using cylindrical projections of the patient anatomy, they showed that the additional degree of freedom introduced by a - by then hypothetical but now real - heavy ion gantry[2] resulted in improved dose distributions.

Moravek et al. (2009) were the first to present a dedicated BAS strategy for particle therapy. They calculated a score which penalizes beam paths traversing OARs and/or tissue heterogeneities for every candidate beam direction. Thereby they tried to combine dosimetric aspects and the criterion of treatment plan robustness for BAS. The score for a beam configuration was given by the sum of the individual scores. An exhaustive search of all beam combinations is performed to identify the best beam configuration. Of course, this exhaustive search was computationally only feasible for a small number of beams. Moravek et al. (2009) did not provide a thorough assessment of the clinical impact of their approach, it is only validated by means of a phantom study and one clinical case.

3.5. Conclusions

3.5.1. Summary

We reviewed more than 40 publications addressing BAS in radiotherapy. All authors advocated the use of an automated BAS strategy. Improvements regarding the quality of treatment plans were observed for all sites, but with different extent. For complex asymmetric treatment geometries, as regularly observed in the abdomen or

[2]The first heavy ion gantry was installed at the Heidelberg Ion-Beam Therapy Center (HIT) in Germany in 2008.

skull, the impact of automated BAS is more pronounced than for very symmetric geometries, such as prostate cases (Stein et al. 1997, Pugachev et al. 2001).

All publications report similar characteristics of the optimized beam configurations. Only for symmetric patient geometries, the ideal beam configuration may be an even distribution of beams around the patient. For asymmetric patient geometries, as regularly observed within the abdomen or skull, the ideal beam ensemble may be highly asymmetric and apply multiple non-coplanar beams (Stein et al. 1997, Pugachev et al. 2001, Woudstra et al. 2005, Potrebko et al. 2008). Interestingly, no publication suggests an improved standard beam configuration for a specific treatment site. The ideal beam ensemble seems to be rather patient specific than site specific.

Given the findings of the reviewed publications, it is remarkable that radiation oncologists still predominantly select beam ensembles for radiotherapy by hand (Orton et al. 2008). We have already speculated about the reasons for this situation in section 1.2. At this point we want to elaborate on the model inherent deficiencies of the published approaches in order to highlight potential starting points for improvement.

Joint beam angle and beamlet weight optimization strategies

Craft (2007) showed that greedy gradient descent strategies are not suited for the optimization of beam ensembles. The BAS problem exhibits multiple local minima and requires a global search to identify the true optimum solution. The first publications focused on simulated annealing techniques for BAS (Bortfeld and Schlegel 1993, Stein et al. 1997, Pugachev et al. 2001); later modern metaheuristics, like genetic algorithms, became a hot topic (Li et al. 2004; 2005).

All reviewed joint beam angle and beamlet weight optimization strategies were valid approaches to the BAS problem but they required prohibitive computation times (Pugachev et al. 2001) or they were not operating at an acceptable level of complexity (Lei and Li 2009). Either the graining in the beam angle space was very sparse or the model for beamlet weight optimization applied extensive approximations.

Heuristic beam ensemble selection strategies

The reviewed heuristics apply prior knowledge about geometric and/or dosimetric aspects of the BAS problem in various ways. Both contained frameworks for BAS and ranking schemes which serve as guidance for a human planner yield beneficial beam configurations.

Some of the heuristic approaches for BAS, however, exhibit the same problem like the joint beam angle and beamlet weight optimization strategies. They may be neglected for clinical application because they operate on an insufficient resolution either in the beam angle or dose distribution space (Rowbottom et al. 1999, Lim et al. 2009, D'Souza et al. 2008).

Furthermore, most heuristics are not contained automated frameworks for BAS. All ranking strategies, for instance, require significant user interaction to establish the final treatment plan (Rowbottom et al. 1998, Braunstein and Levine 2000, Pugachev and Xing 2001b). These BAS algorithms only ease the BAS process but do not automatize it.

The main drawback of the reviewed heuristics is an inadequate representation of the combinatorial problem of BAS. Synergetic effects of different beam orientations, i.e. where one beam exactly compensates for the shortcomings of another beam of the ensemble, are neglected (Pugachev and Xing 2001b, Potrebko et al. 2008). This is especially severe for intensity-modulated fields where different parts of the tumor may receive dose

from different beam orientations. Llacer et al. (2009) presented the first heuristic that tries to address this issue.

The approach of Llacer et al. (2009), however, suffers from another problem that a lot of heuristic BAS strategies share: Most of them are based on very complex and unclear assumptions. Different aspects that have to be incorporated for BAS are compiled into one bulky objective yielding a black box for BAS (Lim et al. 2009, Moravek et al. 2009).

3.5.2. Implications

Based on the mathematical properties of the BAS problem (compare chapter 2) and the reviewed BAS literature, we deduce the following implications for our research on BAS:

- Any new BAS algorithm should incorporate non-coplanar candidate beams (de Pooter et al. 2006).
- The angular resolution in the space of beam angles should be at least 10° (Wang et al. 2003).
- Applications of BAS strategies should focus on complicated cases (i.e. asymmetric patient geometries, convex target volumes, and/or target volumes in proximity to OARs), because these sites promise the biggest clinical impact (Pugachev et al. 2001).
- Any novel BAS heuristic should address the combinatorial problem of BAS within a contained framework, especially with regard to intensity-modulated fields that complement one another.
- Combinatorial optimization strategies that combine beam angle selection and beamlet weight optimization should operate at an adequate resolution in the dose deposition and candidate beam space. They should sample a sufficient number of beam configurations within the vast solution space.
- Dedicated BAS strategies for particle therapy should incorporate the additional criterion of treatment plan robustness (Moravek et al. 2009), as the effect of potential uncertainties during patient setup and treatment planning may be way more severe for particle than for photon treatments.

4. Heuristic beam ensemble selection

This chapter introduces novel concepts for beam ensemble selection based on dosimetric score functions and geometric considerations. The presented methods belong to the second class of BAS algorithms as defined in section 3.1. They exploit prior knowledge about the BAS problem in order to derive beneficial beam ensembles before the optimization of beamlet weights. The approaches are inspired by ranking strategies applying score functions, which were reviewed in section 3.3.3. However, unlike existing heuristics relying on score functions, our approaches represent contained frameworks that attempt to incorporate the combinatorial problem of BAS.

For BAS ranking strategies, the IMRT treatment planning process is split into two independent operations. The beamlet weights are computed after the selection of a beneficial beam ensemble. Hence, this chapter focuses solely on the selection of the beam ensemble. The beamlet weights are optimized afterwards according to the methods introduced in section 2.1.

4.1. Novel concepts for score functions for beam angle selection

4.1.1. Score vectors

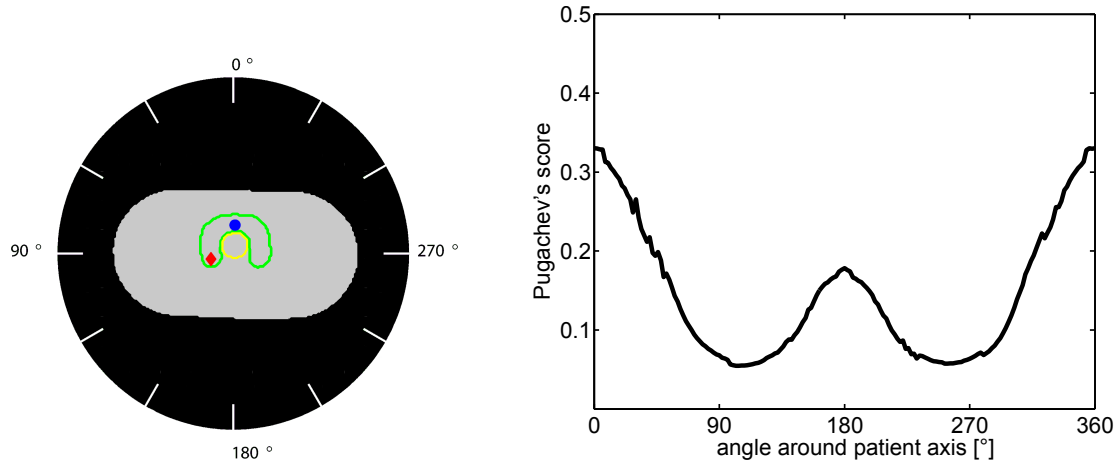
Existing heuristic ranking strategies for BAS assign only a single measure for the expected merit of a candidate beam direction. The shortcomings of this approach can be understood by considering the method introduced by Pugachev and Xing (2001b). They suggested a score function that quantifies the benefit of any given beam direction based on beam's-eye view dosimetrics. First, each candidate beam direction is divided into a discrete grid of beamlets. Second, maximum intensities are determined for all beamlets so that custom tolerance doses for the OARs and normal tissue located in the path of the beamlets are not exceeded. The overall score $\boldsymbol{q}_\beta$ for a candidate beam direction β is given by the accumulated dose within the target volume when irradiating with the previously determined maximum beamlet weighs.

$$\boldsymbol{q}_\beta = \frac{1}{N_{\text{Target}}} \sum_{i \in \text{Target}} \left\{ \frac{d_{\beta i}}{D^{\text{pres}}} \right\}^2 \tag{4.1}$$

N_{Target} denotes the number of target voxels, D^{pres} denotes the prescribed target dose, and $d_{\beta i}$ denotes the dose delivered to target voxel i from beam direction β. According to Pugachev and Xing (2001b), $\boldsymbol{q}_\beta$ captures the main features of a planner's judgment about the quality of a beam. Without a dose limiting OAR except the normal tissue, the quality of a beam direction depends on the depth of the target volume within the patient. A beam that has to travel a longer distance within the patient before reaching the tumor will have a lower score. In case we have an OAR in the beam's path, $\boldsymbol{q}_\beta$ is dominated by the tolerance dose of the OAR and the position of the OAR within the patient: the lower the tolerance of the OAR and/or the closer the OAR to the skin, the lower the dose deliverable to the target volume and the lower the score function value. The approach of Pugachev and Xing (2001b) yields a vector $\boldsymbol{q}$ that lists the score of every candidate beam orientation β.

The behavior of Pugachev's score is demonstrated by means of a phantom case with a C-shaped target volume surrounding an OAR. The phantom, shown in figure 4.1(a), is inspired by the Quasimodo patient model[1] used for IMRT treatment planning studies. Pugachev's corresponding score is displayed in figure 4.1(b). For clarity,

[1] The Quasimodo patient model is available at http://www.daten.strahlentherapie.uni-wuerzburg.de/quasimodo.html



(a) (b)

Figure 4.1.: (a) Patient phantom. The target volume is outlined in green, the OAR is outlined in yellow. The red diamond and the green circle highlight specific sub volumes of the target. (b) Pugachev's score for coplanar candidate beam directions. For its calculation, the relative tolerance doses are $D^{\text{pres}} = 1$, $D^{\text{tol}}_{\text{NT}} = 1$, and $D^{\text{tol}}_{\text{OAR}} = 0.1$. Note that a good beam direction corresponds to a maximum of Pugachev's score.

only coplanar candidate beam orientations are considered.

The characteristics of $\boldsymbol{q}_\beta$ more or less agrees with our intuition. Pugachev's score suggests that 0° is the best irradiation angle for the patient geometry investigated. Beam directions around 0° are associated with higher than average score values. Irradiation from 180° also implies a larger benefit while irradiation from 90° or 270° is considered worst according to Pugachev's score.

Like all other BAS heuristics applying score functions, the approach of Pugachev and Xing has two inherent deficiencies. First, there is no algorithm to compose the final beam ensemble based on the information of the score function. Hence, synergetic effects of different beam angles are neglected. Second, $\boldsymbol{q}_\beta$ does not show that a beam direction β may be beneficial for the irradiation of one part of the target volume while it is unfavorable for the irradiation of another part of the target volume. Consider $\beta = 0°$ for the phantom case. This very beam orientation is probably ideal to irradiate the sub volume of the target highlighted by the red diamond in figure 4.1(a). With an intensity-modulated field you could still spare the OAR from this direction. However, this does not hold for the sub volume highlighted by the blue circle in figure 4.1(a) because the OAR is directly in the path of the beamlet delivering dose to this part of the target volume. For the sub volume highlighted by the blue circle, an intensity-modulated field allows for the best sparing of the OAR from 90° or 270° - actually the two worst beam orientations according to $\boldsymbol{q}_\beta$. If we are to overcome these shortcomings we have to take a novel approach to score functions for BAS.

4.1.2. Score matrices

We suggest to use score matrices to enhance BAS algorithms relying on score functions. Therefore, a radiological score $S_{\beta v}$ is computed for every potential beam angle $\beta \in \mathcal{B}$ and each target voxel $v \in \mathcal{V}$. $\mathcal{B}$ and $\mathcal{V}$ denote the set of candidate beam directions and the set of target voxels, respectively.

$$S_{\beta v} = \frac{d_{\text{NT}} + 100 \cdot d_{\text{OAR}}}{d_{\text{Target}}} \tag{4.2}$$

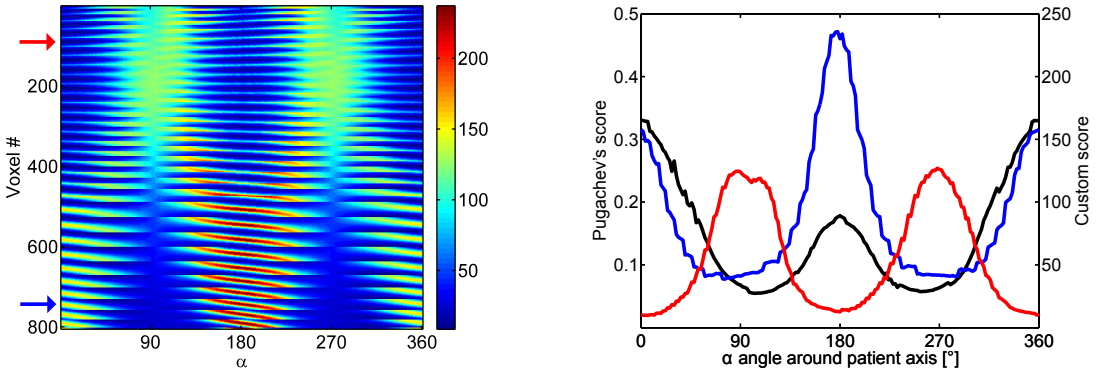


Figure 4.2.: (a) Score matrix S. A row of S lists the radiological quality of all candidate beam directions for one target voxel v. Vice versa, a column of S lists the radiological quality of all target voxels for one candidate beam orientation β. The red and blue arrow indicate the target voxels highlighted in figure 4.1(a). The periodic structure of S emerges from the three-dimensional proximity relationship of the target voxels which is projected to one dimension. (b) Comparison of our custom score $S_{\beta v}$ (red and blue lines) with Pugachev's score (black line) for coplanar irradiation angles. For the calculation of Pugachev's score the relative tolerance doses are $D^{\text{pres}} = 1$, $D^{\text{tol}}_{\text{NT}} = 1$, and $D^{\text{tol}}_{\text{OAR}} = 0.1$. Note that a good beam direction corresponds to a minimum of $S_{\beta v}$ and to a maximum of Pugachev's score (Pugachev and Xing 2001b). It is the central difference between the two approaches that Pugachev assigns a unified score for the entire target, while $S_{\beta v}$ is defined separately for every target voxel. $S_{\beta v}$ is exemplary shown for two target voxels: the blue line corresponds to the score of the target voxel highlighted by the blue circle in figure 4.1(a) and the red line corresponds to the score of the target voxel highlighted by a red diamond in figure 4.1(a). $S_{\beta v}$ indicates that the best irradiation angle depends on the target voxel: while the best irradiation angles for the target voxel highlighted by the blue circle impinge from 90° or 270°, the best irradiation angles for the target voxel highlighted by the red diamond impinge from 0° or 180°.

d_{Target}, d_{NT}, and d_{OAR} denote the doses delivered to the target volume, to the normal tissue, and to potential OARs. $S_{\beta v}$ decreases with increasing dose to the target and decreasing dose to normal tissue and OARs. The smaller $S_{\beta v}$ the better beam orientation β for voxel v. $S_{\beta v}$ implies a twofold nature of the beam selection problem: In absence of OARs, the score is given by the ratio $d_{\text{NT}}/d_{\text{Target}}$; in presence of OARs, the score is predominantly given by $d_{\text{OAR}}/d_{\text{Target}}$, as dose contributions to potential OARs are weighted hundredfold. The weighting factor was found empirically. In a range from $\sim 20 - 500$ it does not have a significant impact on a beam ensemble established within our framework for conventional high energy photon beams. Here, we assign a uniform weight to all OARs, but it may also be feasible to assign individual weights to define custom trade-offs between different OARs.

It is essential that $S_{\beta v}$ may be evaluated individually not only for every candidate beam orientation β but also for every target voxel v. For the computation of $S_{\beta v}$ it is assumed that only target voxel v, not the entire target volume, is irradiated from direction β, i.e. we only simulate the beamlet passing through voxel v from direction β and not a broad beam irradiating the entire target. Hence, the radiological score $S_{\beta v}$ for direction β is the same for target voxels along the axis of the same beamlet, but it varies for off axis voxels within the target. Unlike Pugachev's score $\boldsymbol{q}_\beta$, $S_{\beta v}$ thereby considers that direction β may be a good beam angle for one part of the target, but at the same time, a bad direction for another part of the target that may be in the beam path of an OAR.

Figure 4.2(a) shows the resulting matrix $S_{\beta v}$ which lists the quality for every candidate beam direction β and

target voxel v individually. Pugachev's approach in contrast yields only a vector $\boldsymbol{q}$ that assigns one global measure for the expected merit of every candidate beam orientation for the collective of all target voxels. Figure 4.2(b) contrasts $S_{\beta v}$ and $\boldsymbol{q}$.

Differentiating regarding the expected merit of a candidate beam orientation for different parts of the target volume is a central analogy to the concept of intensity-modulated fields, which deliver non uniform lateral dose profiles from different directions to complement one another. By not integrating over all target voxels and not assigning a uniform score per candidate direction, we attempt to incorporate the concept of intensity-modulated fields, not only during the calculation of the score function, but also during the selection of the beam ensemble.

$S_{\beta v}$ is calculated with a simplified dose algorithm on delineated computed tomography (CT) data. For every beam angle β, water equivalent densities are obtained from the patient CT cube on a precomputed divergent three-dimensional mesh encompassing the entire target volume. For every ray, dose contributions to target, normal tissue, and OAR are computed with a tabulated photon depth dose curve of a 10 mm $\times$ 10 mm photon beam. The required assignment of CT cube voxels to target, OAR, and normal tissue is performed based on a precomputed classification cube with the same resolution and dimensions like the CT cube. Finally, photon scatter is modeled by convoluting dose contributions from central and off axis terms using a spherical Lorentz kernel (Djouguela et al. 2009). This reduces the calculation of $S_{\beta v}$ to two ray tracings and basic matrix operations.

Apparently, the score matrix S provides more information for BAS than the score vector $\boldsymbol{q}$, but the interesting question is how to use this surplus for BAS in an optimal way. It is not possible to directly relate S to the simultaneous optimization problem of beamlet weights and beam orientations as formulated in equation (2.19). Hence, we suggest two alternative formulations of the BAS problem that can be solved optimally with the information contained in S. Section 4.2 introduces a clustering problem of locally ideal beam orientations and section 4.4 introduces a combinatorial optimization problem of an average score. Of course, the two presented formulations are not the only possibilities to exploit the information of S for BAS - additional approaches are thinkable.

Section 4.5 applies the concepts introduced in section 4.2 for geometric BAS.

4.2. Spherical cluster analysis for beam angle selection

BAS by spherical cluster analysis is a two-step process: a spherical data set representing the BAS problem is generated in the first step and analyzed in the second step by a spherical K-means clustering algorithm.

4.2.1. Method

Formulation of beam angle selection as a spherical clustering problem and the set of locally ideal beam orientations

Without loss of generality the isocenter of the patient geometry is defined as the origin of our coordinate system. Thereby, a candidate beam direction β can be represented by the appendant source position given as a point $\boldsymbol{\beta}$ on the three dimensional unit sphere. Based on $S_{\boldsymbol{\beta} v}$ the locally ideal beam angle $\boldsymbol{\beta}_v^*$ for irradiation of voxel v is given by

$$\boldsymbol{\beta}_v^* = \underset{\boldsymbol{\beta} \in \mathcal{B}}{\arg\min}\, S_{\boldsymbol{\beta} v} \tag{4.3}$$

since $\boldsymbol{\beta}_v^*$ maximizes the dose delivered to the target relative to the dose delivered to normal tissue and OARs weighted hundredfold. $\boldsymbol{\beta}_v^*$ is called "locally" ideal because it is only ideal according to a restricted definition of the BAS problem: $\boldsymbol{\beta}_v^*$ is the ideal beam orientation if we aim to irradiate only voxel v considering the effects of the irradiating beamlet on other voxels located within the target, normal tissue, and potential OARs for the score calculation.

For the entire target volume, we may define a set of locally ideal irradiation angles $\mathbb{B}$ which is given by the set of the locally ideal beam angles of all target voxels.

$$\mathbb{B} = \{ \boldsymbol{\beta}_v^* \mid v \in \mathcal{V} \} \tag{4.4}$$

The set of locally ideal beam angles $\mathbb{B}$ is highlighted on the score matrix S in figure 4.3(a) for the phantom case shown in figure 4.1(a). Among others, $\mathbb{B}$ includes a locally ideal beam orientation impinging from $\sim 355°$ for the target voxel indicated by the red diamond and one locally ideal beam orientation impinging from $\sim 280°$ for the target voxel indicated by the blue circle. Figure 4.3(b) depicts a histogram of the locally ideal beam orientations for the phantom case. The histogram shows four accumulation points around 10°, 60°, 300°, and 350°. Candidate beams between 90° and 270° were never accepted into the set of locally ideal beam angles $\mathbb{B}$. According to our model it does not make sense to irradiate the phantom from these angles.

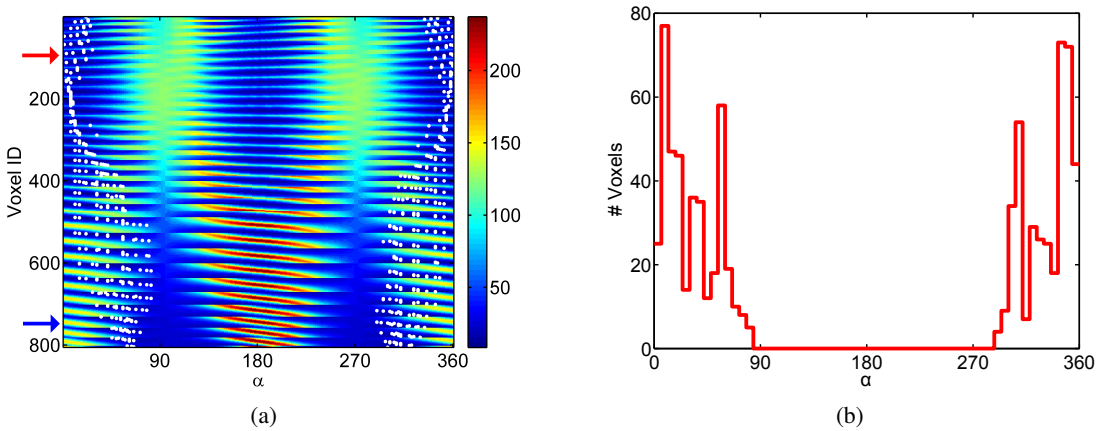
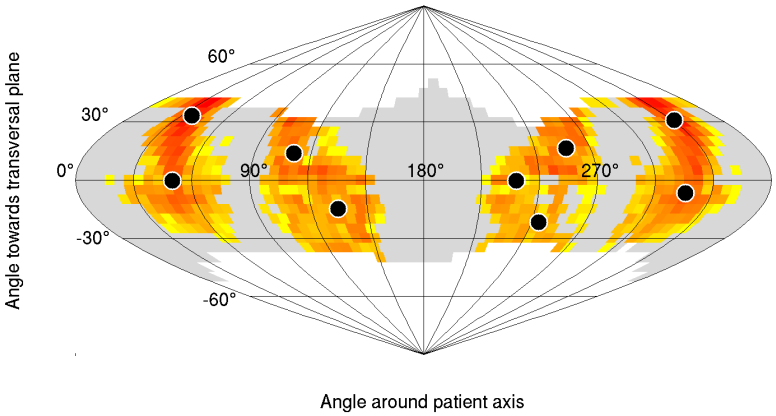
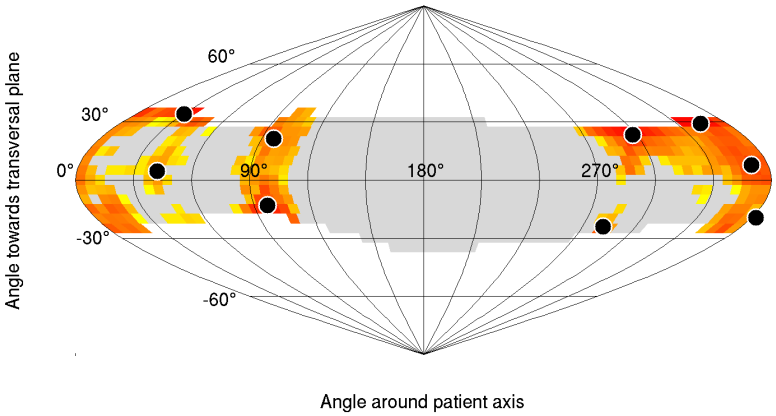


Figure 4.3.: (a) Score matrix S for the phantom case depicted in figure 4.1(a). The white points indicate the locally ideal beam orientations for every voxel, i.e. there is one white point per target voxel/row. (b) Histogram showing the frequency of all candidate beam angles within the set of locally ideal beam angles $\mathbb{B}$.

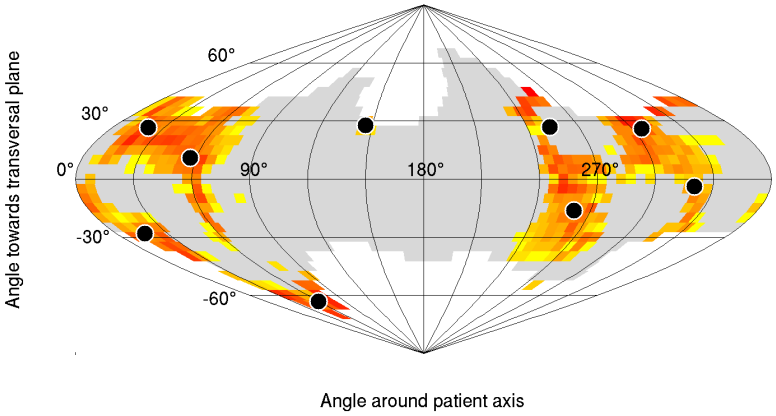
Figure 4.4 visualizes the set of loacally ideal beam orientations $\mathbb{B}$ for three clinical cases considering also non-coplanar candidate beam orienataions. For these visualizations, $\mathbb{B}$, given as a set of points on the 3-sphere, is projected to two dimensions. The relative frequencies of beam orientations within $\mathbb{B}$ is color coded. Just like for the phantom case, not every beam direction $\boldsymbol{\beta} \in \mathcal{B}$ is an element of $\mathbb{B}$, some beam orientations appear more frequent than others, and some may never be the ideal irradiation angle for any target voxel. Like the histogram shown in figure 4.3(b), figure 4.4 clearly exhibits spatial angles where ideal irradiation angles accumulate and regions of beam angles which are not represented in $\mathbb{B}$. The central idea of spherical cluster analysis for BAS is to identify the cluster centroids of the distributions of locally ideal beam angles on the unit sphere and interpret these as beam directions for IMRT treatment planning. This approach implies that an irradiation with the beam ensemble defined by $\mathbb{B}$, i.e. a treatment plan featuring all locally ideal beam orientations, yields a beneficial dose distribution. The clustering algorithm subsequently finds a representation of $\mathbb{B}$ with lower complexity, i.e. a beam ensemble reduced to a number of beams that is applicable in a clinical setting.



(a)



(b)



(c)

Figure 4.4.: Projection of the set of locally ideal beam angles $\mathbb{B}$ (given as points on the 3-sphere) to two dimensions. Red indicates regions of high density of locally ideal beam angles and yellow indicates regions of low density of locally ideal beam angles. (a) Prostate lesion II, (b) pancreas lesion I, and (c) intracranial lesion II. The beam directions of the optimized nine-beam-plan for all three cases are visualized by black dots. The area shaded in gray represents the set of candidate directions $\mathcal{B}$.

By considering only information about a single ideal direction per target voxel for beam selection, information already computed for the matrix $S_{\beta v}$ is neglected during the subsequent beam selection process. Alternative techniques that facilitate the entire information contained in $S_{\beta v}$ for BAS are discussed in section 4.4.

Spherical K-means clustering

To establish an ensemble with η beams, the set of locally ideal beam angles $\mathbb{B}$, depicted in figure 4.4, is divided into η disjunct subsets applying a spherical K-means clustering algorithm. The resulting η centers of the clusters, indicated by black dots in figure 4.4, are interpreted as beam directions for an IMRT treatment plan.

K-means clustering is an established algorithm to classify multi-dimensional data. This section includes a comprehensive recipe of the algorithm; a detailed discussion may be found in the introductory machine learning text books by MacKay (2003) or Alpaydin (2004). Zhong (2005) gives a practical overview of spherical K-means clustering which is very close to our implementation.

In order to create an IMRT treatment plan featuring η beams, η centroids $\boldsymbol{\mu}_k^0 \in \mathbb{B}$ are randomly selected during initialization. The algorithm proceeds by alternating between an assignment step, where every beam direction $\boldsymbol{\beta}_i^* \in \mathbb{B}$ is assigned to its closest centroid, and an update step, where the estimate for a centroid is refined to the mean direction of all beam directions assigned to this centroid. Spherical K-means clustering models the data as a superposition of von Mises-Fisher distributions, which can be considered an analogon to an uncorrelated multivariate Gaussian distribution on a $(D-1)$-dimensional sphere in D-dimensional space. Hence, the assignment c^j during iteration j is performed according to the cosine similarity, i.e. the angle between $\boldsymbol{\beta}_i^*$ and $\boldsymbol{\mu}_k^j$.

$$c^j(\boldsymbol{\beta}_i^*) = \arg\max_k \left\{ \boldsymbol{\beta}_i^{*\mathsf{T}} \boldsymbol{\mu}_k^j \right\} \tag{4.5}$$

The updated estimate for each centroid, i.e. the mean direction of associated directions, is defined as

$$\boldsymbol{\mu}_k^{j+1} = \frac{\bar{\boldsymbol{\beta}}_k^*}{\left\| \bar{\boldsymbol{\beta}}_k^* \right\|} \quad \text{with} \quad \bar{\boldsymbol{\beta}}_k^* = \sum_{c^j(\boldsymbol{\beta}_i^*) = k} \boldsymbol{\beta}_i^* \tag{4.6}$$

By alternating between the assignment and estimation step, spherical K-means clustering converges to a local maximum of the the average cosine similarity L within the η clusters.

$$L = \sum_i \boldsymbol{\beta}_i^{*\mathsf{T}} \boldsymbol{\mu}_{c_j(\boldsymbol{\beta}_i^*)} \tag{4.7}$$

The algorithm terminates if the assignment of beam directions to centroids does not change in two subsequent iterations.

$$c^j(\boldsymbol{\beta}_i^*) = c^{j+1}(\boldsymbol{\beta}_i^*) \; \forall \; \boldsymbol{\beta}_i^* \in \mathbb{B} \tag{4.8}$$

The clustering algorithm is repeated 100 times with different starting conditions in order to exclude inferior solutions trapped in local minima. L facilitates the comparison between different clustering solutions.

4.2.2. Results

The suggested method for BAS is evaluated in a treatment plan comparison for three prostate, pancreas, and intracranial cases. The IMRT treatment plans, as optimized and delivered at the German Cancer Research Center or the University Clinic in Heidelberg, are obtained from our patient data base for each site. All original treatment plans employ equi-spaced coplanar beams. They are optimized by an experienced radiation oncologist with our in-house treatment planning software KonRad (Preiser et al. 1997) using a standard quadratic objective

function (Oelfke and Bortfeld 2001) with individually adjusted maximum and minimum dose constraints and penalties. Thereby, nine sophisticated treatment plans are created, which serve as reference for comparison with plans featuring an optimized beam configuration with the same number of beams or less. Table 4.1 gives a detailed overview of the nine patient cases under investigation.

The calculation of the score matrix S according to equation (4.2) and the spherical K-means clustering algorithm are implemented in MATLAB[2]. The CT data is processed at a $2.62 \times 2.62 \times 2.62$ mm^3 voxel resolution yielding geometries with up to $25,000$ target voxels. The set of candidate directions includes up to $1,200$ non-coplanar beams with a five degree separation between adjacent beams. Directions lacking CT data for the whole beam path are excluded from the set of candidate beam directions. Limitations to accessible gantry and couch angles, which depend on the geometry of every linear accelerator, are not considered during this study. For clinical purposes, however, beam orientations yielding couch-gantry collisions could also be excluded from the set of candidate beam directions. The average computation time for S is $t_{\varnothing} = 358$ s ($t_{\max} = 417$ s, $t_{\min} = 267$ s) on a 2.3 GHz work station. One hundred runs of spherical K-means clustering only take about one second of computation time; the main computational burden arises from the calculation of the score matrix S. After the beam configuration is determined, the beamlet weights are computed with our in-house treatment planning software KonRad (Preiser et al. 1997). Since different beam orientations implicate different trade-offs during inverse planning, maximum dose constraints and penalties of the reference plans are readjusted for the optimized beam configurations in order to establish the best treatment plan possible for every test case. Hence, resulting treatment plans cannot be evaluated by means of objective function values; they are compared based on resulting dose statistics, dose distributions, and dose volume histograms.

The beam configurations established by spherical cluster analysis differ significantly from equi-spaced coplanar beams. Figure 4.4 visualizes the clustering landscapes and the resulting beam configurations for one selected case per site. There are large "forbidden" areas where candidate directions were investigated but never accepted into the set of locally ideal beam angles $\mathbb{B}$. This is particularly pronounced for the pancreas cases (figure 4.4(b)) and for the intracranial cases (figure 4.4(c)), where all treatment beams impinge from the anterior. This highly asymmetric beam configuration mirrors the asymmetric patient geometry (figures 4.6(c), and 4.6(e)). Pugachev et al. (2001) have reported similar findings. The one-sided beam configurations typically spare OARs which are located far away from the target, for example the spine for the pancreas patients, automatically. On the other hand, penalties on normal tissue have to be increased in order to guarantee a sharp dose fall off around the whole target and to avoid hot spots in the normal tissue.

The clinical benefit of the suggested BAS strategy differs for the three sites under investigation and is therefore discussed individually in the following three paragraphs. Detailed statistics is displayed in table 4.1.

Prostate cases

The reference plans for the three prostate cases studied feature nine coplanar equi-spaced beams and are compared to treatment plans with an optimized configuration of nine or seven beams. Homogeneous doses of 76 Gy and 70 Gy are prescribed to the GTV and CTV, respectively. For patient II, inclusive DVHs are given in figure 4.5(a) and 4.5(b). Figures 4.6(a) and 4.6(b) contrast the dose distributions of the equi-spaced coplanar nine beam plan and the optimized nine beam plan. All prostate treatment plans feature clinically equivalent target coverage and the maximum doses for the rectum and bladder differ by less than 1.2 Gy for the optimized and original treatment plans. However, the mean rectum dose of patient I is increased by 7.2 Gy (6.7 Gy) for the optimized plans with nine (seven) beams. As these plans apply maximum dose constraints of 40 Gy to the rectum, there is no incentive for the optimizer to confine medium and low doses. For the same reason, the mean dose to the bladder is increased by 7.6 Gy (10.4 Gy) and 10.6 Gy (9.8 Gy) for the optimized plans with seven

[2]MATLAB R2009b, The MathWorks, Nantick, MA

Prostate

#	Config	GTV	CTV	Rectum	Bladder	Femur (l)	Femur (r)	NT
I	9 equi	76.0	69.6	7.7 (71.1)	30.0 (71.4)	13.5 (32.5)	14.9 (30.4)	3.1
	9 opti	76.0	69.2	14.9 (71.3)	32.8 (69.2)	8.0 (24.0)	10.5 (28.1)	3.8
	7 opti	76.0	69.9	14.4 (71.1)	31.9 (69.5)	6.3 (29.7)	12.4 (36.1)	3.9
II	9 equi	76.0	70.5	34.4 (71.1)	23.1 (73.7)	20.1 (35.4)	19.4 (34.6)	4.3
	9 opti	76.0	70.5	33.4 (70.7)	31.5 (72.7)	16.8 (40.0)	15.0 (34.9)	5.5
	7 opti	76.0	70.2	33.1 (70.8)	33.5 (75.2)	13.9 (39.6)	15.1 (35.4)	5.4
III	9 equi	76.0	70.8	37.1 (70.3)	15.7 (74.7)	5.3 (26.4)	6.9 (27.9)	3.0
	9 opti	76.0	71.2	34.2 (69.4)	26.3 (75.1)	9.0 (27.2)	12.2 (31.9)	3.8
	7 opti	76.0	71.1	34.3 (70.3)	25.5 (75.5)	6.1 (22.4)	11.0 (33.3)	3.8

Pancreas

#	Config	GTV	CTV	Kidney (r)	Kidney (l)	Liver	Spine	Intestine	NT
I	9 equi	54.0	44.9	6.0 (23.0)	6.8 (33.1)	5.4 (44.8)	7.2 (26.7)	13.3 (48.0)	3.0
	9 opti	54.0	44.8	3.8 (16.1)	5.0 (33.3)	4.6 (45.7)	7.4 (22.3)	13.8 (50.1)	3.2
	7 opti	54.0	44.8	4.3 (13.4)	5.2 (35.4)	5.0 (41.6)	7.7 (23.0)	12.8 (48.1)	3.2
II	9 equi	54.0	46.0	9.5 (23.3)	8.7 (18.9)	5.0 (47.9)	9.1 (30.8)	13.7 (49.5)	3.3
	9 opti	54.0	46.3	7.0 (20.2)	5.5 (14.5)	7.1 (47.5)	5.9 (17.1)	13.7 (50.2)	3.5
	7 opti	54.0	46.2	7.6 (21.1)	5.7 (15.3)	6.1 (46.9)	6.2 (16.6)	13.2 (50.2)	3.3
III	9 equi	54.0	46.6	5.7 (22.3)	7.1 (16.1)	6.2 (50.3)	7.7 (31.7)	8.5 (53.3)	3.3
	9 opti	54.0	46.5	5.8 (18.2)	4.0 (13.6)	5.9 (50.1)	6.9 (14.4)	8.5 (53.9)	3.4
	7 opti	54.0	46.7	5.8 (16.1)	4.2 (13.8)	5.9 (50.9)	6.1 (14.4)	8.9 (53.7)	3.5

Intracranial

#	Config	CTV	Brainstem	Eye (r)	Eye (l)	Opticus (r)	Opticus (l)	Chiasm	NT
I	7 equi	50.4	17.2 (29.4)	11.8 (28.3)	12.0 (26.6)	33.9 (48.7)	37.2 (50.4)	29.5 (39.7)	5.5
	7 opti	50.4	13.2 (26.2)	12.4 (23.8)	9.8 (23.4)	32.9 (45.7)	35.4 (45.3)	24.0 (35.7)	4.8
	5 opti	50.4	10.2 (25.5)	10.3 (29.5)	10.2 (25.6)	34.8 (44.0)	36.5 (43.8)	25.7 (39.3)	4.9
II	9 equi	60.0	15.5 (31.6)	15.8 (26.2)	12.2 (20.8)	44.5 (53.5)	33.5 (40.5)	24.0 (50.2)	4.0
	9 opti	60.0	8.0 (29.3)	9.7 (22.2)	9.5 (14.0)	42.8 (51.2)	23.4 (40.0)	21.2 (48.4)	4.2
	7 opti	60.0	4.9 (21.5)	14.8 (25.0)	14.0 (24.0)	43.7 (54.1)	27.7 (37.5)	15.3 (46.0)	3.9
III	9 equi	60.0	22.2 (53.6)	20.3 (35.5)	19.2 (30.0)	48.4 (58.7)	33.8 (47.1)	41.0 (54.1)	4.0
	9 opti	60.0	17.7 (55.5)	16.1 (24.2)	5.4 (22.1)	40.3 (52.0)	12.4 (31.0)	36.3 (55.8)	4.0
	7 opti	60.0	18.0 (55.2)	21.0 (32.1)	5.1 (13.2)	37.6 (50.0)	10.8 (31.5)	35.2 (53.8)	4.1

Table 4.1.: Mean (Max) dose [Gy] of exclusive DVHs for all nine cases under investigation sorted by treatment site. The dose statistics includes all volumes of interest (NT = normal tissue) that define the main conflict during treatment planning of the corresponding site. All plans are normalized to the mean target dose.

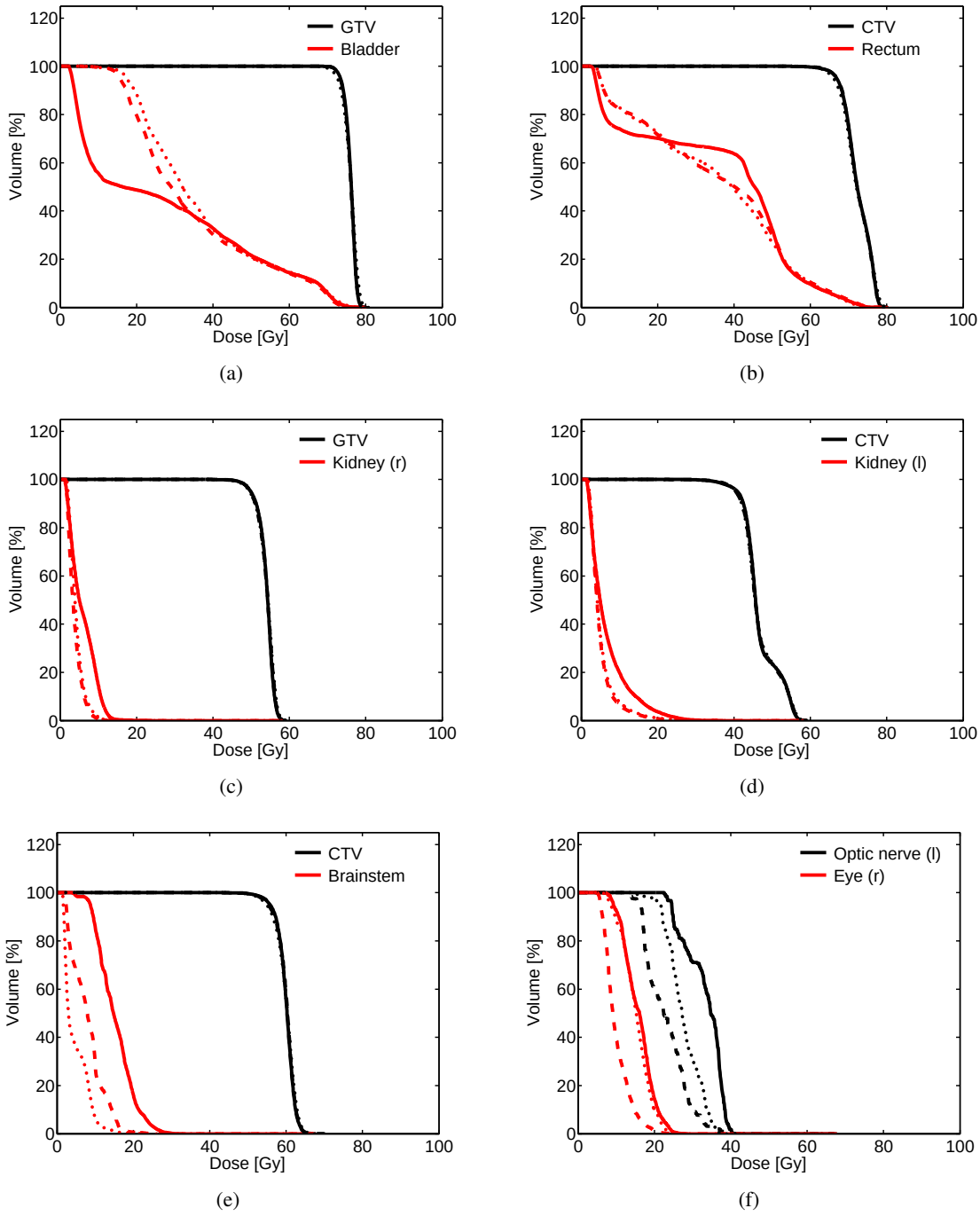


Figure 4.5.: Inclusive DVHs for (a) & (b) Prostate lesion II, (c) & (d) abdominal lesion I, and (e) & (f) intracranial lesion II. Solid lines represent the standard equi-spaced coplanar nine-beam plan, dashed lines represent the optimized non-coplanar nine-beam plan, and dotted lines represent the optimized non-coplanar seven-beam plan. DVHs are shown for selected volumes of interest that define the main conflict during treatment planning of the corresponding case.

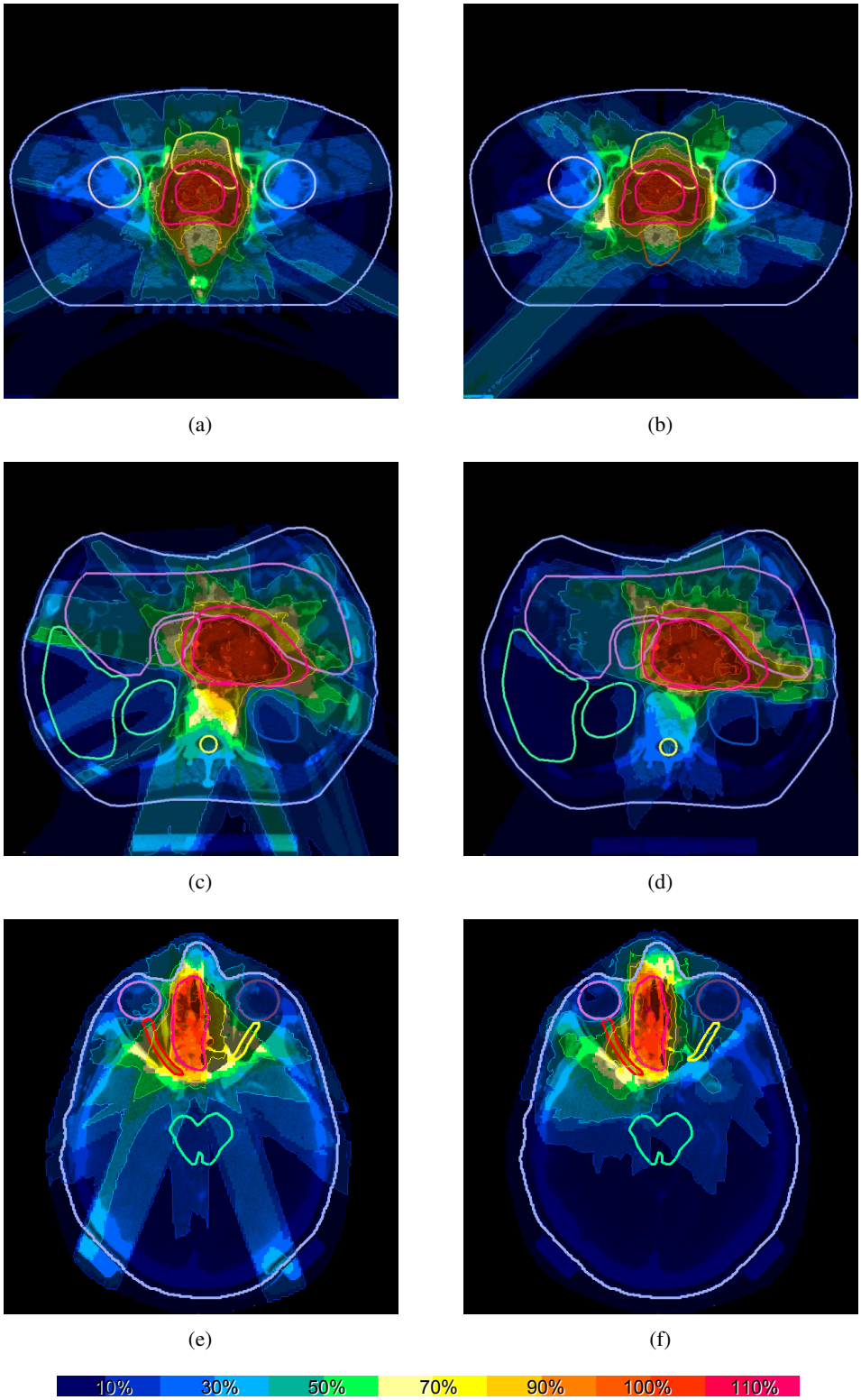


Figure 4.6.: Dose distributions on a selected transversal CT slice for (a) & (b) Prostate lesion II, (c) & (d) abdominal lesion I, and (e) & (f) intracranial lesion II. The dose distributions (a), (c), and (e) are computed with nine equi-spaced coplanar beams, the dose distributions (b), (d), and (f) are computed with nine optimized non-coplanar beams. 100% corresponds to the prescribed target dose.

(nine) beams for patient II and patient III, respectively. For patients I and II, the mean and maximum doses to the right and left femoral head can be reduced with an optimized beam configuration. All three cases exhibit a slightly increased mean dose (0.7 Gy - 1.2 Gy) in the normal tissue, as applying non-coplanar beams enlarges the volume of normal tissue receiving low doses.

Pancreas cases

The reference plans for the three pancreas cases studied feature nine coplanar equi-spaced beams and are compared to treatment plans with an optimized configuration of nine or seven beams. Homogeneous doses of 54 Gy and 45 Gy are prescribed to the GTV and CTV, respectively. For all three patients investigated, the optimized beam configurations result in a reduction of the mean and maximum dose in the kidneys, e.g. a reduction of the mean dose by 3.1 Gy for patient III and a reduction of the maximum dose by 9.6 Gy for patient I. Merely the mean dose to the right kidney of patient III is increased by 0.1 Gy and the maximum dose to the left kidney of patient I is increased by 2.3 Gy. There is no clear trend observed regarding dose contributions to the liver. For the spine, however, the optimized beam configurations, even with a reduced number of beams, result in significantly reduced maximum doses (up to 16.3 Gy for patient III). Doses to the intestine and normal tissue of the optimized and standard treatment plans are clinically equivalent. Figures 4.5(c) and 4.5(d) display DVHs for patient I. Figures 4.6(c) and 4.6(d) illustrate the general observation, that the dose distribution resulting from the optimized beam configuration is more confined to the pancreas.

Intracranial cases

The reference plan for patient I featuring seven equi-spaced coplanar beams is compared to treatment plans with seven and five optimized beams. The reference plans for patient II and patient III featuring nine equi-spaced coplanar beams are compared to treatment plans with nine and seven optimized beams. For patient I, a homogeneous dose of 50.4 Gy and for patients II and III, a homogeneous dose of 60 Gy is prescribed to the CTV. The target volumes of all cases are located in close proximity to the eyes, the optic nerves, or the brainstem. Figures 4.5(e) and 4.5(f) display DVHs of patient II, corresponding dose distributions are visualized by figures 4.6(e) and 4.6(f). For all three cases under investigation, the optimized nine beam configuration yields superior sparing of OARs while guaranteeing equivalent target coverage. Only the maximum dose to the brainstem of patient II is increased by 1.9 Gy. The most remarkable benefits of the optimized nine beam configurations are a reduction of the mean dose by 5.5 Gy for the chiasm of patient I, a reduction of the mean (maximum) dose by 21.4 Gy (16.1 Gy) for the left optic nerve of patient II, a reduction of the mean (maximum) dose by 13.8 Gy (7.9 Gy) for the left eye of patient III, and a reduction of the mean dose by 10.1 Gy (7.5 Gy) for the left optic nerve (brainstem) of patient II. Even the optimized treatment plans with a reduced number of beams result in beneficial OAR sparing when compared to the standard treatment plans. Here, the most remarkable benefits are a reduction of the mean (maximum) dose by 7.0 Gy (3.9 Gy) for the brainstem of patient I, a reduction of the mean (maximum) dose by 10.6 Gy (10.1 Gy) for the brainstem of patient II, a reduction of the mean (maximum) dose by 10.8 Gy (8.7 Gy) for the right optic nerve of patient III, and a reduction of the mean (maximum) dose by 23.0 Gy (15.6 Gy) for the left optic nerve of patient III. For patient I, however, it is not possible to reach the same target confirmation around the whole target volume using only five beams.

4.2.3. Discussion

This is the first attempt to apply spherical clustering analysis for BAS. It is an intuitive heuristic to select a beam ensemble for IMRT. The results of the treatment plan comparisons are encouraging, especially for the pancreas and intracranial cases under investigation. The BAS strategy yields the biggest impact for target volumes located asymmetrically within the patient anatomy. For the prostate cases, clinically comparable IMRT treatment plans are generated by BAS. The coplanar equi-distant nine beam configuration cannot be improved

for a symmetric target located in a central position within the patient.

The computation times of the proposed BAS strategy are competitive in a clinical setting even though our approach includes an explicit dose dependency within the calculation of the score matrix S. It is possible to further decrease the computation time, which predominantly emerges from the calculation of the score matrix S, by using parallel programming techniques and a more efficient programming language. The calculation of S scales linearly with the number of candidate directions $|\mathcal{B}|$ and the number of target voxels $|\mathcal{V}|$. The K-means clustering scales linearly with the number of beams η of the treatment plan and the number of target voxels $|\mathcal{V}|$. Consequently, a higher angular or spatial resolution will not drastically increase computation times. Due to a modular design, the implementation of the presented BAS framework is possible in combination with any existing treatment planning software. However, a more comprehensive delineation of OARs in CT data may be necessary in order to consider dose depositions far away from the transversal isocenter plane by non-coplanar beams.

Cluster analysis on the sphere of candidate directions is an elegant way to deal with the combinatorial issue of beam selection. A solution featuring the best combination of η beams and not the best η beams is determined. Furthermore it is obsolete to penalize solutions with adjacent beams (Moravek et al. 2009) or to remove them from the set of feasible solutions during the beam selection process (Potrebko et al. 2008). These are excluded per se by the minimization of intra-cluster variance and maximization of inter-cluster separation during the K-means algorithm. Unfortunately it is impossible to infer the number of clusters, which corresponds to the number of beams η, from the underlying data with a standard criterion for model selection such as the Bayesian information criterion. Since the set of locally ideal beam orientations $\mathbb{B}$ is given on a discrete grid this learning process requires more sophisticated clustering techniques, as discussed in section 4.3. For BAS by spherical cluster analysis, it is debatable, if the number of clusters really corresponds to the ideal number of beams. It may be beneficial for the plan quality to place two beams within one cluster that stretches over a wide angle, even though a statistical analysis suggests to model this cluster by one centroid only.

Depending on the starting conditions, the K-means algorithm converges to a local, not the global optimum and repeating the K-means procedure for different starting conditions is only an incomplete means to arrive at the global optimum. However, this suboptimal convergence characteristic is not anticipated to compromise the clinical performance of the suggested beam ensemble selection strategy. The spherical K-means algorithm always converges to a valid representation of the set of locally ideal beam orientations and yields improved treatment plans. It is generally open to question whether global convergence is a necessity for BAS. Llacer et al. (2009) and Meedt et al. (2003) show that the beam selection problem exhibits a degeneracy similar to the beamlet weight optimization problem (Alber et al. 2002). Multiple beneficial solutions exist close to the global optimum. For clinical purposes, any BAS yielding a treatment plan with a clear benefit for the patient is valuable.

Even though, BAS is not intertwined with beamlet weight optimization, BAS by spherical cluster analysis includes to some extent the concept of intensity-modulated fields that complement one another to constitute a beneficial IMRT treatment plan. Different parts of the tumor, i.e. different target voxels, are represented by different irradiation angles in the set of ideal beam angles $\mathbb{B}$. Hence, sparing one part of the tumor from one direction but irradiating it from others is considered not only when calculating the score, but also when selecting the beam ensemble. With the score matrix that depends on the individual target voxel, we introduce information about the spatial dependence of dose depositions from different beams inside the target volume to the class of beam ensemble selection strategies that rely on scoring functions. During beam selection, however, it is still neglected that it may be favorable to deliver dose to one target voxel by several beam angles in order to spread the dose over the normal tissue. One target voxel is only represented by one beam angle in the set of ideal beam angles $\mathbb{B}$. Furthermore a dose accumulated by different beams outside of the target volume, potentially

within OARs, is not represented in the beam selection process. The score matrix S does not comprise spatial information about dose deposition in OARs or normal tissue. However, the treatment plan comparison indicates that these two issues do not affect the quality of the final treatment plans as they are both considered ex post during beamlet weight optimization.

Spherical clustering analysis of locally ideal beam angles is not an exact mathematical formulation of the global BAS problem - it is a valid and clinically valuable approximation and provides a flexible framework for BAS. Every scalar function mapping the BAS problem to a spherical clustering problem can be analyzed with the presented method. Existing scoring functions only have to be modified to include individual scores for every target voxel to function within this framework.

4.3. Learning the number of beams and the infinite von Mises-Fisher Mixture Model

In the previous section, we introduced a spherical K-means algorithm to cluster a set of locally ideal beam orientations for BAS. Given the number of clusters η, the K-means algorithm infers the locations of the cluster centers which may be interpreted as beam orientations for IMRT. Here, we are investigating means to also infer the number of clusters η from the set of locally ideal beam orientations. In the picture of spherical cluster analysis, the appropriate number of beams may correspond to the number of clusters of the underlying spherical distribution of locally ideal beam orientations.

This section introduces a more sophisticated clustering algorithm to analyze the spherical data sets of locally ideal beam orientations. It sidesteps the delicate question of finding the right number of clusters by replacing the point estimate of one particular number with a probabilistic belief over the total number of clusters. By adapting previous work on infinite Gaussian mixture models (iGMM) (Rasmussen 2000) in Cartesian spaces to the D-Sphere, we develop the infinite von Mises-Fisher mixture model (iMFMM). While the derivation of the iMFMM in section 4.3.1 may require previous knowledge about probabilistic machine learning, the discussion of the role of the iMFMM for radiation therapy treatment planning in sections 4.3.2 and 4.3.3 does not.

4.3.1. Method

The von Mises-Fisher distribution is a spherical analogon to an uncorrelated multivariate Gaussian distribution in Cartesian space. On the D-sphere, it is defined as

$$\mathcal{F}(\boldsymbol{\beta};\boldsymbol{\mu},\tau) = \frac{\tau^{D/2-1}}{(2\pi)^{D/2} I_{D/2-1}(\tau)} \exp(\tau\, \boldsymbol{\mu}^T \boldsymbol{\beta}) \tag{4.9}$$

with the scalar precision parameter τ and the mean direction $\boldsymbol{\mu}$. I_ν is the modified Bessel function of the first kind and order ν. For $D = 3$, we obtain the special form

$$\mathcal{F}(\boldsymbol{\beta};\boldsymbol{\mu},\tau) = \frac{\tau}{4\pi \sinh(\tau)} \exp(\tau\, \boldsymbol{\mu}^T \boldsymbol{\beta}) \tag{4.10}$$

Figure 4.7 shows $1,000$ samples from a von Mises-Fisher distribution with mean direction $\boldsymbol{\mu} = (1,1,1)/\sqrt{3}$ and precision $\tau = 30$. In the following, we concentrate on the 3-dimensional case, but all derivations can be easily extended to D dimensions.

To construct our model, we assume that the N samples $\boldsymbol{\beta}_i$, i.e. all locally ideal beam orientations, are each generated from a mixture of an unknown and unbounded number K of independent von Mises-Fisher distributions with unknown parameters $\boldsymbol{\mu}_k, \tau_k$.

$$p(\{\boldsymbol{\beta}_i\}; \boldsymbol{\mu}_1, ..., \boldsymbol{\mu}_K, \tau_1, ..., \tau_K, \pi_1, ..., \pi_K) = \sum_{k=1}^{K} \pi_k \prod_{i=1}^{N} \mathcal{F}(\boldsymbol{\beta}_i; \boldsymbol{\mu}_k, \tau_k) \tag{4.11}$$

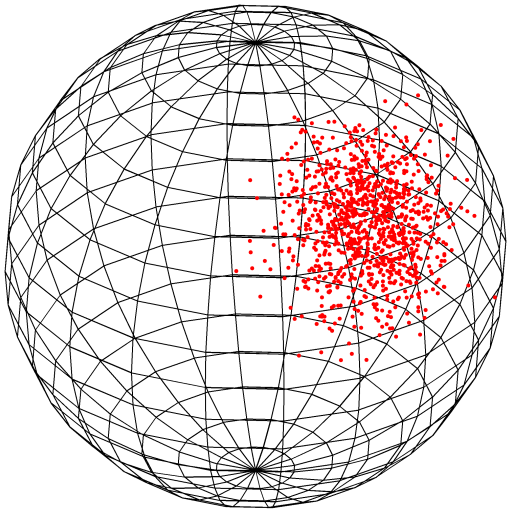


Figure 4.7.: Samples from a von Mises-Fisher distribution with mean direction $\boldsymbol{\mu} = (1, 1, 1)/\sqrt{3}$ and precision $\tau = 30$.

$\pi_k = p(c_i = k)$ is the probability of sample i stemming from cluster k and c_i indicates the assignment of sample i to a cluster k.

First we assume that the number of clusters K is fixed and finite; the limit $K \to \infty$ is introduced later. The derivations are closely modeled on work by Rasmussen (Rasmussen 2000) and Neal (Neal 2000) with regard to Gaussian mixture models. Our contribution is the transformation of their work from Cartesian space to the sphere (i.e. from mixtures of Gaussians to mixtures of von Mises-Fisher distributions).

Given a prior $p(\boldsymbol{\mu}_{1\ldots K}, \tau_{1\ldots K}, c_{1\ldots N})$ on the unknown parameter values, the goal of an inference algorithm on such a model is to track a posterior belief $p(\boldsymbol{\mu}_{1\ldots K}, \tau_{1\ldots K}, c_{1\ldots N}|\boldsymbol{\beta}_{1\ldots N})$ over the parameters of the mixture model given the observed data. In our implementation, inference is performed using Gibbs sampling (Geman and Geman 1984), a widely used Markov chain Monte Carlo scheme. It consists of iteratively sampling values of all parameters of the model individually, conditioned on the current samples from all other parameters. Gibbs sampling is guaranteed to produce samples from the exact posterior in the limit of large numbers of sampling steps.

To keep the computational cost of the sampling scheme manageable, we use conjugate priors for the parameters of the mixture components. A parametric distribution

$$p(z|\boldsymbol{a}) = f(z; \boldsymbol{a}) \tag{4.12}$$

on the variable z with parameters $\boldsymbol{a}$ is called a conjugate prior to a likelihood $p(d|z)$ of z under the data d if the posterior can be formulated in the exact parametric form of the prior using Bayes' rule.

$$p(z|d, \boldsymbol{a}) = \frac{p(d|z)p(z|a)}{\int p(d|z)p(z|a)dz} = f(z; \boldsymbol{a}') \tag{4.13}$$

The von Mises-Fisher distribution forms an exponential family (Bishop 2007). All distributions forming exponential families have conjugate priors for their parameters, and a general construction for these priors exists.

Conjugate prior for $\boldsymbol{\mu}_k$ given τ_k

For the mean parameter $\boldsymbol{\mu}_k$, the conjugate prior is itself a von Mises-Fisher distribution

$$g(\boldsymbol{\mu}_k; \tau_k, \boldsymbol{\beta}_0) = \mathcal{F}(\boldsymbol{\mu}_k; \boldsymbol{m}_0, t_0) \tag{4.14}$$

with two parameters $\boldsymbol{m_0}$ and t_0. In our experiments, we set $t_0 = 0.1$ which leads to a distribution so broadly covering the sphere that the precise value of $\boldsymbol{m_0}$ becomes irrelevant. For the lack of a better option, it was set to the mean of the entire dataset. According to Bayes' theorem, the posterior update is given by the product of prior and likelihood ratio.

$$\begin{aligned} p(\boldsymbol{\mu}_k | \{\boldsymbol{\beta}_{i\in k}\}, \tau_k) &= \mathcal{F}(\boldsymbol{\mu}_k; \boldsymbol{m}_0, t_0) \cdot \prod_{i\in k} \mathcal{F}(\boldsymbol{\beta}_i; \boldsymbol{\mu}_k, \tau_k) \\ &= \mathcal{F}(\boldsymbol{\mu}_k; \boldsymbol{\xi}/|\boldsymbol{\xi}|, |\boldsymbol{\xi}|). \end{aligned} \tag{4.15}$$

with $\boldsymbol{\xi} = t_0 \boldsymbol{m}_0 + \tau_k \sum_{i\in k} \boldsymbol{\beta}_i$. The notation $i \in k$ confines an operation to samples that are assigned to cluster j.

Conjugate prior for τ_k given $\boldsymbol{\mu}_k$

Up to normalization, the conjugate prior on the precision parameter τ_k is

$$f(\tau_k; a, b) \propto \left\{ \frac{\tau_k}{4\pi \sinh(\tau_k)} \right\}^a \exp(\tau_k b) \tag{4.16}$$

with scalar parameters $a > b > 0$. We set $a = 5.0$ and $b = 4.7$ to obtain a suited initial distribution of the precision parameters τ_k for our data. Using Bayes' theorem, the update rule for the posterior given data and $\boldsymbol{\mu}_k$ is

$$\begin{aligned} p(\tau_k; \{\boldsymbol{\beta}_{i\in k}\}, \boldsymbol{\mu}_k) &\propto f(\tau_k; a, b) \prod_{i\in k} \mathcal{F}(\boldsymbol{\beta}_i; \boldsymbol{\mu}_k, \tau_k) \\ &\propto f(\tau_k; a + N_K, b + \sum_{i\in k} \boldsymbol{\mu}_k^T \boldsymbol{\beta}_i) \end{aligned} \tag{4.17}$$

where N_k is the number of members of cluster k. We are not aware of an efficient method to analytically obtain samples from this distribution, but any one-dimensional Markov chain Monte Carlo method can be used to produce samples from this marginal. In our implementation, we use the slice sampling algorithm (Neal 2003), which is a particularly efficient Markov chain Monte Carlo method for one-dimensional distributions.

Prior on the mixing proportions π_k

The joint probability $p(c_1, c_2, ...c_N)$ of the class memberships of the samples $\boldsymbol{\beta}_i$ is a multinomial distribution parameterized by the unknown mixture parameters π_k:

$$p(\boldsymbol{c}|\pi_k) = \prod_{k=1}^{K} \pi_k^{N_k} \tag{4.18}$$

The multinomial distribution is also a member of the exponential family, and the conjugate prior for its parameter vector $\boldsymbol{\pi}$ is the Dirichlet distribution with a K-dimensional parameter vector $\boldsymbol{\alpha}$. If we set all elements $\alpha_k = \alpha/K$ with a scalar constant α, the Dirichlet distribution puts uniform probability mass on all possible values of π_k and has the form

$$\mathcal{D}(\pi; \alpha) = \frac{\Gamma(\alpha)}{\Gamma(\alpha/K)^K} \prod_{k=1}^{K} \pi_k^{\alpha/K - 1}. \tag{4.19}$$

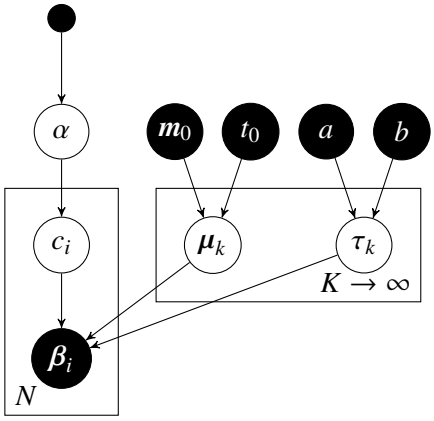


Figure 4.8.: Directed graphical model (Bayesian network) of the iMFMM representing both the generative process used to model the data, and the factorization properties of the joint distribution of all variables in the model. Any node in the graph is conditionally independent of the rest of the graph given values of its parents, its children and parents of its children. The deterministic quantities, such as the data $\boldsymbol{\beta}_i$ and the hyperparameters a, b, $\boldsymbol{m}_0$ and t_0 are depicted by filled circles, while probabilistic (latent) parameters are shown as hollow circles. The inverse Gamma prior on α is shown as a small black circle. Rectangles with label N and K are so-called "plates" representing N and K copies of their contents.

It is a crucial characteristic of the Dirichlet distribution that it is possible to integrate out the values of π_k under the posterior (Rasmussen 2000), leading to a joint distribution for the c_i which is only a function of α, K, and the cluster sizes N_k. It does not depend on the individual values of c_i:

$$p(c_1, ..., c_N; \alpha) = \frac{\Gamma(\alpha)}{\Gamma(N+\alpha)} \prod_{k=1}^{K} \frac{\Gamma(N_k + \alpha/K)}{\Gamma(\alpha/K)} \tag{4.20}$$

During Gibbs sampling, we condition on all but one particular sample. In this case the Gamma functions cancel and we arrive at the simple discrete conditional probability:

$$p(c_i = k | c_{\setminus i}, \alpha) = \frac{n_{\setminus i,k} + \alpha/K}{N - 1 + \alpha} \tag{4.21}$$

where $c_{\setminus i}$ indicates all indices except i and $n_{\setminus i,j}$ is the number of observations, excluding $\boldsymbol{\beta}_i$, that are associated with cluster k.

Also conditioning on the value of sample $\boldsymbol{\beta}_i$ yields the Gibbs sampling probability for c_i of:

$$p(c_i = k; \boldsymbol{c}_{\setminus i}, \alpha)\; p(\boldsymbol{\beta}_i | \boldsymbol{\mu}_k, \tau_k, \boldsymbol{c}_{\setminus i}) = \frac{n_{\setminus i,k} + \alpha/K}{N - 1 + \alpha}\; \mathcal{F}(\boldsymbol{\beta}_i; \boldsymbol{\mu}_k, \tau_k) \tag{4.22}$$

The likelihood for α itself can be derived from equation 4.20. Together with a prior of inverse Gamma shape (Rasmussen 2000), the posterior update for α is

$$p(\alpha; n, N) = \frac{\alpha^{n-3/2} \exp(-\frac{1}{2\alpha})\Gamma(\alpha)}{\Gamma(N+\alpha)} \tag{4.23}$$

As there is no efficient analytical sampling scheme for this distribution, we apply the slice sampling algorithm to obtain updates for α (Neal 2003).

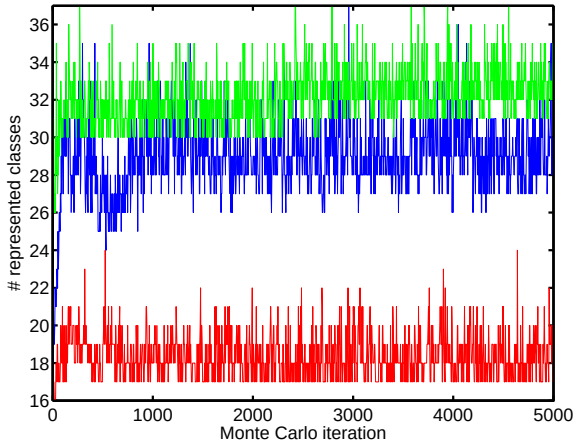


Figure 4.9.: Number of represented classes K, which corresponds to the number of beams, for the initial 5000 iterations for an intracranial (red), pancreas (green), and prostate case (blue).

The infinite limit

So far, we have assumed a constant number of clusters. But, sidestepping a few technicalities (Neal 2000, Rasmussen 2000, Griffiths and Ghahramani 2005), it is intuitively easy to take the limit of equation 4.22 for $K \to \infty$. All clusters containing more than one sample, i.e. $n_{\backslash i,k} > 0$, retain a finite probability

$$\frac{n_{\backslash i,j}}{N-1+\alpha} \cdot p(\boldsymbol{\beta}_i; \boldsymbol{\mu}_k, \tau_k) \tag{4.24}$$

of being chosen. And because the overall probability of choosing any cluster has to be 1 and all clusters have parameters with the same prior distribution $p(\boldsymbol{\mu}, \tau)$, all infinitely many remaining clusters together have the finite probability

$$\frac{\alpha}{N-1+\alpha} \int p(\boldsymbol{\beta}_i|\boldsymbol{\mu}, \tau) \cdot p(\boldsymbol{\mu}, \tau)\, d\boldsymbol{\mu}\, d\tau \tag{4.25}$$

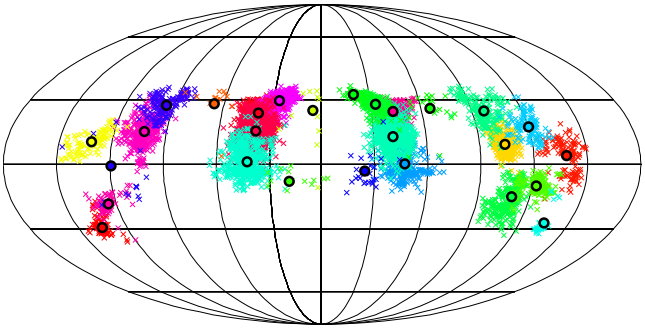
of being chosen. The integral in this equation can be approximated by Monte Carlo integration, i.e. generating l samples from the prior for $\boldsymbol{\mu}$ and τ, summing their likelihood terms $p(\boldsymbol{\beta}_i; \boldsymbol{\mu}, \tau)$, and dividing by l. We found that often even $l = 1$ is sufficient for convergence of the Gibbs sampler.

The resulting probability measure over probability measures, widely known as the "Dirichlet process", controls the number of mixture components. The limit-construction used here is known as the "Chinese Restaurant Process" (Griffiths and Ghahramani 2005).

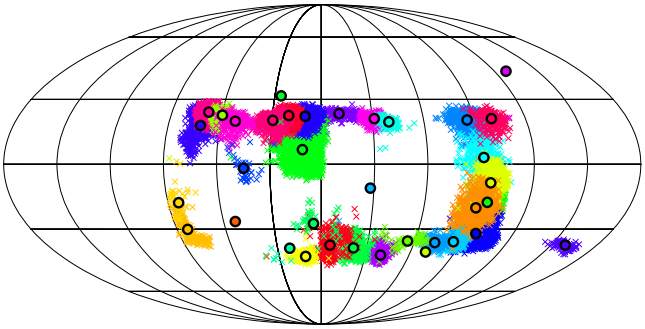
4.3.2. Results

For inference on our spherical beam angle data set, we initialize the iMFMM with K_0 mixture components. K_0 is sampled uniformly from the interval $[10, 20]$. The corresponding starting parameters for $\boldsymbol{\mu}_k$ and τ_k are drawn from the prior and subsequently posterior updates on all model parameters are performed during a large number of Gibbs sampling iterations. The initialization of K_0 reflects our expectation regarding the number of classes underlying the beam angle data set. It reduces the number of iterations needed for burn-in and leaves the long term behavior of the iMFMM unaltered. Here, "burn-in" denotes the drift from the initial Markov chain state to regions of high probability mass.

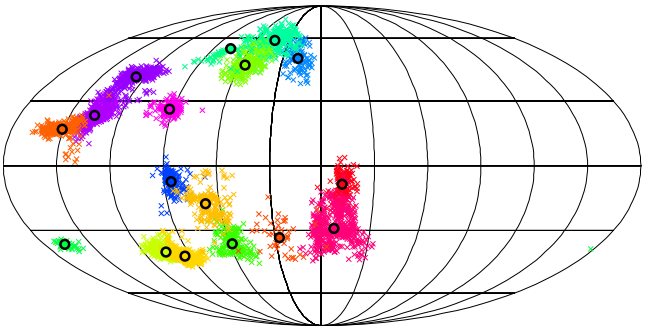
Figure 4.9 shows the number of represented classes K during the first $5,000$ iterations for three patient data sets. This corresponds to the belief of the iMFMM regarding the most probable number of clusters constituting



(a) Prostate lesion



(b) Pancreas lesion



(c) Intracranial lesion

Figure 4.10.: One iMFMM sample for the spherical data sets of locally ideal beam angles of a prostate (a), pancreas (b), and intracranial lesion (c). The longitude corresponds to the angle around the patient axis, the latitude to the angle towards the transversal plane. The centroids of the mixtures $\boldsymbol{\mu}_k$ are indicated by black circles, the color coding corresponds to the current assignment of the data to the mixtures c_i. We did not attempt to visualize the precisions τ_k. Note the relatively large number of clusters required in 4.10(b), as the von Mises-Fisher distribution cannot model directional correlation. Considering 4.10(a) and 4.10(b) it might be straight forward to identify a smooth path on the sphere passing by the cluster centroids and respecting potential physical limitations of the irradiation device for improved arc therapy.

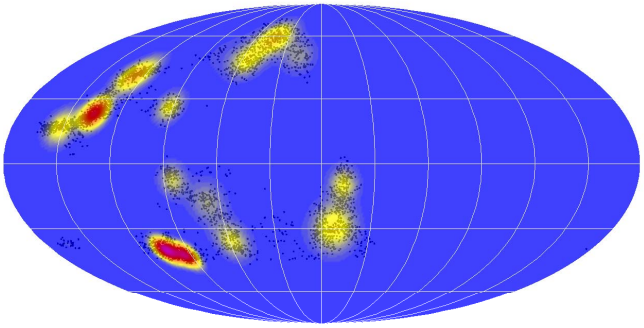


Figure 4.11.: Mollweide projection of a set of locally ideal beam angles $\mathbb{B}$ for an intra cranial lesion. The longitude corresponds to the angle around the patient axis, the latitude to the angle towards the transversal plane. The data, indicated by black dots, is overlaid by a density estimate of the iMFMM averaged over $1,000$ Monte Carlo iterations.

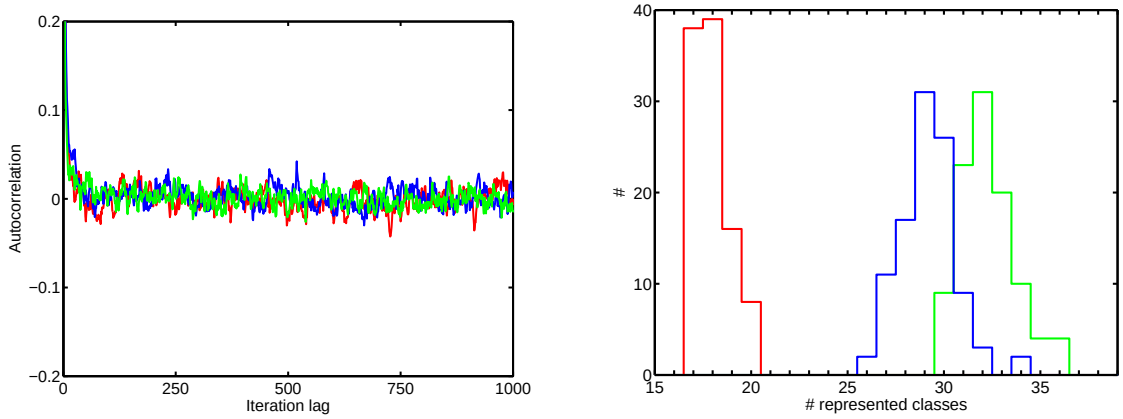
the distribution of locally ideal beam orientations $\mathbb{B}$. The number of classes, i.e. the number of beams of the final IMRT treatment plan, is rapidly adjusted by the iMFMM within the first 100 iterations and subsequently undergoes slight fluctuations for the intracranial and prostate data set. We discard the first $2,000$ iterations for burn-in. The state of the iMFMM in a single iteration after burn-in, i.e. its current estimate of the number of cluster components, their means and precisions, and the assignment of data points to cluster components, is shown in figure 4.10 for three clinical data sets.

If we want to evaluate the full posterior distribution of all samples from the Gibbs scheme, we have to average over a large number of iterations. Figure 4.11 shows the resulting density estimate overlying the set of locally ideal beam orientations obtained by averaging over $1,000$ samples. Simply put, figure 4.11 may be considered an average over multiple instances of figure 4.10(c).

Within a complex Gibbs sampling scheme, it may be possible, that subsequent iterations of the iMFMM are statistically dependent. The autocorrelation of the number of represented classes after $2,000$ iterations is shown in figure 4.12(a). We do not observe a significant correlation for any of the three data sets under investigation. The effective autocorrelation length, computed as the sum of the autocorrelation between an iteration lag of $-1,000$ and $1,000$ (Rasmussen 2000), does not exceed 10 iterations for the three data sets investigated.

For a concrete statement regarding the number of represented classes (which in our case will correspond to the number of treatment beams) we now draw 100 independent iMFMM samples. More precisely, we evaluate 100 samples after burn-in which are each separated by one autocorrelation length. Figure 4.12(b) shows the spectra for K for the three data sets under investigation. As the data does not stem from a mixture of von Mises-Fisher distributions but from our formulation of the BAS problem, there is not a distinct number of components that explain our data. The iMFMM yields a distribution over the probability of the number of represented components. It found the highest probability for the intracranial data set to be generated by a mixture of $K^*_{\text{intra}} = 18$ von Mises-Fisher distributions. The analysis for the pancreas and prostate data sets yields highest probabilities for mixtures of $K^*_{\text{pancreas}} = 32$ and $K^*_{\text{prostate}} = 29$ von Mises-Fisher distributions, respectively.

Given the number of represented components K^*, it is straightforward to infer exact orientations for the treatment beams: Using a finite von Mises Fisher mixture model with K^* components, we maximize the posterior probability of the parameters $\boldsymbol{\mu}_k$, τ_k, and c_i. For the purpose of localizing the treatment beams, only the centroid positions $\boldsymbol{\mu}_k$ will be of interest.



(a) (b)

Figure 4.12.: (a) Autocorrelation of the number of represented classes after 2,000 iterations for an intracranial (red), pancreas (green), and prostate case (blue). (b) Histograms of the number of represented classes for an intracranial (red), pancreas (green), and prostate case (blue) displaying the frequency of the number of represented classes.

4.3.3. Discussion

In extension of previous work on density estimation with Dirichlet process mixture models in Cartesian spaces (Rasmussen 2000, Neal 2000), we introduced the infinite von Mises-Fisher mixture model as a general framework for density estimation on the D-sphere. We constructed conjugate priors for its parameters $\boldsymbol{\mu}$ and τ, and derived a Markov chain Monte Carlo inference algorithm relying on Gibbs sampling scheme for posterior inference from data. Using a nonparametric Dirichlet process prior, the iMFMM infers probability distributions over both the number of clusters and their parameter values. In contrast to clustering methods based on point estimates, such as K-means clustering or factor analysis, the iMFMM, being a fully probabilistic method, does not suffer from the problem of overfitting, and can automatically determine the range of clusters to use.

The iMFMM was applied to infer a treatment beam ensemble for external radiation therapy based on a set of locally ideal beam angles distributed on the unit sphere. For the data sets studied, the iMFMM returns mixtures of $18-32$ beam orientations. This represents a considerable dimensionality reduction from the original data sets containing $\sim 10^4$ locally ideal beam directions. However, in conventional linear accelerator treatments, which typically include $5-11$ beam orientations, $18-32$ beams are not feasible. In order to arrive at a number of beams that is efficient in a clinical setting, heuristic merging of clusters could be used, but as we already have a functional alternative to cluster beam orientations with a predefined number of clusters this would actually not be a real improvement. Extending the infinite mixture model to integrate directional correlation on the sphere, which implies the non trivial transition from von Mises-Fisher distributions to Kent distributions, might only result in a slight decrease of the number of beams (consider figure 4.10(b)). Unfortunately it is impossible to directly bias the probabilistic process of the iMFMM that decides about the number of clusters because this compromises the convergence properties of the iMFMM.

As the iMFMM cannot be easily adjusted to yield less beams, we see two potential applications within radiation therapy treatment planning that may benefit from the analysis of the set of locally ideal beam orientations with the iMFMM. First, the iMFMM might be used with slight modifications for beam ensemble selection for robotic radiosurgery treatments, where typically $50-100$ beam orientations are accessed. Second, the iMFMM beliefs could be used to find beneficial trajectories for arc therapy, where the treatment beam is rotated around the patient during irradiation. Yang et al. (2011) showed that arc therapy may be improved using non-coplanar

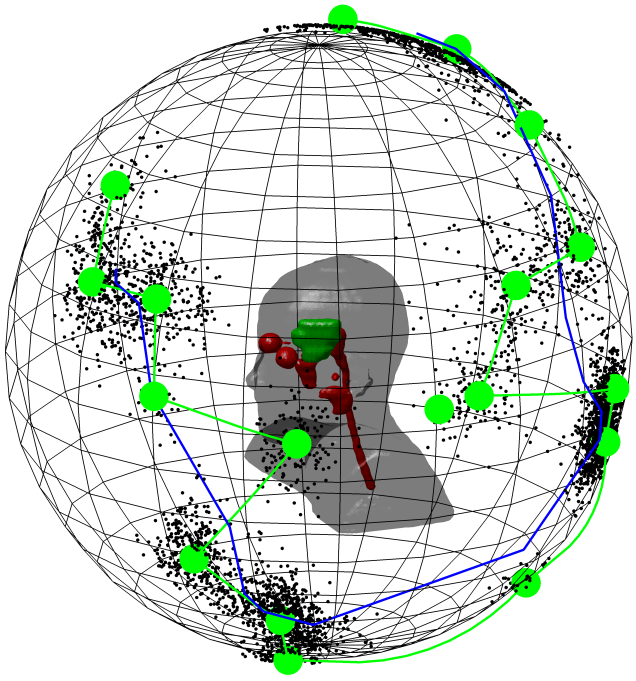


Figure 4.13.: The iMFMM could be used to find optimized trajectories for arc therapy. First, the iMFMM finds a low complexity representation (large green dots) of the set of locally ideal beam orientations (small black dots). Second, based on this low dimensional representation we can find a trajectory that passes by all locally ideal beam orientations (green curve). Third, this trajectory could be smoothened for the delivery with an actual irradiation device (blue curve). The locally ideal beam orientations and the trajectories are shown relative to the target volume (green) and the OARs (red) located within the patient (gray).

rotation trajectories. It is possible to define a path through the centroids of the von Mises-Fisher distributions found by the iMFMM. Such a path "passes by" all elements in the set of locally ideal beam orientations and may potentially yield improved treatment plans. For the visualization shown in figure 4.13, we solved a traveling salesman problem on the sphere with a genetic algorithm (green line) and smoothened the trajectory (blue line) with a von Mises-Fisher kernel. The smooth trajectory may be more convenient for the actual delivery with a linear accelerator. By considering the spatial relation of the target voxels that are associated with the locally ideal beam orientations it may even be possible to resolve potential sequencing issues before beamlet weight optimization.

The focus of this section was on the derivation of the iMFMM. Unfortunately, an application of the iMFMM for conventional IMRT treatment planning is not feasible, but it may be beneficial for robotic radiosurgery or arc therapy. A thorough assessment of the impact of the iMFMM on the quality of treatment plans for arc therapy and robotic radiosurgery is left for future research. The work presented in the previous section 4.2 has already provided strong evidence for the clinical value of the approximation to formulate the search for beneficial beam directions as a clustering problem of locally ideal beam angles on the three-dimensional sphere. The incorporation of the iMFMM into this framework implies only the replacement of the spherical K-means algorithm with a more sophisticated method.

4.4. Beam ensemble selection for particle therapy

Beam configurations for intensity-modulated particle therapy (IMPT) are usually selected based on the experience of human experts - just like beam configurations for conventional IMRT. Consequently, treatment regimes evolved over time for standard patient geometries, while beam configurations for unique and complex patient geometries have to be adopted during a tedious trial-and-error process. At the Francis H. Burr Proton Therapy Center of Massachusetts General Hospital, for example, prostate patients are irradiated with a standard beam configuration featuring a left and right laterally opposed field. An experienced radiation oncologist, however, may need more than 30 minutes for the manual selection of an adequate beam configuration for a complex intracranial lesion.

BAS for particle therapy received only limited attention in scientific journals (Moravek et al. 2009) but first studies presented at international conferences indicate that it is very difficult for a radiation oncologist to consider the non-intuitive interplay of intensity-modulated particle fields (Li et al. 2006, Speer et al. 2009). Currently, the possibilities of IMPT are not fully exploited for the benefit of the patient. The treatment plan quality could be improved by a dedicated BAS strategy - just like in IMRT.

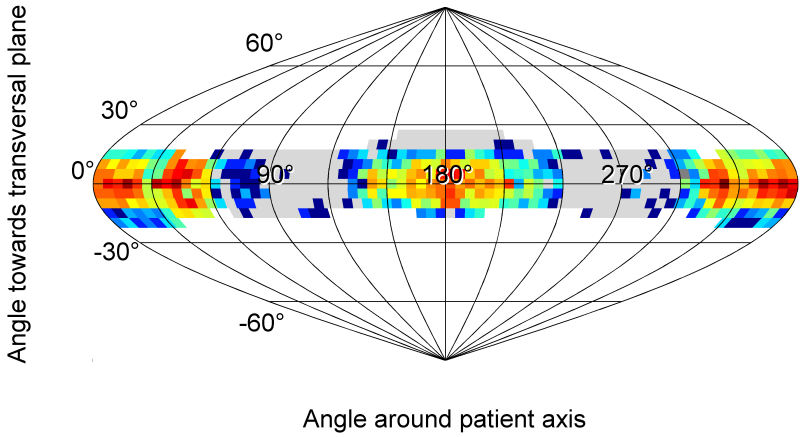


Figure 4.14.: Projection of the set of locally ideal beam angles $\mathbb{B}$ to two dimensions for the phantom case shown in figure 4.1(a) for protons. Dark red indicates regions of high density of locally ideal beam angles and dark blue indicates regions of low density of locally ideal beam angles. The area shaded in gray represents the set of candidate directions $\mathcal{B}$.

In principle, we could use spherical cluster analysis as is in combination with the score matrix S for BAS in particle therapy. Only the depth dose profile has to be adjusted for the calculation of the score S. Figure 4.14 depicts the resulting clustering landscape for the phantom case that is shown in figure 4.1(a).

Treatment plans for IMPT typically comprise two to four beams. With a decreasing number of beams, however, the clusters found by the spherical K-means algorithm span over a larger spatial angle. At some point, this compromises the validity of the low dimensional representation of the set of locally ideal beam orientations found by the clustering algorithm. The exact location of the cluster centroids and likewise the beam orientations become irrelevant. Consequently, an alternative strategy is developed to evaluate the score matrix S. This strategy exploits the full information content of S for BAS, not just the information about one locally ideal beam orientation per target voxel.

4.4.1. Method

A combinatorial beam angle selection framework based on score matrices

Recall that the best beam orientation $\boldsymbol{\beta}_v^*$ for the irradiation of a single target voxel v is given by

$$\boldsymbol{\beta}_v^* = \underset{\boldsymbol{\beta} \in \mathcal{B}}{\arg\min}\, S_{\boldsymbol{\beta} v} \tag{4.26}$$

since $\boldsymbol{\beta}_v^*$ maximizes the dose delivered to the target relative to the dose delivered to normal tissue and OARs weighted hundredfold. The best score s_v^* for target voxel v is

$$s_v^* = \min_{\boldsymbol{\beta} \in \mathcal{B}} S_{\boldsymbol{\beta} v}. \tag{4.27}$$

By summation over the best score values s_v^* for all target voxels v we can define a score for the entire target volume $\bar{s}^*$.

$$\bar{s}^* = \sum_v \min_{\boldsymbol{\beta} \in \mathcal{B}} S_{\boldsymbol{\beta} v} \tag{4.28}$$

According to this definition, the ideal beam ensemble B^* exactly corresponds to the set of locally ideal beam orientations $\mathbb{B}$

$$B^* = \mathbb{B} \tag{4.29}$$

because all beams within the set of locally ideal beam orientations contribute to $\bar{s}^*$. All treatment beams within the set of candidate beam orientations $\mathcal{B}$ may be a member of the ideal beam ensemble B^* resulting in a large number of treatment beams η.

$$\eta = |B^*| = |\mathbb{B}| \leq |\mathcal{B}| \tag{4.30}$$

In a clinical setting, however, we have to restrict the number of beams η and consequently we have to restrict the composition of the score for the entire target volume. If we are to find the best beam ensemble featuring one beam orientation only, we have to calculate the score $\bar{s}_1$ given by the summation over all target voxels v for a fixed beam orientation $\boldsymbol{\beta}$.

$$\bar{s}_1(\boldsymbol{\beta}) = \sum_v S_{\boldsymbol{\beta} v} \tag{4.31}$$

The ideal one-beam ensemble B_1^* is given by

$$B_1^* = \underset{\boldsymbol{\beta} \in \mathcal{B}}{\arg\min} \left(\sum_v S_{\boldsymbol{\beta} v} \right). \tag{4.32}$$

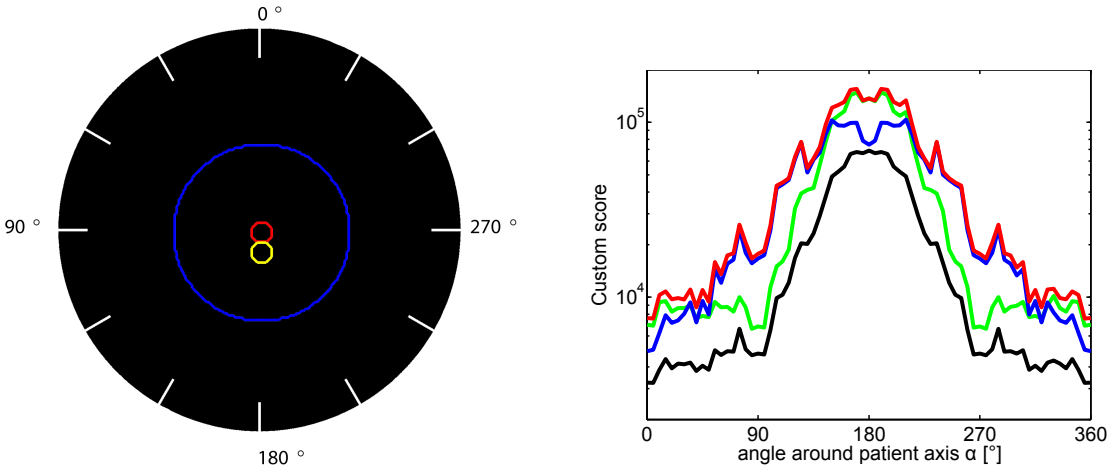
For an arbitrary beam ensemble B featuring η beams the score for the entire target volume $\bar{s}_\eta$ is given by

$$\bar{s}_\eta(B) = \sum_v \left(\min_{\boldsymbol{\beta} \in B} S_{\boldsymbol{\beta} v} \right) \quad \text{with} \quad B \in \mathcal{P}_\eta(\mathcal{B}) \tag{4.33}$$

$\mathcal{P}_\eta(\mathcal{B})$ denotes the set of subsets of $\mathcal{B}$ with η elements. Equation 4.31 is a special case of equation 4.33 for $\eta = 1$. The best η-beam ensemble B_η^* is given by the subset of the set of candidate beam orientations $\mathcal{B}$ with η elements that minimizes $\bar{s}_\eta(B)$.

$$B_\eta^* = \underset{B \in \mathcal{P}_\eta(B)}{\arg\min} \sum_v \left(\min_{\boldsymbol{\beta} \in B} S_{\boldsymbol{\beta} v} \right) \tag{4.34}$$

Equation 4.34 represents an NP-hard combinatorial optimization problem. For a small number of beams $\eta < 4$ it may be solved efficiently by exhaustive search, i.e. a brute force evaluation of all combinations for $B \in \mathcal{P}_\eta(\mathcal{B})$. For a larger number of beams, we apply a genetic algorithm to find near optimal beam configurations. Note that the evaluation of a beam configuration according to equation 4.33 is very fast. It requires only a summation over selected elements of the precomputed score matrix S which typically comprises $1,000 \times 5,000$ elements, i.e. 40 MB in double precision.



(a) (b)

Figure 4.15.: (a) A phantom case. The circular target volume, outlined in red, is located in the center of a circular phantom, outlined in blue. A circular OAR, outlined in yellow, is located close to the target volume. b) Score $\bar{s}_1$ for a treatment plan applying $\eta = 1$ beam for the phantom case depicted in figure 4.15(a). The nominal score is given by the black curve, the worst case score considering setup uncertainties only is given by the blue curve, the worst case score considering range uncertainties only is given by the green curve, and the worst case score considering setup and range uncertainties is given by the red curve. A range and setup uncertainty of 3 mm is assumed for the score calculation. Note that the score function is rather noisy due to sampling artifacts because the target volume comprises only few voxels.

Robustness

IMPT treatment plans are very sensitive to uncertainties in the calculated range of the individual Bragg peaks and errors during patient setup (Lomax 2008a;b). It is essential to incorporate these uncertainties in order to establish robust IMPT treatment plans. Currently, margins and computationally very intensive probabilistic or worst case methods are applied to account for both range and setup uncertainties during beamlet weight optimization (Unkelbach and Oelfke 2004, Pflugfelder et al. 2008b). This section investigates means to account for the criterion of treatment plan robustness already during the selection of the beam ensemble by a straight forward extension of the score matrix S.

Therefore, the dosimetric aspects which are integrated into the score matrix S are reconsidered regarding potential range and setup uncertainties. We do so by computing a worst case score comparable to the worst case dose distribution introduced for robust fluence optimization (Pflugfelder et al. 2008b): $S_{\beta v}$ is not only calculated for the nominal range and patient alignment, but also for a set of worst case configurations featuring Bragg peaks that are displaced laterally and in depth for every target voxel v and candidate beam angle $\boldsymbol{\beta}$. During the subsequent beam ensemble selection process, only the worst, i.e. maximum score, will be considered. This methodology is discussed by means of a phantom case which is depicted in figure 4.15(a). For clarity, the considerations are restricted to the coplanar plane.

If we are to find the best treatment plan with $\eta = 1$ beam, the overall score for every irradiation angle is computed according to equation 4.31. The nominal score, visualized by the black curve in figure 4.15(b), suggests beam orientations impinging between 270° and 90° yield clinically almost equivalent treatment plans. The best beam orientation is given by $\alpha_{\text{nom}} = 0°$. Protons impinging from 0°, however, stop directly in front of the OAR. If the range calculation is not 100% correct, an irradiation from this direction may result in an increased

dose in the OAR. The score incorporating range uncertainties, visualized by the green curve, accounts for these uncertainties. It exhibits two minima at $\alpha_{\text{range}} = 90°$ and $\alpha_{\text{range}} = 270°$. For these beam orientations, "overshooting" does not result in an increased dose in the OAR. If we only account for setup uncertainties, the best beam orientation is given by $\alpha_{\text{setup}} = 0°$, as visualized by the blue curve. Lateral shifts of the patient do not result in an increased dose within the OAR from this direction. The red curve visualizes the score that accounts for both range and setup uncertainties. The best beam orientation according to the worst case score is $\alpha_{\text{wc}} = 0°$. Apparently, a setup error of 3 mm has a bigger impact on the worst case score of the beam orientations around 270° and 90° than a range uncertainy of 3 mm on the beam orientations around 0°.

Unfortunately, it does not suffice to select beam orientations with this methodology to establish robust treatment plans (Bangert and Oelfke 2009a). It is necessary to combine the BAS with techniques for robust fluence optimization, which are computationally very expensive.

Pflugfelder et al. (2008b) report about the characteristics of robust fluence profiles for IMPT treatments. For range uncertainties they observed that the dose gradients between the target and OARs are shaped using the lateral instead of the distal gradient of the Bragg peaks. For setup uncertainties they observed that the beam dose is reduced for parts of the beam which hit the patient at a narrow angle, as there a small shift leads to a large change in radiological depth. We suggest to use the information about the relative quality of a beam orientation $\boldsymbol{\beta}$ for the irradiation of target voxel v, which was already computed for the score matrix S and used for BAS, to enforce such robust fluence profiles a priori. The score matrix S may not only be used to find a beneficial beam ensemble B, but also to identify the beam orientation $\boldsymbol{\beta} \in B$ that is suited best for the irradiation of of a certain spot position within the target.

In contrast to conventional fluence optimization, where all beam spots within the target may be irradiated from every beam $\boldsymbol{\beta}$ of the treatment beam ensemble B, only the best beam orientation for every spot according to matrix S is selected. Therefore all spot positions are mapped to the closest target voxel. The beam orientation $\boldsymbol{\beta}_v^*$ used for irradiation of voxel v is given by

$$\boldsymbol{\beta}_v^* = \underset{\boldsymbol{\beta} \in B}{\arg\min}\, S_{\boldsymbol{\beta} v}. \tag{4.35}$$

This imposes automatic restrictions on the fluence modulation of the individual subfields in order to automatically compute robust treatment plans even though a conventional fluence optimization is applied. The assignment of individual spot positions within the target volume to certain beam orientations of the treatment beam ensemble may be considered an intertwined patch field approach.

The strategy is evaluated for a phantom case featuring a C-shaped target surrounding an OAR. Three different scenarios are compared: Conventional fluence optimization, robust fluence optimization according to the worst case optimization approach introduced by Pflugfelder et al. (2008b), and conventional fluence optimization in combination with the modulation restrictions according to the score matrix S. All treatment plans use the same optimized three-beam ensemble. Corresponding worst case dose distributions are employed to evaluate the robustness of the three scenarios. The study focuses on the influence of range uncertainties.

4.4.2. Results

We calculated a score matrix S that included the nominal scenario and two worst case scenarios modeling an uncertainty of ±3 mm in the calculated range of the individual Bragg peaks. The best three beam configuration according to equation 4.33 was 10°, 255°, and 320°. Figure 4.16 shows the dose deposition of the individual proton fields for the conventional fluence optimization, the robust worst case fluence optimization, and the fluence optimization using only one beam direction per spot position within the target. The conventional fluence optimization exploits the steep distal dose fall-off behind the Bragg peak to shape a conformal dose distribution.

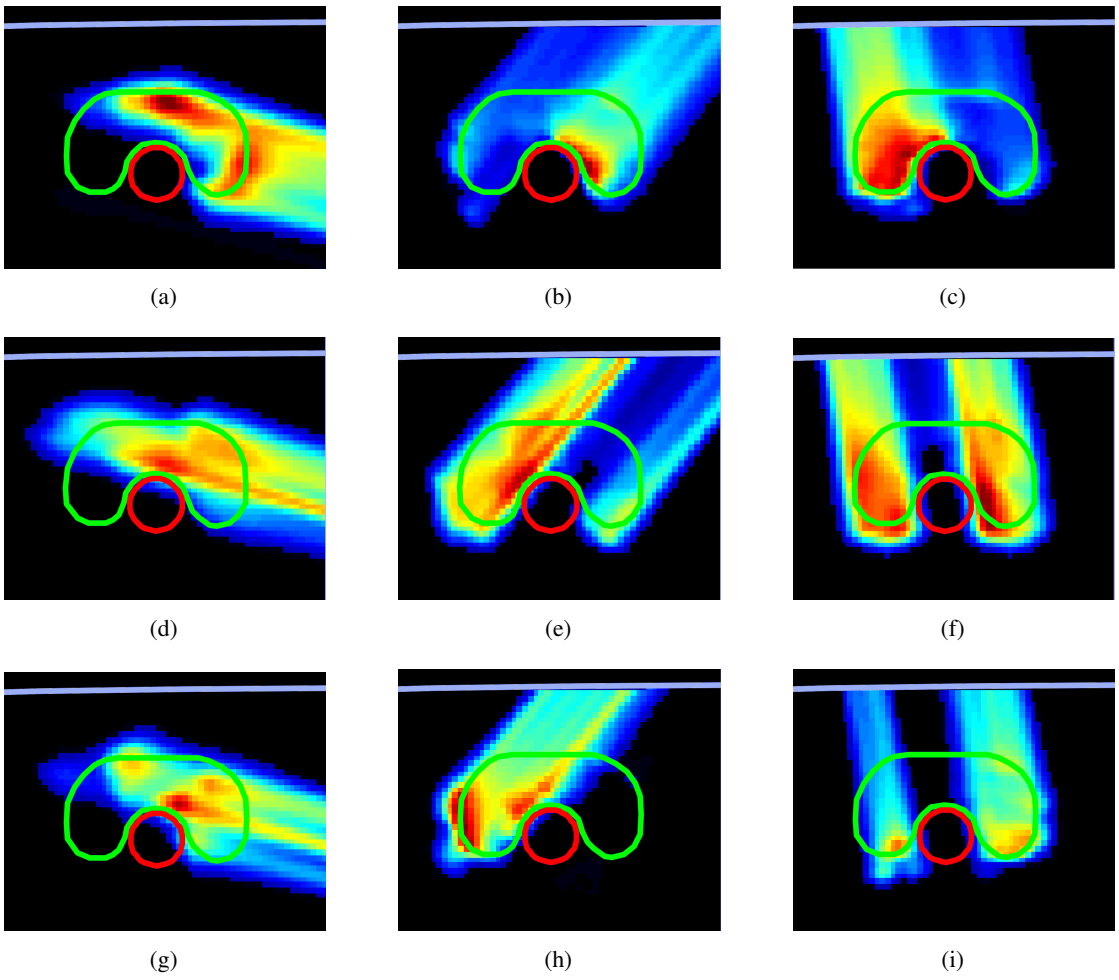
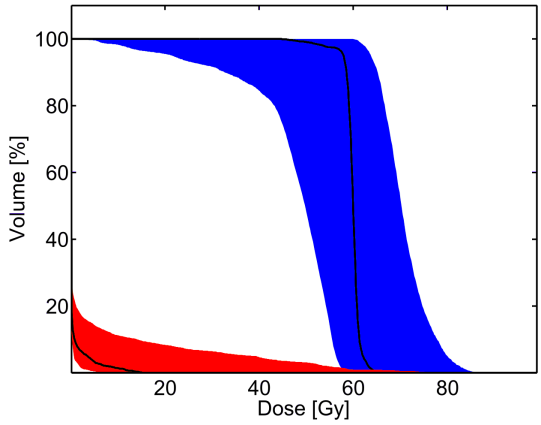


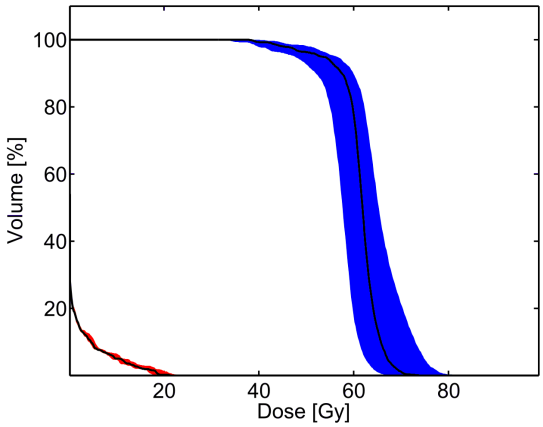
Figure 4.16.: Dose distribution of the individual proton fields on a selected transversal CT slice. The target volume is outlined in green, the OAR is outlined in red, and the normal tissue is outlined in purple. (a), (b), and (c) depict individual proton fields obtained by conventional fluence optimization. (d), (e), and (f) depict individual proton fields obtained by robust worst case fluence optimization (Pflugfelder et al. 2008b). (g), (h), and (i) depict individual proton fields obtained by fluence optimization using only one beam direction per spot position within the target.

Protons stop directly in front of the OAR, as shown in figures 4.16(b) and 4.16(c). The worst case fluence optimization, in contrast, exploits the lateral dose fall-off to shape a conformal dose distribution, as shown in figures 4.16(d), 4.16(e), and 4.16(f). The same holds for the fluence optimization using only one beam direction per spot position within the target. However, figures 4.16(g), 4.16(h), and 4.16(i) show that the fluence modulation within the target volume is more pronounced in comparison to the individual fields of the worst case fluence optimization.

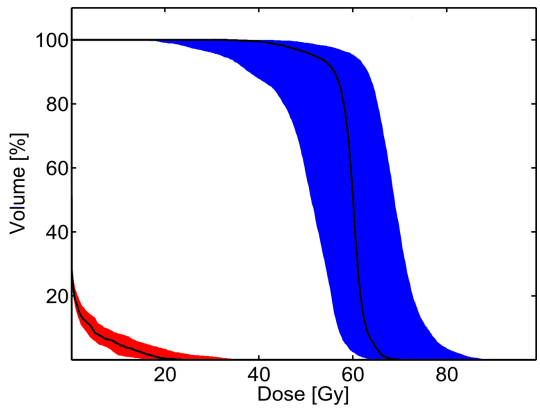
Figure 4.17 shows the corresponding DVHs of the three investigated scenarios. As expected, the treatment plan established by worst case beamlet weight optimization is associated with the smallest uncertainties and the treatment plan established by conventional beamlet weight optimization is associated with the largest uncertainties. The treatment plan established by fluence optimization using only one beam direction per spot position within the target has a significantly reduced uncertainty within the OAR compared to the conventional fluence optimization. For the target volume, however, only marginal improvements are observed.



(a)



(b)



(c)

Figure 4.17.: DVHs for the target volume and OAR for (a) conventional fluence optimization, (b) robust worst case fluence optimization (Pflugfelder et al. 2008b), and (c) fluence optimization using only one beam direction per spot position within the target. The areas shaded in blue and red indicate the uncertainty associated with the nominal DVH data based on a worst case dose distribution as defined by Pflugfelder et al. (2008b) considering a range uncertainty of ±3 mm.

4.4.3. Discussion

In combination with an alternative BAS algorithm, the score matrix introduced in section 4.1 is applicable for BAS in particle therapy. The selected beam orientations are plausible and yield beneficial treatment plans. The choice for our custom score as defined in equation 4.2, however, showed two disadvantages for protons. First, the calculation of the worst case score scenario may become numerically instable if the target dose d_{Target} approaches 0 Gy. This may happen in the case of "under shoot" where proton spot positions on the proximal edge of the target volume get fully retracted from the target volume. Second, a constant weighting factor of 100 may yield an over estimated impact on the score if an OAR receives only dose by the distal fall-off behind the Bragg peak. This effect may be even more severe for carbon ions due to the fragmentation tail after the Bragg peak. Both issues may be resolved with a score that facilitates custom maximum doses as suggested by Pugachev and Xing (2001b) but still assigns an individual score for every target voxel. A thorough investigation of such an approach in the context of particle therapy is left for future research.

The alternative strategy to evaluate the score matrix S exploits its full information content. While spherical cluster analysis incorporates only information about one locally ideal beam orientation $\boldsymbol{\beta}_v^*$ per voxel, the combinatorial approach considers the individual trade-off for every voxel. On the one hand, if a voxel v may be irradiated equally good from every direction according to the score matrix S, changing the beam direction for this very voxel does not have a big impact on the combined score $\bar{s}_\eta$ of the final beam ensemble. Hence, the relative influence of voxel v on the final beam configuration is very limited. On the other hand, if a voxel v may only be irradiated very well from a certain direction $\boldsymbol{\beta}$ according to the score matrix S, there is a large incentive to include this very beam direction into the final beam ensemble in order to minimize the combined score $\bar{s}_\eta$ of the final beam ensemble. For spherical cluster analysis, however, all target voxels are represented by one locally ideal beam orientation and consequently all target voxels have equal influence on the final beam ensemble. In principle, these aspects are also valid for BAS for photon therapy (Bangert and Oelfke 2009b). However, potential conceptual advantages over the spherical cluster analysis of the set of locally ideal beam orientations did not translate into superior treatment plans for photons due to the increased number of beams ($\eta \geq 5$). Furthermore, the performance of the combinatorial approach is compromised by the increased computation times of the genetic algorithm and spherical clustering analysis may be considered a more intuitive approach.

The measures taken to incorporate the influence of uncertainties on BAS yield plausible beam configurations. As expected, BAS based on robustness criteria alone does not suffice to establish robust treatment plans. It is indispensable to incorporate methods for robust beamlet weight optimization. The approach to impose restrictions on the fluence modulation patterns for individual subfields according to the BAS matrix S before beamlet weight optimization yields encouraging yet not satisfying results. The characteristics of robust treatment plans are well-known (Pflugfelder et al. 2008b); consequently it may be worthwhile to investigate the potential of more sophisticated restrictions on the fluence modulation for IMPT treatment planning. Just as it is possible to resolve sequencing issues for IMRT with smoothness constraints on the modulation of individual fields, it might be possible to enforce robustness of IMPT treatment plans with comparable concepts.

4.5. Geometric beam ensemble selection

In a 2008 paper, Potrebko et al. (2008) introduce anatomic beam orientation optimization (A-BOO), a strategy for BAS which is solely based on geometric analysis of the patient anatomy. For A-BOO, triangulated surface mesh data is analyzed to identify beam orientations that impinge tangentially to the surface elements of the target volume. As shown in figure 1.1(a), photon beams exhibit the steepest gradient in lateral direction. Hence, beam orientations that impinge parallel to surface elements of the target result in steep gradients at this surface element.

A-BOO provides a ranking of candidate directions which serves as guidance for a radiation oncologist deciding about the beam ensemble for an IMRT treatment plan. Potrebko et al. were able to show that A-BOO helps to significantly improve the quality of treatment plans for gastric, prostate and oropharynx cases within minutes of computation time. Even though A-BOO is a highly useful and fast algorithm for BAS, it remains in some respect incomplete: A-BOO requires significant user interaction when the radiation oncologist composes a treatment beam ensemble based on the ranking of beams provided by A-BOO.

It is possible to use the technique of spherical cluster analysis, which was introduced in section 4.2 for the analysis of score matrices, to enhance A-BOO. By analyzing the weighted spatial distribution of beneficial beam orientations with a spherical K-means clustering algorithm, treatment beam ensembles may be selected in a fully automated process.

4.5.1. Method

Again, the origin of the coordinate system underlying our considerations is defined as the isocenter of the patient anatomy and a beam direction is represented by a point $\boldsymbol{\beta}$ on the three dimensional unit sphere.

A-BOO is based on a triangulation T of the target surface, i.e. an approximation by triangular surface elements. A triangular surface element $t_i \in T$ may described by the three vertices of the triangle given as vectors $\boldsymbol{a}_i$, $\boldsymbol{b}_i$, and $\boldsymbol{c}_i$. To identify the beam orientations that impinge parallel to the target surface element t_i, the normal vector $\boldsymbol{n}_i$ is computed

$$\boldsymbol{n}_i = \frac{(\boldsymbol{c}_i - \boldsymbol{a}_i) \times (\boldsymbol{b}_i - \boldsymbol{a}_i)}{|(\boldsymbol{c}_i - \boldsymbol{a}_i) \times (\boldsymbol{b}_i - \boldsymbol{a}_i)|}. \tag{4.36}$$

The set of beam orientations P_i that impinge parallel to target surface element t_i are given by

$$P_i = \left\{ \boldsymbol{\beta}^* \;|\; \arccos \frac{\boldsymbol{\beta}^{*\mathsf{T}} \boldsymbol{n}_i}{|\boldsymbol{\beta}^{*\mathsf{T}} \boldsymbol{n}_i|} = \pi \right\}. \tag{4.37}$$

P_i may be considered as the set of locally ideal beam orientations for target surface element t_i. It can be represented by a circle on the unit sphere. Remember, that we worked with a single locally ideal beam orientation per target voxel in section 4.2. Here, we are exercising the equivalent approach with a "circle" of locally ideal beam orientations per target surface element.

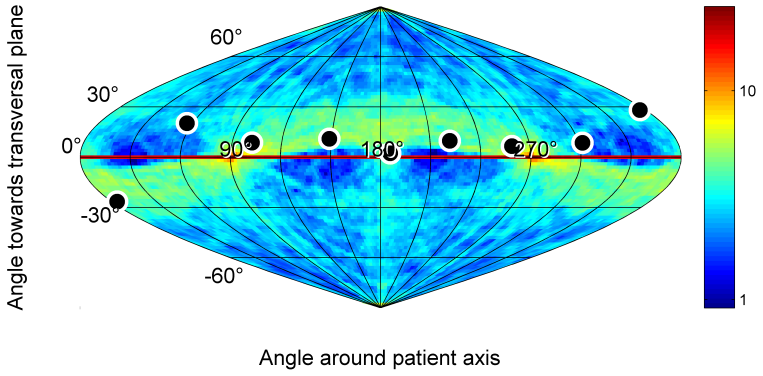
In a discrete representation of the patient anatomy, P_i is given by the candidate beam orientations within $\mathcal{B}$, that are almost perpendicular to the normal vector of the target surface element $\boldsymbol{n}_i$.

$$P_i = \left\{ \boldsymbol{\beta}^* \in \mathcal{B} \;|\; \left|\arccos \frac{\boldsymbol{\beta}^{*\mathsf{T}} \boldsymbol{n}_i}{|\boldsymbol{\beta}^{*\mathsf{T}} \boldsymbol{n}_i|} - \pi\right| < \epsilon \right\}. \tag{4.38}$$

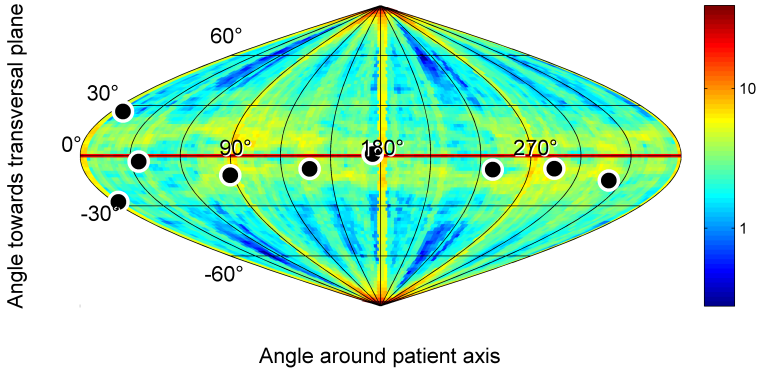
ϵ is set to the angular spacing of candidate beam directions. The set of locally ideal beam orientations for the entire target volume $\mathbb{B}$ is given by the set of all P_i of the individual target surface elements. Note that $\mathbb{B}$ may have duplicate elements.

$$\mathbb{B} = \{P_i \mid i \in T\} \quad \text{where} \quad |\mathbb{B}| = \sum_i |P_i|. \tag{4.39}$$

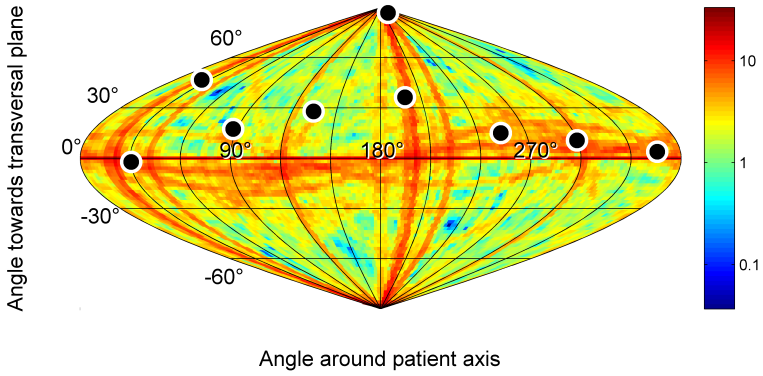
Figure 4.18 visualizes the set of locally ideal beam orientations for three selected clinical cases. Compared to the earlier clustering approach, $\mathbb{B}$ is simply an alternative spherical representation of the BAS problem. It does not rely on the evaluation of a score matrix S but on geometric considerations. We are going to show that it is



(a)



(b)



(c)

Figure 4.18.: Projection of the set of locally ideal beam angles $\mathbb{B}$ (given as points on the 3-sphere) to two dimensions. Red indicates regions of high density of locally ideal beam angles and blue indicates regions of low density of locally ideal beam angles. (a) Prostate lesion I, (b) pancreas lesion III, and (c) intracranial lesion III. The beam directions of the optimized nine-beam-plan for all three cases are visualized by black dots. Note that the density of locally ideal beam orientations relates to a true physical quantity: as the influence of a target surface element is weighted according to its area, the color coding corresponds to the percentage of the area of the target surface that can be irradiated tangentially from this beam orientation.

possible to use the same spherical clustering algorithm with slight modifications to find a beam ensemble for IMRT treatment planning.

In order to create an IMRT treatment plan featuring η beams, η centroids $\boldsymbol{\mu}_k^0 \in \mathbb{B}$ are randomly selected during initialization. As previously discussed, K-means clustering alternates between an assignment step, where every beam direction $\boldsymbol{\beta}_i^* \in \mathbb{B}$ is assigned to its closest centroid, and an update step, where the estimate for a centroid is refined to the mean direction of all beam directions assigned to this centroid.

For spherical cluster analysis for BAS based on geometric data, however, two modifications of the algorithm are necessary. First, the update of the centroid estimate has to be modified because the triangular target surface elements have different areas. In order to guarantee that the influence of a target surface element is proportional to its area, the area A_i of every target surface element t_i is computed

$$A_i = \frac{1}{2} \cdot |(\boldsymbol{c}_i - \boldsymbol{a}_i) \times (\boldsymbol{b}_i - \boldsymbol{a}_i)| \tag{4.40}$$

and the updated estimate of each centroid is given by a weighted average of all associated beam orientations

$$\boldsymbol{\mu}_k^{j+1} = \frac{\bar{\boldsymbol{\beta}}_k^*}{\left\|\bar{\boldsymbol{\beta}}_k^*\right\|} \quad \text{with} \quad \bar{\boldsymbol{\beta}}_k^* = \sum_{c^j(\boldsymbol{\beta}_i^*) = k} A_i \boldsymbol{\beta}_i^*. \tag{4.41}$$

Second, we have to consider that opposing beam orientations are in fact parallel to the same target surface elements (this is why the clustering landscapes are point symmetric to the origin of the coordinate system). Consequently the assignment c^j of beam orientations $\boldsymbol{\beta}^* \in \mathbb{B}$ to its closest centroid has to be invariant regarding parallel and antiparallel beam orientations.

$$c^j(\boldsymbol{\beta}_i^*) = \arg\max_k \left\{ \left| \boldsymbol{\beta}_i^{*\mathsf{T}} \boldsymbol{\mu}_k^j \right| \right\} \tag{4.42}$$

Note that this symmetry regarding parallel or antiparallel beam orientations is also important for the calculation of the updated mean directions $\boldsymbol{\mu}_k^{j+1}$. In order to avoid cancellation effects of antiparallel beam orientations, some beam orientations may have to be "flipped" so that all beam orientations that are associated with centroid k "point" in the same direction. Furthermore, the point symmetry may be exploited to fine-tune the final beam ensemble. As mirroring a beam orientation with respect to the origin of the coordinate system results in exactly the same solution according to the metric of our clustering algorithm, we may flip individual beams to accommodate additional objectives of treatment planning. If a beam would have to travel a long distance through normal tissue or an OAR before hitting the target volume, for example, we could simply use the exactly opposing beam orientation instead. For this study, however, such adjustments are not performed.

Potrebko et al. suggested to additionally weight the influence of a target surface element by its distance to the closest OAR. According to Potrebko et al. (2008), the impact of a target surface element should increase with decreasing distance to an OAR. Our studies showed for the considered cases that the suggested distance weighting does not affect the final beam configurations. Consequently we did not include this weighting into the automated BAS process. We argue that close proximity of an OAR to the target already implies a larger target surface parallel to this OAR. The prostate, for instance, is located between rectum and bladder. Hence, the prostate takes a flattened shape with large interfaces towards the rectum and bladder. The weighting of target surface elements according to their area alone guarantees a dominating impact of this OAR-target interfaces during the BAS process. Equivalent arguments hold for target volumes surrounding the brainstem or the optic nerve.

4.5.2. Results

The suggested method for BAS is evaluated in a treatment plan comparison for the same three prostate, pancreas, and intracranial cases already used in section 4.2. The optimized beam ensembles applying nine, seven, or five beams were compared to beam ensembles applying nine equi-spaced beams. The fluence profiles of all treatment plans were optimized with our in-house treatment planning software KonRad (Preiser et al. 1997) using a standard quadratic objective function (Oelfke and Bortfeld 2001). We used the same maximum and minimum dose constraints and penalties for all treatment plans. Table 4.2 gives a detailed overview of the nine patient cases under investigation.

The triangulations are computed with our in-house treatment planning software VIRTUOS (Bendl et al. 1994). The approximations of the target volumes comprise up to 3,000 target surface elements. The analysis of the triangulated surface data is implemented in MATLAB[3]. The overall BAS computation time considering about 10′000 non-coplanar candidate beam orientations is about 2 s on a 2.3 GHz workstation. Beam orientations that would result in a collision of the treatment head with the patient or the couch are excluded from the set of candidate beam directions. Figure 4.19 shows the graphical user interface for BAS. As the final evaluation of a treatment plan is based on the dose distribution calculated with KonRad, this tool is in principle ready for first clinical testing.

Figure 4.18 visualizes the clustering landscapes and the resulting beam configurations for one selected case per treatment site. The clinical benefit of the suggested BAS strategy is discussed individually for every site in the following three paragraphs. Detailed statistics is displayed in table 4.2.

Prostate cases

The reference plans for the three prostate cases studied feature nine coplanar equi-spaced beams and are compared to treatment plans with an optimized configuration of nine or seven beams. Homogeneous doses of 76 Gy and 70 Gy are prescribed to the GTV and CTV, respectively. For patient II, inclusive DVHs are given in figures 4.20(a) and 4.20(b). Figures 4.21(a) and 4.21(b) contrast the dose distributions of the equi-spaced coplanar nine beam plan and the optimized nine beam plan. All prostate treatment plans feature clinically equivalent target coverage and the maximum doses for the rectum and bladder differ by less than 0.6 Gy for the optimized and original treatment plans. However, the mean rectum dose of patients I and III is increased by up to 3.8 Gy. As these plans apply maximum dose constraints of 50 Gy to the rectum, there is no incentive for the optimizer to confine medium and low doses. For the same reason, the mean dose to the bladder is increased by up to 5.3 Gy. We observe slight mean and maximum dose reductions for the right and left femoral heads with an optimized beam configuration. The three cases exhibit almost the same mean doses to the normal tissue.

Pancreas cases

The reference plans for the three pancreas cases studied feature nine coplanar equi-spaced beams and are compared to treatment plans with an optimized configuration of nine or seven beams. Homogeneous doses of 54 Gy and 45 Gy are prescribed to the GTV and CTV, respectively. For patient I, inclusive DVHs are given in figure 4.20(c) and 4.20(d). Figures 4.21(c) and 4.21(d) contrast the dose distributions of the equi-spaced coplanar nine beam plan and the optimized nine beam plan. The transversal slices of the dose distributions demonstrate that the dose distribution is more conformal for the optimized beam configurations.The three cases investigated exhibit equivalent target coverage. The optimized beam configurations applying nine beams allow for a slight reduction of the mean (maximum) dose in the left or right kidney of up to 1.6 Gy (2.7 Gy). Merely the maximum dose to the left kidney of patient I is increased by 0.7 Gy. There is no clear trend observed regarding dose

[3]MATLAB R2009b, The MathWorks, Nantick, MA

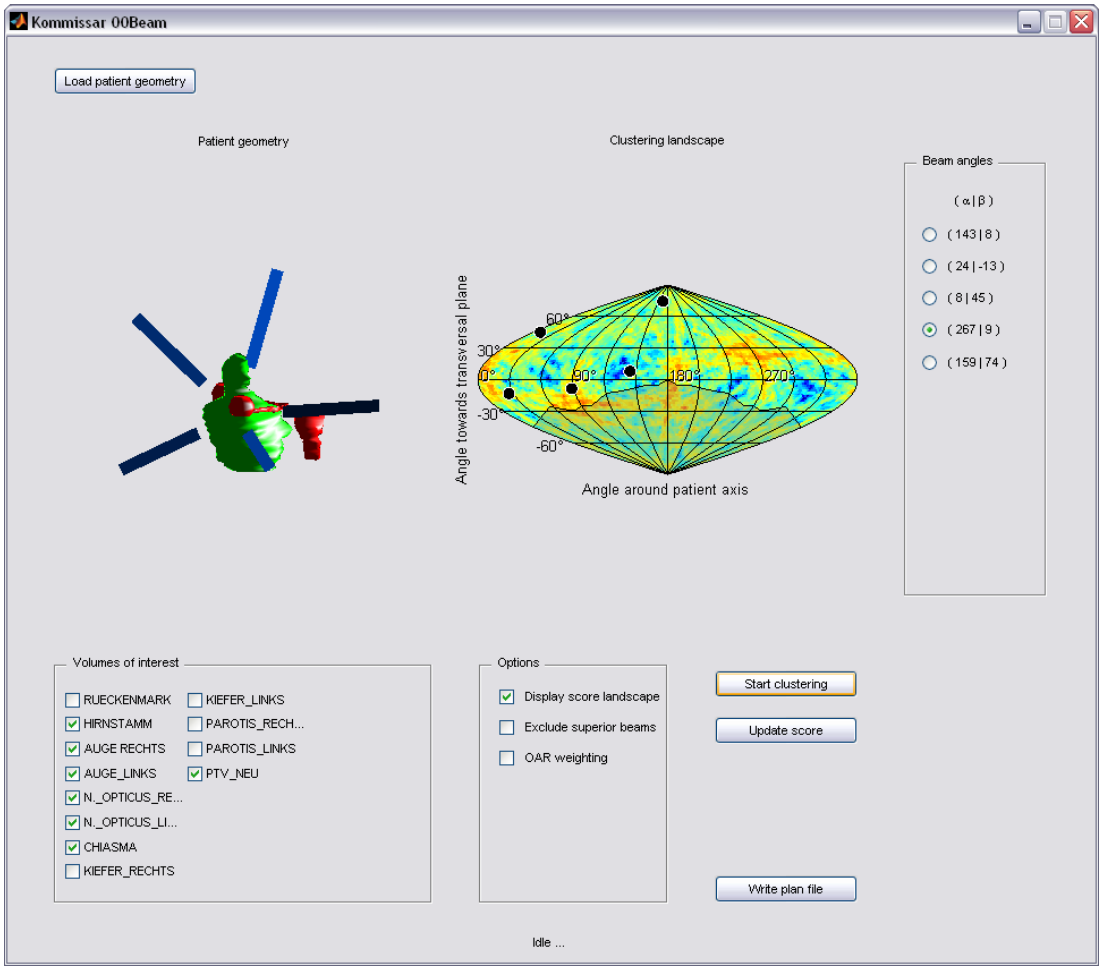


Figure 4.19.: Graphical user interface used for spherical cluster analysis of geometric data for BAS. It is possible to load a patient anatomy, visualize the triangulation of the patient anatomy, turn volumes of interest on and off, display the clustering landscape, exclude superior beams for treatments in the torus, include the OAR weighting proposed by Potrebko et al. (2008), cluster η beam orientations, flip individual beams of the solution, and write plan files for KonRad.

Prostate

#	Config	GTV	CTV	Rectum	Bladder	Femur (l)	Femur (r)	NT
I	9 equi	76.0	72.4	6.8 (75.5)	29.7 (77.2)	9.3 (25.7)	10.9 (27.1)	3.2
	9 opti	76.0	72.4	10.1 (75.8)	33.8 (77.3)	6.2 (23.0)	5.9 (23.4)	3.3
	7 opti	76.0	72.4	9.4 (75.2)	32.8 (77.2)	8.5 (24.2)	7.9 (23.2)	3.1
II	9 equi	76.0	72.3	32.2 (75.9)	25.2 (77.2)	18.5 (34.1)	17.1 (31.5)	4.0
	9 opti	76.0	72.4	33.9 (76.0)	27.1 (77.6)	13.9 (30.3)	13.0 (31.6)	4.2
	7 opti	76.0	72.3	35.2 (76.0)	29.6 (77.4)	18.1 (28.5)	14.7 (32.0)	4.1
III	9 equi	76.0	72.7	36.4 (75.3)	15.3 (77.1)	2.4 (13.8)	4.6 (34.8)	2.7
	9 opti	76.0	72.7	40.2 (74.2)	17.3 (77.4)	2.0 (11.1)	4.6 (20.0)	2.7
	7 opti	76.0	72.7	39.9 (74.7)	20.6 (77.5)	1.5 (9.1)	2.1 (25.7)	2.6

Pancreas

#	Config	GTV	CTV	Kidney (r)	Kidney (l)	Liver	Spine	Intestine	NT
I	9 equi	54.0	47.4	11.3 (26.1)	10.2 (34.9)	5.4 (48.0)	4.7 (23.4)	12.4 (52.7)	4.4
	9 opti	54.0	47.4	9.7 (23.5)	9.9 (35.6)	4.8 (48.0)	5.0 (22.1)	12.4 (54.3)	4.5
	7 opti	54.0	47.4	11.1 (25.7)	10.2 (36.2)	5.4 (47.1)	5.7 (25.0)	12.6 (54.1)	4.6
II	9 equi	54.0	47.8	7.3 (24.1)	7.2 (19.1)	3.5 (47.3)	4.1 (17.9)	11.6 (48.9)	3.3
	9 opti	54.0	47.7	7.3 (23.9)	7.1 (18.6)	4.2 (37.0)	6.0 (19.7)	11.4 (49.1)	3.5
	7 opti	54.0	47.9	8.0 (27.0)	6.3 (23.6)	3.7 (47.3)	5.9 (20.3)	12.1 (48.9)	3.5
III	9 equi	54.0	48.0	4.4 (20.3)	6.5 (18.6)	4.7 (46.6)	3.9 (17.8)	7.0 (50.4)	3.3
	9 opti	54.0	48.0	3.8 (17.6)	5.8 (18.6)	5.0 (47.5)	5.7 (17.3)	7.9 (50.0)	3.7
	7 opti	54.0	48.0	4.5 (22.4)	6.4 (19.7)	4.8 (46.6)	5.2 (18.5)	6.8 (50.9)	3.4

Intracranial

#	Config	CTV	Brainstem	Eye (r)	Eye (l)	Opticus (r)	Opticus (l)	Chiasm	NT
I	7 equi	50.4	14.1 (25.6)	8.8 (30.2)	15.4 (38.1)	30.5 (48.0)	37.8 (48.4)	33.1 (42.1)	5.5
	7 opti	50.4	9.5 (24.8)	9.5 (30.2)	18.6 (42.2)	28.3 (47.5)	39.8 (48.5)	30.3 (45.0)	5.7
	5 opti	50.4	12.4 (23.5)	10.2 (27.0)	21.5 (39.1)	35.1 (49.1)	40.6 (48.7)	27.0 (42.1)	6.6
II	9 equi	60.0	12.8 (31.0)	13.6 (27.1)	10.8 (16.7)	34.9 (49.9)	23.0 (37.5)	16.1 (27.8)	3.9
	9 opti	60.0	9.6 (33.8)	13.0 (30.1)	8.8 (16.3)	29.9 (47.0)	13.9 (28.2)	14.4 (26.3)	4.2
	7 opti	60.0	8.2 (32.6)	13.7 (28.0)	10.9 (17.6)	34.3 (48.6)	16.0 (28.0)	15.7 (27.8)	4.1
III	9 equi	60.0	18.5 (55.6)	14.6 (33.4)	7.6 (12.8)	38.0 (51.4)	13.1 (25.8)	38.5 (56.5)	3.7
	9 opti	60.0	18.0 (55.5)	12.8 (29.3)	5.2 (13.8)	26.1 (54.6)	12.1 (21.1)	36.8 (59.0)	3.9
	7 opti	60.0	16.2 (56.2)	14.6 (36.8)	5.5 (15.1)	32.0 (53.2)	15.5 (25.3)	38.6 (59.4)	4.2

Table 4.2.: Mean (Max) dose [Gy] of exclusive DVHs for all nine cases under investigation sorted by treatment site. The dose statistics includes all volumes of interest (NT = normal tissue) that define the main conflict during treatment planning of the corresponding site. All plans are normalized to the mean target dose.

contributions to the liver or the spine. Doses to the intestine and normal tissue of the optimized and standard treatment plans are clinically equivalent.

Intracranial cases

The reference plan for patient I featuring seven equi-spaced coplanar beams is compared to treatment plans with seven and five optimized beams. The reference plans for patient II and patient III featuring nine equi-spaced coplanar beams are compared to treatment plans with nine and seven optimized beams. For patient I, a homogeneous dose of 50.4 Gy and for patients II and III, a homogeneous dose of 60 Gy is prescribed to the CTV. The target volumes of all cases are located in close proximity to the eyes, the optic nerves, or the brainstem. Figures 4.20(e) and 4.20(f) display DVHs of patient III, corresponding dose distributions are visualized by figures 4.6(e) and 4.6(f). The mean doses to the brainstem of patient I and II could be reduced by up to 3.6 and 3.3 Gy, respectively. There is no clear trend observed regarding the doses to the eyes. The mean doses to the chiasm are slightly decreased (up to 3.1 Gy) while the maximum doses are slightly increased (up to 2.9 Gy). For patient II (III) it was possible to reduce the dose to the left (right) optic nerve by 9.1 Gy (11.9 Gy).

4.5.3. Discussion

Spherical cluster analysis of geometric data for BAS assumes that beam orientations are advantageous for IMRT if and only if they impinge tangentially to surface elements of the target. All other aspects of the BAS problem are neglected. The results of the treatment planning study showed only marginal improvements for the optimized beam configurations for prostate and pancreas lesions. For the three intracranial cases investigated, however, spherical cluster analysis based on geometric data provided superior treatment plans as compared to coplanar equi-spaced beam orientations. The dose distributions were more confined to the complex outlines of the target volumes, as shown in figure 4.21(f), and provided superior sparing of OARs. However, compared to spherical cluster analysis of dosimetric data, which was presented in section 4.2, spherical cluster analysis of geometric data yields inferior treatment plans. Dosimetric information provides additional valuable input for BAS. Nevertheless spherical cluster analysis of geometric data is attractive for clinical application for complex intracranial treatments. It is a contained framework for BAS which requires only about 2 s of computation time. The algorithm comprises only basic matrix algebra such as the computation of vector and dot products. It scales linear with the number of triangular target surface elements $|T|$ and the number of beams η.

The spherical clustering algorithm based on geometric data incorporates the combinatorial problem of intensity-modulated fields that complement one another into the automated BAS process - but only to some extend. Our earlier approach, which is based on dosimetric data, has a clear correspondence between one target voxel v and one locally ideal beam orientation $\boldsymbol{\beta}_v^*$: every target voxel is represented by one element in the set of locally ideal beam orientations $\mathbb{B}$, i.e. one point on the unit sphere, and eventually in the BAS process. For spherical cluster analysis of geometric data, however, this does not hold. Here, a target surface element t_i is represented by a set P_i, i.e. a circle on the unit sphere. Consequently, some target surface elements may have a stronger influence on the result of the K-means algorithm than others; certain features of the target surface may disappear or dominate during the BAS process. This will become more clear during the next paragraph.

The presented method is prone to deficiencies in the delineation of the target volume. E.g. if the target volume is delineated in one transversal plane of the computed tomography data set and abruptly ends in the next one, this results in a completely flat "lit" of the target volume within the transversal plane. Figures 4.18(a) and 4.18(b) show that this lit of the target surface element is projected to one high density band of locally ideal beam orientations at the equator of the clustering landscapes. As discussed in the previous paragraph, this single feature has a disproportionately high influence on the entire beam configuration. It "pulls" all beam orientations of the ensemble towards the equator. We observed these irregularities in 50% of the patient cases.

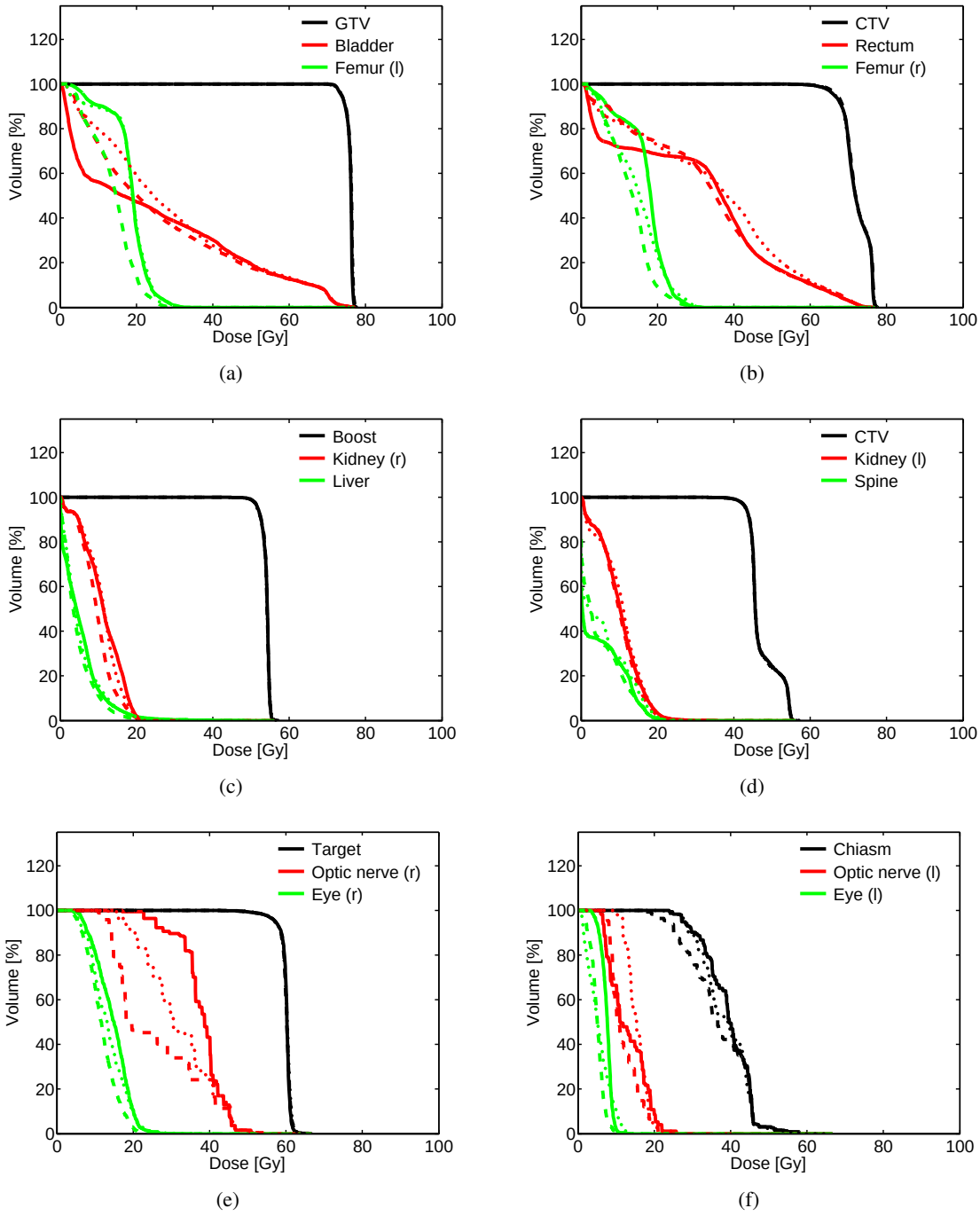


Figure 4.20.: Inclusive DVHs for (a) & (b) Prostate lesion II, (c) & (d) abdominal lesion I, and (e) & (f) intracranial lesion III. Solid lines represent the standard equi-spaced coplanar nine-beam plan, dashed lines represent the optimized non-coplanar nine-beam plan, and dotted lines represent the optimized non-coplanar seven-beam plan. DVHs are shown for selected volumes of interest that define the main conflict during treatment planning of the corresponding case.

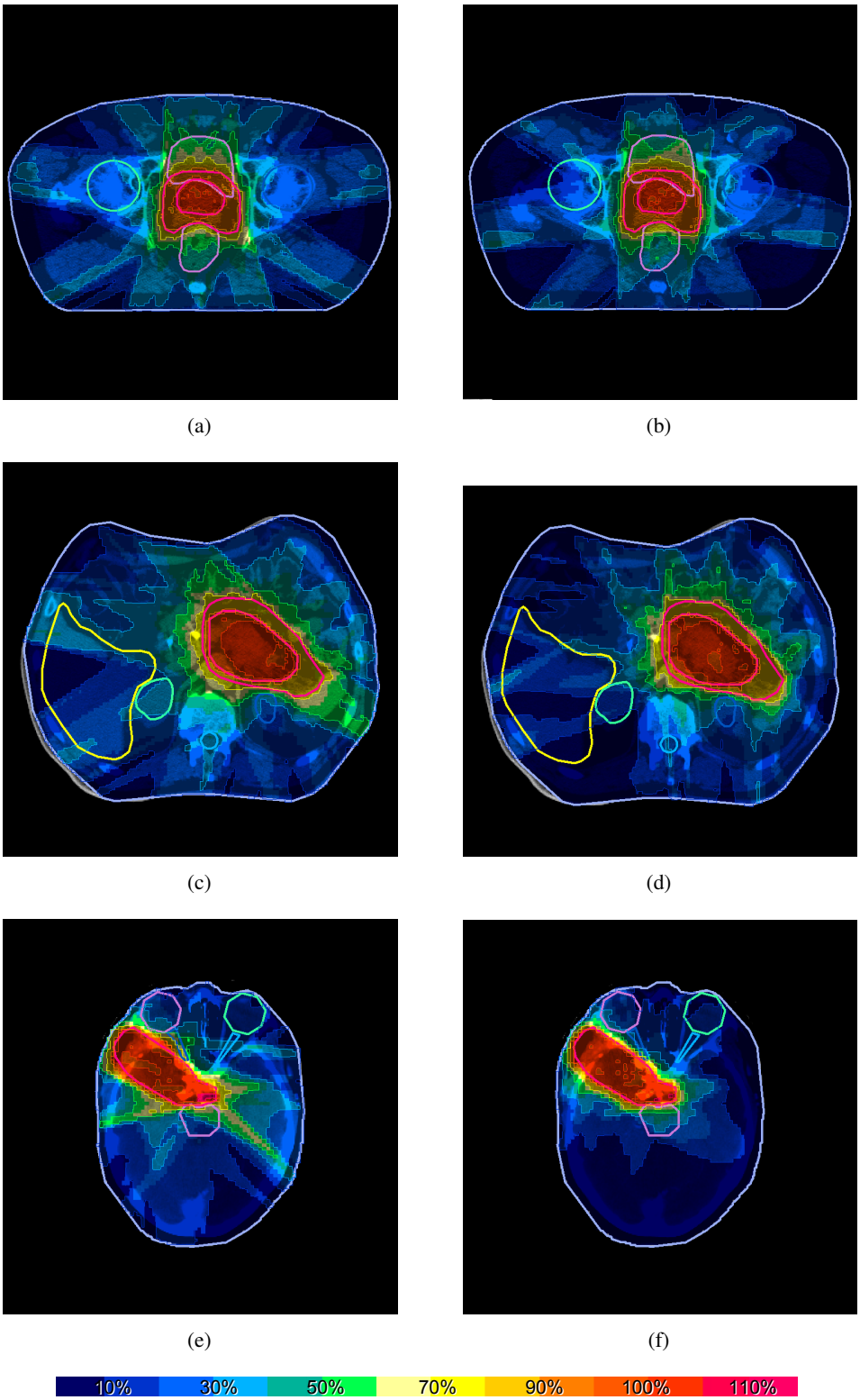


Figure 4.21.: Dose distribution on a selected transversal CT slice for (a) & (b) Prostate lesion II, (c) & (d) abdominal lesion I, and (e) & (f) intracranial lesion III. The dose distributions (a), (c), and (e) are computed with nine equi-spaced coplanar beams, the dose distributions (b), (d), and (f) are computed with seven optimized non-coplanar beams. 100% corresponds to the prescribed target dose.

Apparently, the deliniation process is strongly biased because the MD draws the volumes of interest in the transversal plane of the computed tomography data set. For a human planner, it is very difficult to model a smooth three-dimensional object with two-dimensional contours. It is debatable whether these deficiencies of the deliniation implicitly favour coplanar beam ensembles because coplanar beam orientations impinge per se tangentially to the coplanar plane of the computed tomography data set.

The sets of locally ideal beam angles $\mathbb{B}$ shown in figure 4.18 all exhibit the same structure. We observe closed bands of higher density of locally ideal beam orientations because the set of locally ideal beam orientations P_i for a single target surface element t_i is given by a closed circle. Figures 4.18(a), 4.18(b), and 4.18(c) exhibit such bands that wrap around the equator. These geometric features may represent alternative beneficial trajectories for arc therapy. Compared to spherical cluster analysis of dosimetric data for BAS (section 4.2) we observed a different characteristic of the resulting beam ensembles. The configurations going back to the geometric data sets are not as asymmetric as the configurations going back to the dosimetric data. Figure 4.18 shows that the beam orientations are distributed all around the patient but they are locally refined to coincide with beam orientations that have a higher density of locally ideal beam orientations. As the target volume has a closed surface, beam orientations impinging from all over 4π are tangential to at least a small fraction of the target surface. Unlike the set of locally ideal beam orientations going back to the dosimetric data sets (shown in figure 4.4), the set of locally ideal beam orientations going back to the geometric data sets does not exhibit "forbidden" areas (shown in figure 4.18). All candidate beam orientations are a member of the set of locally ideal beam orientations. This may be considered another indication that the higher the number of beams the higher the conformity of an IMRT treatment plan.

A-BOO seems inappropriate for application in particle therapy. Due to the influence of range and setup uncertainties, a dose calculation is necessary to decide whether beams that are tangential or perpendicular to target surface elements are beneficial for irradiation.

5. Combinatorial beam ensemble optimization

This chapter introduces concepts for beam ensemble selection based on combinatorial optimization techniques. All presented methods belong to the first class of BAS algorithms as defined in section 3.1. In contrast to the previous chapter, BAS is now intertwined with the optimization of beamlet weights. The methods are explicitly designed to minimize the objective function F of inverse planning with respect to both the beamlet weights and beam orientations. Consequently, a beam angle optimization (BAO) is performed according to the subtle distinction made in section 2.2.8.

A formal definition of the objective function F of the inverse planning problem including BAS was already derived in section 2.2.1. Using the standard quadratic form of the beamlet weight optimization problem (Oelfke and Bortfeld 2001), it is given by

$$\{B^*, \boldsymbol{w}^*\} = \underset{B \in \widetilde{\mathcal{B}},\, \boldsymbol{w} \in \mathbb{R}_+^{n_b}}{\arg\min} \quad F(B, \boldsymbol{w}) \tag{5.1}$$

$$= \underset{B \in \widetilde{\mathcal{B}},\, \boldsymbol{w} \in \mathbb{R}_+^{n_b}}{\arg\min} \quad (I(B)\,\boldsymbol{w} - \boldsymbol{d}^{\mathrm{p}})^{\mathrm{T}}\; P\; (I(B)\,\boldsymbol{w} - \boldsymbol{d}^{\mathrm{p}}). \tag{5.2}$$

Here, a greedy iterative search algorithm and three different metaheuristics are applied to solve equation 5.2: a simulated annealing algorithm, a genetic algorithm, and a cross-entropy algorithm. The strategies are evaluated with regard to their algorithmic efficiency and the quality of the resulting treatment plans. Furthermore they are compared to the heuristic BAS strategies introduced in the previous chapter.

5.1. Method

Solving the BAS problem by minimizing equation 5.2 is an NP-hard combinatorial problem in a non-convex, exponentially growing search space, as discussed in section 2.2. Even though there is no algorithm that guarantees global convergence in polynomial runtime for NP-hard problems, it is possible to efficiently derive near optimal solutions with modern metaheuristics.

All these metaheuristics rely on the same approach: during a large number of iterations, candidate solutions are sampled to identify beneficial sub regions within the search space. Over time, the sampling process focuses more and more on those beneficial sub regions. For BAS, the sampling process corresponds to the computation of the objective function value $F(B)$ for different beam ensembles B. Based on the objective function values obtained by a full beamlet weight optimization, it is possible to identify sub regions within the space of possible beam configurations which promise better objective function values for subsequent iterations. Over time, more and more beam ensembles stemming from those beneficial sub regions are evaluated.

Apparently, such approaches depend on a large number of beamlet weight optimization processes for different beam configurations, i.e. solving equation 2.9 for a large number of fixed beam ensembles. Using conventional treatment planning software, the beamlet weight optimization process for one beam configuration may require more than 5 min on a state-of-the-art workstation for ”large“ cases. Within 24 hours we could only compare 288 beam configurations. Hence, it is crucial to accelerate the beamlet weight optimization process significantly for BAS algorithms. Earlier approaches pursuing a joint optimization of beamlet weights and beam orientations, which were reviewed in section 3.2, (1) apply a sparse resolution in the dose deposition space, (2) use a large

spacing of candidate beam orientations, and/or (3) sample only a limited number of beam configurations to arrive at acceptable computation times.

We are convinced, however, that the BAS problem in IMRT treatment planning has to be addressed with (1) a sufficient precision of the underlying dose calculation, as small changes of the dose distribution may already be critical for the treatment outcome (e.g. consider the sparing of the optic nerves), (2) a sufficient set of non-coplanar candidate beams, as non-coplanar beams may provide superior dose distributions, and (3) a sufficient number of evaluated beam configurations, as the solution space of the BAS problem increases exponentially with the number of beams.

Consequently, we employ a full precision large-scale BAS framework that does not introduce simplifications of the clinical inverse planning program KonRad. It exploits high performance parallel computing architectures in combination with efficient algorithmic concepts to accelerate the beamlet weight optimization. The framework comprises two modules:

- **The parallel beamlet weight optimization module** computes the objective function values for predefined beam configurations.
- **The beam ensemble sampling module** controls the sampling process of the beam configurations.

The communication between the two modules is limited to the exchange of objective function values and beam configurations. The parallel beamlet weight optimization module is introduced in section 5.1.1 and the beam ensemble sampling module is introduced in section 5.1.2.

5.1.1. The parallel beamlet weight optimization module

The parallel beamlet weight optimization engine is the heart of the combinatorial BAS framework. It applies the mathematical concepts introduced in section 2.1.4 to compute ideal beamlet weights for predefined beam configurations. In comparison to the clinical version of our in-house inverse planning software KonRad (Preiser et al. 1997, Nill 2001), the most important refinements of the mathematical concepts are the use of a L-BFGS two loop recursion algorithm to approximate the inverse Hessian for the Quasi Newton algorithm and the use of a proper line search to adjust the step length (Nocedal and Wright 1999). Our experiments with photons indicate that KonRad requires about three times more iterations to converge to the same objective function values. For particles, the modifications may even yield a speed-up factor of six (Pflugfelder et al. 2008a).

If we want to evaluate several thousand beam configurations per patient within a couple of hours, however, we would need a speed-up factor that is about two orders of magnitude larger. This cannot be achieved by an improved mathematical approach alone. Hence we apply parallel high performance computing architectures combined with improved algorithmic concepts. Figure 5.1 gives a schematic overview of the underlying idea.

Using an inverse planning approach which is based on precomputed dose influence data, the computation of the objective function F and its gradient ∇F requires only basic arithmetic operations, as shown in table 2.1.4. Ziegenhein et al. (2008) showed that the arithmetic load does not limit the runtime of the optimization of beamlet weights. Due to the design of modern computers it is the transportation process of the dose influence data from the storage hardware to the processor that limits the runtime. This phenomenon is well-known among computer scientists as the von Neumann bottleneck (Backus 2007).

For BAS an efficient data transportation process is especially important because we have to deal with an increased amount of dose influence data compared to standard inverse planning. All $|\mathcal{B}|$ candidate beams have to be stored in the main memory - not only the dose influence data of η beams. The CPU may have to ”run

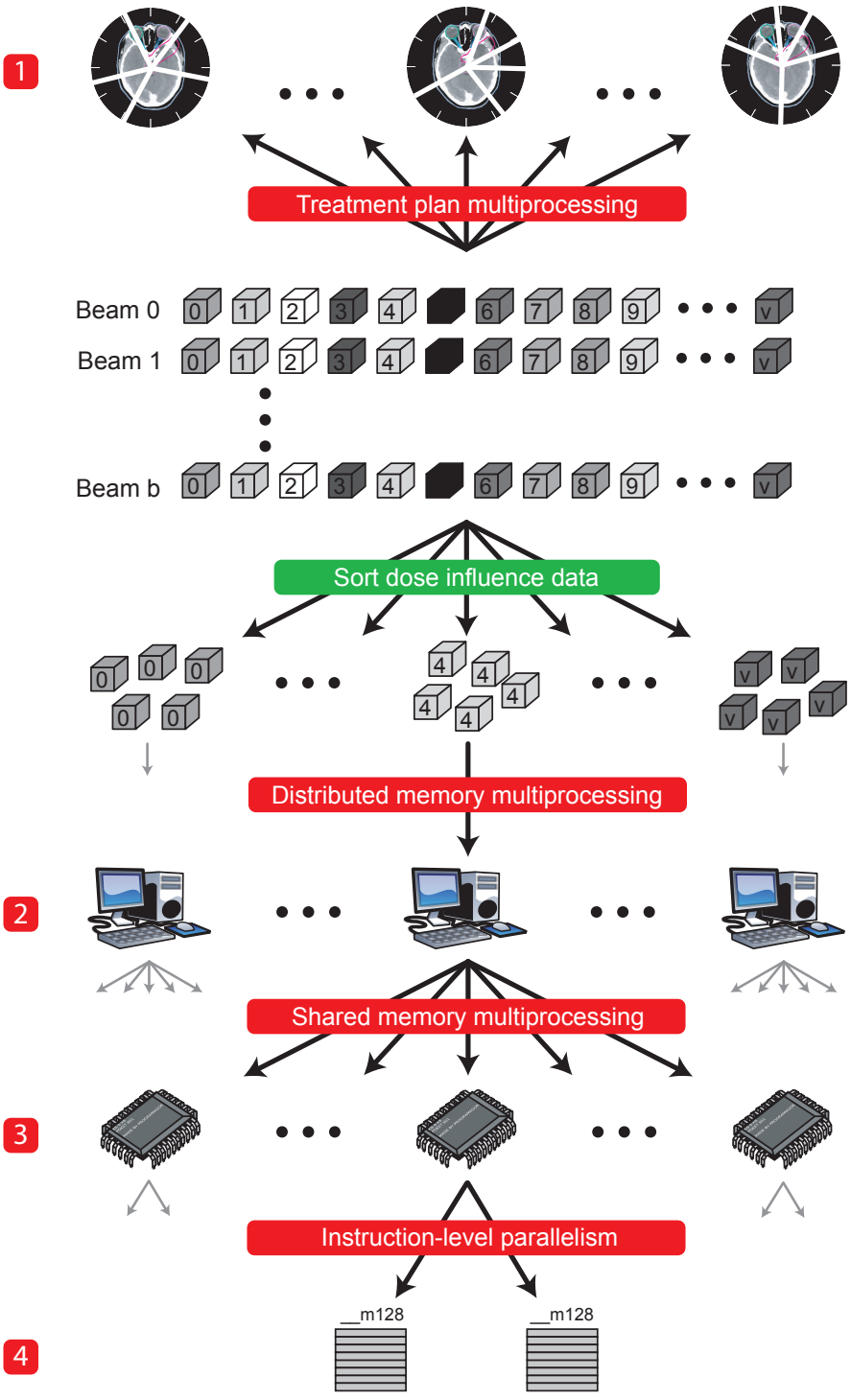


Figure 5.1.: Schematic workflow of the fourfold parallel optimization framework. t treatment plans with different beam configurations are optimized in parallel (1). The dose influence data is resorted and distributed per voxel on a parallel computer cluster (2). In one iteration, each node of the cluster evaluates the contribution of the assigned voxels to the t objective functions and t gradients at once using thread parallel (3) and instruction parallel (4) operations.

over“ the dose influence data of $|\mathcal{B}| >> \eta$ beams to access the relevant base data to calculate the objective function and the gradient for certain beam configurations. Consequently the CPU may be idle for a long time while waiting for the relevant base data. We address this issue with a custom sorting and distribution scheme of the dose influence data in combination with the parallel computation of several treatment plans at once. We intend to accelerate inverse planning by both distributing the computation on multiple CPUs and increasing the individual CPU load.

The dose influence data of all candidate beams is therefore rearranged by voxel, as shown in figure 5.1. Hence, every voxel element ”knows“ its dose contributions of all beamlets of all candidate beams and the dose influence data may be distributed per voxel on a high performance computer cluster. This sorting and distribution scheme allows for the independent calculation of the contribution of a group of voxels to the objective function and gradient on a single node of the cluster. The shared memory multiprocessing communication between different nodes is implemented using the MPI standard[1].

One node performs all calculations voxel after voxel. For the calculation of the dose contribution of beam β_a to voxel i, the processor fills the cache memory with the corresponding base data. Besides the dose influence data of beam β_a of voxel i, the cache memory is filled up with the dose influence data of subsequent beams β_{a+1}, $\beta_{a+2}, \ldots, \beta_{a+m}$ of voxel i. Hence, it is possible to calculate the dose in voxel i for t different beam configurations at zero additional memory transportation cost. This is an elegant implementation of the principle of locality of reference (Denning 2005).

The algorithmic treatment plan multiprocessing of t different beam configuration is the first and the distributed memory multiprocessing on the high performance computer cluster is the second level of parallelism of the beamlet weight optimization module, as depicted in figure 5.1. Shared memory multiprocessing using OpenMP[2] is the third level of parallelism. Here, multiple CPUs or cores of one node share the main memory and calculate the dose contributions of different beam configurations and beamlets to one voxel in parallel. Instruction-level parallelism using Streaming SMID Extensions[3] is the fourth level of parallelism. Here, two arithmetic operations in double precision are executed at the same time by one CPU.

Basically, we are solving t optimization problems on a joint data set at the same time. The parallel computation of the objective function values and gradients for t beam configurations has to be intertwined with a parallel line search and a parallel L-BFGS update of the inverse Hessian. Furthermore, a parallel convergence test has to be implemented because the beamlet weight optimization process does not converge after a constant number of iterations for t different beam configurations. The repeated optimization of the beamlet weights of the same beam may be accelerated by initializing the beamlet weights of a candidate beam β uniformly with $\boldsymbol{w}_\beta^{\text{init}} = 1$ if it is used for the first time during the combinatorial optimization process. After an optimization, the beamlet weights are stored in a global beamlet weight initialization database $\boldsymbol{w}_\beta^{\text{BWIDB}} = \boldsymbol{w}_\beta^{\text{opt}}$. If candidate beam β is used again during the combinatorial optimization, it is initialized according to the beamlet weight initialization database with $\boldsymbol{w}_\beta^{\text{init}} = \boldsymbol{w}_\beta^{\text{BWIDB}}$. Treatment plan parallelism implies modifications of the entire optimization process and bears many potential pitfalls.

Note that it would be naive to distribute the optimization processes for different beam configurations on different nodes of the cluster. In such an implementation every node would require the dose influence data of all candidate beams. This would restrict the number of candidate beams to the memory of one node. As we are distributing the dose influence data per voxel we can exploit the full memory of the entire cluster without redundant copies of the base data. Furthermore we cannot account for the principle of locality of reference in an equally efficient

[1]The message passing interface standard http://www.mcs.anl.gov/research/projects/mpi
[2]Open multiprocessing http://openmp.org
[3]Streaming single instruction multiple data extensions developed by Intel, Santa Clara, CA.

way if every node operates on a copy of the entire base data.

5.1.2. The beam ensemble sampling module

Simply put, the beam ensemble sampling module asks the parallel beamlet weight optimization module to optimize the beamlet weights for one beam configuration B_i or a set of different beam configurations $\{B_i\}$ in every iteration k. Based on the resulting objective function values $F(B_i)$, the beam ensemble sampling module refines its sampling process, i.e. the process deciding which beam configurations should be evaluated next, and asks the parallel beamlet weight optimization module to optimize the beamlet weights for a set of new beam configurations during iteration $k+1$. We implemented four different algorithms within the beam ensemble sampling module. The following subsections provide a brief yet comprehensive description of all algorithms. Readers requiring a broader discussion find references to suited publications.

Iterative beam angle selection algorithm

The iterative BAS algorithm, which is inspired by the work of Woudstra and Storchi (2000), is the most basic algorithm implemented for the BAS navigation module. It assumes that the beam configuration of a beneficial treatment plan featuring $\eta - 1$ beams is a subset of the beam configuration of a treatment plan featuring η beams. Consequently, the optimization process of equation 5.2 corresponds to an iterative construction of the best η-beam plan based on the best $(\eta - 1)$-beam plan.

To find a beneficial beam ensemble B^*_η featuring η beams, the iterative BAS algorithm computes the objective function values of all possible treatment plans featuring one beam $\beta \in \mathcal{B}$ during the first iteration. The beam orientation β^*_1 for the best 1-beam plan is given by the minimizer of the objective function.

$$\beta^*_1 = \underset{\beta \in \mathcal{B}}{\arg\min}\, F(\beta) \tag{5.3}$$

$$B^*_1 = \left\{\beta^*_1\right\} \tag{5.4}$$

Given the best beam configuration $B^*_{\eta-1}$ featuring $\eta - 1$ beams, the additional beam β^*_η for the η-beam plan is found by evaluating all combinations of $B^*_{\eta-1}$ with the remaining $|\mathcal{B}| - (\eta - 1)$ candidate beams.

$$\beta^*_\eta = \underset{\beta \in \mathcal{B}\setminus B^*_{\eta-1}}{\arg\min}\; F(\left\{\beta, B^*_{\eta-1}\right\}) \tag{5.5}$$

$$B^*_\eta = \left\{\beta^*_\eta, B^*_{\eta-1}\right\} = \left\{\beta^*_\eta, \beta^*_{\eta-1}, ..., \beta^*_1\right\} \tag{5.6}$$

The number of beam configurations n_I that has to be evaluated to derive B^*_η is not a free parameter of the iterative BAS algorithm. It is defined by the number of candidate beams $|\mathcal{B}|$ and the number of beams η of the final treatment plan.

$$n_I = \sum_{i=0}^{\eta-1} |\mathcal{B}| - i \tag{5.7}$$

Algorithm panel 5.1 gives an overview of the corresponding pseudocode.

Algorithm 5.1 Iterative beam angle selection algorithm

$B^* = \{\,\}$;
for $(k = 0;\ k < \eta;\ i++)$ **do**
$$B^* = \left\{B^*, \underset{\beta \in \mathcal{B}\setminus B^*}{\arg\min}\, F(\{\beta, B^*\})\right\};$$
end for

Simulated annealing beam angle selection algorithm

Simulated annealing is a metaheuristic for global optimization (Kirkpatrick et al. 1983). It is inspired by the annealing process in metallurgy where a material is heated and afterwards gradually cooled so that misaligned atoms may escape local energy minima and arrange in larger crystals.

Simulated annealing was included in the BAS navigation module because it is the predominant approach to solve the joint optimization problem of beamlet weights and beam orientations in radiation therapy treatment planning, as discussed in section 3.2.2. It is used as a benchmarking algorithm for other strategies.

The simulated annealing BAS algorithm is initialized with a random beam configuration B^*. The starting temperature T_0 is set to be 10% higher than the largest value of the objective functions calculated for 10 sets of random beam ensembles (Pugachev et al. 2001). During test runs we observed that the cooling rate λ has no significant influence on the convergence properties of the annealing process in range from 0.95 to 0.999. We chose $\lambda = 0.99$ as we did not find any reference in the literature.

During iteration k of the simulated annealing process, one beam $\beta \in B^*$ is randomly replaced with another beam $\beta' \in \mathcal{B} \setminus B^*$ to yield a new candidate solution B^k. The new beam configuration B^k replaces B^* if $F(B^k) \leq F(B^*)$. In case $F(B^k) > F(B^*)$, the new beam configuration is accepted with probability

$$p = e^{-\frac{F(B^k)-F(B^*)}{T_k}}. \tag{5.8}$$

In every iteration, the temperature is adjusted according to an exponential cooling schedule (Pugachev et al. 2001).

$$T_k = \lambda \cdot T_{k-1} \tag{5.9}$$

This yields an optimization process, that accepts almost all changes of the beam configurations in the beginning, but becomes more and more greedy as the temperature decreases. Algorithm panel 5.2 gives an overview of the corresponding pseudocode.

Algorithm 5.2 Simulated annealing beam angle selection algorithm

```
initialize one beam configuration B* ∈ P_η(B) at random;
for (k = 0; k < n_I; k++) do
    T_k = λ · T_{k−1};
    B^k = B*;
    replace one beam β ∈ B^k randomly with β' ∈ B \ B^k;
    r = randDouble[0, 1];
    p = exp{(F(B*) − F(B^k))/T_k};
    if (r ≤ p) then
        B* = B^k;
    end if
end for
```

Genetic beam angle selection algorithm

Genetic algorithms (Mitchell 1998) mimic an evolution process for optimization. The average fitness of a population of candidate solutions is improved by simulating natural processes like inheritance, selection, crossover, and mutation. Among others, genetic algorithms proved useful for applications in engineering, operations research, and bioinformatics. Recently, genetic algorithms were successfully applied for BAS (Hou et al. 2003,

Li et al. 2004, Nazareth et al. 2009).

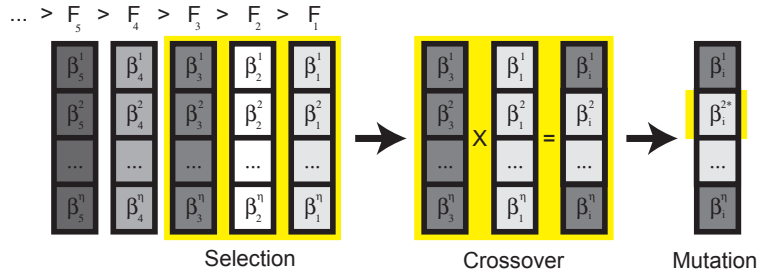


Figure 5.2.: Schematic work flow of the genetic BAS algorithm. The γ quantile of the current generation is selected for procreation by cross over. Eventually some beams will be exchanged according to a mutation probability.

For the genetic BAS algorithm, a beam ensemble B_i is represented by a chromosome whose genes $\{\beta_i^1, \beta_i^2, ..., \beta_i^\eta\}$ represent the individual beams of the ensemble, as shown in figure 5.2. Before the evolutionary process starts, the initial generation $\mathcal{G}^0 = \{B_i\}$ is initialized with a random set of n_E beam ensembles. In every iteration k, the objective function values are computed for all beam configurations to identify the fitness of the current generation. In order to simulate evolutionary pressure, only the γ quantile of the current generation $\mathcal{G}^k$, according to the objective function values $F(B_i)$, forms the set of procreating beam configurations Π^k and may generate offspring in the next generation. This is called hard selection.

Algorithm 5.3 Genetic beam angle selection algorithm

initialize a set of random beam configurations $\mathcal{G}^0 = \{B^i \mid B_i \in \mathcal{P}_\eta(\mathcal{B})\}$;
for $(k = 0;\ k < n_I;\ k++)$ **do**
  calc obj func $\forall\ B_i \in \mathcal{G}^k$;
  // Selection
  Select the γ quantile of $\mathcal{G}^k$ for Π^k;
  for $(j = 0;\ j < n_E;\ j++)$ **do**
    draw two random beam configurations B_a and B_b from Π^k;
    for $(l = 0;\ l < \eta;\ l++)$ **do**
      // Crossover
      $r_1 =$ `randDouble`$[0, 1]$;
      if $(r_1 \leq \chi)$ **then**
        $\beta_j^l = \beta_a^l$;
      else
        $\beta_j^l = \beta_b^l$;
      end if
      // Mutation
      $r_2 =$ `randDouble`$[0, 1]$;
      if $(r_2 \leq \mu)$ **then**
        replace β_j^l with a random beam $\beta \in \mathcal{B} \setminus B_j$;
      end if
    end for
  end for
end for

In contrast to random walk techniques, a genetic algorithm applies a crossover operation to create a new generation. Therefore two parenting chromosomes B_a, $B_b \in \Pi^k$ are selected to exchange genes according to a crossover probability χ to generate a new beam configuration, as shown in figure 5.2. Finally, some genes of the new chromosome will experience a mutation according to a mutation probability μ. Algorithm panel 5.3 gives an overview of the corresponding pseudocode.

The parameters of the genetic algorithm were established in a series of convergence test runs. We set the number of beam ensembles per generation $n_E = 50$, the cross over probability $\chi = 0.4$, the mutation probability $\mu = 0.05$, and the selection quantile $\gamma = 0.1$. We found that none of these parameters has a critical influence on the genetic algorithm. Over a broad range we observed comparable convergence characteristics.

Cross-entropy beam angle selection algorithm

The cross-entropy method is a generic approach for combinatorial optimization and rare event simulation (Rubinstein and Kroese 2004). It has been successfully applied for many standard combinatorial optimization problems in operations research like the traveling salesman problem, the max cut problem, and the Knapsack problem (De Boer et al. 2005), but its performance on the BAS problem has not yet been studied.

The cross-entropy method was originally developed to efficiently estimate the probability of rare events in Monte Carlo simulations. With slight modifications, however, it may also be used for combinatorial optimization. Therefore, a probabilistic process is established that controls the generation of random samples. Generating the optimal sample corresponds to a rare event according to the initial probability distribution of the random sample generator. The cross-entropy method "learns" how to adopt the probabilistic sampling process to increase the probability for the generation of the optimal sample. Readers who are interested in the formal derivation of the cross-entropy method and the general algorithm for the refinement of the probabilistic sampling process are referred to Rubinstein and Kroese (2004). Here, we focus on the implementation of the cross entropy method for BAS.

It is possible to use the cross-entropy method almost exactly as applied for the traveling salesman problem (De Boer et al. 2005) also for BAS. Therefore the probabilistic sampling of different beam configurations is modeled by a multinomial distribution combined with a modified urn model where the generation of an ensemble featuring η beams corresponds to drawing η marbles from a hypothetical "BAS urn". This urn contains marbles with $|\mathcal{B}|$ different colors, i.e. the number of colors corresponds to the number of candidate beam orientations. Apparently, the probability of selecting candidate beam orientation β for beam ensemble B_i is proportional to the relative number of marbles with color β in the urn. After a marble with color β, i.e. beam orientation β, was drawn from the urn, not only the drawn marble but all marbles with color β have to be removed from the urn. Hence, we avoid drawing duplicate beams for beam ensemble B_i. Of course, we have to start over for the generation of the next beam ensemble B_{i+1} and refill the BAS urn.

In order to produce improved beam ensembles over time, the numbers of marbles with different colors is refined in every iteration. During the first iteration a uniform probability p_β is assigned to all candidate beam orientations β. All candidate beam orientations or colors are represented in the BAS urn by the same number of marbles.

$$p_\beta = \frac{1}{|\mathcal{B}|}. \tag{5.10}$$

We use these probabilities for a multinomial distribution combined with the modified replacement strategy and draw n_E beam ensembles B_i from the BAS urn. The corresponding beamlet weights are then optimized by the parallel beamlet weight optimization module. The γ-quantile according to the resulting objective function

values $F(B_i)$ of all beam ensembles B_i passes the so called γ-criterion. The update formula for the probabilities of the multinomial distribution, i.e. the adjustment of the numbers of marbles with different colors in the hypothetical BAS urn, is given by

$$p_\beta = (1-\rho)\cdot p_\beta + \rho \cdot \frac{\sum_i I_\gamma(B_i)\, I_\beta(B_i)}{\sum_i I_\gamma(B_i)} \tag{5.11}$$

with the learning rate ρ and the indicator functions I_γ and I_β.

$$I_\gamma(B_i) = \begin{cases} 1 \text{ if } B_i \text{ passes } \gamma\text{-criterion} \\ 0 \text{ else} \end{cases} \qquad I_\beta(B_i) = \begin{cases} 1 \text{ if } \beta \in B_i \\ 0 \text{ else} \end{cases} \tag{5.12}$$

Apparently the probability p_β gradually drifts towards the relative frequency of β within the beam ensembles passing the γ-criterion. Algorithm panel 5.4 gives an overview of the corresponding pseudocode.

Algorithm 5.4 Cross-entropy beam angle selection algorithm

```
// Initialize multinomial sampling process with uniform probabilities
for (j = 0; j < |B|; j++) do
    p_j = 1 / |B|
end for
for (k = 0; k < n_I; k++) do
    // Draw beam ensemble from the BAS urn
    for (i = 0; i < n_E; i++) do
        Draw random beam ensemble B_i from a multinomial distribution with probabilities p;
        Calculate objective function value F(B_i)
    end for
    // Update probabilities of the multinomial sampling process
    for (j = 0; j < |B|; j++) do
        p_j = (1 − ρ) · p_j + ρ · (Σ_i I_γ(B_i) I_β(B_i)) / (Σ_i I_γ(B_i))
    end for
end for
```

The parameters of the cross-entropy algorithm were established in a series of convergence test runs. We set the number of beam ensembles per generation $n_E = 50$, the learning rate $\rho = 0.15$, and the quantile $\gamma = 0.075$. We found that none of these parameters has a critical influence on the cross-entropy algorithm. Over a broad range we observed comparable convergence characteristics.

5.2. Results

We study the performance of the full precision large-scale BAS framework in a treatment planning study for three intracranial cases. Therefore, the parallel beamlet weight optimization module and the beam ensemble sampling module are implemented in a joint C++ program running on a research cluster which comprises ten nodes that communicate via InfiniBand links. Every node features two Intel Xeon CPUs with six cores. Every node has 10 GB of main memory.

During initialization, the patient CT, the segmentation, the maximum and minimum doses, the penalties and the dose influence data are read from binary files. We apply a resolution of 2.62 mm × 2.62 mm × 2.62 mm

for the patient CT and the associated segmentation. The dose influence data is pre-computed with our in-house treatment planning software KonRad (Preiser et al. 1997). The dose calculation applies 5 mm × 5 mm beamlets and takes about 2 s per candidate beam orientation. The dose influence data comprises 7.2 - 14.0 GB for 392 - 404 candidate beam directions. Note that the KonRad dose calculation applies importance sampling by default to reduce the memory requirements of the dose influence data (Thieke et al. 2002). All treatment plans of a patient apply the constraints that were fine tuned for the reference plans with coplanar equi-spaced beams.

5.2.1. Beamlet weight optimization runtimes

We observed that the computation times for a full beamlet weight optimization depend strongly on the amount of dose influence data that has to be handled in the main memory. Optimizing a 7-beam plan requires only about 0.1 s on two nodes if we only load seven beams. Compared to the GPU implementation for IMRT treatment plan optimization by (Men et al. 2009) this corresponds to a speed-up factor of more than 20. If we load about 400 beams, however, a beamlet weight optimization for the same beam ensemble requires about 2.5 s on two nodes.

Unfortunately, the communication between multiple nodes is not stable on our research cluster. Due to message synchronization issues, we do not observe an additional speed-up of the beamlet weight optimization if we distribute the calculation of the objective function and its gradient on more than two nodes. This could be easily resolved on a high performance cluster that uses professional message passing software.

Furthermore, we do not observe the expected acceleration of the beamlet weight optimization by calculating multiple treatment plans at once. This issue could probably be resolved with a more low level implementation that explicitly controls the filling of the cache memory.

5.2.2. Convergence properties of the combinatorial optimization strategies

Figure 5.3 visualizes the navigation of the four BAS strategies through the search space. The iterative strategy evaluates the search space in a predefined order (diagonal stripes) and iteratively adds beams to the ideal beam ensemble (horizontal stripes), as shown in figure 5.3(a).

Simulated annealing exhibits randomness during its navigation through the search space, as shown in figure 5.3(b). While the beam configuration undergoes strong fluctuations within the first 500 iterations, afterwards only single beams of the ensemble are replaced and we observe long horizontal stripes.

For the genetic and cross-entropy algorithm, which are visualized in figures 5.3(c) and 5.3(d), a candidate beam may be represented multiple times within one iteration because multiple beam ensembles are evaluated in parallel during one iteration. Again, we observe long horizontal stripes representing beams that prevail over several iterations. In addition, we observe short horizontal stripes representing candidate beam orientations that were evaluated during a few iterations but eventually discarded. This effect is more prominent for the genetic algorithm than for the cross-entropy algorithm.

As expected, the four combinatorial strategies do not yield exactly the same beam configurations. However, figure 5.3 shows that they share certain features. For instance, all modalities exhibit beams around candidate beam # 2 and candidate beam # 349.

Figure 5.4 shows the evolution of the objective function values versus the number of evaluated beam configurations for the four different beam ensemble sampling techniques for the three patient cases. The objective function values of the equi-spaced coplanar beam ensemble and the beam ensemble selected by spherical cluster analysis, which was introduced in section 4.2, are included for comparison. Note that one iteration of the

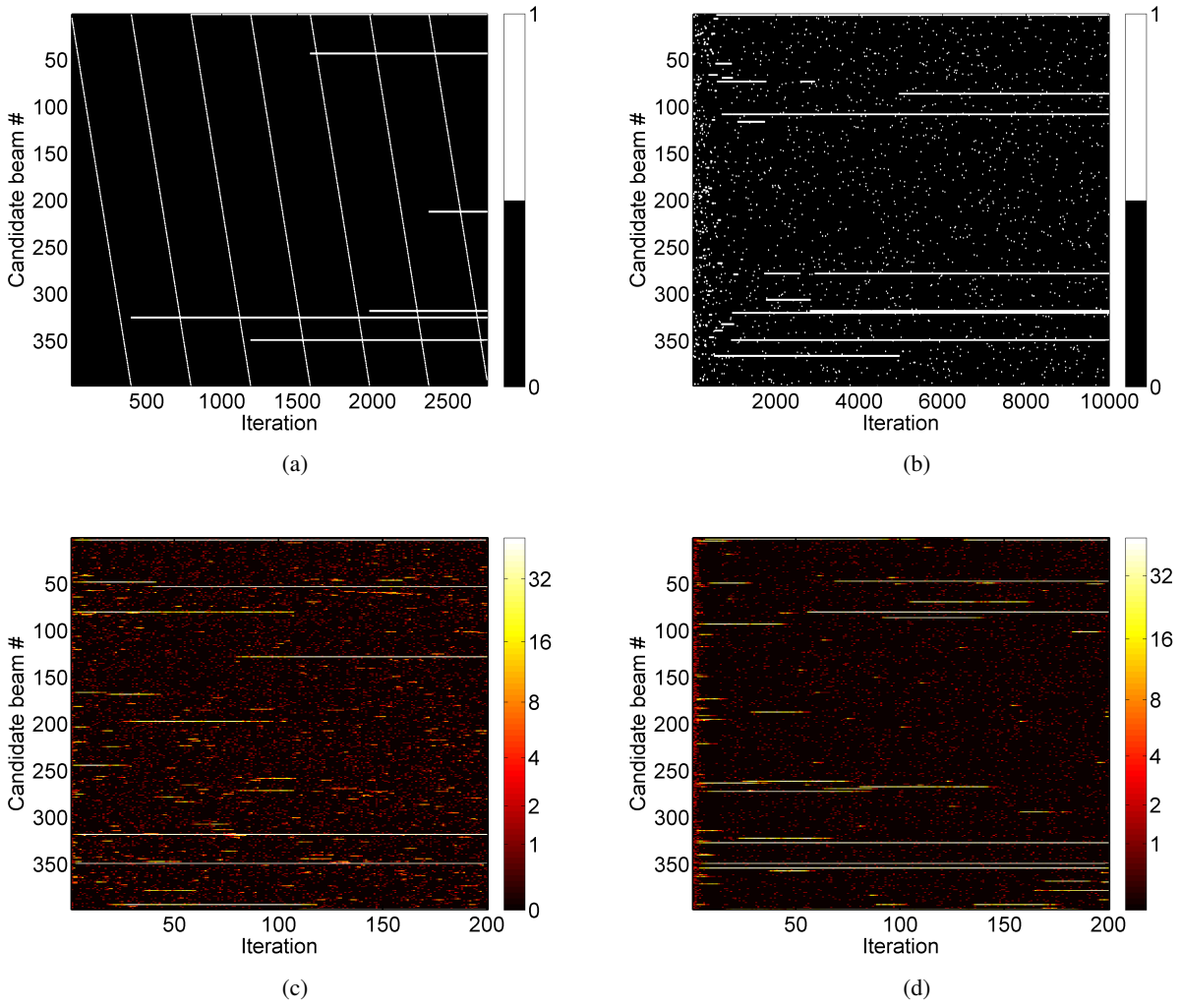
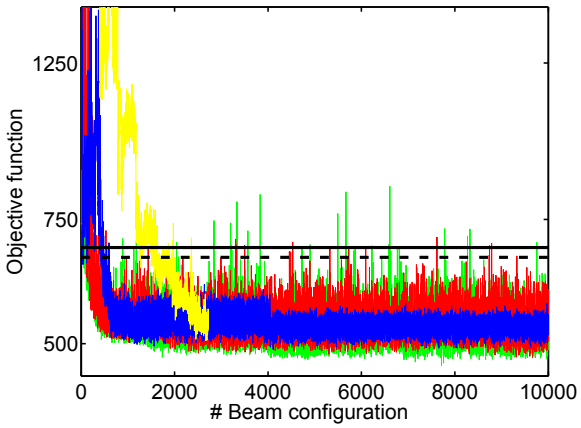
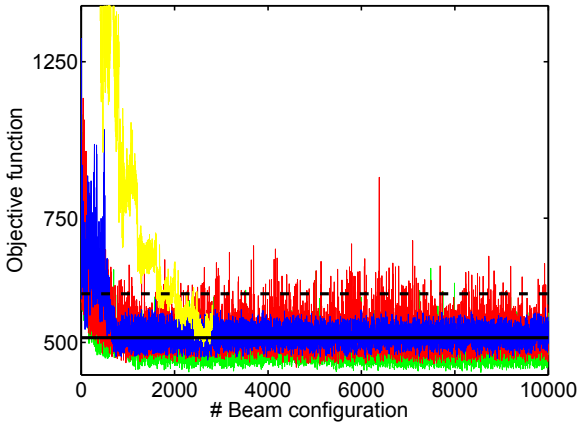


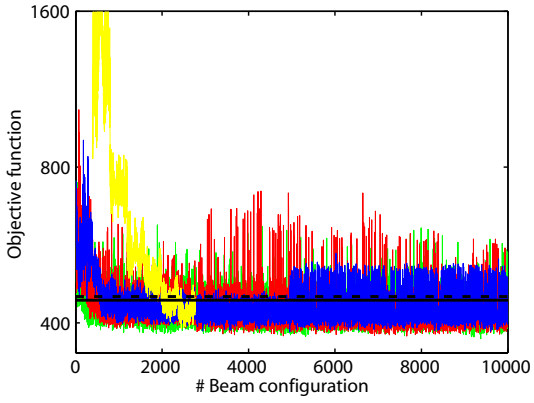
Figure 5.3.: Frequency of the candidate beams during the iterations of (a) the iterative algorithm, (b) the simulated annealing algorithm, (c) the genetic algorithm, and (d) the cross-entropy algorithm for patient case III. Note that the iterative algorithm and the simulated annealing algorithm evaluate only one beam ensemble at a time yielding a binary image. The genetic and cross-entropy algorithm, however, evaluate 50 beam ensembles per iteration. Consequently, a candidate beam may be represented multiple times within one iteration.



(a)



(b)



(c)

Figure 5.4.: Objective function versus number of beam configurations for (a) patient case I, (b) patient case II, and (c) patient case III. Green line: genetic algorithm, red line: cross-entropy algorithm, blue line: simulated annealing algorithm, yellow line: iterative algorithm, solid black line: equi-spaced coplanar beam configuration, dashed black line: beam ensemble selected by spherical cluster analysis. Note that the number of beam configurations corresponds to the number of iterations for the simulated annealing algorithm and the iterative algorithm. For the genetic algorithm and the cross-entropy algorithm 50 beam configurations constitute one iteration.

genetic algorithm and the cross-entropy algorithm corresponds to 50 evaluated beam configurations.

For all three patient cases, the genetic algorithm outperforms the cross-entropy algorithm and the simulated annealing algorithm. Especially during the first 500 iterations, simulated annealing yields higher objective function values. The iterative strategy eventually reaches objective function values comparable to the simulated annealing algorithm. By design, however, the iterative strategy first evaluates treatment plans with less beams and consequently higher objective function values. All strategies eventually yield lower objective function values than the equi-spaced coplanar treatment plan and the treatment plan applying the beam ensemble selected by spherical cluster analysis.

5.2.3. Treatment plan comparison

Table 5.1 summarizes all BAS modalities and patient cases that were investigated for this study. It lists detailed dose statistics for all volumes of interest defining the main conflicts during treatment planning. Figure 5.5 shows dose volume histograms for selected combinatorial beam ensemble optimization modalities in comparison to coplanar equi-spaced beam ensembles and the beam ensembles selected with spherical cluster analysis for all patient cases. Figure 5.6 shows dose distributions of all BAS strategies for patient case III on a selected transversal CT slice.

Table 5.1 shows that all automated BAS selection strategies tend to yield a benefit for the patient. Slightly increased mean and maximum doses compared to the equi-spaced coplanar beam ensembles were observed for some volumes of interest but they may be explained by inadequate tolerance doses in the majority of cases. For the left eye of patient III, for example, there is no incentive for the optimizer to restrict doses below 30 Gy. Consequently, we observe an increased maximum dose of up to 6.1 Gy for the genetic algorithm. However, this explanation is not valid for the right optic nerve of patient case III, where we observe slightly increased maximum doses to the right optic nerve even though the maximum dose is set to 50.0 Gy. Likewise, the maximum dose to the left eye of patient case II is increased by 2.5 Gy for the simulated annealing algorithm and the maximum dose to the right eye (right optic nerve) of patient case I is increased by 1.0 Gy (0.4 Gy) for the genetic algorithm even though the maximum dose constraints are ”active“.

Among the combinatorial optimization strategies, the genetic algorithm yields the best results. The most significant improvements include a reduction of the mean (maximum) dose to the left optic nerve of patient case III by 5.3 Gy (9.4 Gy), a reduction of the mean (maximum) dose to the brainstem of patient case II by 4.2 Gy (6.6 Gy), and a reduction of the mean (maximum) dose to the chiasm of patient case I by 5.9 Gy (2.6 Gy). However, the simulated annealing algorithm and the cross-entropy algorithm also provide superior treatment plans. The mean (maximum) dose to the left optic nerve of patient case II could be reduced by 7.0 Gy (6.0 Gy) with the beam ensemble found by the cross-entropy algorithm and the mean (maximum) dose to the chiasm of patient case I could be reduced by 9.6 Gy (6.2 Gy) with the beam ensemble found by the simulated annealing algorithm. On average, the beam ensembles found with the iterative algorithm resulted in the worst treatment plans out of the four combinatorial optimization strategies. Nevertheless we observed sporadic yet striking improvements compared to the coplanar equi-spaced beam ensembles. For patient case III, for instance, the mean (maximum) dose to the left optic nerve could be reduced by 6.1 Gy (7.9 Gy).

Interestingly, the beam ensembles found by spherical cluster analysis yield very good treatment plans compared to the combinatorial optimization strategies. They are especially beneficial regarding the reduction of low doses to OARs, as visualized in figure 5.5(c) for patient case II and in figure 5.6 for patient case III.

Figure 5.6 also shows that the beam ensembles which were selected with spherical cluster analysis or optimized with one of the four combinatorial strategies yield more conformal dose distributions. Furthermore, the mod-

I

Config	PTV	Opticus (r)	Opticus (l)	Eye (r)	Eye (l)	Chiasm	NT
7 equi	50.3	28.4 (42.1)	30.5 (42.8)	7.1 (17.7)	7.8 (27.1)	37.9 (40.4)	4.8 (55.7)
7 IT	50.3	26.1 (42.1)	30.5 (42.8)	5.6 (18.8)	6.9 (26.8)	32.1 (36.6)	5.2 (57.1)
7 SA opti	50.3	26.4 (41.6)	33.4 (42.2)	5.0 (19.3)	6.9 (25.0)	28.3 (34.2)	5.2 (54.0)
7 GA opti	50.3	24.1 (42.5)	30.7 (42.5)	5.1 (18.7)	6.5 (27.1)	32.0 (37.8)	5.3 (56.1)
7 CE opti	50.3	21.6 (41.4)	31.9 (42.8)	4.9 (17.4)	6.7 (26.5)	28.3 (35.6)	5.2 (57.5)
7 CL opti	50.3	30.5 (43.9)	35.5 (43.4)	6.6 (21.9)	7.9 (24.2)	34.3 (40.0)	4.8 (60.2)
Max dose	-	40.0	40.0	15.0	15.0	40.0	15.0

II

Config	PTV	Opticus (r)	Opticus (l)	Eye (l)	Brainstem	Chiasm	NT
9 equi	59.9	33.6 (52.0)	17.1 (22.0)	8.3 (13.4)	22.5 (41.9)	28.0 (39.8)	3.8 (62.3)
7 IT	59.9	34.7 (50.5)	10.7 (15.9)	8.1 (15.6)	15.7 (36.3)	30.7 (39.9)	4.1 (63.1)
7 SA opti	59.9	29.2 (51.7)	13.4 (16.6)	8.9 (15.9)	20.4 (44.1)	25.9 (38.1)	3.8 (61.8)
7 GA opti	59.9	32.9 (51.5)	15.1 (18.0)	5.5 (10.8)	18.3 (35.3)	28.1 (38.2)	3.8 (62.3)
7 CE opti	59.9	31.9 (51.5)	10.1 (16.0)	6.5 (12.7)	21.0 (39.5)	30.0 (44.5)	3.8 (62.9)
7 CL opti	59.9	40.8 (50.0)	23.8 (31.0)	8.9 (14.4)	6.2 (30.1)	12.1 (31.6)	3.7 (65.5)
Max dose	-	52.0	45.0	10.0	54.0	52.0	15.0

III

Config	PTV	Opticus (r)	Opticus (l)	Eye (l)	Brainstem	Chiasm	NT
9 equi	59.9	40.9 (54.0)	20.9 (36.7)	10.3 (16.8)	26.0 (52.6)	42.5 (50.2)	3.6 (60.9)
7 IT	59.9	28.4 (55.1)	14.8 (28.8)	5.7 (11.2)	21.7 (53.6)	43.7 (50.1)	3.3 (60.9)
7 SA opti	59.9	32.7 (55.8)	18.6 (33.9)	6.2 (19.9)	19.4 (53.4)	40.5 (50.2)	3.4 (61.5)
7 GA opti	59.9	36.5 (50.3)	15.6 (27.3)	5.9 (22.9)	23.3 (53.1)	40.5 (50.1)	3.5 (61.7)
7 CE opti	59.9	36.7 (55.7)	20.6 (33.3)	7.4 (22.5)	20.3 (53.6)	41.6 (50.2)	3.6 (61.4)
7 CL opti	59.9	31.7 (54.6)	11.6 (37.4)	3.8 (14.2)	19.1 (52.8)	40.7 (50.1)	3.7 (62.8)
Max dose	-	50.0	45.0	30.0	52.0	50.0	15.0

Table 5.1.: Mean (Max) dose [Gy] of exclusive DVHs for all three patient cases under investigation. The dose statistics includes all volumes of interest (NT = normal tissue) that define the main conflict during treatment planning of the corresponding case. The last rows list the maximum doses used for th optimization of beamlet weights.

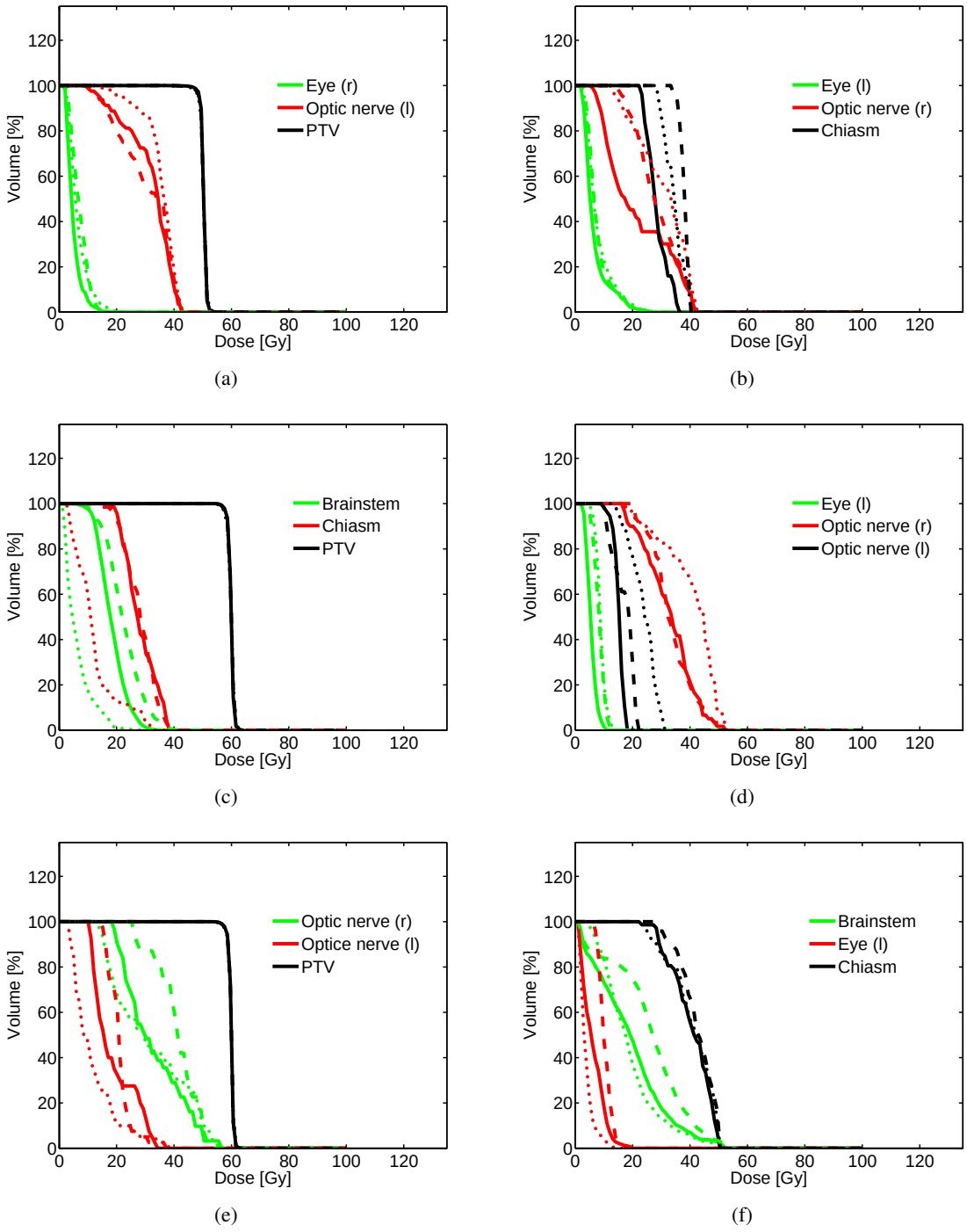


Figure 5.5.: Inclusive DVHs for for (a) & (b) patient case I, (c) & (d) patient case II, and (e) & (f) patient case III. For (a) & (b), the solid lines represent the beam ensemble optimized by the cross-entropy algorithm. For (c) & (d), the solid lines represent the beam ensemble optimized by the genetic algorithm. For (e) & (f), the solid lines represent the beam ensemble optimized by the simulated annealing algorithm. For patient case I, the dashed lines represent the equi-spaced coplanar 7-beam ensemble. For patient case II & III, the dashed lines represent the equi-spaced coplanar 9-beam ensemble. The dotted lines represent the 7-beam ensembles that were selected by spherical cluster analysis, as introduced in section 4.2.

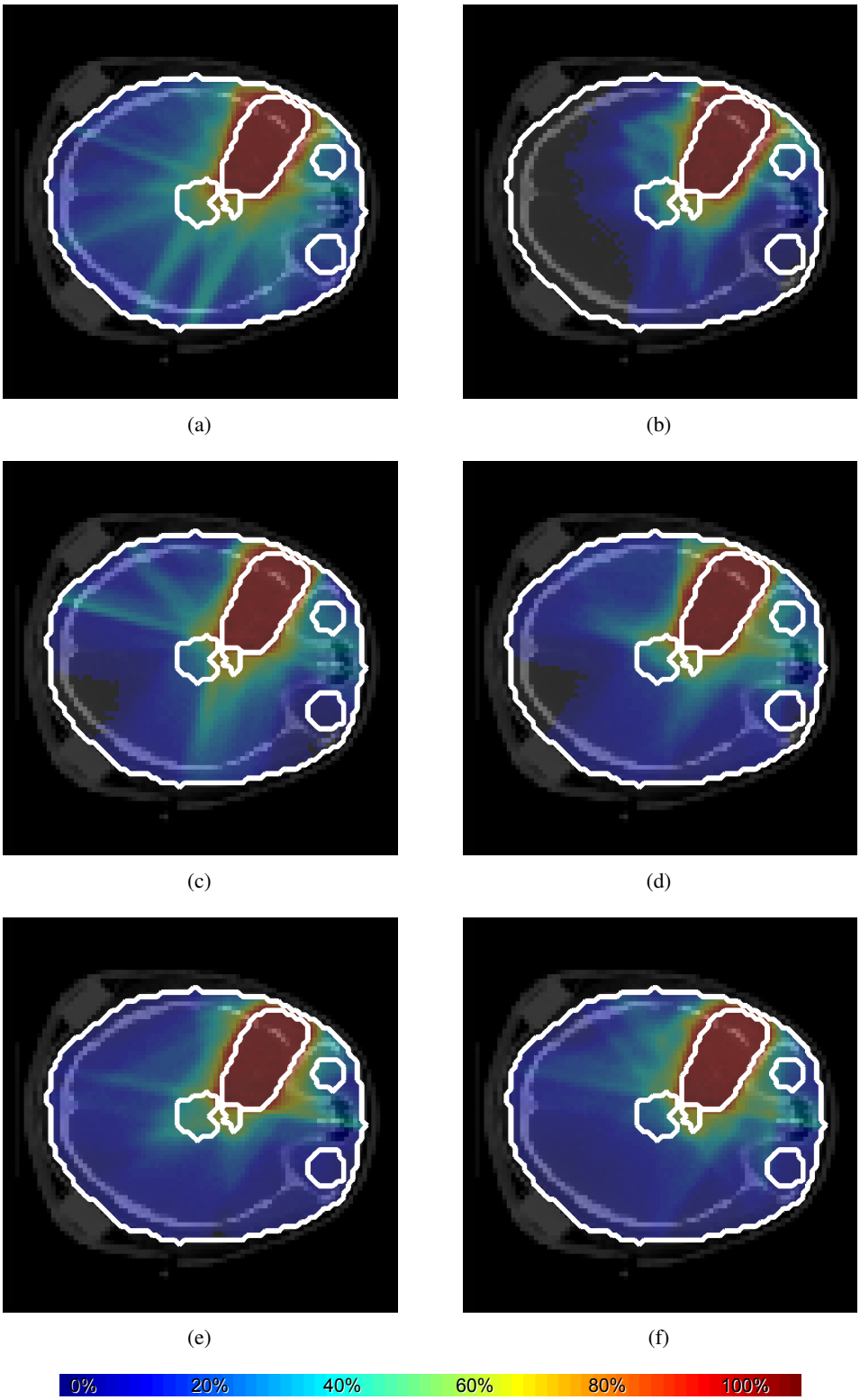


Figure 5.6.: Dose distributions on a selected transversal CT slice for (a) nine equi-spaced coplanar beams, (b) seven beams selected by spherical cluster analysis, (c) seven beams selected with the iterative algorithm, (d) seven beams selected with the simulated annealing algorithm, (e) seven beams selected with the genetic algorithm, and (f) seven beams selected with the cross-entropy algorithm for patient case III. 100% corresponds to the prescribed target dose (60 Gy).

ulation of the individual fields is less pronounced for the optimized beam ensembles. The high dose stripes, which are visible in figure 5.6(a) almost disappear for the optimized beam ensembles for the three patient cases investigated.

5.3. Discussion

5.3.1. Parallel beamlet optimization module

The parallel beamlet weight optimization module yields a significant speed-up of the inverse planning process. For conventional treatment planning problems featuring $\eta < 11$ beams the beamlet weight optimization process is accelerated by more than two orders of magnitude compared to our clinical inverse planning software. The acceleration, however, is significantly compromised if the amount of dose influence data in the main memory is very large. When handling about 7 GB of dose influence data on one node we observed optimization times of 2.5 s for a treatment plan featuring seven beams. Nevertheless, this allows for the calculation of $10,000$ beam configurations in 7 h.

The development of the parallel beamlet weight optimization algorithm is work in progress. As pointed out earlier, we still have issues regarding the controlled filling of the cache memory and the message synchronization between multiple nodes. An additional significant increase in speed is anticipated by resolving these two problems. Solving the message synchronization issues is also critical if we want to increase the number of non-coplanar candidate beams beyond the current limitations.

The application of a high performance beamlet weight optimization engine is not limited to the BAS problem. It may be a general tool to boost new approaches to radiation therapy treatment planning. The simultaneous optimization of multiple treatment plans may allow for novel and fast concepts for robust treatment planning for particle therapy. Multicriteria optimization and direct aperture optimization may be accelerated and tackled with techniques that require a lot of beamlet weight optimization processes - just like BAS. Furthermore the high performance beamlet weight optimization engine may contribute to a more efficient optimization process for modulated arc therapy and robotic radiosurgery.

5.3.2. Combinatorial beam ensemble optimization

All four combinatorial BAS strategies yield a significant objective function decrease compared to the standard coplanar equi-spaced beam configuration. The genetic algorithm showed the best convergence properties. All strategies yield only marginal improvements of the objective function values after evaluating about $2,500$ beam ensembles indicating that the algorithms reached convergence even though only a vanishing fraction of the solution space has been searched. This may be considered as additional evidence for the degeneracy of the BAS problem.

It may be possible to improve the efficiency of both the genetic algorithm and the cross-entropy algorithm with a more sophisticated fine tuning of the optimization parameters, but it is debatable whether this translates into a measurable clinical benefit. We doubt that the simulated annealing algorithm may be improved by incorporating spatial information during the beam ensemble sampling process as suggested by Mageras and Mohan (1993) due to the severe non-convexity of the search space.

All four combinatorial BAS strategies yield a significant improvement of the resulting dose distributions compared to the standard coplanar equi-spaced beam configuration for the three patient cases under investigation. The genetic algorithm yields slightly superior treatment plans compared to the cross-entropy algorithm and simulated annealing algorithm. The iterative strategy provides surprisingly good treatment plans compared to

the three other strategies even though it is the least flexible BAS algorithm. Once a candidate beam is accepted into the beam ensemble it cannot be replaced by a better beam.

Compared to the combinatorial optimization strategies, the beam ensembles selected by spherical cluster analysis yield very good treatment plans at a significantly lower computational cost. Apparently, our heuristic approach captures a lot of the attributes of a beneficial beam ensemble a priori.

It will be interesting to investigate the performance of the combinatorial BAS strategies for alternative treatment sites within the abdomen and pelvis. We do not expect a different relative performance of the combinatorial optimization strategies, but it may be possible that they yield improvements for symmetric patient geometries that were not observed for the heuristics discussed in chapter 4. Furthermore, the combinatorial beam ensemble optimization framework could be used directly for BAS for particle therapy. It is only necessary to provide the dose influence data of the desired radiation modality. Of course, the current implementation does not account for potential uncertainties in the calculated range of particles or for errors during patient immobilization. This is left for future research.

The choice of suited constraint and penalty settings is crucial for the BAS process with our full precision large-scale framework. The optimizer exploits any shortcomings of the prescription to arrive at a lower objective function value. It is very difficult to select the constraints and penalties before the optimization process, because not all trade-offs of the optimization are clear a priori and they may have to be identified on a patient specific basis. A data base which provides dose statistics of comparable patient geometries, as suggested by Moore et al. (2011), might be useful for the planner to realize what prescription is feasible. Unfortunately we did not have the time to investigate the role of different penalty and constraint settings for the patient cases under investigation. As shown in table 5.1, the dose constraints on the OARs were ”inactive“ in some cases and the decreased mean and maximum doses were only induced indirectly by a tight maximum dose constraint on the surrounding normal tissue. It may be possible to further improve the dose distributions with targeted OAR constraints.

Ultimately we should experiment with different simplifications of the combinatorial beam ensemble optimization process. Using the results of the full precision large-scale BAS framework as a gold standard, we can identify which approximations compromise the quality of the resulting treatment plans and which approximations do not have a negative effect on the resulting treatment plans. Hence, the complexity may eventually be reduced in order to alleviate a transition of the full precision large-scale BAS framework to clinical application.

6. Conclusions

6.1. Summary

We introduced BAS as an essential part of the overall treatment planning process in radiation therapy in chapter 1. The relevant characteristics of photon and particle beams for BAS were illustrated and potential reasons for the limited clinical acceptance of existing automated BAS algorithms were discussed.

Section 2.1 gave a detailed overview of the conventional inverse planning problem in radiation therapy. We provided an in-depth analysis of the mathematical characteristics of the beamlet weight optimization problem and explained efficient algorithms for the solution of the beamlet weight optimization problem that apply an L-BFGS update of the inverse Hessian in combination with an Armijo line search. Section 2.2 introduced BAS as a formal extension of the conventional inverse planning problem and discussed the challenges of the BAS problem in detail. We elaborated on the difficulties to obtain derivatives of the objective function with respect to beam orientations, the characteristics of the solution space, namely its pronounced non-convexity, its exponential NP-hard complexity, and its degeneracy, as well as the delicate question regarding the ideal number of beams. We commented on the prerequisite to incorporate the influence of uncertainties into a dedicated BAS framework for particle therapy and made a clear distinction between beam angle selection and beam angle optimization strategies.

Within chapter 3, we identified two different classes of BAS algorithms in the literature: while the first class intertwines BAS and the optimization of beamlet weights, the second class derives beneficial beam ensembles before the optimization of beamlet weights. Based on a thorough review of published BAS approaches we deduced two main implications for improved BAS strategies: the realization of higher complexity in existing BAS models of the first class and the conceptual advancement of BAS models of the second class.

Section 4.1 introduced new concepts for BAS algorithms relying on score functions. Based on a custom radiological score, we developed a score matrix S that introduces spatial information about the dose deposition of a candidate beam within the target volume. Section 4.2 showed how to use this additional information to overcome certain limitations of existing BAS algorithms and to establish a contained framework that considers the combinatorial problem of BAS. Therefore the BAS problem was translated to an intuitive clustering problem of locally ideal beam orientations on the unit sphere. We were able to demonstrate a clear benefit of our approach based on a treatment planning study for nine patient cases. Our model had the biggest impact for complex patient geometries where the target volume was located asymmetrically within the patient. In section 4.3 we investigated the potential to infer the ideal number of beams from the spherical distribution of locally ideal beam orientations. Therefore we developed the iMFMM as a general framework for density estimation on the D-sphere. The numbers of beams derived with the iMFMM were clinically prohibitive for conventional IMRT treatments, but an application of the iMFMM to improve treatment planning for robotic radiosurgery or modulated arc therapy may be feasible in the future. Section 4.4 applied the score matrix S for BAS in proton therapy and introduced an alternative concept to translate the information of S into a beneficial beam ensemble. In order to incorporate the influence of range and setup uncertainties on the BAS, worst case score scenarios were considered within our framework and we introduced restrictions on the modulation of the individual proton fields. First experiments yielded plausible beam configurations but the measures taken to induce treatment plan robustness are only satisfying regarding OAR sparing. Within section 4.5 we picked up an idea

by Potrebko et al. (2008) and enhanced their geometric BAS algorithm with the concept of spatial target information. The resulting fully automated BAS algorithm required only 2 s to establish a beam ensemble. A planning study including nine patient cases showed that treatment plans for very complex lesions within the skull may be improved significantly with this technique. Interestingly, the resulting beam configurations exhibited a fundamentally different structure compared to the beam ensembles selected by spherical cluster analysis in combination with the radiological score: while the beam configurations resulting from the geometric approach were equally distributed around the patient and only locally refined, the beam configurations resulting from the radiological approach were strongly asymmetric yet both yielded beneficial treatment plans.

Chapter 5 dealt with the brute force solution of the extended inverse problem, i.e. the simultaneous optimization of beamlet weights and beam orientations. Therefore, a fast beamlet weight optimization engine was implemented on a high performance cluster that decreased the optimization time for standard patient cases to less than one second. This allowed for the application of modern metaheuristics that sample a large number of beam ensembles within the solution space at an unprecedented precision to identify a beneficial solution. First experiments with three intracranial patient cases showed that a genetic algorithm outperforms a greedy iterative algorithm, a simulated annealing algorithm, and a cross-entropy algorithm. We found that the prescription, i.e. the patient specific, nontrivial setting of tolerance doses and penalties, is critical for the outcome of the beam ensemble optimization. The beam ensembles selected by spherical cluster analysis yielded clinically comparable treatment plans at a significantly lower computational cost.

6.2. Innovation

The main contribution of this work is the advancement of BAS strategies relying score functions. With the extension of score vectors that list a radiological quality for every candidate beam orientation to score matrices that list the radiological quality for every candidate beam and every target voxel, we introduce spatial information about the dose deposition of individual beamlets within the target to the class of heuristic BAS strategies. The translation of the BAS problem either to a clustering problem on the sphere or to a combinatorial optimization problem of an average score are intuitive and efficient concepts to establish a fully automated BAS process. Furthermore, the idea of considering a worst case score scenario is one of the first approaches to induce treatment plan robustness for particle therapy treatments through smart BAS.

Our full precision large-scale BAS framework is one of the fastest beamlet weight optimization engines and it can handle the dose influence data of over 400 candidate beam orientations. With this framework, we compare different metaheuristics for combinatorial optimization and evaluate a cross-entropy algorithm for BAS for the first time. In comparison to a simulated annealing algorithm and a genetic algorithm, however, we found that the quality of the treatment plans does not depend critically on the metaheuristics used for combinatorial optimization.

6.3. Recommendations and outlook

6.3.1. Photons

We have introduced conceptual advancements of BAS heuristics and investigated joint beamlet weight and beam orientation optimization strategies at an unprecedented precision. In agreement with the findings of earlier work, we have confirmed a clear clinical benefit of dedicated BAS frameworks.

Based on our experience with different BAS strategies, we think that future work on BAS for photons should focus on an efficient transition of existing BAS algorithms into the clinic. For the benefit of the patient it is

critical to answer the question why treatment planners still do not use automated BAS strategies.

We found that a smart heuristic, like spherical cluster analysis of locally ideal beam orientations based on radiological information, may yield significant improvements of the treatment plans of very complex patient geometries at an acceptable computational cost. To advance this tool into clinical testing, only minor modifications would be necessary. In contrast, an ad hoc application of the full precision large-scale BAS framework used in chapter 5 in the clinic may be more difficult yet not impossible. Even though an interactive BAS is infeasible, beneficial beam ensembles for more than ten patients could be calculated in one night using the current implementation. The BAS tool performing spherical cluster analysis based on geometric information may be very attractive in a clinical setting because it requires less than 5 s to derive a beneficial beam configuration. A stand-alone MatLab implementation is ready-for-use.

During the transition process, it will be essential to have the acceptance of the treatment planners and incorporate their needs into the software. Therefore, the BAS algorithms may have to be more flexible to accommodate individual preferences regarding the beam orientations. The planner should be able to quickly adjust a beam ensemble according to her/his ideas. Unfortunately this is currently impossible with our in-house treatment planning software KonRad, where a beam ensemble is absolutely fixed for a given treatment plan. With regard to elongated treatment times it may be desirable to combine a set of equi-spaced coplanar beams with one, two, or three optimized non-coplanar beams. Such combined approaches could be a first step towards a clinical application of automated BAS at our center. A clinical application of the heuristics presented in this thesis may be encouraged by first making the clustering landscapes available to treatment planners as guideline during the manual selection of a beam ensemble. Maybe it is necessary to start with such intermediate steps in order to gain the planners' confidence in a fully automated BAS algorithm in the long run.

Note that there are no legal or quality assurance issues associated with an immediate clinical test of different BAS strategies because the final treatment plan is still calculated with a certified inverse planning program. In principle, any beam ensemble derived with an automated BAS algorithm might have been chosen by the human planner. As long as the final treatment plan is approved it does not matter how the underlying beam ensemble was established.

Besides the transition of existing BAS strategies into the clinic we suggest research regarding the application of BAS beyond conventional IMRT treatment planning. Finding ideal rotation paths for modulated arc therapy and finding ideal beam orientations for robotic radiosurgery are two potential directions.

6.3.2. Particles

Even though the selection of suited beam orientations appears to be more intuitive for particle therapy than for conventional photon therapy, we think that a dedicated BAS strategy for particles may provide a clinical benefit for complicated cases.

For the future we see two potential pathways for research projects related to BAS for particle therapy. First, it may be possible to enhance the concepts for BAS based on a score matrix S and accommodate the peculiarities of particle therapy. As the choice of our custom score proved somewhat unfortunate for particles in retrospect, we suggest to experiment with modified formulations of our score. However, we vehemently advocate to keep the matrix structure of the score for particles because here the quality of a candidate beam orientation may vary even more for different target voxels than for photons. Second, the fast beamlet weight optimization module may even allow for a straight forward robust fluence optimization in combination with a combinatorial optimization of beam orientations. The a priori definition of robust fluence profiles of individual fields as brought up in section 4.4 may be a cheap alternative to a costly robust beamlet weight optimization.

Custom-tailored BAS strategies for more basic particle therapy techniques could be developed in a first step. As of 2011, using spread-out Bragg peaks on passive beam lines is still the predominant irradiation technique in therapy centers around the world[1]. Here, the quantification of the influence of uncertainties on the resulting treatment plans is not as complex as for full-fledged intensity modulated particle therapy treatment plans.

[1] See http://ptcog.web.psi.ch/ptcentres.html for a list of particle therapy facilities in operation.

Appendix

A. Coordinate system

Figure A.1 shows a sketch of the custom coordinate system, which is used throughout this manuscript to describe beam orientations. This coordinate system represents a beam orientation by specifying the angle α around the patient axis and the angle β towards the transversal plane. For us, this coordinate system is more intuitive than the standard IEC system that uses couch and gantry angles to describe beam orientations.

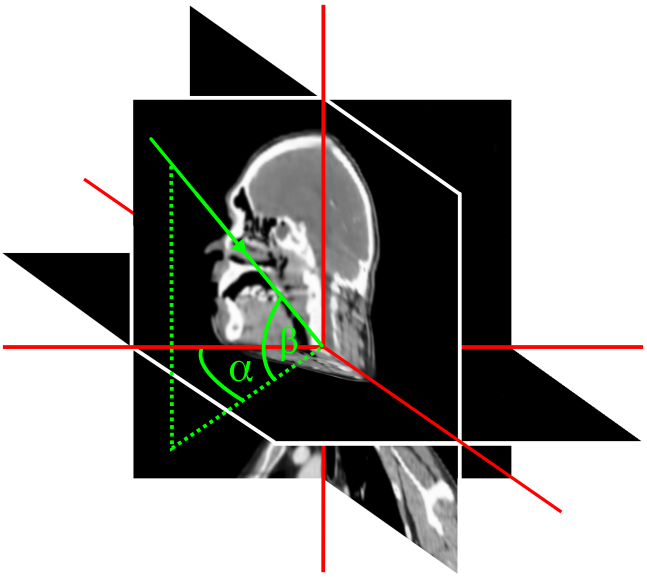


Figure A.1.: Sketch of the coordinate system used throughout this manuscript to describe beam orientations. α denotes the angle around the patient axis, β denotes the angle towards the transversal plane.

Bibliography

M. Alber and G. Meedt. On the visualization of universal degeneracy in the IMRT problem. *Radiation Oncology*, 1(1), 2006. 19

M. Alber and F. Nüsslin. Optimization of intensity modulated radiotherapy under constraints for static and dynamic MLC delivery. *Physics in Medicine and Biology*, 46, 2001. 23

M. Alber, G. Meedt, F. Nusslin, and R. Reemtsen. On the degeneracy of the IMRT optimization problem. *Medical Physics*, 29(11), 2002. 19, 55

E. Alpaydin. *Introduction to Machine Learning (Adaptive Computation and Machine Learning)*. The MIT Press, 2004. 49

H. Andreyev. Gastrointestinal problems after pelvic radiotherapy: the past, the present and the future. *Clinical Oncology*, 19, 2007. 13

J. Backus. Can programming be liberated from the von Neumann style?: a functional style and its algebra of programs. In *ACM Turing award lectures*. ACM, 2007. 84

M. Bangert and U. Oelfke. A novel framework for beam angle optimization in intensity modulated hadron therapy. *Posterpresentation for the 48th Meeting of the Particle Therapy Co-Operative Group in Heidelberg, Germany.*, 2009a. 9, 68

M. Bangert and U. Oelfke. Fast optimization of non-coplanar beam angle sets for intensity modulated radiation therapy. In *World Congress on Medical Physics and Biomedical Engineering, September 7 - 12, 2009, Munich, Germany*, volume 25/I of *IFMBE Proceedings*. Springer, 2009b. 9, 71

M. Bangert and U. Oelfke. Clustering analysis of geometric data for beam ensemble selection in radiation therapy. In *Proceedings of the XVIth International Conference on the Use of Computers in Radiation Therapy (ICCR)*. J.-J. Sonke, 2010a. 9

M. Bangert and U. Oelfke. Robust IMPT treatment planning by reducing the degeneracy of the optimization problem in combination with beam angle optimization. *Posterpresentation for the 49th Meeting of the Particle Therapy Co-Operative Group in Gunma, Japan.*, 2010b. 9

M. Bangert and U. Oelfke. Spherical cluster analysis for beam angle optimization in intensity-modulated radiation therapy treatment planning. *Physics in Medicine and Biology*, 55(19), 2010c. 9

M. Bangert, P. Hennig, and U. Oelfke. Using an infinite von Mises-Fisher mixture model to cluster treatment beam directions in external radiation therapy. In *Machine Learning and Applications, Fourth International Conference on*, volume 0, Los Alamitos, CA, USA, 2010. IEEE Computer Society. 9

R. Bendl, J. Pross, A. Hoess, M. Keller, K. Preiser, and W. Schlegel. VIRTUOS–A program for VIRTUal radiOtherapy Simulation and verification. In *Proc 11th Int Conf on the Use of Computers in Radiation Therapy, Manchester*, 1994. 75

S. Bentzen, L. Constine, J. Deasy, A. Eisbruch, A. Jackson, L. Marks, R. Ten Haken, and E. Yorke. Quantitative Analyses of Normal Tissue Effects in the Clinic (QUANTEC): an introduction to the scientific issues. *International Journal of Radiation Oncology* Biology* Physics*, 76(3), 2010. 13

C. M. Bishop. *Pattern Recognition and Machine Learning*. Springer, Berlin, new edition edition, 2007. 57

T. Bortfeld. IMRT: a review and preview. *Physics in Medicine and Biology*, 51(13), 2006. 13, 23

T. Bortfeld. The number of beams in imrt investigations and implications for single-arc IMRT. *Physics in Medicine and Biology*, 55(1), 2010. 31

T. Bortfeld and W. Schlegel. Optimization of beam orientations in radiation therapy: some theoretical considerations. *Physics in Medicine and Biology*, 38(2), 1993. 15, 25, 26, 36, 41

M. Braunstein and R. Y. Levine. Optimum beam configurations in tomographic intensity modulated radiation therapy. *Physics in Medicine and Biology*, 45(2), 2000. 39, 41

G. T. Chen, D. R. Spelbring, C. A. Pelizzari, J. M. Balter, L. C. Myrianthopoulos, S. Vijayakumar, and H. Halpern. The use of beam's eye view volumetrics in the selection of non-coplanar radiation portals. *International Journal of Radiation Oncology* Biology* Physics*, 23(1), 1992. 39

B. C. Cho, W. H. Roa, D. Robinson, and B. Murray. The development of target-eye-view maps for selection of coplanar or noncoplanar beams in conformal radiotherapy treatment planning. *Medical Physics*, 26(11), 1999. 39

T. Cormen. *Introduction to algorithms*. The MIT press, 2001. 29

L. Cozzi, A. Fogliata, A. Lomax, and A. Bolsi. A treatment planning comparison of 3D conformal therapy, intensity modulated photon therapy and proton therapy for treatment of advanced head and neck tumours. *Radiotherapy and Oncology*, 61(3), 2001. 13

D. Craft. Local beam angle optimization with linear programming and gradient search. *Physics in Medicine and Biology*, 52(7), 2007. 15, 23, 25, 38, 41

S. Das, T. Cullip, G. Tracton, S. Chang, L. Marks, M. Anscher, and J. Rosenman. Beam orientation selection for intensity-modulated radiation therapy based on target equivalent uniform dose maximization. *International Journal of Radiation Oncology* Biology* Physics*, 55(1), 2003. 35

P. De Boer, D. Kroese, S. Mannor, and R. Rubinstein. A tutorial on the cross-entropy method. *Annals of Operations Research*, 134(1), 2005. 90

W. De Gersem, F. Claus, C. De Wagter, B. Van Duyse, and W. De Neve. Leaf position optimization for step-and-shoot IMRT* 1. *International Journal of Radiation Oncology* Biology* Physics*, 51(5), 2001. 23

J. de Pooter, A. Méndez Romero, W. Jansen, P. Storchi, E. Woudstra, P. Levendag, and B. Heijmen. Computer optimization of noncoplanar beam setups improves stereotactic treatment of liver tumors. *International Journal of Radiation Oncology* Biology* Physics*, 66(3), 2006. 38, 42

J. A. de Pooter, A. M. Romero, W. Wunderink, P. R. M. Storchi, and B. J. M. Heijmen. Automated non-coplanar beam direction optimization improves IMRT in SBRT of liver metastasis. *Radiotherapy and Oncology*, 88 (3), 2008. 38

P. Denning. The locality principle. *Communications of the ACM*, 48(7), 2005. 86

R. V. der Laarse and J. Strackee. Pseudo optimization of radiotherapy treatment planning. *The British Journal of Radiology*, 49(581), 1976. 35

D. Djajaputra, Q. Wu, Y. Wu, and R. Mohan. Algorithm and performance of a clinical IMRT beam-angle optimization system. *Physics in Medicine and Biology*, 48(19), 2003. 37

A. Djouguela, D. Harder, R. Kollhoff, S. Foschepoth, W. Kunth, A. Ruhmann, K. Willborn, and B. Poppe. Fourier deconvolution reveals the role of the lorentz function as the convolution kernel of narrow photon beams. *Physics in Medicine and Biology*, 54(9), 2009. 46

W. D. D'Souza, H. H. Zhang, D. P. Nazareth, L. Shi, and R. R. Meyer. A nested partitions framework for beam angle optimization in intensity-modulated radiation therapy. *Physics in Medicine and Biology*, 53(12), 2008. 40, 41

M. Ehrgott, A. Holder, and J. Reese. Beam selection in radiotherapy design. *Linear Algebra and its Applications*, 428(5-6), 2008. 15, 24, 31

K. Engel and E. Tabbert. Fast simultaneous angle, wedge, and beam intensity optimization in inverse radiotherapy planning. *Optimization and Engineering*, 6(4), 2005. 36

S. Geman and D. Geman. Stochastic relaxation, gibbs distributions and the bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6(6), 1984. 57

A. Griewank and A. Walther. *Evaluating derivatives: principles and techniques of algorithmic differentiation*. Society for Industrial and Applied Mathematics (SIAM), 2008. 25

T. L. Griffiths and Z. Ghahramani. Infinite latent feature models and the indian buffet process. *IN NIPS*, 18, 2005. 60

O. C. Haas, K. J. Burnham, and J. A. Mills. Optimization of beam orientation in radiotherapy using planar geometry. *Physics in Medicine and Biology*, 43(8), 1998. 38

C. S. Hope and O. Cain. A computer program for optimised stationary beam treatment planning using score functions. *Computer Programs in Biomedicine*, 2(3), 1972. 35

C. S. Hope, J. Laurie, J. S. Orr, and K. E. Halnan. Optimization of x-ray treatment planning by computer judgement. *Physics in Medicine and Biology*, 12(4), 1967. 35

Q. Hou, J. Wang, Y. Chen, and J. M. Galvin. Beam orientation optimization for IMRT by a hybrid method of the genetic algorithm and the simulated dynamics. *Medical Physics*, 30(9), 2003. 30, 37, 88

O. Jäkel and J. Debus. Selection of beam angles for radiotherapy of skull base tumours using charged particles. *Physics in Medicine and Biology*, 45(5), 2000. 40

S. Kirkpatrick, C. D. Gelatt, and M. P. Vecchi. Optimization by simulated annealing. *Science*, 220(4598), 1983. 36, 88

J. Lei and Y. Li. An approaching genetic algorithm for automatic beam angle selection in IMRT planning. *Computer Methods and Programs in Biomedicine*, 93(3), 2009. 37, 41

H. Li, C. Fox, H. Romeijn, and J. Dempsey. SU-FF-T-04: 3D Intensity Modulated Proton Therapy with Minimal Beam Number. *Medical Physics*, 33, 2006. 65

Y. Li and J. Lei. A feasible solution to the beam-angle-optimization problem in radiotherapy planning with a DNA-based genetic algorithm. *IEEE Transactions on Bio-Medical Engineering*, 57(3), 2010. 37

Y. Li, J. Yao, and D. Yao. Automatic beam angle selection in IMRT planning using genetic algorithm. *Physics in Medicine and Biology*, 49(10), 2004. 37, 41, 89

Y. Li, D. Yao, J. Yao, and W. Chen. A particle swarm optimization algorithm for beam angle selection in intensity-modulated radiotherapy planning. *Physics in Medicine and Biology*, 50(15), 2005. 37, 41

G. Lim, A. Holder, and J. Reese. A clustering approach for optimizing beam angles in IMRT planning. In *Proceedings of the IIE Annual Conference*, 2009. 40, 41, 42

R. Liu, J. Buatti, T. Howes, J. Dill, J. Modrick, and S. Meeks. Optimal number of beams for stereotactic body radiotherapy of lung and liver lesions. *International Journal of Radiation Oncology* Biology* Physics*, 66 (3), 2006. 32

J. Llacer, S. Li, N. Agazaryan, C. Promberger, and T. D. Solberg. Non-coplanar automatic beam orientation selection in cranial IMRT: a practical methodology. *Physics in Medicine and Biology*, 54(5), 2009. 30, 38, 42, 55

A. Lomax. Intensity modulated proton therapy and its sensitivity to treatment uncertainties 1: the potential effects of calculational uncertainties. *Physics in medicine and biology*, 53, 2008a. 32, 67

A. Lomax. Intensity modulated proton therapy and its sensitivity to treatment uncertainties 2: the potential effects of inter-fraction and inter-field motions. *Physics in Medicine and Biology*, 53, 2008b. 32, 67

D. J. C. MacKay. *Information Theory, Inference and Learning Algorithms*. Cambridge University Press, 2003. 49

G. Mageras and R. Mohan. Application of fast simulated annealing to optimization of conformal radiation treatments. *Medical physics*, 20, 1993. 99

M. Matuszak, E. Larsen, and B. Fraass. Reduction of IMRT beam complexity through the use of beam modulation penalties in the objective function. *Medical physics*, 34, 2007. 23

M. Matuszak, E. Larsen, K. Jee, D. McShan, and B. Fraass. Adaptive diffusion smoothing: a diffusion-based method to reduce IMRT field complexity. *Medical physics*, 35(4), 2008. 23

G. Meedt, M. Alber, and F. NÃ$\frac{1}{4}$sslin. Non-coplanar beam direction optimization for intensity-modulated radiotherapy. *Physics in Medicine and Biology*, 48(18), 2003. 30, 38, 55

C. Men, X. Gu, D. Choi, A. Majumdar, Z. Zheng, K. Mueller, and S. Jiang. GPU-based ultrafast IMRT plan optimization. *Physics in medicine and biology*, 54, 2009. 92

J. Meyer, S. M. Hummel, P. S. Cho, M. M. Austin-Seymour, and M. H. Phillips. Automatic selection of non-coplanar beam directions for three-dimensional conformal radiotherapy. *British Journal of Radiology*, 78 (928), 2005. 39

M. Mitchell. *An Introduction to Genetic Algorithms*. The MIT Press, third printing edition, 1998. 37, 88

K. Moore, R. Brame, D. Low, and S. Mutic. Experience-Based Quality Control of Clinical Intensity-Modulated Radiotherapy Planning. *International Journal of Radiation Oncology* Biology* Physics*, 2011. 100

Z. Moravek, M. Rickhey, M. Hartmann, and L. Bogner. Uncertainty reduction in intensity modulated proton therapy by inverse monte carlo treatment planning. *Physics in Medicine and Biology*, 54(15), 2009. 40, 42, 55, 65

L. C. Myrianthopoulos, G. T. Chen, S. Vijayakumar, H. J. Halpern, D. R. Spelbring, and C. A. Pelizzari. Beam's eye view volumetrics: an aid in rapid treatment plan development and evaluation. *International Journal of Radiation Oncology* Biology* Physics*, 23(2), 1992. 39

D. P. Nazareth, S. Brunner, M. D. Jones, H. K. Malhotra, and M. Bakhtiari. Optimization of beam angles for intensity modulated radiation therapy treatment planning using genetic algorithm on a distributed computing platform. *Journal of Medical Physics / Association of Medical Physicists of India*, 34(3), 2009. 37, 89

R. M. Neal. Markov chain sampling methods for dirichlet process mixture models. *Journal of Computational and Graphical Statistics*, 9(2), 2000. 57, 60, 63

R. M. Neal. Slice sampling. *The Annals of Statistics*, 31(3), 2003. 58, 59

S. Nill. *Development and application of a multi-modality inverse treatment planning system*. PhD thesis, Universitätsbibliothek Heidelberg, 2001. 84

J. Nocedal and S. Wright. *Numerical optimization*. Springer Verlag, 1999. 17, 19, 20, 22, 23, 84

U. Oelfke and T. Bortfeld. Inverse planning for photon and proton beams. *Medical Dosimetry*, 26(2), 2001. 17, 50, 75, 83

C. G. Orton, T. R. Bortfeld, A. Niemierko, and J. Unkelbach. The role of medical physicists and the AAPM in the development of treatment planning and optimization. *Medical Physics*, 35(11), 2008. 15, 35, 41

D. Pflugfelder, J. Wilkens, S. Nill, and U. Oelfke. A comparison of three optimization algorithms for intensity modulated radiation therapy. *Zeitschrift für Medizinische Physik*, 18(2), 2008a. 84

D. Pflugfelder, J. J. Wilkens, and U. Oelfke. Worst case optimization: a method to account for uncertainties in the optimization of intensity modulated proton therapy. *Physics in Medicine and Biology*, 53(6), 2008b. 32, 67, 68, 69, 70, 71

P. S. Potrebko, B. M. C. McCurdy, J. B. Butler, A. S. El-Gubtan, and Z. Nugent. Optimal starting gantry angles using equiangular-spaced beams with intensity modulated radiation therapy for prostate cancer on RTOG 0126: a clinical study of 5 and 7 fields. *Radiotherapy and Oncology*, 85(2), 2007. 36

P. S. Potrebko, B. M. C. McCurdy, J. B. Butler, and A. S. El-Gubtan. Improving intensity-modulated radiation therapy using the anatomic beam orientation optimization algorithm. *Medical Physics*, 35(5), 2008. 39, 41, 55, 71, 72, 74, 76, 102

K. Preiser, T. Bortfeld, K. Hartwig, W. Schlegel, and J. Stein. A new program for inverse radiotherapy planning. In *Proc. of the XIIth International Conference on the Use of Computers in Radiation Therapy*. Medical Physics Publishing, 1997. 49, 50, 75, 84, 92

A. Pugachev and L. Xing. Pseudo beam's-eye-view as applied to beam orientation selection in intensity-modulated radiation therapy. *International Journal of Radiation Oncology* Biology* Physics*, 51(5), 2001a. 39

A. Pugachev and L. Xing. Computer-assisted selection of coplanar beam orientations in intensity-modulated radiation therapy. *Physics in Medicine and Biology*, 46(9), 2001b. 26, 39, 41, 43, 44, 45, 71

A. Pugachev and L. Xing. Incorporating prior knowledge into beam orientation optimization in IMRT. *International Journal of Radiation Oncology* Biology* Physics*, 54(5), 2002. 40

A. Pugachev, J. G. Li, A. L. Boyer, S. L. Hancock, Q. T. Le, S. S. Donaldson, and L. Xing. Role of beam orientation optimization in intensity-modulated radiation therapy. *International Journal of Radiation Oncology* Biology* Physics*, 50(2), 2001. 36, 38, 41, 42, 50, 88

A. B. Pugachev, A. L. Boyer, and L. Xing. Beam orientation optimization in intensity-modulated radiation treatment planning. *Medical Physics*, 27(6), 2000. 36

C. E. Rasmussen. The infinite gaussian mixture model. *Advances in Neural Information Processing Systems 12*, 12, 2000. 56, 57, 59, 60, 62, 63

A. T. Redpath, B. L. Vickery, and D. H. Wright. A new technique for radiotherapy planning using quadratic programming. *Physics in Medicine and Biology*, 21(5), 1976. 35

L. Richard. *Burden, J. Douglas Faires, Numerical analysis*. PWS Publishing Co., Boston, MA, 1988. 25

C. G. Rowbottom, S. Webb, and M. Oldham. Improvements in prostate radiotherapy from the customization of beam directions. *Medical Physics*, 25(7 Pt 1), 1998. 39, 41

C. G. Rowbottom, S. Webb, and M. Oldham. Beam-orientation customization using an artificial neural network. *Physics in Medicine and Biology*, 44(9), 1999. 38, 41

C. G. Rowbottom, V. S. Khoo, and S. Webb. Simultaneous optimization of beam orientations and beam weights in conformal radiotherapy. *Medical Physics*, 28(8), 2001a. 36

C. G. Rowbottom, C. M. Nutting, and S. Webb. Beam-orientation optimization of intensity-modulated radiotherapy: clinical application to parotid gland tumours. *Radiotherapy and Oncology*, 59(2), 2001b. 37

R. Rubinstein and D. Kroese. *The cross-entropy method: a unified approach to combinatorial optimization, Monte-Carlo simulation, and machine learning*. Springer-Verlag New York Inc, 2004. 90

E. Schreibmann, M. Lahanas, L. Xing, and D. Baltas. Multiobjective evolutionary optimization of the number of beams, their orientations and weights for intensity-modulated radiation therapy. *Physics in Medicine and Biology*, 49(5), 2004. 37

D. Shepard, M. Earl, X. Li, S. Naqvi, and C. Yu. Direct aperture optimization: a turnkey solution for step-and-shoot IMRT. *Medical Physics*, 29, 2002. 23

S. Soderstrom and A. Brahme. Selection of suitable beam orientations in radiation therapy using entropy and Fourier transform measures. *Physics in Medicine and Biology*, 37, 1992. 26, 39

S. Speer, J. Karg, M. Schmidt, and R. G. Müller. Beam angle optimization in particle therapy with optic. In *World Congress on Medical Physics and Biomedical Engineering, September 7 - 12, 2009, Munich, Germany*, volume 25/I of *IFMBE Proceedings*. Springer, 2009. 65

A. M. Steadham, H. H. Liu, C. H. Crane, N. A. Janjan, and I. I. Rosen. Optimization of beam orientations and weights for coplanar conformal beams in treating pancreatic cancer. *Medical Dosimetry: Official Journal of the American Association of Medical Dosimetrists*, 24(4), 1999. 37

J. Stein, R. Mohan, X. H. Wang, T. Bortfeld, Q. Wu, K. Preiser, C. C. Ling, and W. Schlegel. Number and orientations of beams in intensity-modulated radiation treatments. *Medical Physics*, 24(2), 1997. 36, 41

A. S. A. Sultan. *Optimization of beam orientations in intensity modulated radiation therapy planning*. PhD thesis, Deutsche Nationalbibliothek, 2006. 29

C. Thieke, S. Nill, U. Oelfke, and T. Bortfeld. Acceleration of intensity-modulated radiotherapy dose calculation by importance sampling of the calculation matrices. *Medical Physics*, 29, 2002. 92

S. Ulrich, S. Nill, and U. Oelfke. Development of an optimization concept for arc-modulated cone beam therapy. *Physics in Medicine and Biology*, 52, 2007. 32

J. Unkelbach and U. Oelfke. Inclusion of organ movements in IMRT treatment planning via inverse planning based on probability distributions. *Physics in Medicine and Biology*, 49, 2004. 32, 67

R. Vaitheeswaran, V. K. S. Narayanan, J. R. Bhangle, A. Nirhali, N. Kumar, S. Basu, and V. Maiya. An algorithm for fast beam angle selection in intensity modulated radiotherapy. *Medical Physics*, 37(12), 2010. 26, 39

C. Wang, J. Dai, and Y. Hu. Optimization of beam orientations and beam weights for conformal radiotherapy using mixed integer programming. *Physics in Medicine and Biology*, 48(24), 2003. 37, 42

X. Wang, X. Zhang, L. Dong, H. Liu, Q. Wu, and R. Mohan. Development of methods for beam angle optimization for IMRT using an accelerated exhaustive search strategy. *International Journal of Radiation Oncology* Biology* Physics*, 60(4), 2004. 36, 40

X. Wang, X. Zhang, L. Dong, H. Liu, M. Gillin, A. Ahamad, K. Ang, and R. Mohan. Effectiveness of noncoplanar IMRT planning using a parallelized multiresolution beam angle optimization method for paranasal sinus carcinoma. *International Journal of Radiation Oncology* Biology* Physics*, 63(2), 2005. 36, 40

S. Webb, D. Convery, and P. Evans. Inverse planning with constraints to generate smoothed intensity-modulated beams. *Physics in Medicine and Biology*, 43, 1998. 23

Wikipedia. Facility location — wikipedia, the free encyclopedia, 2010. URL http://en.wikipedia.org/w/index.php?title=Facility_location&oldid=386576379. [Online; accessed 25-March-2011]. 29

E. Woudstra and B. J. M. Heijmen. Automated beam angle and weight selection in radiotherapy treatment planning applied to pancreas tumors. *International Journal of Radiation Oncology* Biology* Physics*, 56 (3), 2003. 38

E. Woudstra and P. R. M. Storchi. Constrained treatment planning using sequential beam selection. *Physics in Medicine and Biology*, 45(8), 2000. 38, 87

E. Woudstra, B. J. Heijmen, and P. R. Storchi. Automated selection of beam orientations and segmented intensity-modulated radiotherapy (IMRT) for treatment of oesophagus tumors. *Radiotherapy and Oncology*, 77(3), 2005. 38, 41

R. Yang, J. Dai, Y. Yang, and Y. Hu. Beam orientation optimization for intensity-modulated radiation therapy using mixed integer programming. *Physics in Medicine and Biology*, 51(15), 2006. 37

Y. Yang, P. Zhang, L. Happersett, J. Xiong, J. Yang, M. Chan, K. Beal, G. Mageras, and M. Hunt. Choreographing Couch and Collimator in Volumetric Modulated Arc Therapy. *International Journal of Radiation Oncology* Biology* Physics*, 2011. 32, 63

S. Zhong. Efficient online spherical k-means clustering. In *Proceedings of IEEE International Joint Conference on Neural Networks.*, volume 5, 2005. 49

P. Ziegenhein, J. Wilkens, S. Nill, T. Ludwig, and U. Oelfke. Speed optimized influence matrix processing in inverse treatment planning tools. *Physics in medicine and biology*, 53, 2008. 84

Acknowledgments

I would like to acknowledge the contribution and support of

- first and foremost my supervisor Prof. Uwe Oelfke,
- my second referee Prof. Wolfgang Schlegel,
- the members of my thesis advisory committee Prof. Jan Wilkens and Prof. Rolf Bendl,
- the jack of all trades Simeon Nill and the entire E0401 group ”Physical Models“ at DKFZ,
- especially the lords of computer programming Peter Ziegenhein, Martin Siggel, and Corijn Kamerling,
- the machine learning expert Philipp der Alte Mann Hennig,
- my travel companion Florian Schimanski Büttner,
- the fantastic crew of office D150a: Thomas König, Malte Frese and Martin Fast, and
- Eva.

Vielen Dank.