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

Diplom-Physicist:	Tobias Friedrich Spalke
Born in:	Frankfurt am Main

Oral examination: 17th June, 2009

Application of multiobjective optimization concepts in inverse radiotherapy planning

Referees: Prof. Dr. Uwe Oelfke

Prof. Dr. Wolfgang Schlegel

Sei geduldig mit allen Fragen in deinem Herzen,
und versuche, die Fragen an sich zu schätzen...

Rainer Maria Rilke

Zusammenfassung

Anwendung multikriterieller Optimierungskonzepte in der inversen Strahlentherapieplanung

Die Intensitätsmodulierte Strahlentherapie (IMRT) gehört zu den modernsten Methoden der Krebsbehandlung. Aufgrund der hohen Zahl der Freiheitsgrade wird das inverse Problem der IMRT als skalares Optimierungsproblem formuliert. Der konventionelle Ansatz einer Skalarisierung der Zielfunktion entspricht einem a priori Kompromiss zwischen den verschiedenen Planungszielen und kann zu einem zeitaufwändigen Planungsprozess führen. Neue multikriterielle Optimierungsansätze versuchen, die Nachteile der konventionellen Planung zu umgehen. Dazu wird eine Datenbank Pareto optimaler Bestrahlungspläne berechnet, aus der der Planer im Anschluss den besten Plan auswählen kann. In dieser Arbeit wird ein neues multikriterielles Planungssystem evaluiert. Wir untersuchen drei verschiedene auf der generalized equivalent uniform dose (gEUD) basierende Modellierungsansätze und die Sensitivitäten der dazugehörigen Modellparameter. Für die Charakterisierung von ganzen Pareto optimalen Datenbanken werden Qualitätsinidikatoren entwickelt und auf klinische Fälle angewandt. In restropektiven Planungsstudien wird gezeigt, dass das neue Planungssystem die gleiche Planqualität wie das klinische Referenzsystem erreicht und eine erhebliche Verringerung der Gesamtplanungszeit für Prostatafälle erzielt werden kann. Im letzten Teil entwickeln wir eine Methode zum Auffinden der zwingenden Kompromisse in einer Plandatenbank und verwenden Techniken der linearen und nicht-linearen Dimensionsreduktion. Diese erlauben die aussagekräftige Visualisierung von hochdimensionalen Paretofronten und eine Offenlegung der zugrundeliegenden Kompromisse zwischen den beteiligten Planungszielen.

Abstract

Application of multiobjective optimization concepts in inverse radiotherapy planning

Intensity Modulated Radiotherapy (IMRT) belongs to the most advanced techniques in cancer treatment. Due to the large number of degrees of freedom the inverse problem of IMRT is formulated as a scalar optimization problem. The conventional scalarization approach of the objective function represents an a priori trade-off between the planning goals and can lead to time consuming planning process. New multiobjective approaches try to overcome the difficulties by a precalculation of a set of Pareto optimal treatment plans from which the planner can subsequently choose the best suited plan. In this work we evaluate a new multiobjective treatment planning system. We investigate three different generalized euqivalent uniform dose (gEUD) based modeling approaches and study the sensitivities of the corresponding model parameters. Quality measures for entire Pareto optimal databases are developed and applied to clincal cases. In retrospective planning studies we show that the new system is compatible with a clinical reference treatment planning system and that the total planning time for prostate cases can be significantly reduced. In the last part we develop a method to detect the imperative trade-offs in a plan database and apply techniques from linear as well as non-linear dimensionality reduction. They allow meaningful visualizations of high dimensional Pareto fronts and an insight into the underlying trade-offs between the planning goals.

Contents

1 Introduction **1**

2 Fundamentals **3**
2.1 Mathematical Basics . . . 3
2.2 Pareto Optimality . . . 5
2.3 Intensity modulated Radiotherapy . . . 6
2.3.1 Basic Concept . . . 6
2.3.2 Optimization in Radiotherapy - A General Framework . . . 8
2.3.3 Notation and Terminology . . . 9
2.4 Multiobjective Planning Systems . . . 10
2.4.1 MIRA . . . 11
2.4.2 PGEN . . . 12

3 Choosing Models and Parameters **13**
3.1 Modeling Approaches . . . 13
3.1.1 Motivation . . . 13
3.1.2 Methods . . . 13
3.1.3 Results . . . 16
3.1.4 Discussion . . . 23
3.1.5 Conclusions . . . 24
3.2 Quality Indicators for Efficient Sets . . . 25
3.2.1 Motivation . . . 25
3.2.2 Methods . . . 25
Cardinality and Hypervolume Indicator . . . 26
Uniformity and Hyperspheres . . . 28
Coverage error . . . 31
3.2.3 Results . . . 32
3.2.4 Discussion . . . 36
3.2.5 Conclusion . . . 37

4 Clinical Application **39**
4.1 Automatic Localization of a Reference Plan . . . 39
4.1.1 Motivation . . . 39
4.1.2 Material and Methods . . . 39
4.1.3 Results . . . 40
4.1.4 Discussion and Conclusion . . . 41
4.2 Clinical Evaluation . . . 43
4.2.1 Motivation . . . 43

4.2.2 Methods . . . 43
4.2.3 Results . . . 43
4.2.4 Discussion . . . 48
4.2.5 Conclusions . . . 49

5 Understanding Pareto Optimal Planning Databases 51
5.1 Sensitivities of the Efficient Set . . . 51
5.1.1 Motivation . . . 51
5.1.2 Methods . . . 53
5.1.3 Results . . . 54
5.1.4 Discussion . . . 57
5.1.5 Conclusion . . . 57
5.2 Identifying the Main Trade-Offs in Pareto Optimal Databases . . . 58
5.2.1 Motivation . . . 58
5.2.2 Materials and Methods . . . 58
Terminology . . . 58
Principal Components Analysis . . . 59
Isomap Method . . . 59
Phantoms and Clinical Cases . . . 61
5.2.3 Results . . . 61
Phantoms . . . 61
Brain Case . . . 65
Pancreas Case . . . 65
Effective Dimension . . . 72
5.2.4 Discussion . . . 73
5.2.5 Conclusion . . . 75

6 Summary and Outlook 77

7 Acknowledgment 79

Bibliography 81

1 Introduction

Radiotherapy is besides surgery the most important treatment option in oncology. The central goal is to achieve a high tumor control probability while keeping the normal tissue complication probability as low as possible. In order to achieve this goal one aims to deliver a homogeneous high dose to the tumor while ensuring an adequate sparing of the organs at risk (OARs) and the surrounding healthy tissue.
The development of Intensity Modulated Radiotherapy (IMRT) allows to create highly conformal dose distributions (Bortfeld, 2006), i.e. restricting the high dose areas to the target volume, using multileaf collimators and high energetic photons. The task of deriving a set of treatment parameters from a desired dose distribution is called the inverse problem of IMRT. Because of the large number of degrees of freedom from the discretization of the patient geometry into voxels and the beams into beamlets the problem is formulated as a constrained optimization problem. If the irradiation geometry, i.e. number of beams and the beam directions, is fixed the problem yields a fluence map optimization (Oelfke and Bortfeld, 2001). The quality of a treatment plan is usually modeled by the value of a single scalar objective function which has to be minimized. Consequently an entire treatment plan is characterized by a single number.

The inverse planning of IMRT involves inherent trade-offs. The conflicting goals of a high homogeneous target dose and the protection of healthy tissue have to be balanced against each other. Traditionally all the different planning goals or *objectives* are aggregated into the objective function, which is parameterized by weight factors that allow to influence the outcome of the optimization and place emphasis on the individual goals. The selection of the weight factors represents an *a priori* trade-off. As the sensitivity of the optimal solution (therewith the dose distribution) with respect to the weight factors is not known the planning process can become a cumbersome task, involving several human interactions during the optimization to adjust the weights. In addition the weight factors have no clinical meaning.

To overcome these difficulties multiobjective planning concepts for IMRT planning have been proposed in the last years (see e.g. Yu (1997); Cotrutz et al. (2001); Küfer et al. (2000); Thieke et al. (2002); Craft et al. (2006)). Within a multiobjective framework each planning

goal is considered as an independent entity. The optimality criterion is generalized to Pareto optimality. Pareto optimal plans cannot be improved in one objective without worsening another. As a result a whole set of plans is optimal in a multiobjective sense, instead of only one optimal plan for conventional planning. The set of all Pareto optimal plans constitutes the Pareto front or the efficient set. During the multiobjective planning process a finite number of Pareto optimal plans is calculated. These plans are stored in a plan database and should capture the essence of the clinically relevant part of the Pareto front. Subsequently the planner can choose the best suited plan from the database with the help of decision support systems such as interactive navigation tools (Monz et al., 2008).

At present multiobjective planning systems are not yet introduced into clinical practice but are still in a pre clinical stage. A general question is what are the realistic expectations of the new class of planning systems? Furthermore new challenges arise by the increased amount of information due to the Pareto optimal databases, because we are confronted with a whole set of plans instead of only one optimal plan. So we are in a situation unknown to planners in IMRT planning so far.

In this thesis we evaluate the potentials of a new multiobjective planning system MIRA. In addition we develop methods to characterize the Pareto optimal databases and to efficiently gain information that increases the understanding of a case and the planning process in general.

This thesis is organized as follows. In chapter 2 we give a compact introduction to the basic underlying mathematics and the inverse planning in IMRT. A basic knowledge in IMRT treatment planning is assumed and the presentation is kept rather general. For more detailed and practical sources the reader is referred to Webb (2001); Palta and Mackie (2003); J. Bille (2002). The three remaining chapters are devoted to multiobjective modelling approaches and quality measures for efficient sets (chapter 3), the clinical evaluation of the system (chapter 4) and the extraction of information from Pareto optimal plan databases (chapter 5). The thesis concludes with a summary and an outlook.

2 Fundamentals

In this chapter we will first introduce the mathematical basic definitions which are employed in the following. Then we will develop the concept of Pareto optimality. Afterwards we give a short overview over standard and multiobjective radiotherapy planning. The chapter concludes with the presentation of two prototypic multiobjective planning systems.

2.1 Mathematical Basics

In this section we will introduce the fundamental mathematical concepts and the basic notation which will be used throughout the subsequent chapters. In the beginning definitions and notations are given and a general description of the multiobjective optimization framework is presented.

One of the fundamental properties which is exploited in mathematical optimization and in this thesis is convexity. The idea of convexity can apply to sets or functions. The according definitions are as follows:

Definition 1 (Convexity of a set)**.** Let $C \subset \mathbf{R}^n$ be a subset of the Euclidian space. C is convex, if for any $x_1, x_2 \in C$ we have $y = \lambda x_1 + (1-\lambda)x_2 \in C$, with $0 \leq \lambda \leq 1$

and

Definition 2 (Convexity of a function)**.** A function $f : \mathbf{dom}\, f \subset \mathbf{R}^n \to \mathbf{R}$ is convex, if it's domain $\mathbf{dom}\, f$ is a convex set and for any $y, x \in \mathbf{dom}\, f$ and any λ with $0 \leq \lambda \leq 1$ we have

$$f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda)f(y)\,. \tag{2.1}$$

We will also need the term of the convex hull $\mathrm{conv}(\cdot)$ of a subset $S \subset \mathbf{R}^n$, which describes the "smallest" convex set containing S.

Definition 3 (Convex hull)**.** Let $S \subset \mathbf{R}^n$ a subset of the Euclidian space. The convex hull of S is given as $\mathrm{conv}(S) = \{\sum_{i=1}^{N} \alpha_i \cdot s_i \,|\, s_i \in S, N \in \mathbb{N}, \sum_{i=1}^{N} \alpha_i = 1, \alpha_i \geq 0\}$.

So the convex hull consists of all finite convex combinations of elements from the subset S.

Another important concept that will be needed in the context of multiobjective optimization is the one of a cone. A set K is called a cone, if for every $x \in K$ and $\lambda \geq 0$ we have $\lambda x \in K$. If K is also convex it is a convex cone, which means that with $x_1, x_2 \in K$ and $\lambda_1, \lambda_2 \geq 0$ for the conic combination we have

$$\lambda_1 x_1 + \lambda_2 x_2 \in K. \tag{2.2}$$

Each proper cone induces a generalized inequality $x_1 \preccurlyeq_K x_2 \Leftrightarrow x_1 - x_2 \in K$ (Boyd and Vandenberghe, 2004). So x_1 is smaller or equal to x_2 with respect to "$\preccurlyeq_K$" if the difference lies in K. Proper cones will form the basis for the material presented in the following section 2.2.

We are often faced with the situation that we want to choose the "best" option out of a set of alternatives or make the "best" decision for a set of policies. In order to put our choice on a quantitative basis we usually have to model our preferences as a mathematical model which has to be optimized. A scalar optimization problem has the general form

$$\begin{array}{ll} \text{minimize} & f_s(x) \\ \text{subject to} & g_i(x) \leq 0, \quad i = 1, \ldots, m \end{array} \tag{2.3}$$

where $f_s : \mathbf{dom}\, f_s \subset \mathbf{R}^n \rightarrow \mathbf{R}$ is a scalar objective function which is to be minimized, $x \in \mathbf{R}^n$ is the vector of optimization variables we have to choose and $g_i : \mathbf{R}^n \rightarrow \mathbf{R}$ are the corresponding constraint functions or just "constraints". The constraint functions can be further subdivided into equality and inequality constraints. According to the properties of the objective and constraint functions an optimization problem can be classified. Well-known examples would be linear programming (where the objective function and constraints are linear functions of x), quadratic programming or general non-linear programming. For each class of optimization problems there exists a variety of (mostly iterative) algorithms. The classes can considerably differ in their computational complexity. In order to be tractable many real world applications require a convex formulation.

One important theorem of convex programming states that if f_s and g_i are convex functions and if the feasible set $\mathcal{X} = \{x \,|\, g_i(x) \leq 0\}$ is convex then all local optima are also global optima (Bertsekas, 2004).

Considering terminology, each assignment of values to x is called a solution. If it satisfies the constraints it is called a feasible solution and if the objective function has a local (global) minimum at x^*, it is referred to as a local (global) optimal feasible solution. For the sake of simplicity we will however just use the term solution instead of optimal feasible solution.

Often one is faced with more than one single objective to optimize. In this case a common method is to include these objectives in one objective function by scalarization. In many applications with more than one objective a simple scalar model of the problem is not appropriate (Zeleny, 1982). As there is no canonical way to balance the different objectives, we are rather interested in considering all objectives independently. This leads to a substantially new class of *multiobjective* optimization problems. The formulation the multiobjective program (MOP) where all k objectives have to be minimized *simultaneously* (Miettinen, 1999) reads

$$\begin{array}{ll} \text{minimize} & f(x) = (F_1(x), \ldots, F_k(x))^T \\ \text{subject to} & g_i(x) \leq 0, \qquad\qquad i = 1, \ldots, m \end{array} \tag{2.4}$$

$f : \mathbf{dom}\, f \subset \mathbf{R}^n \to \mathbf{R}^k$ is now a vector valued function and each component F_i corresponds to one objective. How a MOP can be solved and what the implications for the term optimality are will be treated in the next section.

2.2 Pareto Optimality

In a standard scalar optimization problem each set of optimization variables is evaluated by a scalar objective function $f_s : \mathbf{R^n} \to \mathbf{R}$. Thus each solution can be characterized by a single number which reflects the quality of the respective solution. This implies that all solutions can be compared by means of the standard order relation "$\leq$". In a multiobjective framework the corresponding objective function is a vector valued function $f : \mathbf{R}^n \to \mathbf{R}^k$, hence all solutions are vectors in $\mathbf{R}^k$. The absence of a total order in $\mathbf{R}^k$ makes it necessary to extend the optimality criterion, because two solutions are not necessarily comparable anymore. This leads to the definition of Pareto optimality. The concept of Pareto optimality was developed by Vilfredo Pareto (Pareto, 1896) and first applied to problems in an economical context. Formally we can define the optimality criterion as follows:

Definition 4 (Pareto optimality)**.** Let $y \in \mathcal{Y}$ be a point with $\mathcal{Y} \subset \mathbf{R}^k$. y is Pareto optimal or efficient if and only if there is no other point $y_0 \in \mathcal{Y}$ with $y_0 \preccurlyeq_K y$, where $\preccurlyeq_K$ is the partial order relation induced by the proper cone K.

We will also say that a point is not dominated by another one, if it is efficient. The choice of the cone K will in general affect the set of solutions which are regarded as efficient (Monz, 2006). In this work K will always be the convex cone of the positive orthant $K = \mathbf{R}^k_+$, so we skip the explicit notation of the cone and just write $\preccurlyeq$ instead of $\preccurlyeq_{\mathbf{R}^k_+}$. This leads to a componentwise comparison between two solutions. That means if for all components i of

two solutions y^1 and y^2 we have $y_i^1 < y_i^2$ then y^1 dominates y^2. We can describe the efficient solutions on the Pareto front also verbally (and a little less precisely):

A solution is Pareto optimal or efficient, if it cannot be improved in one objective without worsening at least one other objective.

The set of all Pareto optimal points is called the Pareto front or the efficient set. Note that two solutions are not necessarily comparable anymore. In particular all efficient solutions are not comparable by the order relation "$\preccurlyeq$". The whole concept is illustrated in figure 2.1.

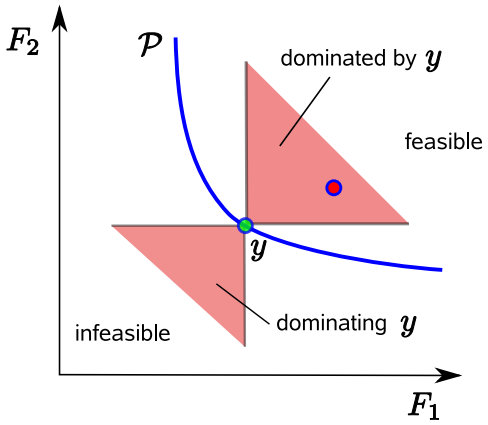

Figure 2.1: Efficient solution (green) on the Pareto front (blue) for a bicriterial problem. The solution in red is dominated by the green one, because it lies in the red cone.

With the so defined extended optimality criterion the goal of an MOP becomes to find the efficient set of the problem in equation 2.4. However, firstly the number of efficient solutions is in general infinite, secondly for most application we are not interested in the whole Pareto front, but rather in a part which includes only relevant trade-offs with respect to the application. Therefore in a practical context we want to find a finite representation of the Pareto front, which captures all relevant trade-offs.

2.3 Intensity modulated Radiotherapy

2.3.1 Basic Concept

The basic concept of IMRT can be illustrated by looking at the well known example by Brahme (Brahme et al., 1982). A two dimensional circular OAR is surrounded by a ring-shaped tumor volume (see figure 2.2). If we perform a rotation therapy by rotating a uniform

radiation source around the central axis we can perfectly spare the OAR by a central block. Due to the geometrical configuration the dose distribution in the target will not be homogeneous.

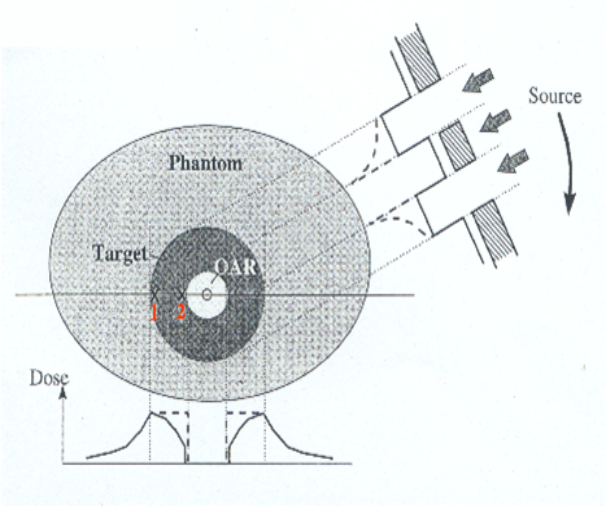

Figure 2.2: The example by Brahme with a rotating source and a centrally blocked beam. Only a mudulated beam profile (dashed lines) leads to a homogeneous dose distribution in the target. Figure from (Brahme et al., 1982).

The key insight is that a homogeneous distribution in the target can be achieved by *inhomogeneous* beam instensity profile. For simple geometries there exist analytical solutions for the intensity profiles, for arbitrarily shaped targets the profiles usually have to be found numerically.
The extension to three dimensions leads to two dimensional intensity maps. These maps are discretized into the so called beamlets. Each beamlet represents a small subunit of the beam and can be mudulated individually. Given a set of beams and the corresponding incident directions one searches for the spatially non-uniform intensity maps of all beams which (when superimposed) lead to the desired dose distribution in the target and an adequate sparing of the OARs.
Due to the physical reasons of a positive energy deposition and scattering effects IMRT can obviously not achieve an 'ideal' dose distribution with a perfect OAR sparing. So we always have to accept a certain irradiation of the critical structures which has to be balanced against the goal of a homogeneous target dose.
Note however, that if there would exist a radiation type with a negative dose contribution we would be able to achieve any desired dose distribution. This fact was shown by Birkhoff (Birkhoff, 1940), who proved that any black-and-white drawing can be obtained by a superposition of black (pencil) and white (ruler) lines of different thickness.

2.3.2 Optimization in Radiotherapy - A General Framework

The optimization process in radiotherapy can be described in a general framework. From an abstract point of view we can think of the inverse planning as a composition of two mappings, relating three vector spaces (Kuefer et al., 2005). The first space is the so called parameter space $\mathbf{X}$, containing the optimization variables. All feasible points $\mathcal{X}$ in the parameter space are then mapped into the design space, which represents the "system design" $\mathcal{D}$ resulting from the chosen parameters. This design is then evaluated by a vector valued function and results in set of feasible objective vectors $\mathcal{F}$. Each component of one objective vector measures the quality of one objective.

If we apply this general framework to IMRT we can identify the parameters with the beamlets (see figure 2.3). This is a simplification because in general we would have to consider all relevant treatment parameters including beam angles, beam energies, etc. leaving us with a non convex problem (Bortfeld and Schlegel, 1993). However we will restrict ourselves to this simplification and assume that we have a convex MOP (so the feasible set $\mathcal{X}$, all objective Functions F_i and constraints g_i are convex). So each point in the parameters space can be identified with a set of beamlets defining one treatment plan. This point is mapped into the design space which corresponds to a dose calculation for the specific beamlets. The calculated dosedistribution is finally evaluated by the respective objective functions and thus mapped into the objective space. Each single objective function F_i is assigned to one structure of interest, typically one function per OAR and one or two functions per target structure.

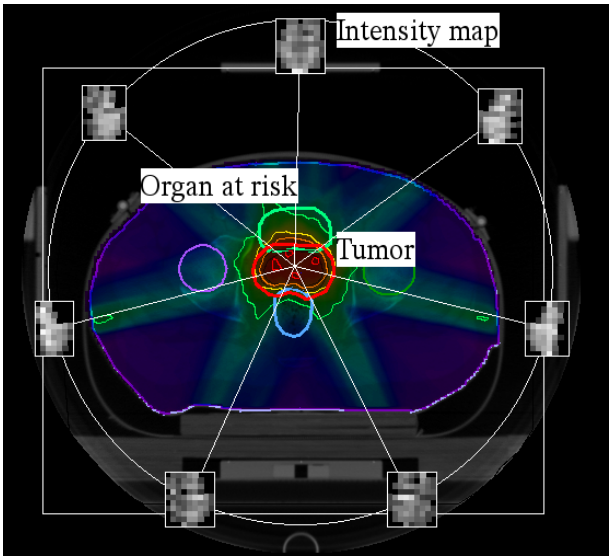

Figure 2.3: The tumor (red) is irradiated with high energetic photons from seven directions. The irradiation fields with spatially non-uniform intensity maps are superimposed to achieve the prescribed dose in the tumor and a sufficient sparing of crticial structures.

Our goal is to find all parameter vectors in the parameter space that correspond to efficient

solutions in the objective space. For practical reasons we want to find at least a meaningful representation of the efficient set from which the planner can then select the best suited treatment plan for the patient.

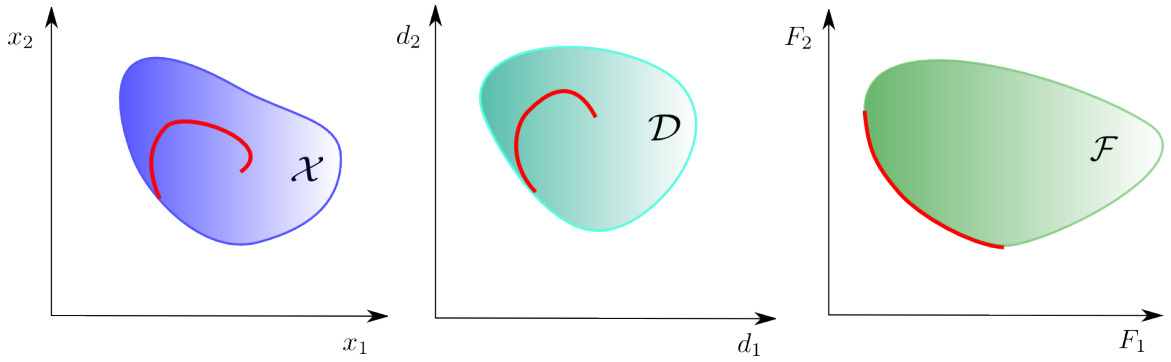

Figure 2.4: All feasible solutions in $\mathcal{X}$ from the parameter space (left) are mapped to the design space (middle). The set of efficient solutions is symbolized in red. The solutions in the design space are evaluated by a vector valued function leading to a representation in the objective space. Here the efficient solutions are located at the boundary of the set $\mathcal{F}$ of feasible objective vectors.

We are only interested in Pareto optimal solutions, otherwise we could improve a solution without worsening any other objective. The set of all Pareto optimal solutions constitutes the Pareto frontier $\mathcal{P}$ or efficient set. The efficient set is a subset of the boundary of all feasible solutions in the objective space. For a convex MOP it is connected (Miettinen, 1999) and the set $\mathcal{F}$ is convex (Romeijn et al., 2004). Each efficient solution can be identified with a k-dimensional objective vector in the objective space and will be denoted as a solution or synonymously as a plan.

It is worth noticing that Pareto optimal solutions can be very different and the optimality of a solution alone does not imply that the corresponding treatment plan is of clinical interest. Giving no dose at all, thereby sparing all OARs and the normal tissue perfectly but having no dose in the target volume, might be a Pareto-optimal "treatment plan". Therefore one has to make sure, that the solutions in the objective space are located within a predefined *planning horizon*, covering the range of clinical relevance.

2.3.3 Notation and Terminology

For the subsequent sections we will introduce the basic notation. We assume that the patient geometry and the radiation field are discretized into N voxels and n beamlets. We will consider a fluence map optimization, where the irradiation geometry is given, thus leaving us with a convex optimization problem. The vector $x \in \mathbf{R}^n_+$ denotes all beamlets, and x_i is the intensity (weight) of beamlet i. The dose distribution $d \in \mathbf{R}^N_+$ can then be calculated

with the influence matrix $D \in \mathbf{R}_+^{N \times n}$ via the linear relationship $d(x) = D \cdot x$, where D_{ij} describes the dose from beamlet j deposited in voxel i under unit fluence (Gustafsson et al., 1994).

The dose distribution $d(x)$ is evaluated by a vector valued function $f(d) \subset \mathbf{R}^k$ where k is the number of independent criteria or objectives, typically one per organ at risk and at least one per target structure. So each set of beamlets x representing the optimization variables is mapped from the n-dimensional parameter space $\mathbf{X}$ into the k-dimensional objective space $\mathbf{Y}$.

The fluence map optimization is formulated as a MOP, meaning that all individual criteria have to be minimized simultaneously subject to the physical constraints $x_i \geq 0$ enforcing non-negative fluence amplitudes. So each individual objective function F_i represents the quality of the corresponding structure for a specific solution. All F_i as well as the feasible set $\mathcal{X}$ with $x \in \mathcal{X}$ are assumed to be convex. The Pareto front will then be a connected hypersurface in the objective space.

2.4 Multiobjective Planning Systems

So far there is no clinical usable multiobjective treatment planning system available. Standard treatment planning systems mostly rely on a single scalar objective function, resulting in one optimal solution. If the solution doesn't match the planner's goals, he has to adapt the objective function (that means either the weight factors representing the relative importance of a structure or the model itself) and start another optimization. Thus, the planning of complex cases can become a time consuming process. A multiobjective planning system aims to elegantly overcome these difficulties by providing the planner with a relevant part of the efficient set. The efficient set can then be interactively explored in real time, helping to select the best treatment plan for the patient. So the multiobjective treatment planning system enables the planner to make an informed choice, and can be viewed as a decision support system.

In the following section we will introduce the multiobjective planning system MIRA (Multiobjective Interactive Radiotherapy Assistant) developed by the ITWM-Kaiserslautern (Küfer et al., 2006) which is the main system used for the computations. The computations in chapter 5.2 were done with another system called PGEN (Pareto Generator) (Craft et al., 2006) which is briefly described afterwards.

2.4.1 MIRA

The Multiobjective Interactive Radiotherapy Assistant MIRA is a pre-clinical multiobjective treatment planning system consisting of two main parts: First, a nonlinear numeric solver, which calculates a discrete representation of the Pareto front and stores the solutions in a plan database. To meet the high computational effort an adaptive clustering strategy is used, where groups of voxels are combined to clusters which are adaptively refined (Scherrer et al., 2005). The second part is an interactive navigation tool (Monz et al., 2008) that allows to explore the calculated plan database through an intuitive user interface. The user can manipulate the individual objectives by sliders. An efficient interpolation mechanism between the single solutions allows a real time update of the shown dose distribution and dose volume histogram information (see figure 2.5). The input data for the MIRA system is calculated within the clinical software environment of the DKFZ: all influence matrices are calculated with the research version of the KonRad planning system (Nill, 2001) and the patient contour data is generated within VIRTUOS (Höss et al., 1995).

The MIRA solver pursues a three phase strategy to generate a representative set of Pareto optimal solutions:

- In phase one a starting solution is calculated. The user will specify a set of treatment goals for each planning structure. The solver then aims to satisfy as many goals as possible to get a balanced starting solution. No fine-tuning is needed for the starting solution because it's only purpose is to define a clinically relevant domain within the objective space. After the initial solution is calculated the planner defines tolerance limits for each planning structure, thereby excluding possibly efficient but meaningless solutions. This confines the area of interest on the Pareto front and establishes the *planning horizon* of the problem.
- In the second phase the planning horizon is filled with additional solutions. The aim is to set up the corner posts of the area of interest, i.e. to calculate solutions enclosing (and filling) the planning horizon. This is achieved by optimizing all subsets of planning goals, while keeping the remaining objectives under the tolerance levels. As a result we get $2^k - 1$ *extreme compromises* that span the planning horizon.
- During the last phase additional *intermediate solutions* are calculated to improve the approximation. For a systematic placement of intermedideate solutions the number of intermediates would be too large. Therefore a fixed number of intermediates in calculated and stochastically distributed within the planning horizon.

The efficient solutions are stored in a plan database. The database can be explored afterwards by a *Pareto-Navigation* (Monz et al., 2008) to interactively select the best suited treatment

plan (see figure 2.5).

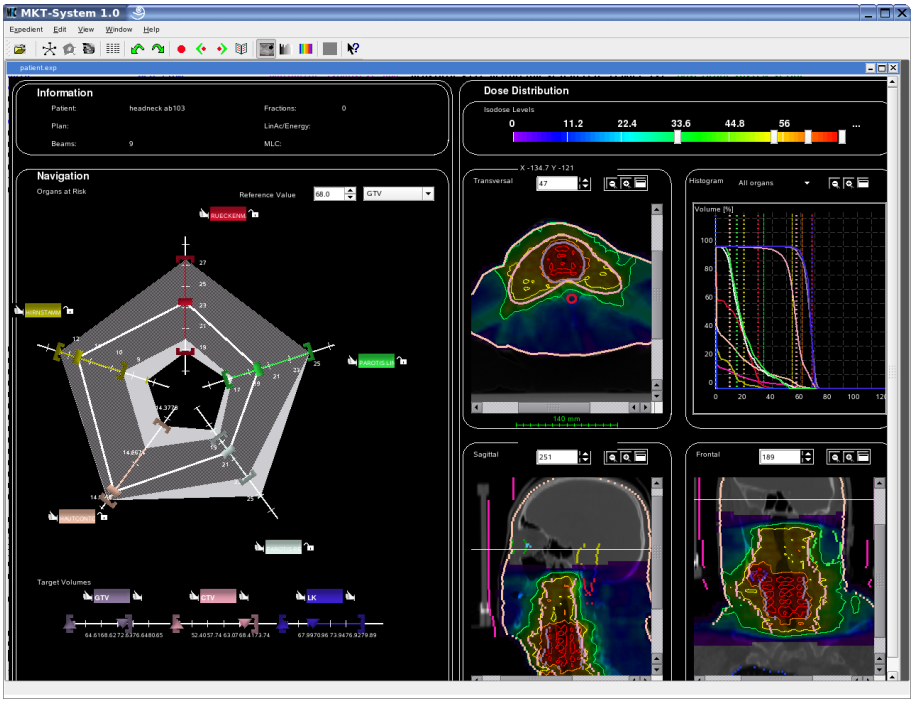

Figure 2.5: Graphical user interface of the navigator. Each objective is represented by a slider (left) and can be manipulated interactively. The corresponding dose distribution and DVH information is updated in real time.

2.4.2 PGEN

The Pareto Generator PGEN also calculates a discrete representation of the efficient set. It persues a strategy different from the MIRA system though. On the one hand a linear modelling approach is chosen which allows for a linear program formulation and the use of efficient numerical solvers. On the other hand the additional solutions are calculated differently. First a set of k anchor points is calculated for a case with k independent objectives, by only considering exactly one objective per optimization. Then additional intermediate solutions are calculated, whereas each solutions aims to improve the inner approximation of the Pareto front as much as possible. This is done by making use of a sandwich-algorithm based on lower and upper bounds of the Pareto front. As a price for the intelligent placement of the additional solutions convex hull computations are used, which currently restricts the system to problem dimensions of about eight.

3 Choosing Models and Parameters

First we investigate different modelling approaches and study the effect of the corresponding model parameters. Then we will introduce quality indicators for efficient sets, from which we can draw conclusions about the choice of the number of intermediate solutions.

3.1 Modeling Approaches

3.1.1 Motivation

To set up a computable optimization problem we have to build a model of the tumor and the OARs, i.e. we must translate our clinical goals into mathematical equations which can be handled by a numerical solver. In this section we will introduce the different mathematical models implemented in the MIRA system.

The main open questions are: Which models should be chosen for which cases? Do some models result in superior clinical treatment plans? What values should be chosen for the model parameters? How do the solutions change if the model parameters are changed? Or in other words, what is the sensitivity of the model parameters for the different optimization problems?

3.1.2 Methods

Two of the considered models are based on the generalized equivalent uniform dose (gEUD), which is a generalization of the EUD, introduced by Niemierko (1997). The EUD was initially proposed for the modelling of target structures and then generalized to model targets and OARs. For an inhomogeneous dose distribution the EUD describes the value of a homogeneous dose distribution that would lead to the same biological effect as the previous one. In that sense the gEUD is a biologically motivated objective, instead of pure physical ones like e.g. the maximum dose of an OAR. The gEUD is given as

$$\mathrm{gEUD}_p(d) = (\frac{1}{N_j}\sum_{i=1}^{N_j} {d_i}^p)^{1/p}, \tag{3.1}$$

where N_j is the number of voxels in the volume of interest (VOI) j. The organ parameter p can be chosen according to the type of the OAR (serial or parallel or inbetween), because the p-norm has the property that $\mathrm{gEUD}(d) \to d_{max}$ for $p \to \infty$ and $\mathrm{gEUD}(d) \to d_{mean}$ for $p \to 1$, with the mean dose d_{mean} and the maximum dose d_{max} of the distribution. When modelling target volumes with the gEUD the organ parameter is chosen to be negative $p \leq 0$.

There are three different models under consideration.

1. In the first model (pq-model) each OAR is modeled by a combination of gEUDs.

$$F_{OAR}(d) = \mu \cdot \mathrm{gEUD}_p(d) + (1-\mu)\mathrm{gEUD}_q(d) \tag{3.2}$$

   with $0 \leq \mu \leq 1$. Mathematically speaking this is just a convex combination of a p- and a q-norm. For target structures the under- and overdosage are considered separately and penalized according to the functions

$$\begin{aligned} F_{low}^T &= (\tfrac{1}{N_j}\textstyle\sum_{i=1}^{N_j} \max(d_{low}^T - d_i, 0)^{p_T})^{1/p_T} \\ F_{up}^T &= (\tfrac{1}{N_j}\textstyle\sum_{i=1}^{N_j} \max(d_i - d_{up}^T, 0)^{q_T})^{1/q_T} \end{aligned} \tag{3.3}$$

   Here d_{up} and d_{low} are the dose values from where on a violation of the under- or overdosage is penalized.

2. In the second model (pt-model) each OAR is evaluated by a tail-penalty function

$$F_{up} = (\frac{1}{N_j}\sum_{i=1}^{N_j} \max(d_i - d_{up}, 0)^{p_t})^{1/p_t} \tag{3.4}$$

   so only voxels which exceed the bound d_{up} will contribute to the function. The target is treated in the same way as in the pq-model.

3. The third model (std-model) uses for the OARs the same function class as the pq-model. For the target however the target mean dose d_{mean} is fixed and simultaneously the standard deviation of the distribution is minimized:

$$F_{std} = (\frac{1}{N_j}\sum_{i=1}^{N_j} (d_i - d_{mean})^2)^{1/2} \tag{3.5}$$

Note that the std-model can also be given a biological motivation as the EUD for targets can be approximated in a first order expansion (of the central moments) as EUD $= d_{mean} - \frac{1}{2}\alpha\sigma^2$ (U. Oelfke, 2002) with the radiobiological sensitivity parameter α and the standard deviation σ.

For each objective upper bounds on the function values can be specified which restrict the particular objective and ensure a reasonable extent of the planning horizon. In addition for each OAR j with an gEUD objective one has to set an aspired value F_j^{asp}, which represents the "ideal" function value one wishes to achieve for the corresponding structure. In the optimization the relative differences of the OAR objectives from the aspired values are minimized. Note that all functions are convex functions (given $p \geq 1$ for all OARs), which allows for a convex MOP formulation.

One difficulty in the comparison of the results from the different models is that the solutions lie in different objective spaces which have no natural connection. Furthermore we are always dealing with a whole set of plans. If we do comparisons of different treatment plans there is no obvious way to determine which plans should be chosen (remember that efficient solutions from one model are not comparable by definition). We therefore have to set up a fixed reference point. So for clinical plan comparisons we will look at plans that offer the same plan quality for the target and compare the remaining OARs.

Another way of comparing the results of the different models is to reevaluate a whole database with the other model's objective functions. That means one efficient set $\mathcal{P}_2$ is mapped from it's objective space $\mathbf{Y}_2$ to another one $\mathbf{Y}_1$. It is obvious that the solutions from $\mathcal{P}_2$ will in general not be efficient in $\mathbf{Y}_1$. However, we can use this fact to quantify the changes in the solutions induced by the change of the model. As the shape of the whole Pareto front will change we can use the Hausdorff metric to give a meaning to the distance of the efficient sets $\mathcal{P}_1$ and $\mathcal{P}_2$. As we are always dealing with compact sets (finite sets of points), the Hausdorff metric is guaranteed to be finite and is given by

$$d_H(X,Y) = \max\{ \max_{x\in X} \min_{y\in Y} d(x,y),\ \max_{y\in Y} \min_{x\in X} d(x,y) \}, \tag{3.6}$$

where X and Y denote subsets of the Euclidean space and $d(\cdot,\cdot)$ is chosen to be the ordinary Euclidean metric. A graphical illustration is given in figure 3.1.

Strictly speaking we are interested in the distance of the efficient sets, but we do not have a complete representation available, but discrete subsets $\mathcal{D}_{1,2} \subset \mathcal{P}_{1,2}$. If we use the finite sets $\mathcal{D}_1$ and $\mathcal{D}_2$ of solutions as a basis for the computation of the Hausdorff metric the result will depend on the actual placement of the points on the Pareto front. Hence we will use the first order approximation (cf. section 3.2.2) of convex combinations, which is also used for

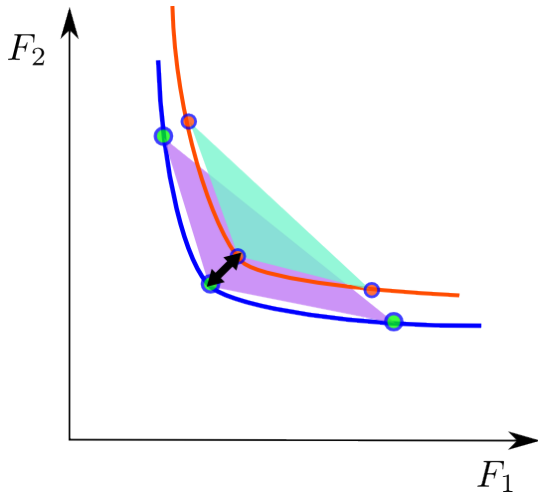

Figure 3.1: Schematic illustration of the Hausdorff distance. Solutions from another model (red) are evaluated in the objective space. The Hausdorff distance between the first order approximations is marked with an arrow.

the navigation. So we calculate $d_H(\text{conv}(\mathcal{D}_1), \mathcal{D}_2)$. The objectives will be scaled to 1 so that the distances are not dominated by objectives with the highest order of magnitude.

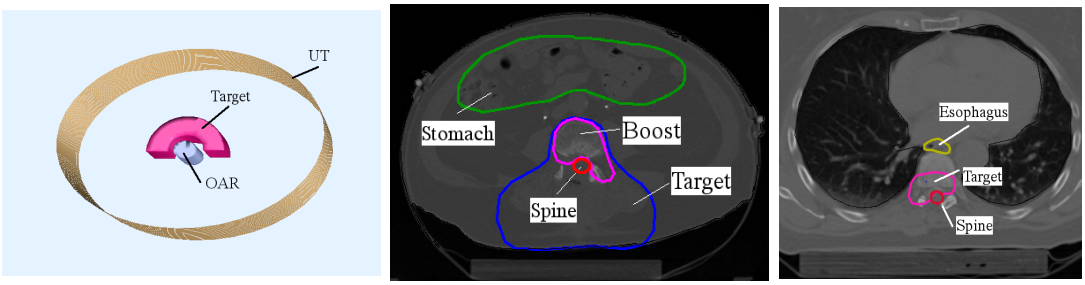

Figure 3.2: The geometries of three cases. A horseshoe phantom case (left) with one OAR and a horseshoe shaped target volume, a paraspinal tumor (paraspinal I) with one boost volume (middle) and a small paraspinal case (paraspinal II) (right).

We will consider three different cases: a simple phantom case with a horseshoe shaped target and a single OAR and two clinical paraspinal cases (see figure 3.2).

3.1.3 Results

We first consider the influence of the target parameters p_T and q_T which affect the penalization of the target under- and overdosage. Two starting solutions are calculated for the horseshoe phantom with different settings for p_T and q_T. As shown in figure 3.3 and 3.4 the influence of these parameters can differ depending on the actual position on the Pareto front. In a region with a high target quality as in figure 3.3 the change of the parameters p_T and q_T has a rather small influence on the shape of the target DVH, changing the minimum target

dose d_{min}^{T} by 3.9% and the maximum target dose by 5.1%. Note also that the maximum OAR dose increases by 7.0% (from left to right) which means that the decrease in the target quality could still be compensated for by improving the OAR through navigation. However, if we investigate a different part of the Pareto front with a good OAR sparing (figure 3.4) the influence of the parameter change on the target quality is stronger. The minimum target dose decreases by 17.4% (from left to right), while the OAR changes by less than 1% and the mean dose of the unclassified tissue increases by 1.6%

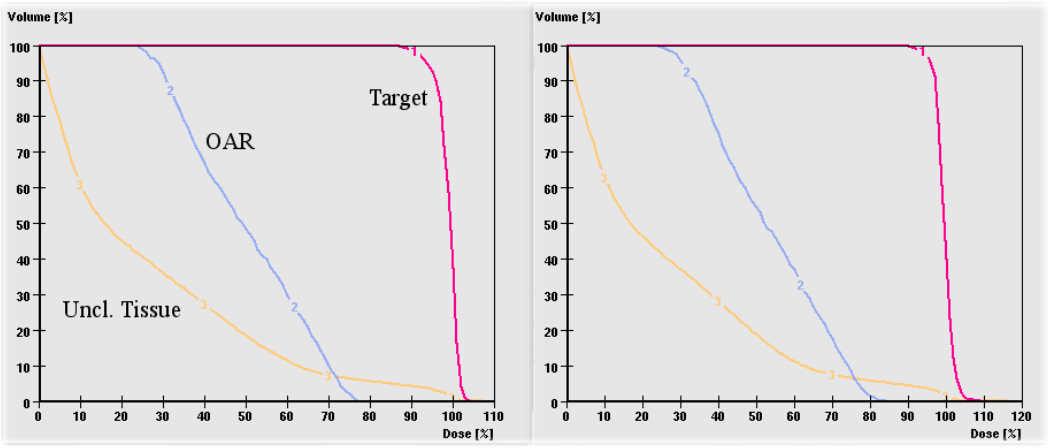

Figure 3.3: Dose volume histograms of the horseshoe phantom. The parameters p_T and q_T are the only parameters which are modified. On the left $p_T = 1$ and $q_T = 8$, setting a high priority on the target overdosage. On the right we have $p_T = 8$ and $q_T = 1$. The aspired objective value for the OAR was $\text{gEUD}_8^{asp} = 30\,\text{Gy}$.

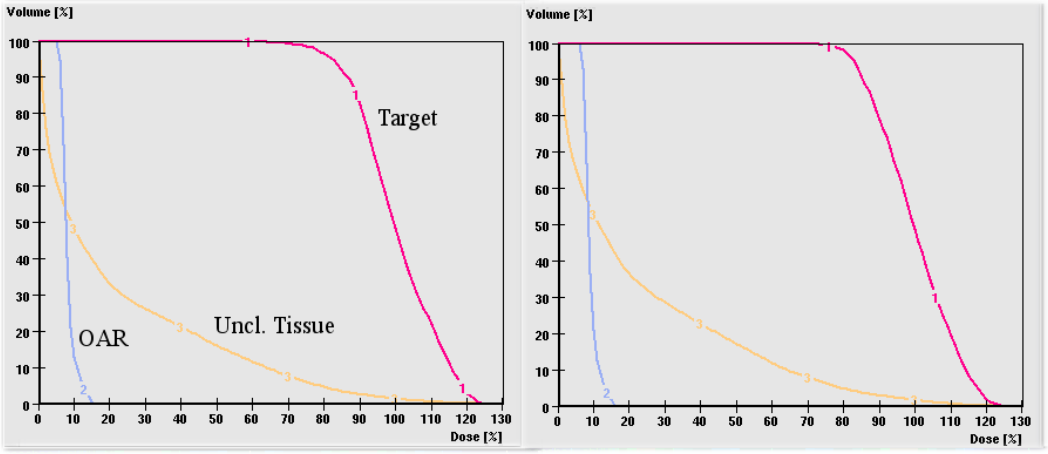

Figure 3.4: DVHs of two solutions from two different databases that only differ in the setting of the parameters p_T and q_T. On the right $p_T = 8$ and $q_T = 1$, setting a high priority on the target underdosage. On the left we have $p_T = 1$ and $q_T = 8$. Here the aspired objective value for the OAR was set to $\text{gEUD}_8^{asp} = 10\,\text{Gy}$.

The effect of the organ parameter p in the pq-model is shown in figure 3.5. While p is increasing from 2 to 8 ($p = 2, 4, 8$ from left to right), the maximum OAR dose decreases

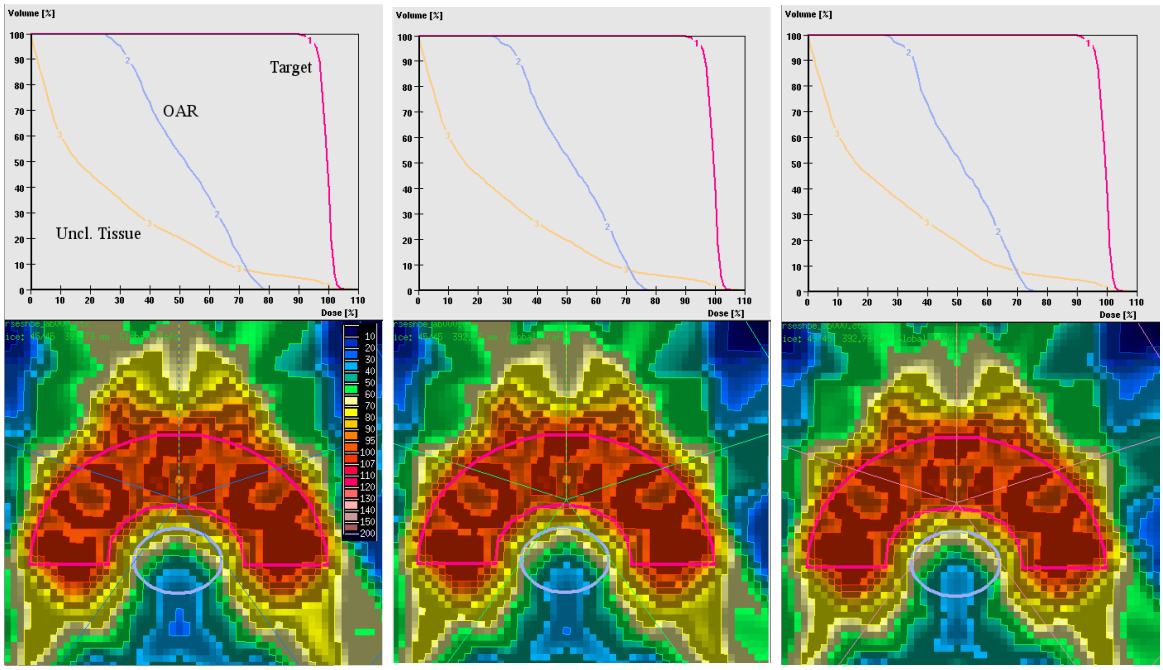

Figure 3.5: Influence of the organ parameter p on the efficient solutions. All three solutions are optimized with the pq-model and the organ parameter is varied from $p = 2$ to $p = 8$ ($p = 2, 4, 8$ from left to right). The maximum OAR dose decreases in total by 3.7% ($d_{max}^{OAR} = 79.5, 77.6, 75.8$).

by only 3.6%. Also, the shape of the DVH for the OAR stays nearly the same. The target minimum and maximum doses as well as the mean dose are not affected to more than 1%.

The parameter with the largest effect on the optimal solution is the aspired objective value F^{asp}. This fact is shown in figure 3.6 where a change from $F_{OAR}^{asp} = 10\,\text{Gy}$ to $40\,\text{Gy}$ results in a change of maximum OAR dose of more then 40%.

Concerning the parameters for the pt-model the results are rather similar as for the pq-model. The parameter p_t which penalizes the tail of the distribution exceeding a dose value of d_{up} has only a small effect on the solution characteristic, whereas modifying d_{up} results in distinctive changes of the dose distribution (see figure 3.7).

The findings for the phantom studies also apply for the clinical cases. Two databases are calculated with the pq-model for a paraspinal case, the planning horizon for the spine was not restricted to have a larger extent of the database. The organ parameter for the spine are $p_1 = 3$ and $p_2 = 8$ with a prescribed dose of $60\,\text{Gy}$ for the target and $70\,\text{Gy}$ for the boost volume. In figure 3.8 two solutions from each of the databases are shown. For very similar target distributions the maximum spine dose changes from $48\,\text{Gy}$ to $42\,\text{Gy}$. Although not clinically relevant, we see that the solutions with the best target coverage both have a maximum spine dose of $63\,\text{Gy}$.

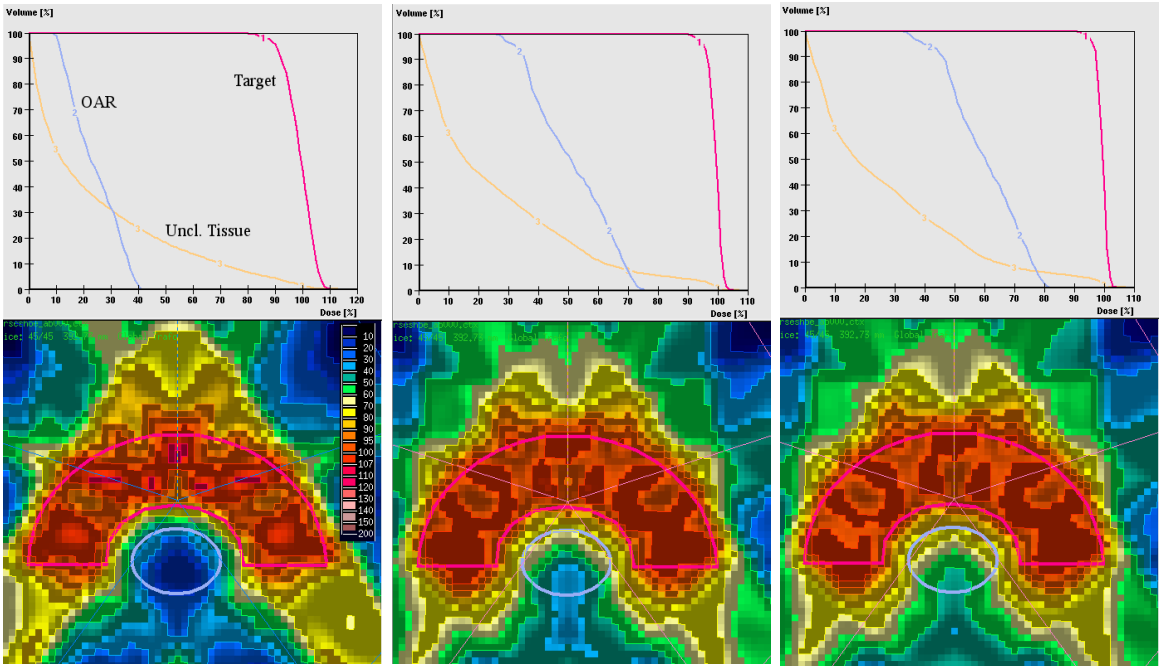

Figure 3.6: Effect of changing the aspired function value for the OAR. The aspired function value for the OAR is varied between $\text{gEUD}_8^{asp} = 10\,\text{Gy}$ (left) $30\,\text{Gy}$ (middle) and $40\,\text{Gy}$ (right). The maximum OAR dose changes by more than 40%.

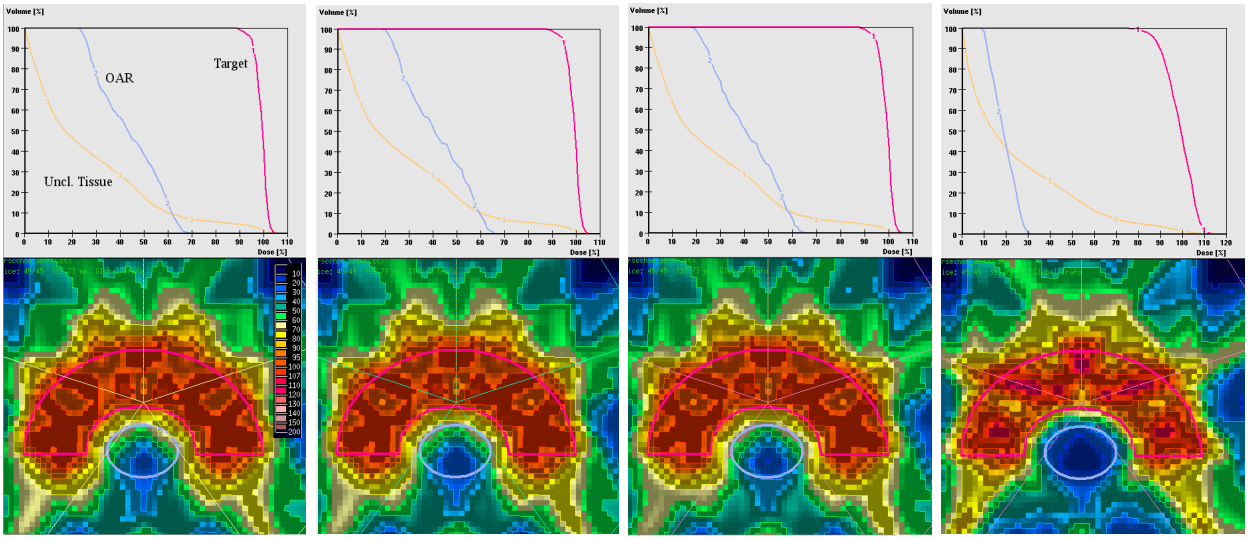

Figure 3.7: The horseshoe phantom is optimized with the pt-model. The first three solutions correspond to different values of p_t ($p_t = 2, 4, 8$ from left to right), while in the solution on the right d_{up} was changed from $30\,\text{Gy}$ to $10\,\text{Gy}$.

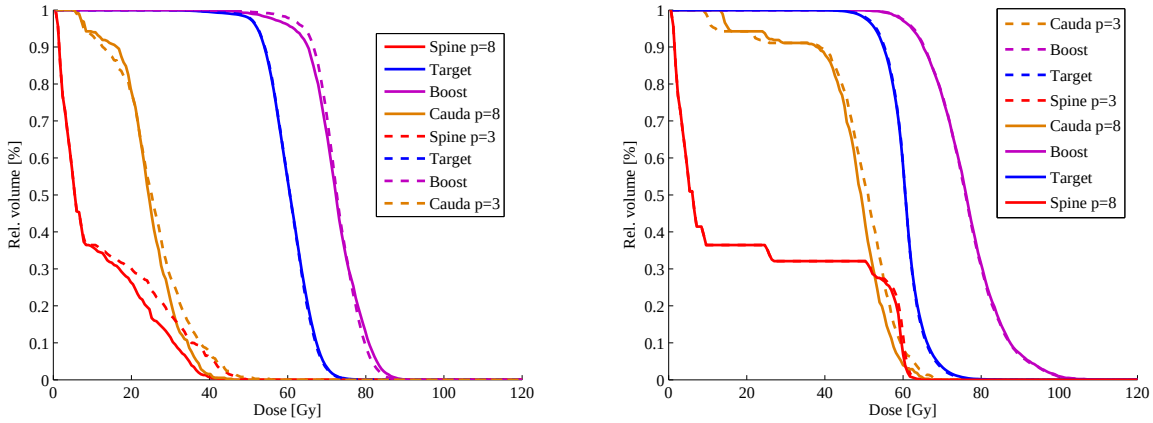

Figure 3.8: Influence of the parameter p for the spine. Two solutions with $p = 3$ (dashed lines) and two solutions with $p = 8$ (solid lines) are shown. While the maximum dose of the spine changes by 6 Gy on the left hand side there is nearly no difference on the right.

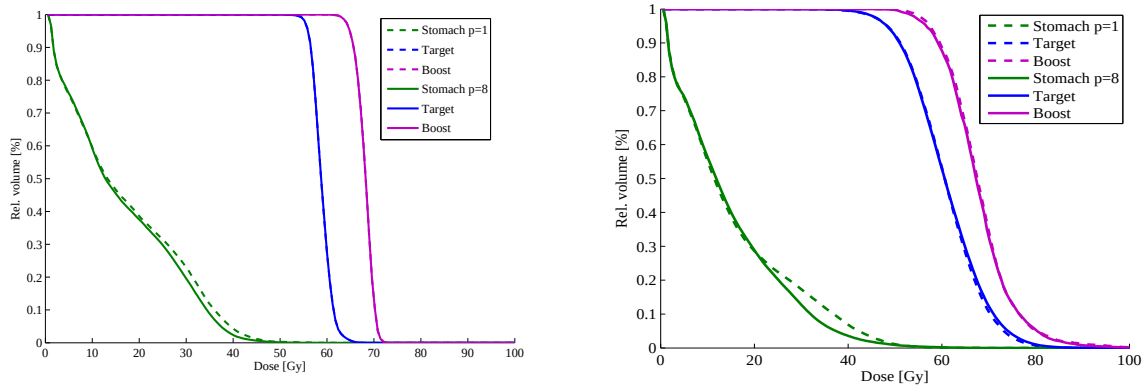

Figure 3.9: Influence of the parameter p for the stomach. All other OARs are virtually the same.

To study the effect of the model parameters for a large organ which is not directly involved in the trade-off with the target we investigate the influence of p for the stomach. As shown in figure 3.9 the influence is marginal.

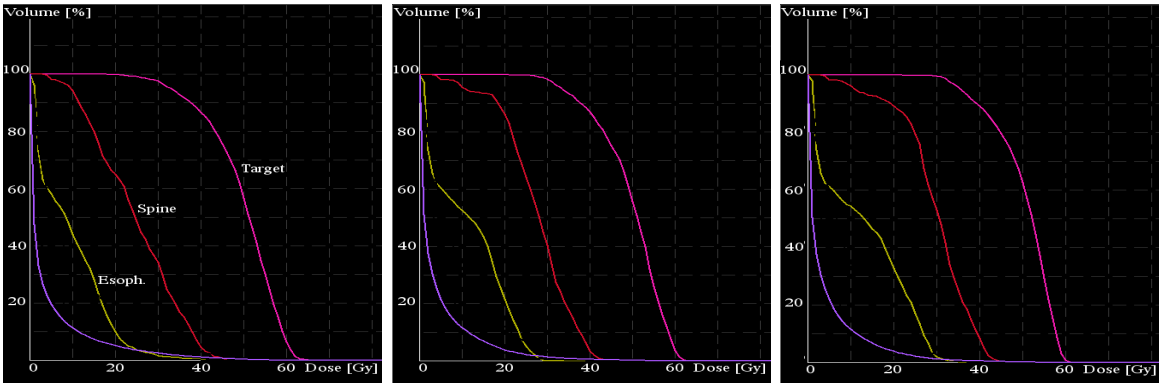

Figure 3.10: A paraspinal case is optimized with three different models. One solution from each database is shown, the pq-model (left), the pt-model (middle) and the std-model (right).

We now investigate the potentials of the different models. For the paraspinal case II three databases are calculated with the three different models. The corresponding solutions from each database show a few differences that are depicted in figure 3.10. The target quality with regard to under- and overdosage improves from left to right, however, the esophagus mean dose also increases. The maximum dose of the spine is below 45 Gy for all solutions, the main differences between the DVH curves of the spine arises in the medium and low dose regions, where the pq-model has significant lower dose values. The corresponding dose distributions in figure 3.11 show that the solutions are not substantially different.

The Hausdorff distances d_H for the horseshoe phantom and the paraspinal II case are shown in table 3.1. Exemplarily the distances for all solutions are depicted in figure 3.12 for the phantom case and the model comparison pt vs pq.

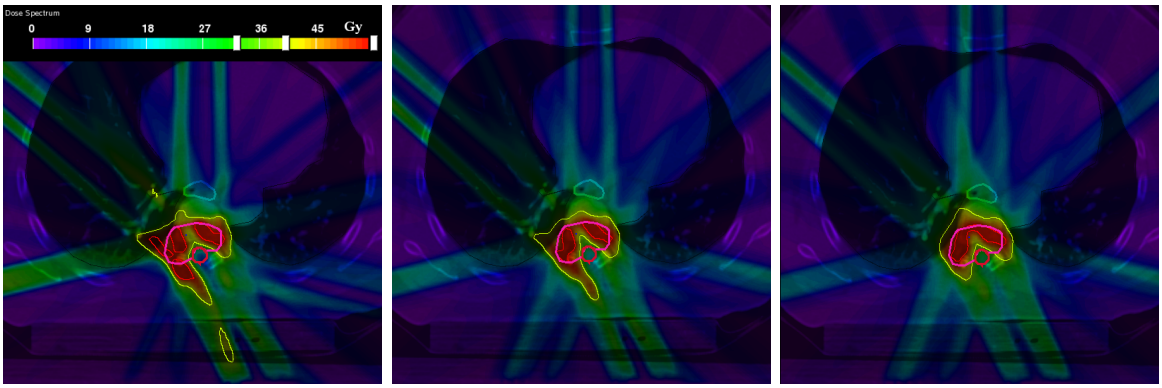

Figure 3.11: Dose distributions for the three optimized with the pq-model (left), the pt-model (middle) and the std-model (right).

Table 3.1: Hausdorff distances $d_H = (\mathrm{conv}(\mathcal{D}_1), \mathcal{D}_2)$ based on the first order approximation of the efficient sets for the three models. The values in brackets correspond to the values of $\max_{y \in \mathcal{D}_2} \min_{z \in \mathrm{conv}(\mathcal{D}_1)} d(y, z)$.

	std vs pq	std vs pt	pt vs pq
Paraspinal	0.36 (0.11)	0.59 (0.09)	0.79 (0.11)
Phantom	0.64 (0.15)	0.82 (0.18)	0.61 (0.17)

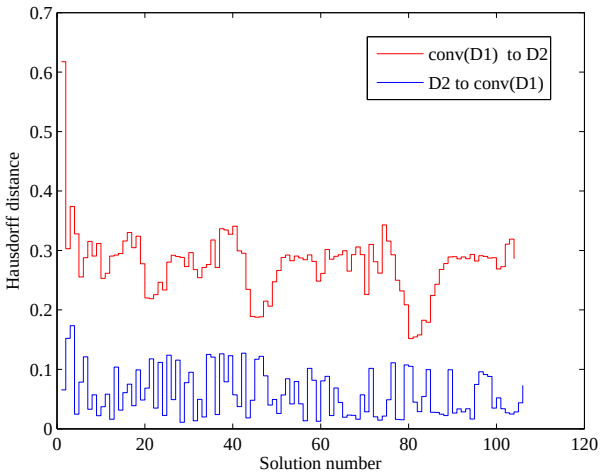

Figure 3.12: Maximal distances for the solutions from $\mathcal{D}_2$ to the convex hull of $\mathcal{D}_1$ (blue) and from the convex hull of $\mathcal{D}_1$ to the solutions from $\mathcal{D}_2$ (red).

3.1.4 Discussion

Concerning the sensitivity of the model parameters it was found that the influence of the penalization parameters p, p_t and p_T is small for most of the cases. This is interesting because in simple scalar optimization or approximation problems the choice of the penalty parameters (or rather the norm) can have a significant effect on the optimal solutions and the residuals. As the influence of p is small, the corresponding influence of μ is also of minor importance. Although the ability of shaping the DVH curves via the penalization parameters is limited, the influence depends on the particular solution. If the target has a homogeneous dose distribution there is not much freedom of further manipulating the remaining planning structures and the whole dose distribution is rather fixed and the actual value of the other model parameters is not of great importance. If the target quality is relaxed however, the effect of the parameters can increase.
The by far largest effect on the solution have the parameters for d_{up}, d_{up}^T, d_{low}^T, d_{mean} and F^{asp}. They allow for large "horizontal" doseshifts in the DVHs and the corresponding dose distributions are substantially affected. Similar results are reported by Krause et al. (2008) where a 2d phantom case is studied by means of elasticity theory.

If the model parameters (like the parameter p) have an interpretation in a biological model, one should obtain their actual values from fits to clinical data. None the less the question remains what are the sensitivities of these parameters, to be able to answer questions such as "What is the effect on the solutions if one uses a "wrong" parameter value?" or "Is it important if the parameter value has a large uncertainty?". Moreover in the pq-model the parameters p and q are combined. The intention is to be able to shape the solutions according to one's own preferences. With this approach however the biological meaning of the parameters vanishes and they become model parameters one has to choose.

Concerning the different optimization models, no model is found to be superior in general. So there is no clinical dominance for all plans produced with one model when compared with another one. Nevertheless there are cases where one model can produce favorable solutions compared to another model. For example the pt-model better translates a relaxation of the target quality into an reduction of high dose region for serial OARs near the target. Whereas the pq-model also for high values of penalty parameters often uses the gained freedom to reduce the intermediate and low dose regions of the OAR which are of minor clinical relevance. This means that in the navigation process the high dose regions can lack controllability. The obvious cure of increasing the parameter p fails, because p is restricted due to numerical stability issues in the optimization algorithm. So for serial OARs there can be a slight advantage when using the pt-model. Nevertheless, at some point the pt-modell will not further improve the OAR, because the low dose regions don't enter the objective,

whereas the pq-modell will also "see" the low dose regions and can therefore improve the OAR still further. These situation will probably arise only for VOIs which are not involved in the direct trade-off with the target volume.
The std-modell results in solutions similar to the pt-model. As the mean dose is fixed there is only one quality indicator per target structure, so that under- and overdosage cannot be controlled separately.

The computations of the Hausdorff distances show that there can occur considerable changes in efficient sets with the change of models. That is the maximal change in a particular solution can be pronounced. Although the practical meaning of d_H might seem limited we can give it an interpretation. If we obtain a particular value of $d_H = v$ after the change of the models we get the following information: there are solutions in the second database that can differ in one objective from the achievable values in the first database by up to $v\,\%$. That happens if the entire change occurs in the direction of one objective (an even split to all objective directions would yield a change of $v/\sqrt{k}$ respectively). Note that many solutions can still be very similar, because we detect only the "maximal" change in the underlying Pareto fronts. This is also seen in figure 3.12 where many solutions change by less than 10%. No solutions has a value of 0 however, which means that one cannot obtain the second solutions by a convex combination of the first ones.

In general a more flexible modelling framework is desirable, where one can arbitrarily assign functions to the respective planning structures. To improve the control over specific parts of a DVH curve one possibility would be to include DVH-constraints, leading to a non-convex optimization problem and a mixed-integer formulation (Halabi et al., 2006). Another approach is to use the existing modelling framework and include a search for specific DVH-information into the navigation process, which is a topic of ongoing research.

3.1.5 Conclusions

We studied the different modeling approaches for objective functions and analyzed the sensitivity of the corresponding model parameters. No modeling approach is found to be superior in all cases. For serial OARs in a direct trade-off with the target the pt-model can be advantageous. The penalization parameters in all models have little effect on the optimal solutions, but the influence depends on the particular location on the Pareto front.

3.2 Quality Indicators for Efficient Sets

In this section we will introduce indicators which allow to evaluate the quality of an approximation of an efficient set.

3.2.1 Motivation

As the solution of a MOP is not unique one is faced with the problem of finding a *good* (discrete) representation of the Pareto front. Since the actual structure of a database is not known to the planner he doesn't know if the underlying treatment plans form a good or less good representation of the Pareto front on which he will base his decision. It is therefore desirable to have additional information on the quality of the representation. We will define several quality measures in the subsequent sections.

3.2.2 Methods

First we discuss the concepts to determine the approximation quality of efficient sets.

The number of efficient solutions for a convex MOP is in general infinite and there is no closed form, describing the efficient set. Consequently we have to consider an approximation of the Pareto front. Different approaches can be used:

- The most elementary approach is the 0-th order approximation: a finite number of Pareto optimal points is calculated and used as an approximation of the whole efficient set.
- The 1-st order methods use piecewise linear approximations. Inner approximations use convex combinations of single efficient points (Chernykh, 1995) while outer approximations make use of supporting hyperplanes of the efficient set (Fruhwirth et al., 1989).
- Higher order approximations have been proposed (see Ruzika and Wiecek (2003)), but will not be further discussed here.

In general the quality can be evaluated using three different criteria (Sayin, 2000):

1. **Cardinality.** The number of discrete points used to represent the connected Pareto front is described by the cardinality. If the points are not wisely placed two representations can differ in their cardinality but carrying the same information about the efficient set. In general we are interested in a small number of discrete points, which

can be calculated in a reasonable amount of time, without losing too much structural information of the Pareto front.

2. **Uniformity.** Uniformity describes the distribution of the points in the objective space. The points should cover the whole region of interest. Underrepresented areas or gaps in the distribution resulting in large uncertainties of the Pareto front are undesirable. Furthermore the distribution should not contain any clusters or overrepresented areas including redundant points that don't contribute to the information about the Pareto front.
3. **Coverage.** For each point on the Pareto front there should be an element in the approximation which represents this point reasonably well. So all areas of the Pareto front have to be adequately covered, keeping the error introduced by the approximation small.

These three issues are addressed in the the following sections.

Cardinality and Hypervolume Indicator

We will first discuss the cardinality, to address the question how many solutions have to be calculated to obtain a good approximation of the Pareto front. As the Pareto front is a hypersurface in the objective space the number of solutions needed for an appropriate coverage will in general increase exponentially (Küfer et al., 2000) with the dimension k of the problem. However if the shape of the Pareto front locally resembles a plane then a 1-st order approximation with convex combinations of single solutions will yield a satisfactory description. Hence less solutions would be needed. Large errors will only occur in areas with a high "curvature", that is where the tangent plane doesn't resemble the Pareto front well.

The number of intermediate solutions to be calculated can be controlled directly within the MIRA system. To address the question how many solutions have to be calculated to obtain an adequate approximation of the Pareto front, we propose a measure based on convex hull volumina. It makes use of the fact that every improvement in the approximation involves an increase of the convex hull volume of all discrete points. Moreover an additional point which adds no significant information (e.g. it is too close to an existing point or it is located in a plane region of the Pareto front) will not contribute to the measure.

Definition 5 (Utopia point)**.** The *utopia point* $f^* \in \mathbb{R}^k$ is defined as the vector of the global minima F_i^* of all individual objective functions F_i, so $f^* = (F_1^*, ..., F_k^*)^T$.

For a nontrivial multiobjective optimization problem the utopia point will (as the name implies) never be reached.

Given a set of Pareto optimal points in the objective space we will define the hypervolume indicator as follows:

Definition 6 (Hypervolume indicator)**.** Let $\mathcal{D}$ be a set of discrete Pareto optimal points in $\mathbf{R}^k$ and $\mathcal{D}_i \subset \mathcal{D}$ with $|\mathcal{D}_i| \geq k+1$ and f^* the utopia point. The hypervolume indicator S_i is defined as

$$S_i = \frac{\text{vol}(\text{conv}(\mathcal{D}_i))}{\text{vol}(\text{conv}(\mathcal{D} \cup f^*)} \tag{3.7}$$

where conv$(\cdot)$ is the convex hull and vol$(\cdot)$ denotes the k-dimensional hypervolume.

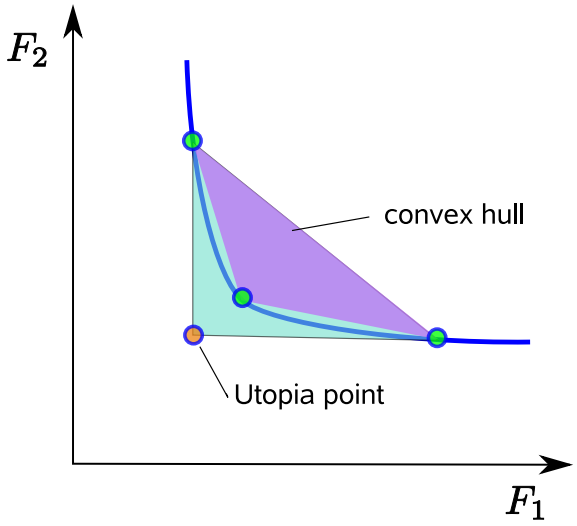

Figure 3.13: Schematic picture of the hypervolume indicator.

The indicator is illustrated in figure 3.13. It allows to examine the improvement of the approximation as a function of increasing number of solutions as follows: First we will set $\mathcal{D}_0$ to be the set consisting of all extreme compromises, yielding a value for S_0. Then each intermediate solution y_i is successively added to $\mathcal{D}_{i-1}$ resulting in a sequence $S_0, S_1, ..., S_M$ where M is the number of intermediate solutions in the database. So we have a monotone increasing sequence of indicator values. Each indicator increase corresponds to an improvement of the discrete approximation, whereas a non-varying step means, that the additional solution doesn't add new information about the Pareto front. The range of the indicator is between 0 and 1, whereas a value of zero corresponds to a flat Pareto front. We also note that S_i is shift invariant, in a sense that the absolute value of S_i only depends on the shape of the Pareto front and not on it's location in the objective space. If the approximation can't be further improved then there won't be a significant increase in S_i anymore, resulting in a saturation. Hence we have a necessary condition for an adequate number of solutions.

Uniformity and Hyperspheres

In this this section we will address the question of uniformity, as we are interested in a uniform distribution of solutions in the objective space we have to be able to detect clusters and gaps within the distribution.
Guarantying an evenly distributed set of solutions in higher dimensions is a non-trivial task. For low dimensional problems methods have been proposed (Das, 1999; Messac et al., 2003), but they cannot be easily extended to higher dimensions. Therefore the MIRA system uses a stochastic mechanism for calculating intermediate solutions (Küfer et al., 2003) to approach a satisfactory approximation also in higher dimensions.

As we are interested in a uniform distribution we have to be able to detect clusters and gaps. Clusters will appear if there is no restriction to the minimal distance between two points so that there are solutions that are too close together. To provide an insight into the spread of the solutions for each solution the Euclidean distance to the nearest neighbour is calculated. An alternative formulation would be: for each solution y_i find the smallest hypersphere that touches y_i and exactly one other solution (implying that there are no solutions within the interior of the hypersphere). The distribution of the radii of smallest hyperspheres enables us to detect clusters in the distribution.

The exact value of the nearest neighbour distance for an ideal uniform distribution depends on M and k. We can try to find a measure that is independent of M and k. For an even distribution of M points on a k-dimensional unit-cube it is proportional to $(\sqrt[k]{M}-1)^{-1}$ (when $M = m^k$ with $m \in \mathbb{N}$). Furthermore the volume of a standard n-simplex α_n is given as $\mathrm{vol}(\alpha_n) = \frac{s^n}{n!}\sqrt{\frac{n+1}{2^n}}$ (Sommerville, 1958), where s is the edge length of the simplex. For the standard n-simplex in $\mathbf{R}^{n+1}$ defined be the vertices $\{e_1, \ldots, e_{n+1}\}$ of the unit vectors e_i we have $s = \sqrt{2}$. So, assuming a flat Pareto front with M evenly distributed solutions as a reference, we will calculate as a characteristic number the following:

$$H_s = \bar{d_s} \cdot (\sqrt[k-1]{M} - 1) \cdot \{ \frac{(k-1)!}{\sqrt{2^{(k-1)}}} \cdot \sqrt{\frac{2^{(k-1)}}{(k-1)+1}} \}^{1/k} = \bar{d_s} \cdot (\sqrt[k-1]{M} - 1) \cdot \{ \frac{(k-1)!}{\sqrt{k}} \}^{1/k} \quad (3.8)$$

$\bar{d_s}$ is the mean of the nearest neighbour values. For an ideal uniform distribution on a flat Pareto front H_s will be 1. If the solutions tend to cluster H_s will decrease and eventually approach 0.

Note that if there are some solutions having a significantly higher value in the nearest neighbour distance, we are also able to identify gaps in the discrete representation. However an even nearest neighbour statistic is not sufficient to guarantee that there are no gaps in

the solution distribution (cf. the left distribution in figure 3.15 where one can remove any solution and the nearest neighbour distribution will remain constant).

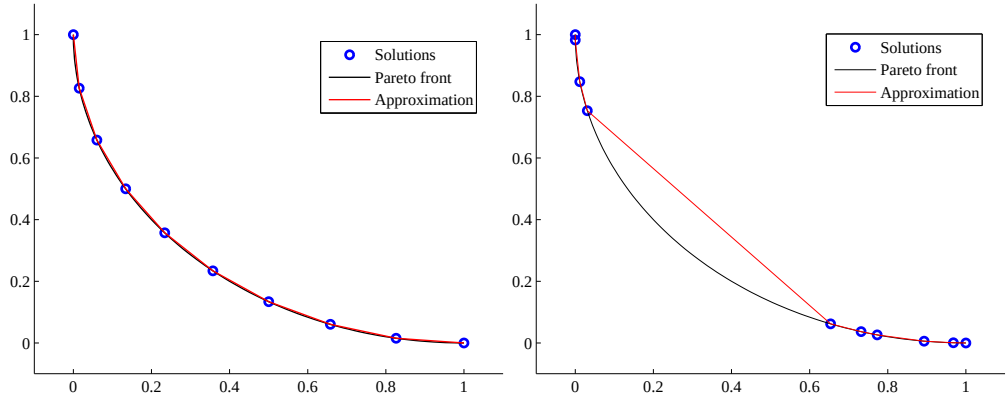

Figure 3.14: Example of two dimensional efficient sets. On the left successive solutions (blue) have the same arc length. The inner approximation (red) is close to the Pareto front (black). On the right hand side the random placement of the same number of solutions can result in large gaps between the approximation and the efficient set. The corresponding indicator values for H_s are $H_s = 1.09$ (left) and $H_s = 0.33$ (right).

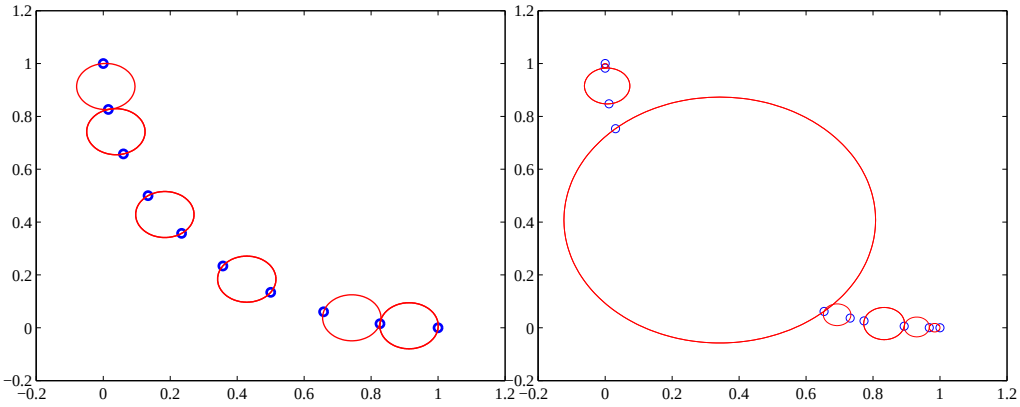

Figure 3.15: For the solutions above the largest hyperspheres are calculated. The gap in the distribution on the right results in a large hypersphere. For the hypersphere indicators we have $H_l = 0.00$ (left) and $H_l = 1.38$ (right).

To detect all gaps another indicator is needed. Similar to the indicator concept above for each solution y_i we will now look at the *largest* hypersphere touching y_i and exactly one other solution, while having no other solutions in the interior of the hypersphere (cf. figure 3.16). We can set up the following discrete optimization problem to find the corresponding hyperspheres. Let the c_{ij} denote the distances between solutions y_i and y_j, and $y_{ij}^m = (y_i + \frac{1}{2}(y_j - y_i)) = \frac{1}{2}(y_i + y_j)$ the middle point of a hypersphere connecting y_i and y_j, then we

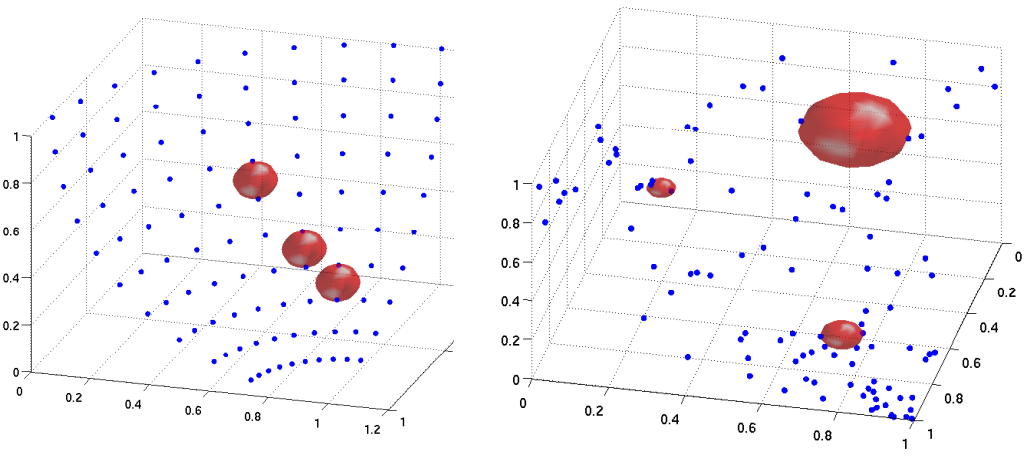

Figure 3.16: Example of a two dimensional Pareto front in three dimensions. Hyperspheres for three arbitrary solutions are shown. The even distribution results in very similar spheres (left) with $H_s = 1.15$ and $H_l = 0.03$ while for the stochastically placed solutions the size of the radii can differ ($H_s = 0.65$ and $H_l = 0.53$).

can introduce the binary decision variables $x_{ij} \in \{0, 1\}$ indicating if there is a hypersphere placed at y_{ij}^m. For each i we then solve

$$\begin{array}{lllll} \text{maximize} & f(x_{ij}) = \sum_j c_{ij} \cdot x_{ij} & & & \\ & & & & \\ \text{subject to} & \sum_j x_{ij} = 1 & & & \text{(select one hypersphere)} \\ & x_{jj} = 0 & & \forall j & \text{(non degenerated)} \\ & (y_l - y_{ij}^m)^2 \geq (\frac{1}{2} c_{ij})^2 \cdot x_{ij} & & \forall j, l & \text{(empty interior)} \\ & x_{ij} \in \{0, 1\} & & & \text{(binary).} \end{array} \tag{3.9}$$

Note that both the objective function and all constraints are linear in the decision variables, leading to a simple integer program.

To characterize the distribution of the hypersphere radii we calculate

$$H_l = \frac{\sigma_l}{\bar{d}_l}. \tag{3.10}$$

also known as the coefficient of variation. $\bar{d}_l$ is the mean of the radii of the hyperspheres and σ_l is the corresponding standard deviation. H_l will approach zero for an even distribution. Any gaps will increase H_l (as long as $\bar{d}_l$ is not too close to zero) and separated solutions can be identified.

Only the approach based on the largest hyperspheres guarantees to find all underrepresented areas on the Pareto front.

Coverage error

An important information about the approximation quality is the coverage error. It measures the distance between the efficient set $\mathcal{P}$ and the approximation $\mathcal{D}$. Commonly the measure is based on the maximum orthogonal line segment from the approximation to $\mathcal{P}$ (cf ref covError). In general we can define the coverage error ε as:

$$\varepsilon = \max_{y \in \mathcal{P}} \min_{z \in \mathcal{D}} d(y, z). \tag{3.11}$$

Here, $d(\cdot, \cdot)$ denotes the Euclidean metric, $\mathcal{P}$ is the efficient set and $\mathcal{D}$ denotes the approximation. This definition doesn't depend on the order of the approximation. In the subsequent section though, we will assume the 1-st order approximation based on convex combinations of single efficient solutions, as used with the MIRA system.

As the Pareto front is in general unknown, we cannot calculate ε directly, but have to consider an upper bound on the coverage error.

We will now describe a procedure to calculate an upper bound on the coverage error for a given set of discrete efficient points:

1. Triangulation. First we will calculate convex hull of all points, resulting in a polyhedron consisting of a set of facets and the corresponding normal vectors. All normal vectors have to be aligned in the same way, pointing outwards (or inwards) the convex hull.
2. Determine the efficient facets. The triangulation will contain facets that are dominated by others and therefore don't belong to the approximation of the efficient set. We will reject all facets whose normal vector has positive components because they are in general not efficient (see proof [1] below). In this way we can determine all efficient

[1] *Proof:* Suppose we have an efficient facet with at least one positive component in the corresponding normal vector. Without loss of generality we assume all normal vectors n are adjusted that they point outwards the convex hull. For each facet with the vertices y_j^{vertex} we look at a point on the facet, for example the middle point given as $y_{mid} = \sum_{j=1}^{k} \frac{1}{k} y_j^{vertex}$. A direction s of a step that stays on the facet is determined by the condition $n_{facet} \cdot s = 0$. If all components of n_{facet} are negative then a non-trivial direction will include positive *and* negative components to fullfill the above condition, thereby conform to the requirement Pareto optimality of the facet. If on the other hand there is a normal vector which has at least one strictly positive component $(n_{facet})_j > 0$, then we can choose a direction $\tilde{s}$ with $(\tilde{s})_j < 0$ and $(\tilde{s})_l < 0$ (for a suited index l) and zero for all other components which satisfies the orthogonal condition above. As we are in the middle of the facet there is a $t > 0$ (t small enough) for which $y_{mid} + t \cdot s$ is still on the facet. Thus we found a step that stays on the facet and improves *two* objective functions without worsening another one, contradicting the assumption of Pareto optimality. □

facets of the triangulation.

3. Identify an upper bound. The utopia point of each facet and the facet itself define a simplex. Due to the convexity of the MOP this simplex consists of upper and lower approximations of the Pareto front, so the front will entirely be contained in the simplex. The simplex has an analytical description, consequently we are able to calculate an upper bound of ε.
4. Restrict the bounds. As the lower bound introduced by the utopia point of each facet is a very rough estimate, we can narrow the lower bounds by neighbouring hyperplanes. Once again we can exploit the convexity of the MOP. The simplex can be restricted by the cut of neighbouring hyperplanes defined by adjacent facets, whereas adjacent facets are facets which share at least one vertex with the current facet.
5. Calculate an upper bound. We can now set up a simple linear program to calculate an upper bound of ε for each facet. The program reads

$$\begin{aligned} &\text{minimize} && f(y) = -n_f^T \cdot y && \\ &\text{subject to} && I \cdot y \succeq f^* && \text{(dominated by utopia point)} \\ &&& n_f^T \cdot y \leq b_f && \text{(bounded by inner approximation)} \\ &&& A \cdot y \succeq b_r && \text{(restrictions)} \end{aligned} \tag{3.12}$$

Here I denotes the $k \times k$ identity matrix, y is the variable point within the simplex. n_f describes the normal vector of the current facet and b_f is the corresponding scalar which stands for the distance to the origin of the hyperplane defined by the facet. The rows of the matrix A contain the normal vectors of the neighbouring hyperplanes used to restrict the simplex and the components of the vector b_r are the corresponding distances to the origin of these hyperplanes. The linear program is easily solved with the LP-solver GLPK (Makhorin, 2009).

3.2.3 Results

We consider a phantom case with two OARs one target volume and the unclassified tissue (cf. chapter 4 figure 4.5) which is optimized with the pq-model, resulting in five objectives. The hypervolume indicator in figure 3.18 shows a large jump after solution 200, which means that a solution adding substantial information on the Pareto front is found. The upper bound on the relative coverage error is $\approx 12\%$ while the mean coverage error is $\approx 0.1\%$. Figure 3.19 shows the corresponding distributions for the nearest neighbour and the hyperspheres.
Another database for the same phantom case with 800 solutions results in a steady curve for

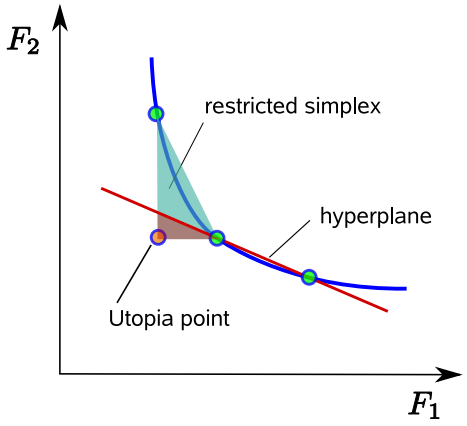

Figure 3.17: The simplex (bluish) formed by a facet and the utopia point for that facet is restricted by the hyperplane (red) defined by a neighbouring facet.

the hypervolume indicator (figure 3.20). Here the main shape of the Pareto front is already captured at the beginning and the approximation gradually improves.
The corresponding coverage error for the database is $\approx 11\%$ and the mean coverage error is $\approx 0.8\%$. Figure 3.21 shows the hypervolume indicator of the phantom case for two databases with 800 and 1200 solutions. After 600 solutions another plan that extends the approximation is found for the second database.

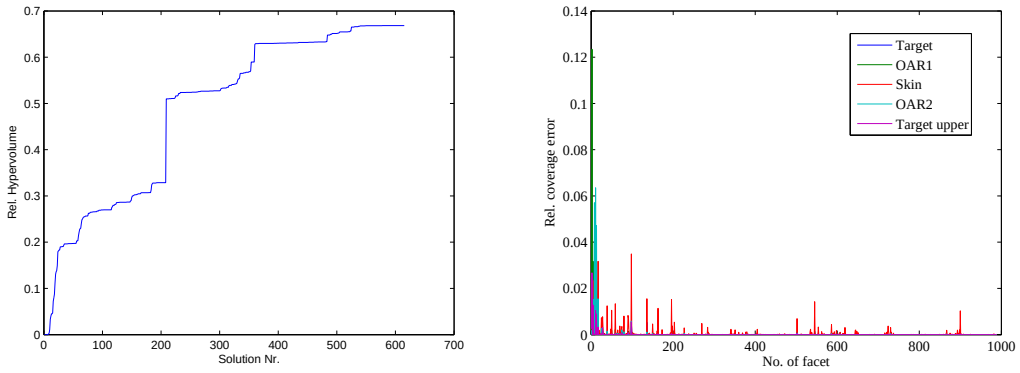

Figure 3.18: The hypervolume indicator for the phantom with two OARs and a database of 615 solutions (left). The upper bound of the coverage error for the database is shown on the right.

We consider the results for a clinical case are shown for a prostate case optimized with the pq-model resulting in a eight-dimensional trade-off. The hypervolume indicator and the corresponding bound on the coverage error is shown in figure 3.22. The database consists of 112 solutions and a value of over 0.7 for S_i is already reached at the solution $i = 38$. The corresponding ϵ for the database is $\epsilon \approx 19\%$ with a mean coverage error of $\approx 1.8\%$.

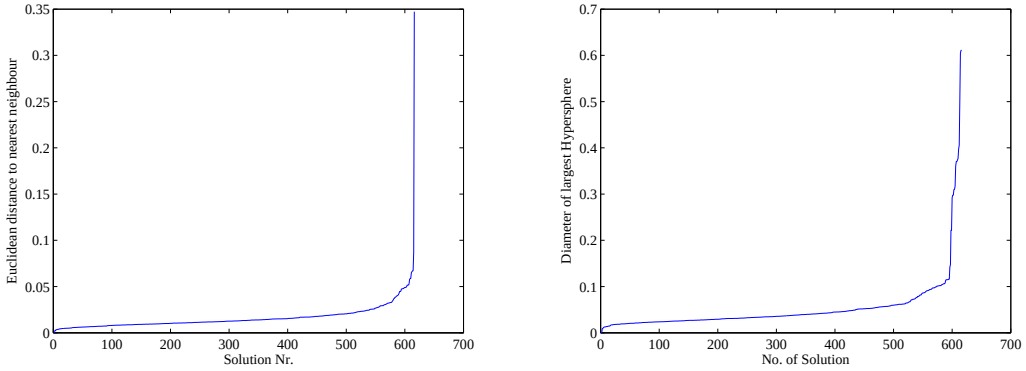

Figure 3.19: Distributions of the nearest neighbour distances (left) and the largest hyperspheres (right) for the phantom case. The corresponding indicator values are $H_s = 0.69$ and $H_l = 1.25$.

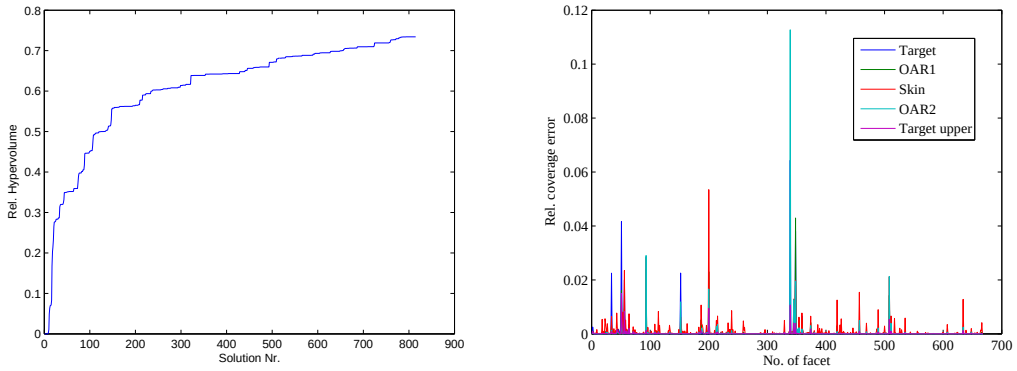

Figure 3.20: Hypervolume indicator of the phantom case with a database of 800 solutions. The upper bound of the coverage error for the database is shown on the right.

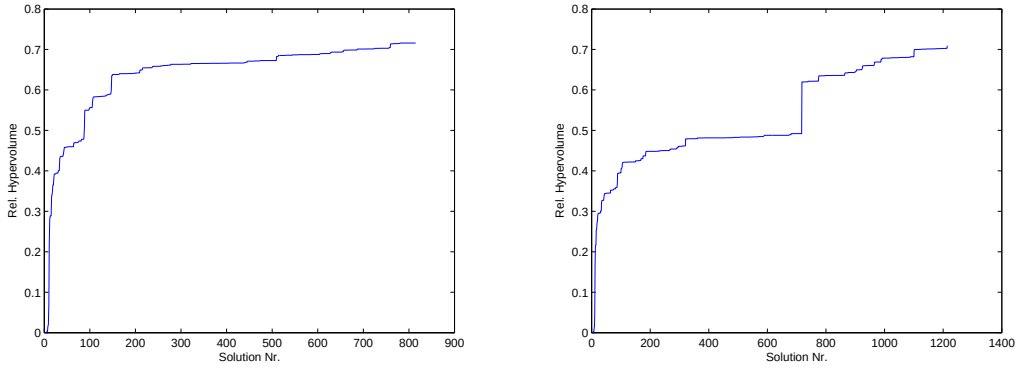

Figure 3.21: Hypervolume indicator for two different databases with 800 and 1200 solutions for the same phantom case.

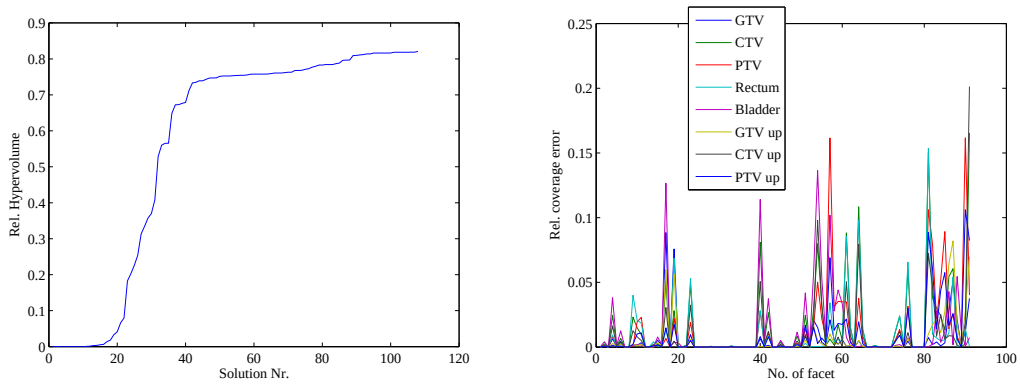

Figure 3.22: Hypervolume indicator (left) and the upper bounds on the coverage error (right) for a prostate database.

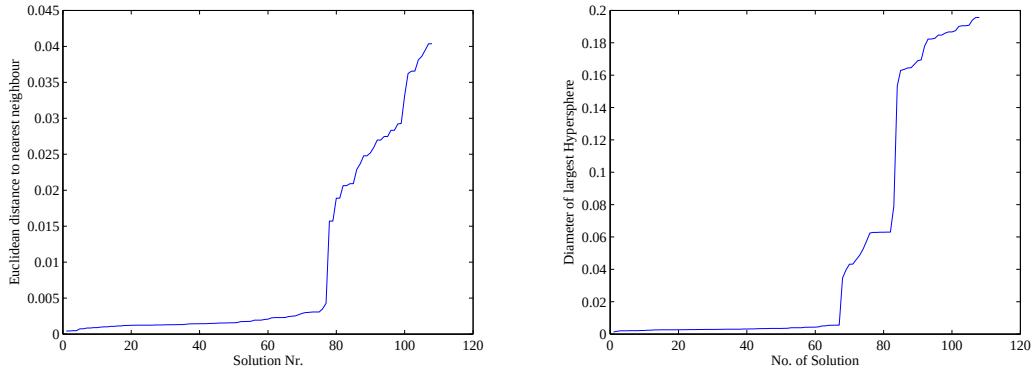

Figure 3.23: For a prostate case with a 110 solutions the nearest neighbour distribution of the solutions (left, $H_s = 0.59$) and the distribution of the largest hyperspheres (right, $H_l = 1.07$) are shown.

3.2.4 Discussion

The introduced measure S_i allows for an analysis of the appropriate number of solutions for a given database. It was found that the number of solutions needed for an appropriate approximation (regarding S_i) can differ from case to case. It also became evident that solutions adding substantial information to the Pareto front can be found in a late phase of the database calculation (cf. figure 3.21). Such a situation would also occur if suddenly a new solution outside the initial planning horizon is produced. For the case in figure 3.21 the jump could be anticipated because S_i only had a value of ≤ 0.5 and another database for the same case had the value of $S_i \geq 0.7$. Although S_i is of course case dependent and related to the curvature of the Pareto front, in most analyzed cases we find a value of $S_i \approx 0.7$ or above.
We also see due to the non-varying steps that many solutions in the later calculation phase don't add new information. Thus a lot of computing time goes into the calculation of solutions which turn out to be not relevant for the approximation. This is in contrast to the PGEN algorithm which tries to steadily improve the approximation based on convex hull information. But this property of the MIRA system comes as a price one has to pay for the stochastical placement of intermediate solutions, that on the other hand allows to handle high dimensional cases. Although the computation time is not a crucial limiting factor at the moment there is still a potential to speed up the database generation.
As the computation of S_i involves convex hull calculations it can't be applied to arbitrary dimensions but is restricted to dimensions up to about ten. However a large class of cases is covered within this dimensional range and more complex cases can always be simplified by grouping some individual structures.
For many studied clinical cases the approximation is improving fast in the beginning phase of the database generation. That means a moderate number of solutions (in the order of 100) can reproduce the main characteristic of the Pareto front also for high dimensional cases. In the later phase however the approximation improves at a much slower rate so that many solutions are needed (several hundred for high dimensional cases) to achieve a reliable representation of the Pareto front.

The introduced hypersphere-based methods allow for a characterization of the distribution of all Pareto optimal points. They show general underlying characteristics for the studied cases.
Firstly, the solutions tend to cluster, which can be seen by the large fraction of solutions with a low nearest neighbour value and which is confirmed in the largest hypersphere distribution, where also a large fraction of solutions shows significant lower values in the hypersphere radii (although they try they cannot grow). Note that if there are kinks in the Pareto front (which can happen), then these points will be found prominently by weighted scalarization methods

with several different settings of weight factors, hence many solutions will be located at these kinks.
Secondly, there are often isolated solutions that can be identified from the nearest neighbour statistic. These turn out to be the extreme compromises which is also apparent in the navigation process where there can suddenly occur large jumps towards the extremal value of an objective during the navigation.
Thirdly, there are large gaps as seen from the maximal values of the hypersphere diameters that are much higher (up to a factor of 5) than the maximum nearest neighbour values.

The indicators H_s and H_l summarize these findings in a single number. H_s is never close to 1 and H_l is never close to 0. Although the exact values of the indicators lack of interpretability they allow for a quick description of the distribution characteristics and a comparison of different databases.

The calculation of upper bounds for the coverage error yields a direct statement of the approximation quality. The mean upper bound on the coverage error for each structure tends to zero for low dimensional ($k = 3$ to 5) cases. Single facets however can still exhibit a large error. For high dimensional ($k = 7$ and above) cases with a moderate number of solutions there can occur still large values for the coverage error, but we have to keep in mind that these are values for single facets which might never be visited in the navigation and moreover it is an upper bound so the actual error will always be smaller.

3.2.5 Conclusion

The presented quality measures allow for a detailed characterization of a Pareto optimal set of plans. They provide an insight into the approximation quality of the Pareto front. This additional information can detect weaknesses of the discrete approximation which would otherwise not be available in the calculation and navigation process, thus offering a quality assurance of the output of the MIRA system. Furthermore an appropriate choice of the number of intermediate solutions can be derived. Finally different databases even with different objective functions can be compared.

4 Clinical Application

In this chapter we evaluate the practical potentials of the planning system. Section 4.1 describes a way to locate a clinical reference plan from another planning system within a Pareto optimal database. In section 4.2 we perform retrospective planning studies and illuminate the practical benefits and limitations for the system.

4.1 Automatic Localization of a Reference Plan

4.1.1 Motivation

In a multiobjective treatment planning framework we are confronted with a whole set of efficient solutions instead of one single optimal solution in conventional planning. A representative discrete subset of the efficient set is stored in a database and is presented to the treatment planner who then chooses the best suited compromise. This raises the question how a conventionally obtained reference plan from an accredited planning system relates to the solutions in the database of the new multiobjective system. In general it is important that the solutions in the database are situated within a clinically relevant domain, preferably around the reference plan, which represents a trusted option. Therefore it is beneficial to have a quick confirmation if a reference plan is contained in the database. In this section we describe how to automatically identify a reference plan within a database of Pareto optimal solutions.

4.1.2 Material and Methods

Solving the MOP results in a database of Pareto optimal treatment plans with $2^k - 1 + M$ solutions. In addition we have a clinically reference plan from another planning system, which was conventionally optimized with different objectives by a planner. To relate the reference plan to the database we can proceed as follows. For each reference plan that is characterized by its dose distribution there exists a representation in the objective space

with respect to the set of objective functions F_i. The coordinates in the objective space are given by the values of the individual objective functions. The question if a reference plan is "contained" in a stored database can also be put in the following way: Is it possible to find a convex combination of solutions from the database that approximates the reference plan reasonably well? As the combination lies within the convex hull, feasibility of the new plan is ensured. Formally we can set up the constrained optimization problem:

$$\begin{aligned} \text{minimize} \quad & f(\lambda) = \| \textstyle\sum_i \lambda_i \cdot y_i - r \|_2 \\ \text{subject to} \quad & \textstyle\sum_{i=1} \lambda_i = 1 \\ & \lambda_i \geq 0 \end{aligned} \tag{4.1}$$

Where λ_i are interpolation coefficients, y_i are the coordinate vectors of the solutions in the database and r is the coordinate vector of the reference plan and the norm denotes the Euclidean norm. The constraints make sure that the combination stays within the convex hull. If the reference plan is contained in the convex hull then the minimal value of 4.1 is zero. Equation 4.1 is easily transformed into an Quadratic Program in standard form and can then be solved by any Quadratic-Program-solver.

For a clinical test case of a paraspinal tumor (see figure 4.1) a discrete representation of the Pareto front was calculated with the MIRA software and stored in a database. Structures considered in the optimization were the spinal cord, the esophagus and the target using the pq-model, consequently there are four independent evaluation criteria and the objective space is four dimensional. For this case a corresponding plan was optimized by the clinical routine version of KonRad, which uses a common quadratic objective function, and chosen as the reference plan. Afterwards the KonRad plan was evaluated by the multiobjective functions and thus mapped into the objective space. The obtained coordinate vector of the reference plan was then used to solve the Quadratic Program of equation 4.1.

4.1.3 Results

A database of 27 Pareto optimal plans was generated and stored. The dose volume histograms of the different solutions which illustrate the different trade-offs are depicted in figure 4.1.

The Quadratic Program for determining the interpolation coefficients terminated with a value of zero, indicating that the reference plan is representable through a convex combination of the solutions in the database. Five interpolation coefficients were found to be non zero (cf. table 1) so only five solutions were necessary to generate the reference plan.

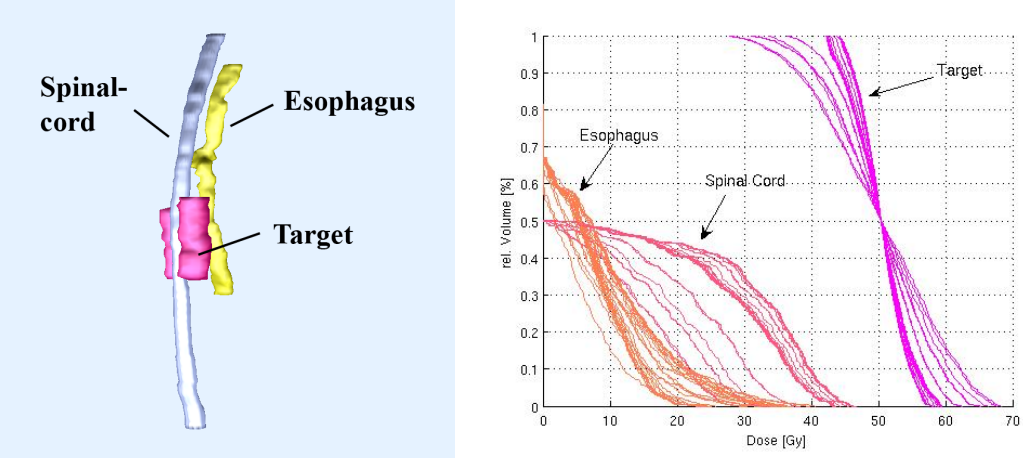

Figure 4.1: Geometry of the vois for the paraspinal case (left) and the DVHs from the Pareto optimal database.

Table 4.1: Values of the interpolation coefficients and the corresponding solution numbers after solving the quadratic program.

Value of interpolation coefficient λ_i	0.016	0.001	0.702	0.052	0.229
Number of solution	1	2	10	14	25

The reference plan and the solution generated with the interpolation coefficients were compared in VIRTUOS. The corresponding DVHs are shown in figure 4.2. Although the reference plan and the solutions in the database were optimized with different optimization models, the resulting DVHs were found to be very similar, indicating that the clinically relevant domain is covered by the database. The convex combination of solutions yields a good approximation of the reference plan.

4.1.4 Discussion and Conclusion

Although the method is based on objective values it is capable to generate new solutions which are very similar to the reference plan, provided it is contained in the database. For simple cases we are mostly able to generate a convex combination of solutions which resemble the reference plan. If the reference plan is rather different the method will return a non-zero value. This happens if either the planning horizon of the database is not well placed and the clinically relevant domain is not covered, or if due to the different optimization models the solutions are so different that they will not match. In the first case we can use this as an indicator that the planning horizon should be extended and the database should be filled with additional solutions. In the second case one could alternatively search in the design space instead of the objective space (e.g. by considering a Chebichev problem which can be

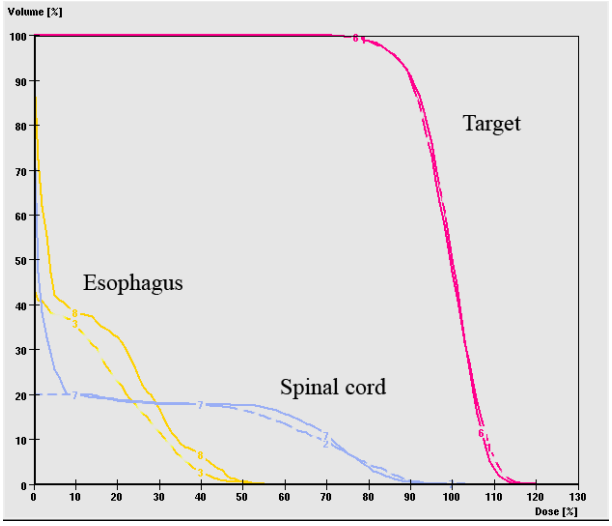

Figure 4.2: The KonRad reference plan compared to the the convex combination of five solution from the database (dashed lines).

formulated as a linear program).

The method can be used to analyze the quality and the benefits of the MIRA system via clinical evaluation. It can assist the quality assurance for the system and allows for a quick identification if a reference plan is contained in a database or if an enhancement of the planning horizon is needed.

4.2 Clinical Evaluation

4.2.1 Motivation

The objective of the studies is to evaluate the clinical potential of the MIRA system and to compare the plan quality with the clinically approved KonRad planning system. On the one hand one has to assure that the MIRA systems meets the KonRad standards. On the other hand one also has to verify that cases conventionally planned with the KonRad system are competitive to the multiobjective planning system which takes into account all objectives simultaneously.

4.2.2 Methods

Pareto optimal plan databases for three different cases are calculated with the MIRA system: a phantom case, a paraspinal case and different prostate cases. They are navigated and compared with a KonRad treatment plan which was conventionally optimized with a quadratic objective function. The navigation process is done within VIRTUOS which offers an interface to incorporate the navigation principle from MIRA (Tesarczyk, 2006). For the prostate cases we also consider a larger retrospective study with ten patients. Besides the plan quality we examine the planning efficiency, i.e. the overall planning time the planer has to spend to arrive at the actual treatment plan. For IMRT of prostate cancer the macroscopic prostate is defined as GTV, a 5 mm margin excluding bladder and rectum is added to obtain the CTV, and another 5 mm are added for the PTV. A total median dose of 78 Gy is delivered in 2 Gy fractions to the GTV with 6 MeV photons using a setup of 7 coplanar, equidistant beams.

4.2.3 Results

The phantom case is optimized with the pt-model and compared to a reference plan from KonRad. We find that the plan quality for both planning systems is rather similar. The DVHs for three different solutions from the database are shown in figure 4.4. The different treatment options available are easily accessible with the MIRA system, e.g. both OARs can be traded independently. Also the cost of the important trade-offs (e.g. target quality vs. OAR maximum doses or countour mean dose) can be explored interactively.

As shown in figure 4.6 (left) the plan quality for the pt-model is similar to KonRad. For another plan of the same phantom optimized with the std-model we can find solutions in the

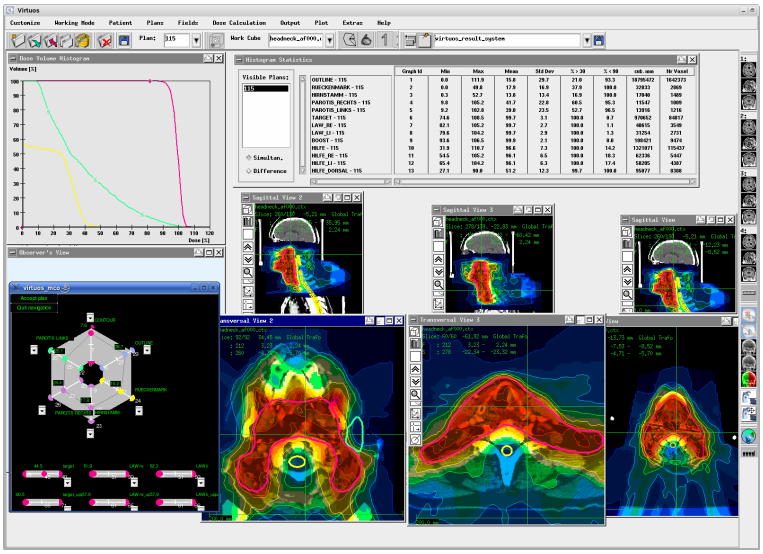

Figure 4.3: Incorporation of the navigation principle into the routine planning system VIRTUOS at the DKFZ.

database which are slightly superior to KonRad, both maximum doses of the OARs as well as the mean dose for the contour are smaller, while the target quality is virtually identical.

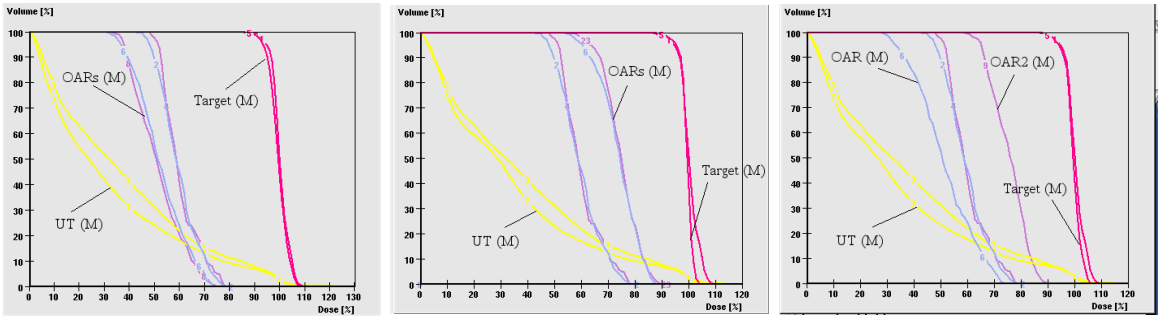

Figure 4.4: Three different trade-offs from a database compared to a KonRad plan. On the left the sparing of both OARs is achieved by a worse target coverage. In the middle the coverage is improved and both OARs are inferior in return. On the right both OARs are decoupled and can be traded independently.

For the prostate case the plans are similar for both systems, again different trade-off options are available with the MIRA system (see figure 4.7).

In the planning study with ten prostate cases, MIRA could process the clinical data sets without further modification. Defining the optimization parameters so that the resulting database really included the clinical relevant part of the solution space took approx. 1 hour for the first patient. For all 9 remaining patients the same parameters could be used as a template. The unattended database generation took approx. 3 hours per patient (between 100 and 150 IMRT plans, voxel resolution 2.6 mm^3, bixel size 1 cm^2). The best IMRT plan was found by interactive navigation in approx. 5 minutes. Realtime interpolation of plans

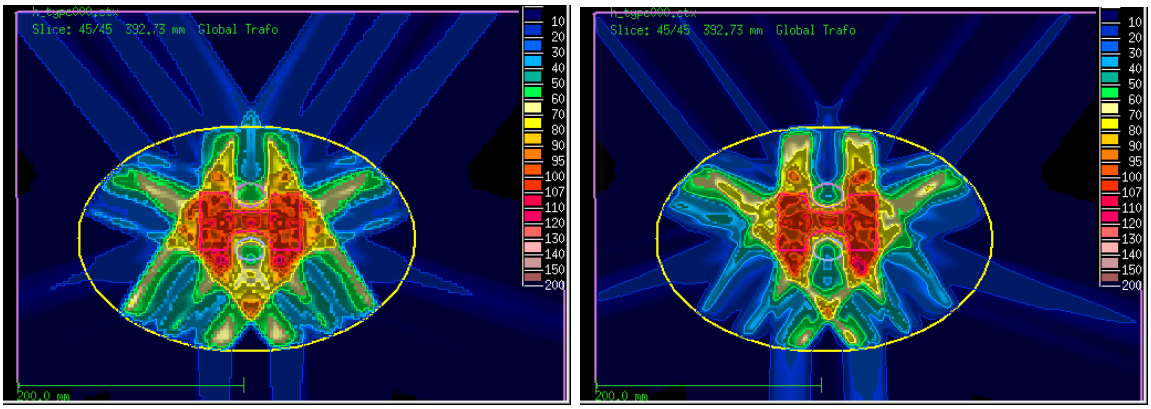

Figure 4.5: Dosedistributions of a KonRad plan (left) and a plan from the database (right) corresponding to the left DVH in figure 4.4.

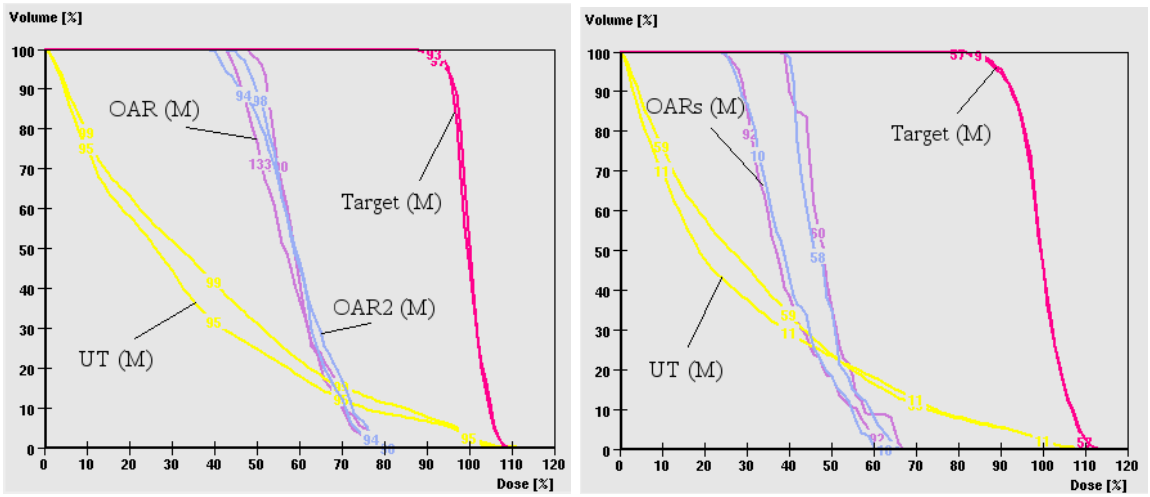

Figure 4.6: The plan from a database optimized with the pt-model is similar to the KonRad plan (left). In a database optimized with the std-model a clinically slightly superior plan compared to KonRad can be found.

ensured that the navigation was smooth and that far more than the 100-150 precalculated plans were accessible. Within the approx. 5 mins, the planner could also experience the sensitivity to changes of the plan, e.g. how much a certain improvement to the GTV coverage increased the dose to the rectum. In KonRad the total planning time was 30-60 minutes per patient during which the planner was occupied, and the planning process was not interactive because each single optimization run took 1-3 minutes. (Considering the uncertainty due to different dose calculation algorithms, the plan quality eventually achieved by KonRad and MIRA was virtually identical for each of the 10 patients. One example is shown in figure 4.7.) The results are summarized in tables 4.2 and 4.3.

Table 4.2: Results for the prostate planning study. The range covered by the database is shown and the values of the corresponding clinical reference plan are reported in brackets.

case	1	2	3	4	5
Max. rectum dose [Gy]	62-94 (77)	64-72 (78)	73-76 (75)	69-79 (77)	70-89 (79)
Rectum volume > 68 Gy [%]	0-15 (2)	0-1 (3)	0-2 (2)	2-39 (3)	1-20 (2)
Mean bladder dose [Gy]	78-86 (80)	67-82 (78)	77-80 (77)	81-85 (80)	77-93 (77)
Max. bladder dose [Gy]	27-59 (42)	17-24 (24)	18-30 (36)	17-31 (23)	26-52 (42)
GTV volume < 70.2 Gy [%]	0-13 (3)	0-7 (1)	0-8 (2)	0-9 (1)	0-2 (2)

Table 4.3: Results for the prostate planning study.

case	6	7	8	9	10
Max. rectum dose [Gy]	68-90 (76)	68-76 (73)	70-110 (83)	57-79 (77)	69-75 (72)
Rectum volume > 68 Gy [%]	0-12 (2)	0-4 (1)	0-22 (3)	0-3 (2)	0-5 (1)
Mean bladder dose [Gy]	89-91 (79)	78-94 (82)	87-93 (85)	78-81 (81)	79-91 (83)
Max. bladder dose [Gy]	33-51 (44)	18-32 (33)	18-28 (26)	12-28 (29)	22-35 (30)
GTV volume < 70.2 Gy [%]	0-7 (2)	0-8 (3)	0-11 (3)	1-17 (1)	0-9 (2)

For a paraspinal case (cf. chapter 3 figure 3.2) we find a MIRA database plan that fails to match the target quality in both target structures. This indicates two different problems. Firstly, the selection of the planning horizon. As the values for the planning horizon have to be specified in the form of objective values they are not necessarily intuitive, that means it can be difficult to guarantee certain desirable solutions in the database as it is difficult to relate objective values to DVH-shapes. Secondly, the normalization problem plays a role, as all plans are normalized to the median target dose, which introduces a perturbation of the efficient set. In the worst case an improvement through the navigation slider will not result in a real improvement for that plan (see discussion of this chapter).

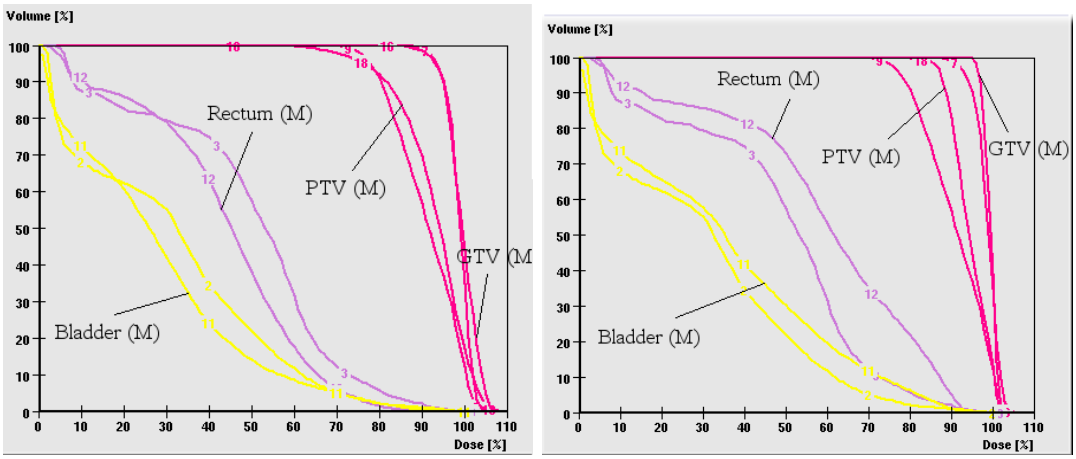

Figure 4.7: DVHs for a prostate case. Two solutions from a database are compared to the KonRad plan.

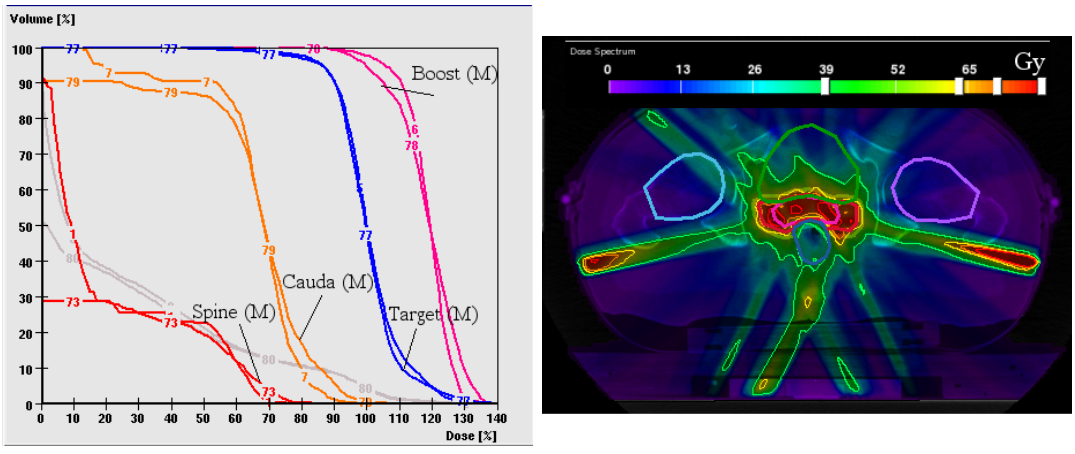

Figure 4.8: For a paraspinal case a clinical inferior plan can be found in the database (left). Both OARs exibit larger maximum doses and the coverage of the boost volume is inferior. For a prostate case there can occur hot spots (right) in the normal tissue.

4.2.4 Discussion

The studies of the phantom cases showed that the MIRA system is able to produce treatment plans which are competitive with the clinical reference system. In some cases there can even be a slight superiority of the MIRA system. This situation can occur if the optimization algorithms use different stopping criteria, as the Gradient- or Newton-based descent algorithms are not run to optimality and the stopping criteria are always arbitrary.

In the prostate planning study it could be demonstrated that MIRA is able to reduce the time for the treatment planning process when many cases are considered. After finding a starting configuration of the optimization parameters it was possible to calculate all subsequent plans with the same parameter setting as a template. So all databases then where located in a suitable part of the Pareto front and could be generated without further user interaction leading to a significant reduction of planning time. However, the initial parameter setting has to be found manually, because the values of the parameters (like the aspired objective values and the planning horizon) can be non intuitive. E.g. to guarantee that solutions with an gEUD value of 40 Gy for an OAR are contained in a database the corresponding parameter for the aspired value could be 10 Gy. As the sensitivity of the solution with respect to the parameters is not constant the search for a suitable set of starting parameters can also take several optimization runs, but there is no need to fine-tune the solution, because the only goal is to roughly position the database on the relevant part of the Pareto front. (If we could automatically generate a starting solution that satisfies all planning goals, we would have revolutionized the conventional planning procedure and there would be no real need for MCO). So for prostate cases and other standard case classes templates can be used to generate the databases, for more complex cases like head and neck cancer several runs for the starting solutions might be necessary before the calculation of the database can take place.

We will shortly discuss the normalization problem in the following. In some cases we find that there are clinically inferior solutions in the database. This seems paradoxical at first, because the solutions are Pareto optimal and the model accurately embodies the clinical goals then there should be no clinical inferior plans. The problem originates from the normalization to the median target dose, which is done for all clinical treatment plans (in some hospitals in the US other DVH-point normalizations to 90 % of the target dose are routinely used). The normalization to a DVH-point destroys the convexity of the Pareto front as all solutions will in general have different median target doses. That means that efficient solutions can loose their property of Pareto optimality after the normalization (depending on the scaling properties of the functions), but they are the solutions the planner chooses from. In the worst case a navigation step which improves a particular objective can turn out be be worse in

that objective after the normalization. This is an inherent problem because the solutions are only Pareto optimal with respect the particular objective functions which allow for different median doses. Note that also in the std-model this normalization effect can occur, because the median and the mean target dose can still differ if the distribution has a high non-zero skewness (the third centralized moment of the distribution), which is often the case. As the median and the skewness of a distribution are non-convex functions they cannot be easily incorporated into the optimization framework which is essentially based on convexity, for example to restrict the skewness of the dosedistribution for target. This problem normally doesn't attract attention in conventional planning, because there is potentially only one normalization for the optimal solution which is then used as the optimal one, independent of the actual scaling. Although this problem is still unsolved it only occurs for some solutions and is not emerging for the whole range of the efficient set.

To maintain the dose conformality in the target the unclassified tissue can be compromised leading to severe hot spots in the dose distribution along the beam entrance directions. The hot spots can be partly controlled by navigating the objective of the unclassified tissue. In some cases however the local characteristic of an hot spot is not controllable via the global objective value for the whole structure.

4.2.5 Conclusions

The MIRA system is able to produce IMRT plans that are competitive with the clinical reference system. There is still some "parameter tweeking", that means a database generation with relevant plans cannot be fully automized and might still need some interaction in the beginning (although the parameters have clinically interpretable meaning in contrast to the conventional weight factors). As demonstrated in the prostate study, a significant reduction in human planning time can be achieved in the long run.
The whole planning process becomes interactive which is a large benefit, thus the planner can focus his time on choosing a plan instead of designing it.
For a first clinical application however an accredited quality assurance is necessary. The normalization problem can hinder the practical navigation process and the planner actively has to check for hot spots.

5 Understanding Pareto Optimal Planning Databases

In this chapter we will present methods that are enable us to gain an understanding of the efficient planning databases as a whole. In section 5.1 we develop a method that allows us to identify the "imperative trade-offs" in a database. Section 5.2 presents ways to identify the main planning options and the relations between the participating planning structures.

5.1 Sensitivities of the Efficient Set

5.1.1 Motivation

Although one can navigate the Pareto optimal databases and get a realtime response from the system it can be difficult to overlook high dimensional trade-offs. Evaluating a five dimensional trade-off for example can become quite difficult if many objectives change at the same time. As visualization techniques are usually limited to dimensions ≤ 3 it is desirable to have some easily accessible and interpretable information about the content of a Pareto optimal database of an arbitrary dimension.

Getting information on the extension of a database is easily achieved by inspecting the extreme compromises. We can also get some information on the distribution of the solutions by looking at scatter plots from the projection of the solutions (see figure 5.1 and 5.2), however the information can be difficult to judge. The matter of interest is the sensitivity of the problem in the objective space, i.e. what is the change in the objectives, if one particular objective is improved by a certain amount (so what does the planner get in return for giving something). If we would have an analytical expression for the efficient set we could employ methods from differential geometry and look at the tangent space of the Pareto front. But firstly the mathematical apparatus would become quite complex and secondly we only have a discrete representation of the efficient set and any smooth approximation to it contains unknown approximation errors. For this reason we develop a simple procedure to study the

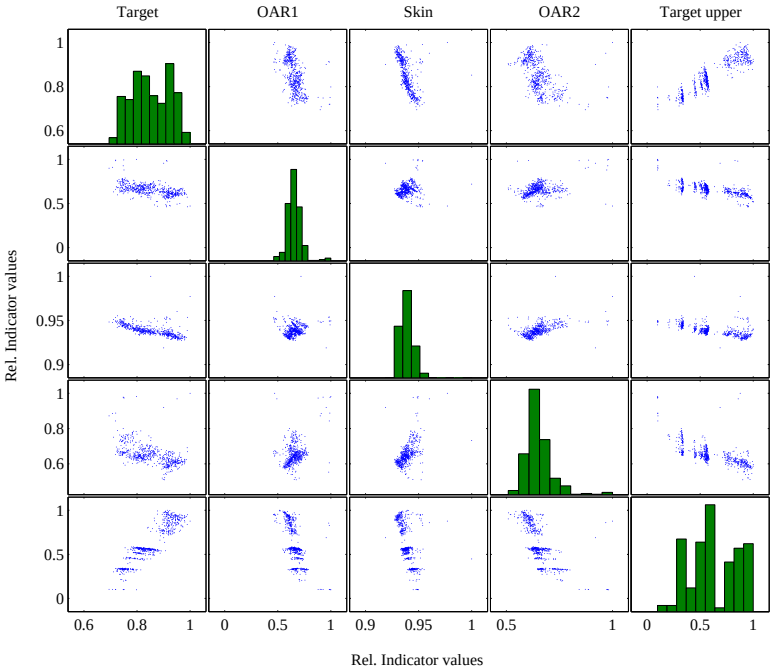

Figure 5.1: Scatter plots of the projected solution for a phantom case with 500 solutions. The figure (i, j) in the matrix corresponds to the projection of all solutions onto the j, i-plane. The diagonal (i, i) shows the histograms of the solution when projected onto the axis of objective i.

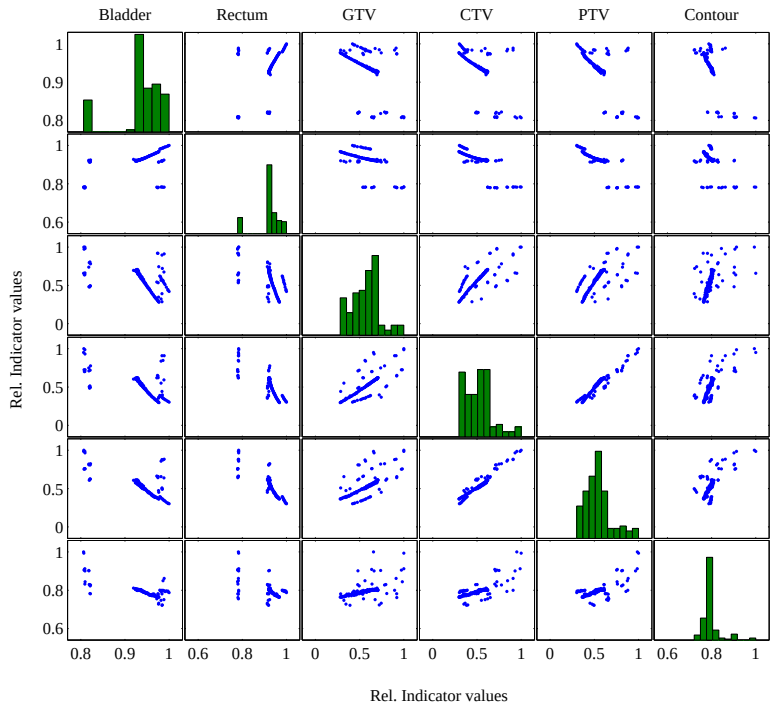

Figure 5.2: Scatter plots of the projected solution for a prostate case.

sensitivity of the Pareto front. The approach uses successive hyperplane restrictions of the efficient set.

5.1.2 Methods

We will investigate how the objectives change if we improve one particular objective k_0. Let $\mathcal{P}$ be the efficient set. We assume without loss of generality that the range of all objectives is between 0 and 1, so we can set up the following halfspace equation

$$h_i = \{y \,|\quad e_{k_0} \cdot y \leq \alpha_i,\, 0 \leq \alpha_i \leq 1\} \tag{5.1}$$

where e_{k_0} is the unit vector in the direction of the objective k_0, y denotes a vector in the objective space and $0 \leq \alpha_i \leq 1$.
Now consider a decreasing sequence of α_i. For each α_i we will look at the intersection of the efficient front and the halfspace $h_i \cap \mathcal{P}$ thereby restricting the efficient front more and more. At each step for all points in the intersection we will determine the minimal achievable objective value in each objective. These values can be plotted against alpha and allow for a visual inspection of the sensitivity of the efficient front. The concept is illustrated in figure 5.3.
We will demonstrate the method for three different cases, including a simple horseshoe phantom case with three objectives (cf. chapter 3 figure 3.2) a clinical prostate case with six objectives and a more complex head and neck case with eight independent objectives (see figure 5.4).

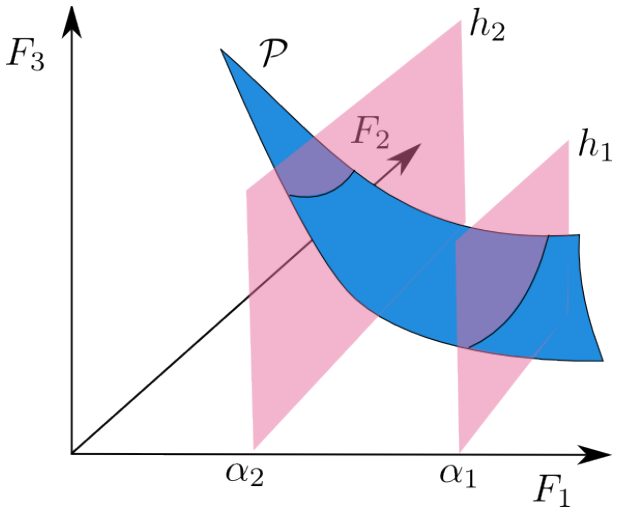

Figure 5.3: For the efficient set $\mathcal{P}$ two restricting hyperplanes are shown.

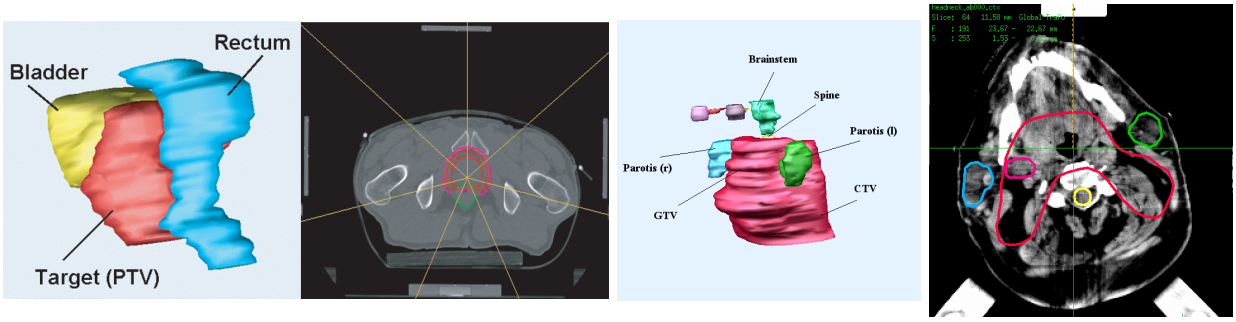

Figure 5.4: Geometries of two clinical cases of a prostate case (left) and a head and neck case (right) with the corresponding CTs.

5.1.3 Results

As a first simple phantom case we consider the three dimensional horseshoe phantom case optimized with the std-model. Figure 5.5 shows the dependency of the target and unclassified tissue objectives (std-deviation and mean dose) from the OAR objective ($gEUD_8$). The unclassified tissue is very little affected and also shows a small range of objective values over the whole database. The obvious "opponent" of the OAR is the target. We can see how the target has to change if the OAR is improved further. For values of alpha between 1 and ≈ 0.7 the dependency of the target is on the OAR is only moderate. If the OAR is improved further the target objective value now has to increase rapidly.

Figures 5.6 and 5.7 show the sensitivities for a prostate case optimized with the std-model when the bladder or the rectum are improved. In both cases the GTV, PTV and CTV have very similar curves due to the geometric formation of these structures. For the rectum the trade-off is imperative over the range of the whole efficient set, that means that the target quality has to be deteriorated whenever the rectum is improved. The target curves for the bladder are flat in the middle which shows that the trade-off is not imperative, so there exists at least one solution with good objective values for the bladder and the target at the same time. Both figures also reveal that the bladder and the rectum are not direct opponents but can be improved simultaneously (then of course at the cost of the target).

The sensitivities for a head and neck case with eight different objectives are shown in figure 5.8. The objective values of the spine are restricted. Except for small values of alpha, all curves of the other objectives are nearly flat. This means that the trade-off between the spine and the other objective is not imperative, but there is a freedom of choice which structure is traded against the spine.

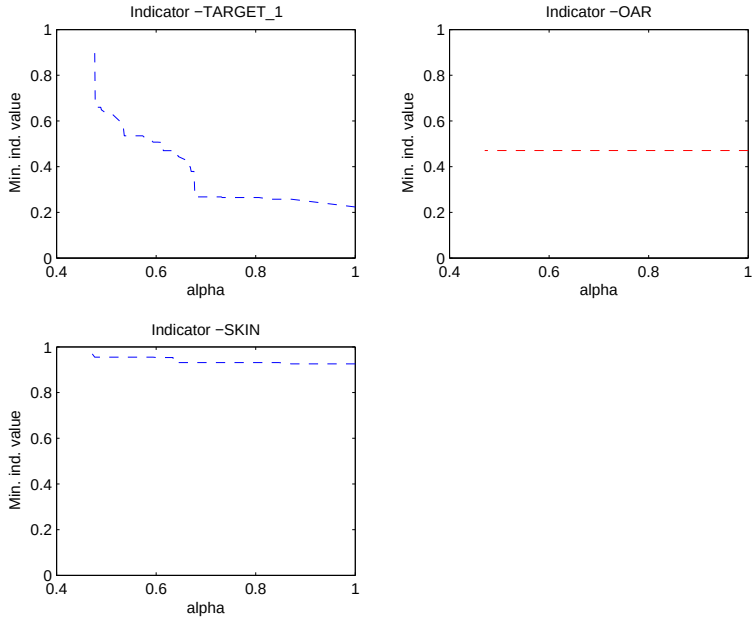

Figure 5.5: For the horseshoe phantom with three objectives the sensitivities for the objectives is shown when the OAR (red) is restricted.

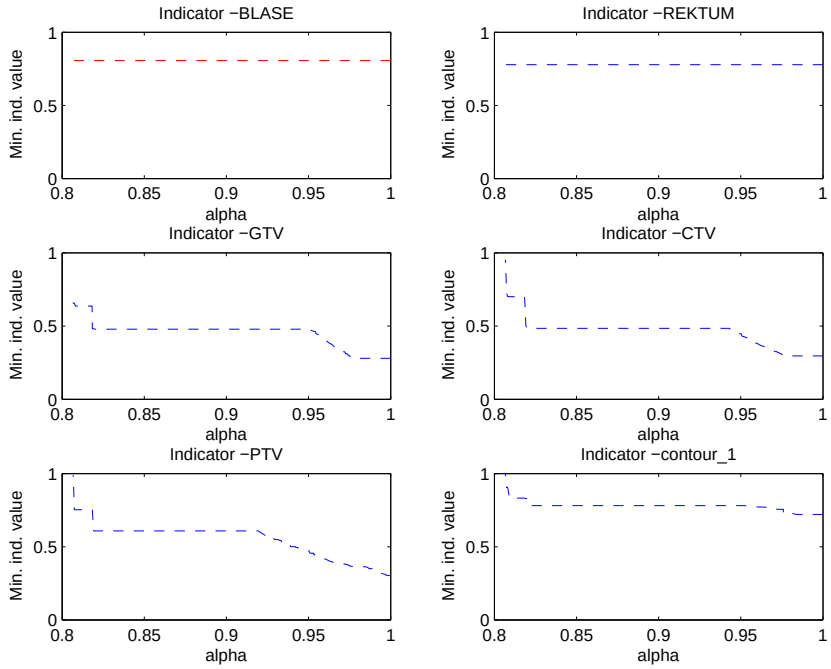

Figure 5.6: The sensitivities of the objectives for a prostate case where the objective values for the bladder are restricted.

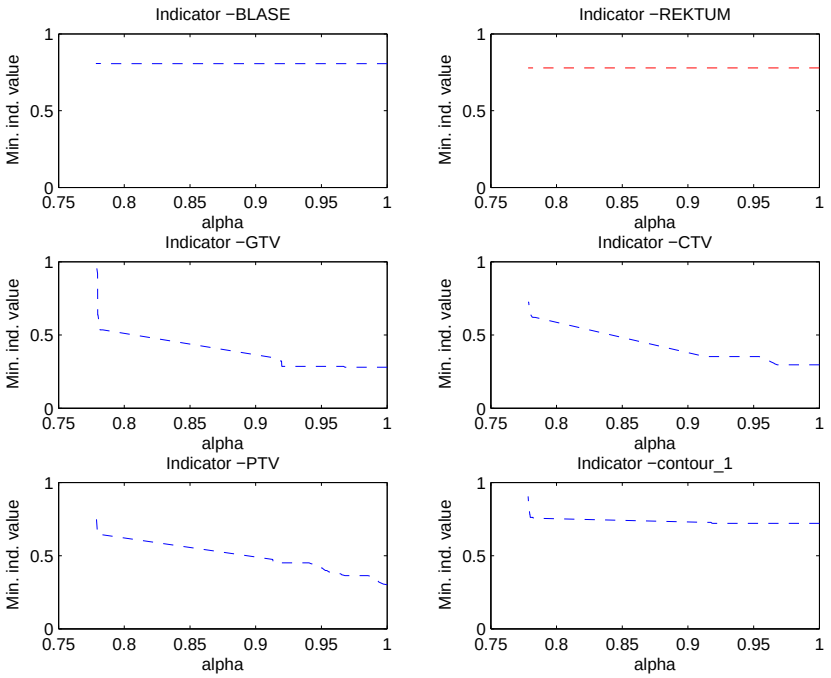

Figure 5.7: Restriction of the objective values of the rectum and the corresponding sensitivities for the other objectives.

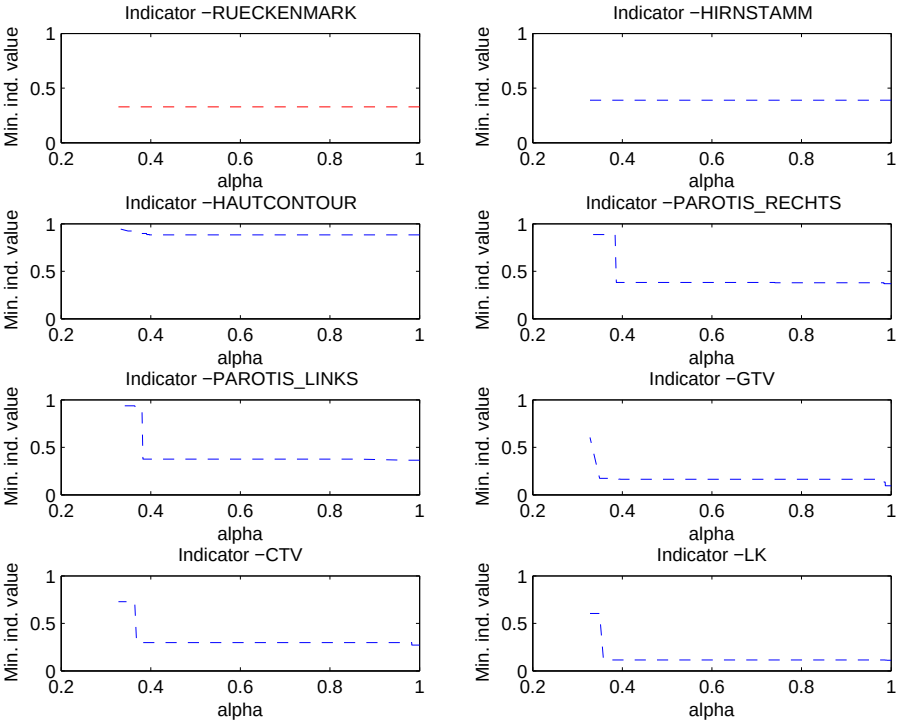

Figure 5.8: A head and neck case with eight independent objectives. The sensitivities of the objectives is shown when the spine (red curve) is restricted.

5.1.4 Discussion

For simple cases we find the obvious relations between the volumes of interest which are also evident from the geometry of the cases (for example whenever an OAR is close to a target structure it will be an obvious "opponent" of the target quality). But with the presented method we are now able to answer more sophisticated questions. E.g. are the rectum and the bladder real opponents in a prostate case (so are we only shifting dose from one side of the target to the other)? Which objective has a stronger effect on the target quality? And "where" in a database is this effect strong?

For the presented (and other) prostate cases we see that the bladder and the rectum are no direct opponents, and the influence of the rectum on the target is larger then for the bladder. However the bladder *becomes* a strong opponent of the target at a particular point of the objective values. So it is possible to identify different "regimes" for each objective. In each regime there can be different correlations between the objectives.

For the eight dimensional head and neck case we find that the curves are much more flat over a large extend of the database. Note that the plots do not necessarily represent the change in another objectives when the criterion k_0 is improved, because this change is of course dependent on the chosen path on the Pareto front. In fact the plots show what is maximal achievable for each objective when k_0 is improved. Note also that a flat curve does not contradict the convexity of the set $\mathcal{F}$ or the Pareto optimality of the solutions (it would in two dimensions, where the method just reproduces the Pareto front). It only states that the trade-offs between the structures is not imperative.

We see that with an increasing complexity of a case the trade-offs can become more "independent" and further planning options become available.

Two things make the whole concept easy to grasp. First no sophisticated mathematical formalism is involved, second one has to compare only two things at the same time.

5.1.5 Conclusion

We presented a simple procedure to visualize the sensitivity of an efficient set of an arbitrary dimension. The plots for each indicator show what is maximal achievable for each objective when a particular objective k_0 is improved. So we can reveal if a certain trade-off between two structures is really imperative.

The analysis of the cases also shows that for simple cases the objectives can be strongly correlated which restricts the variety of the planning options in contrast to more complex cases where there exists more freedom in the plan design.

5.2 Identifying the Main Trade-Offs in Pareto Optimal Databases

5.2.1 Motivation

The shape of the Pareto surface and therefore the structure of a database is not known *a priori* and depends on the geometry of the clinical case and the formulation of the optimization problem. In the course of a better understanding and an accurate characterization of multiobjective databases there are several arising questions:

- What are the main underlying trade-offs?
- What is the complexity for a given database and how we can quantify it?
- What is the variety of planning options available?

We will address these questions by studying the efficient set in the space of IMRT "beamlets", which is a high-dimensional vector space. We are interested to see if we can "boil down" the results by use of dimensionality reduction methods to gain a better understanding of the underlying tradeoffs. Two different methods for dimensionality reduction will be investigated: principal components analysis (linear) and the (nonlinear) isomap method (Tenenbaum et al., 2000). Both methods will reveal insight into the topology of the efficient set and will increase the understanding of the underlying trade-offs present in the database of a specific case.

5.2.2 Materials and Methods

Terminology

We consider an IMRT fluence map optimization with fixed beam angles. The optimization problem is formulated as a convex MOP, meaning the feasible set of the beamlets is convex and all objectives for the different anatomical structures are convex functions. For OARs we use the max and mean model (Thieke et al., 2002), and for targets we use linear ramp functions (Craft et al., 2007). One treatment plan is defined by the values of all beamlets. This plan can be interpreted as a point in high dimensional beamlet space (in which the dimension equals the number of beamlets). So a plan will be often referred to as a point and the set of plans in the database as the data.

Principal Components Analysis

Principal components analysis (PCA) is a well known method for multidimensional data reduction. It finds an ordered set of orthogonal directions along which the data primary lies. Each direction is a one dimensional subspace and defines one principal component (PC). The dimensionality reduction is achieved by considering only a small number of subspaces. The PCs are the eigenvectors of the covariance matrix of the data and are sorted with respect to the corresponding decreasing eigenvalues. The first PC (the one with the largest eigenvalue) offers the largest variance of the data points when projected on this subspace. In other words, the data primarily lie along this direction. The computation of the PCs was done with MATLAB using the function `svd` to compute the singular value decomposition. PCA belongs to the class of linear spectral dimensionality reduction methods, where it is assumed the data lie on a linear manifold.

Isomap Method

In general the efficient set does not necessary lie on a linear manifold but instead is a curved structure. The feasible set in beamlet space is convex and the efficient set will - depending on the formulation of the optimization problem- lie on boarder of this convex set. This motivated us to investigate a method capable of describing nonlinear low dimensional structures embedded in a higher dimensional space, such as the isomap method. Figure 5.9 shows an example of a curved set, where PCA fails to detect the true dimensionality but the isomap method is successful.

Isomap allows for an estimate of the manifold dimensionality (see 5.2.2) and for low dimensional representation of the data. It is able to detect nonlinear characteristics in the data. As input N datapoints in n dimensional space are used, which represent a sampling of the underlying manifold.

The isomap algorithm consists of four parts (for a detailed description see Tenenbaum et al. (2000)). First a distance matrix Δ is calculated from the N points. Each entry Δ_{ij} is the Euclidean distance between points i and j. Second, a number k of nearest neighbours is chosen and a weighted graph G is constructed using the underlying data as nodes. For each node, k edges to the k nearest neighbors (distance between nodes is given by Δ_{ij}) are created and these edges are weighted by the Euclidean distances. Third, geodesic distances between every pair of points is estimated by solving a shortest path problem between the two points in the graph G. These geodesic distances take the geometry of the manifold into account and incorporate the nonlinear shape of the manifold. They approximate the distance between the points *on* the manifold. Last, a simple metric multi dimensional scaling

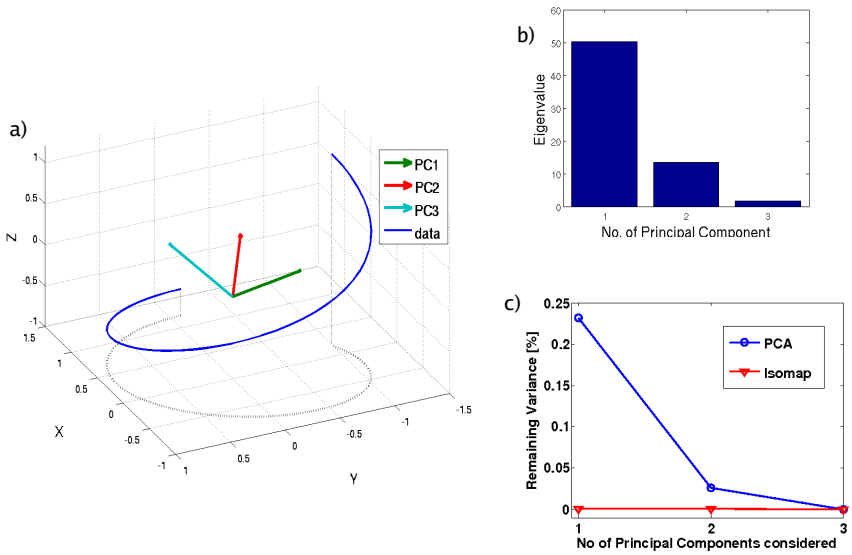

Figure 5.9: Example of a one dimensional curve embedded in three dimensions and the directions of the PCs (a). The PCA yields three nonzero eigenvalues (b) and is not able to discover the true dimensionality due to the non-linear structure. The isomap method unambiguously reveals the true dimension (c).

is performed with the geodesic distances as inputs. The multidimensional scaling finds, for a set of pairwise similarities between objects, a spatial low dimensional configuration of points, whose distances best resemble these similarities. As a result we get a $p < n$ dimensional approximation of the data, which is referred to as a p-dimensional embedding. For $p \leq 3$ the output can be visualized by plotting the isomap coordinate values of each point. The MATLAB implementation of isomap provided by Tenenbaum was used for the calculations. The isomap method is a graph based dimensionality reduction method.

A standard way of evaluating the quality of a low dimensional approximation is to plot the remaining variance, which is calculated with the help of eigenvalues Haerdle and Simar (2003) and has a range from zero to one. If all of the data can be explained with a small number of components then the remaining variance will be close to zero and additional components do not add any significant information.

The plots of the remaining variance allow for an estimate of the manifold dimension which we will call the effective dimension. We have to determine the point in the curve where it stops decreasing. At this point additional dimensions do not add much explanatory power and this number gives the estimate of the effective dimension. The more pronounced the kink is, the better is the estimate. Only one curve with a specific nearest neighbor number k is used for the estimate. A typical set of curves for different values of k is shown in Figure

5.23, where one sees that for a broad range of values for k, the effective dimension is stable.

It should be noted that the dimension of the PCA and the isomap method relate to two different description concepts, in analogy to Euclidean and Riemann geometry. The dimension for a PC approximation describes the dimension of the subspace formed by the considered PCs, while the effective dimension in the isomap method is a better representation of the number of degrees of freedom needed to describe the data. As an analogy one can think of the generalized coordinates in Lagrangian mechanics.

Phantoms and Clinical Cases

We demonstrate the techniques on a two dimensional horseshoe phantom with multiple OARs (see Figure 5.10) and two clinical cases. All phantom cases have five equidistant beams with 16 bixels per beam and a total number of 2300 voxels. For all OARs the mean dose is chosen as the respective objective while the target and the unclassified tissue are modeled with linear ramp functions (for the target the ramp function penalizes dose under the prescription value, and for the unclassified tissue the ramp function penalizes doses above a certain level). The prescribed dose for the target is 70 Gy. Hard constraints of 60 Gy minimum dose are set for the target, as well as 50 Gy maximum dose constraints for all OARs and 70 Gy for the unclassified tissue.

Two clinical examples, a brain and a pancreas case, are also considered. The prescribed target doses are 59.4 Gy for the brain case and 50.4 Gy for the pancreas case. Hard constraints for the minimum target doses are set to 55 Gy and 45 Gy, respectively. In the optimization of the brain case five independent objectives are considered: target under- and overdosage, brain stem, optical nerve and the chiasm. The pancreas case was optimized with respect to five different objectives: target underdosage, stomach, both kidneys, liver and the unclassified tissue. For all OARs the mean dose was considered in the optimization. All databases are calculated with PGEN (Craft et al., 2006). The phantom databases contain 160 solutions for the first case (one OAR) and 107 solutions for the second (two OARs), and both clinical cases have 60 solutions, which is well over the number of cases required for a close representation of the Pareto surface (Craft and Bortfeld, 2008).

5.2.3 Results

Phantoms

The first phantom case has one OAR (OAR1), so four objectives are optimized independently: the OAR, target under- and overdosage, and the unclassified tissue. A PCA is performed

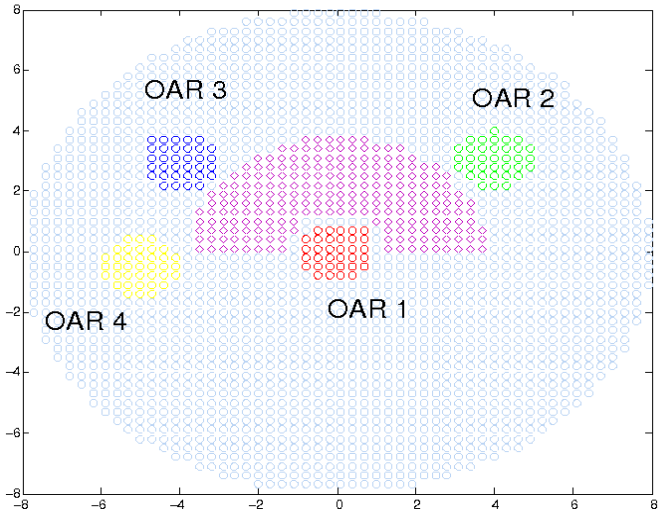

Figure 5.10: Geometry of the phantoms. OAR1 and OAR2 are used for the isomap and PCA analysis, whereas up to four OARs are added to study the effective dimension.

and, to display the trade-offs exposed by the principal tradeoffs, new points are created along the directions of the first and second PC (see Figures 5.11 and 5.12). Along the first PC the main trade-off is between OAR1 and the unclassified tissue, while the target coverage is nearly unchanged. In the direction of the second PC both the OAR and the unclassified tissue are now balanced against the target coverage.

The second phantom has one additional OAR (OAR2), which yields five independent objectives. The PCA shows that both OARs are effected along PC1 and PC2 (see Figure 5.13 and 5.14). In the direction of PC1 the OAR1 increases in the mean dose while OAR2 is spared more and more. Also the target underdosage slightly improves. Along the second PC both OARs show a decreasing mean dose while the unclassified tissue dose increases. It is also possible to decouple the two OARs by combining both PCs.

The phantom with one OAR is analyzed by the isomap method. An estimate of the dimensionality is obtained by looking for the "elbow" in the graph of the remaining variance. Clearly, a two dimensional structure is suggested (Figure 5.15).

Next the isomap method is applied to the phantom with two OARs. For this phantom the isomap method also yields an effective dimension of two. The corresponding 2d embedding in Figure 5.16 shows that all inherent trade-offs can still be captured by considering only two degrees of freedom. Clearly two main trade-off directions can be identified. They can be interpreted as the two diagonals of the plot. In the upper left corner the best solutions for OAR2 are located. In opposition to this in the lower right are solutions with a good unclassified tissue sparing and no overdosage in the target. The other diagonal consists of

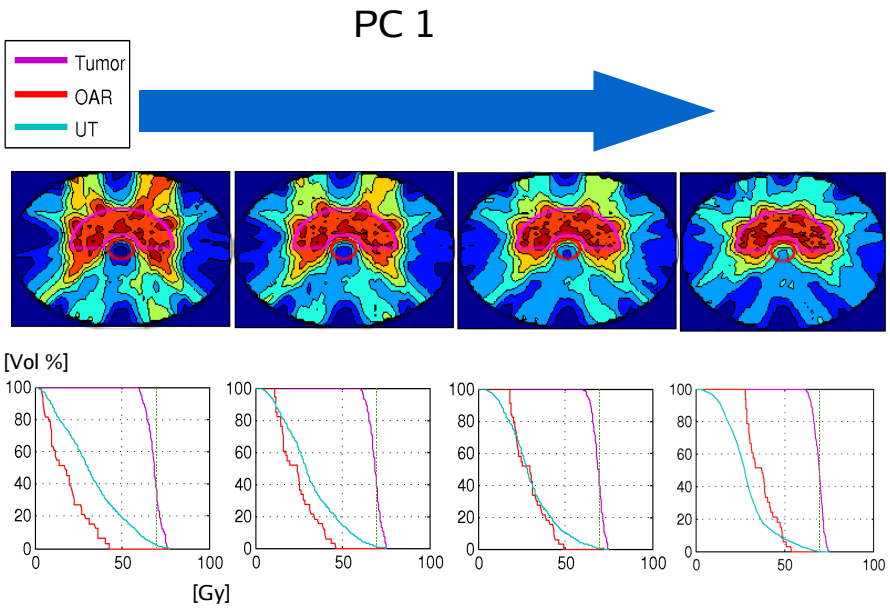

Figure 5.11: Solutions along the first PC for the phantom with one OAR. The prescribed dose of 70 Gy is marked.

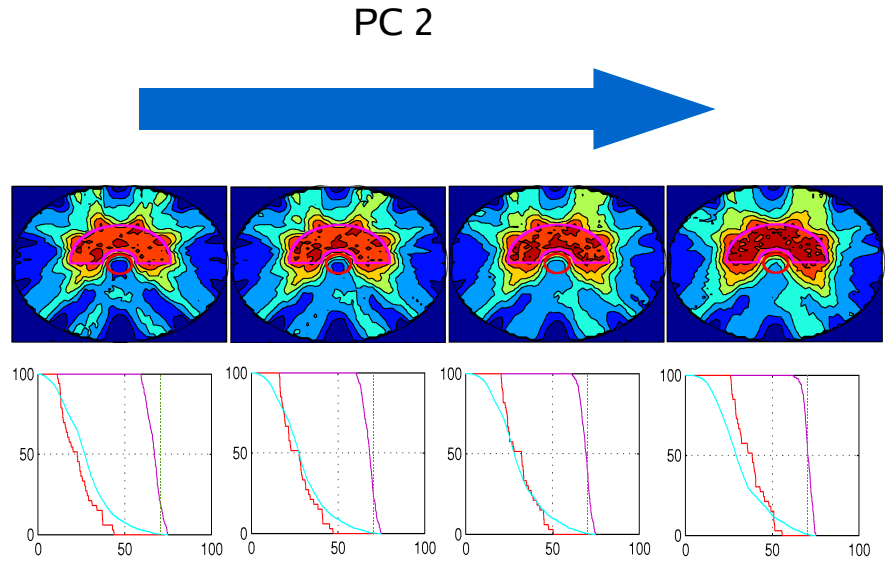

Figure 5.12: Solutions along the second PC for the phantom with one OAR.

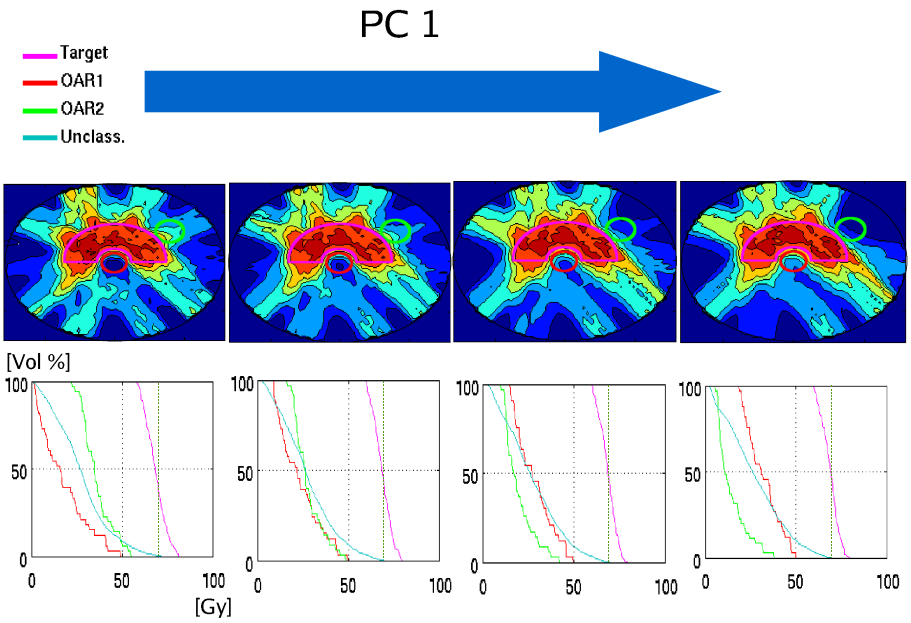

Figure 5.13: Solutions along the first PC for the phantom with two OARs.

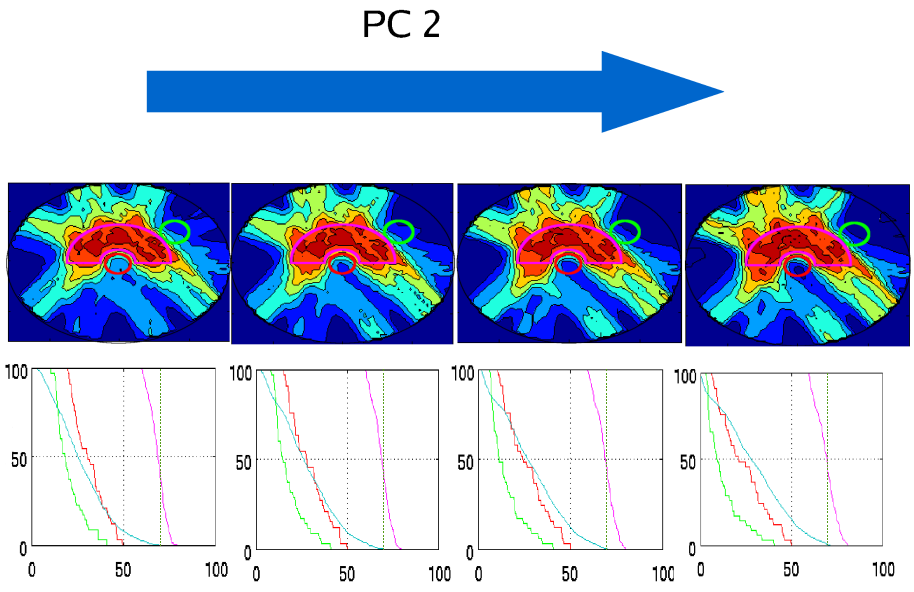

Figure 5.14: Solutions along the second PC for the phantom with two OARs.

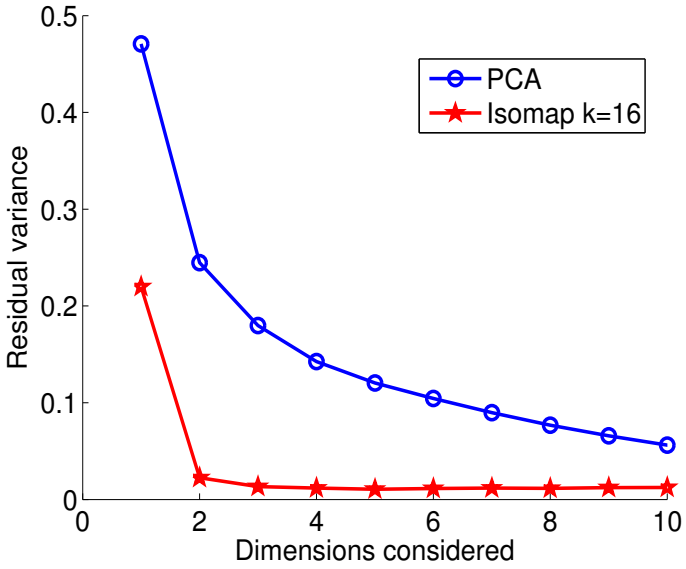

Figure 5.15: The remaining variance for the PCA and the isomap method is shown for the phantom with one OAR.

the extremes of the best solutions for the OAR1 (upper right) and solutions with no target underdosage. Alternatively we can interpret the vertical direction as OAR sparing vs target quality and the horizontal direction as OAR1 vs OAR2.

Brain Case

The brain case is optimized with respect to five objectives: target under- and overdosage, brain stem, chiasm and the optical nerve. A PCA is performed. PC1 shows the main trade-off between the optic nerve and the brain stem while the target quality remains almost constant. The second PC balances the overdosage in the target against the optic nerve and the chiasm.

The brain case shows an effective dimension of two (see Figure 5.19), thus only two degrees of freedom are needed to describe the data.

Pancreas Case

For the pancreas case five objectives are considered including target underdosage, stomach, liver, both kidneys and the unclassified tissue. A PCA shows that the dominating trade-off is between the liver together with the stomach and the two kidneys together with a slight target underdosage. In the second PC the kidneys are not coupled anymore. The right

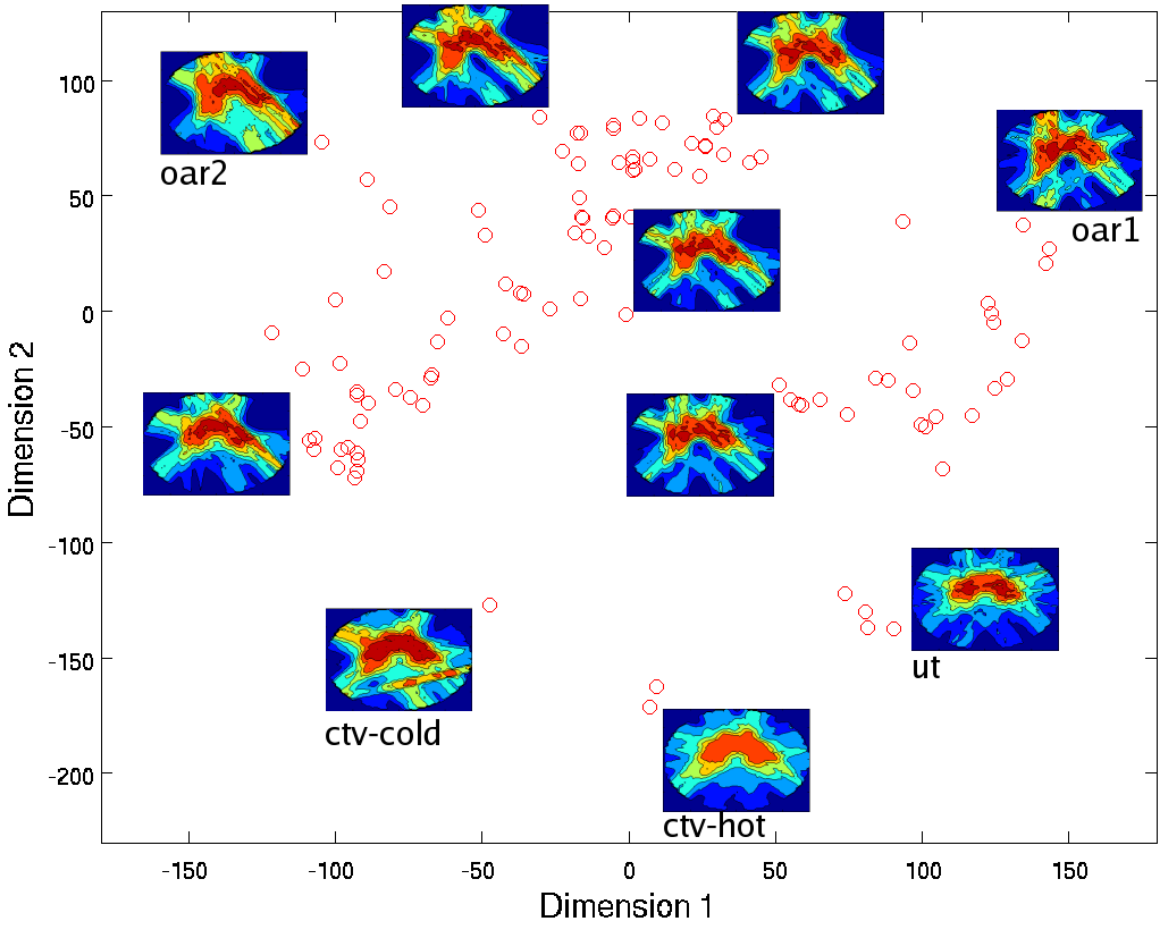

Figure 5.16: Two dimensional embedding of all solutions of the phantom case with two OARs. The anchor point solutions for both OARs, the target under- and overdosage (ctv-cold, ctv-hot) and the unclassified tissue (ut) are labeled.

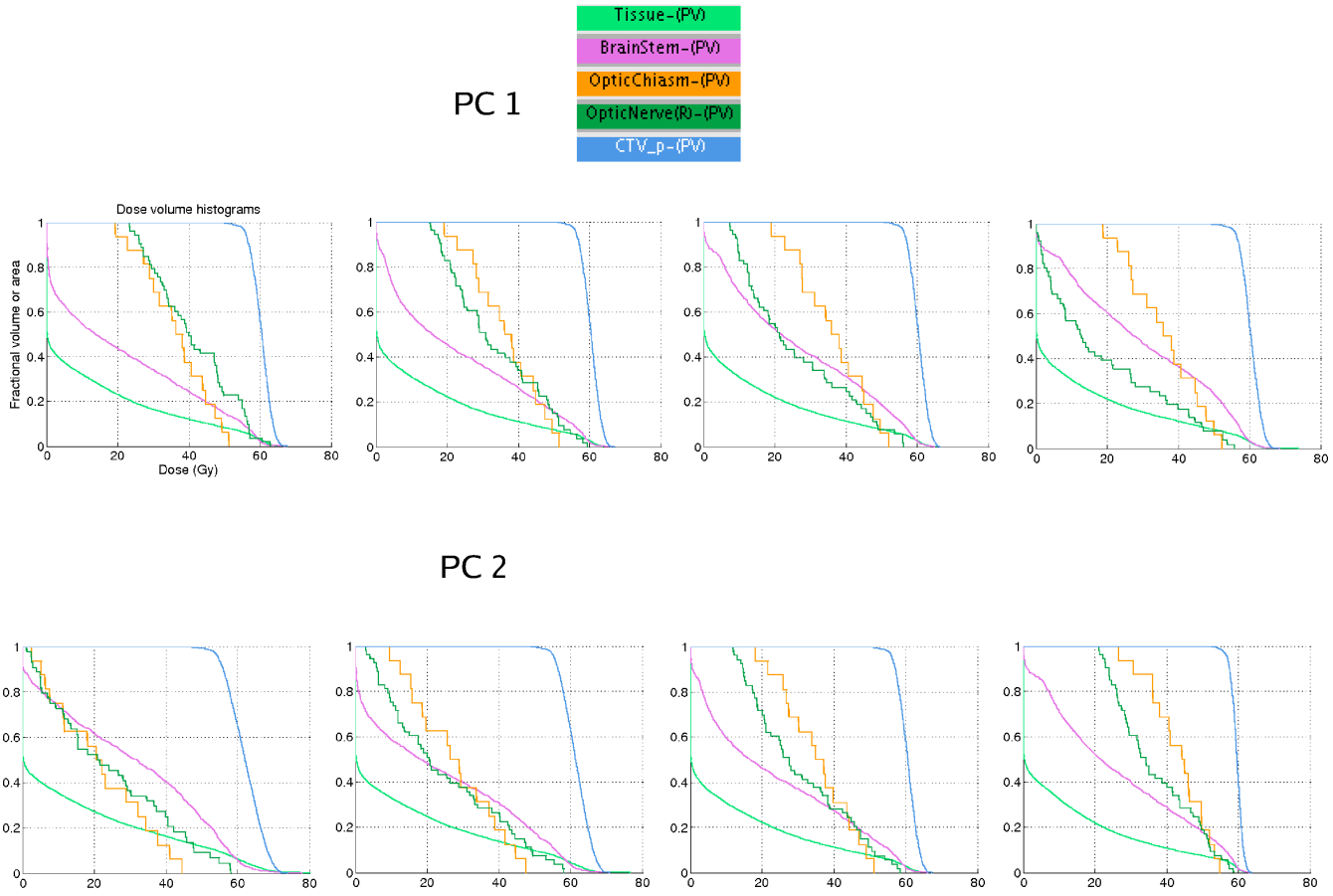

Figure 5.17: DVHs of the brain case for solutions along the first and the second PC.

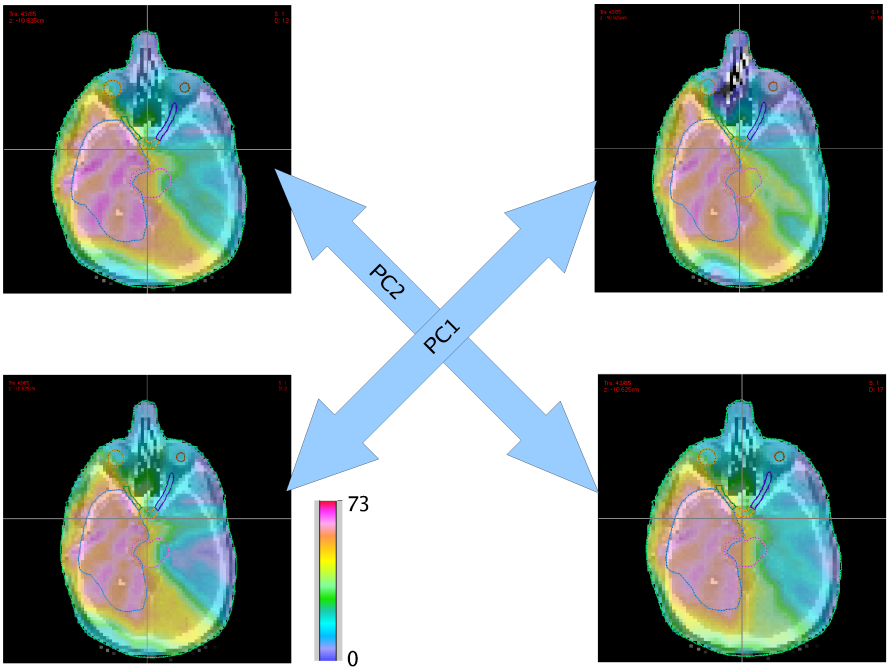

Figure 5.18: Dose distributions of the brain case along the first PC (lower left to upper right) and the second PC (upper left to lower right).

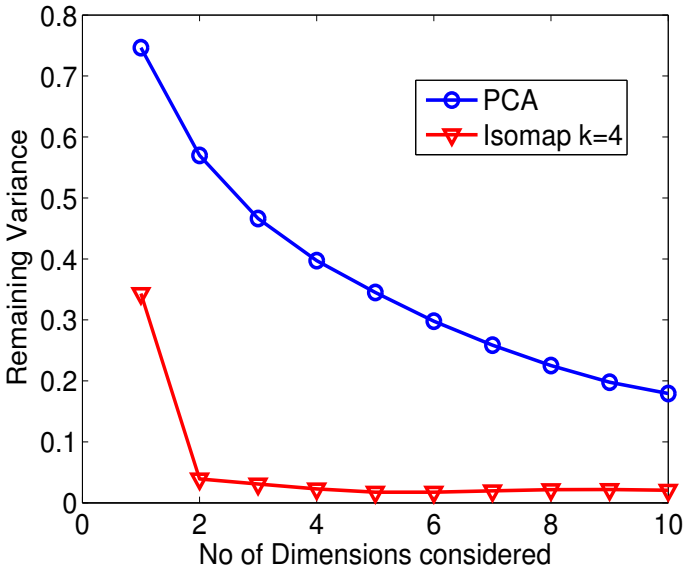

Figure 5.19: The remaining variance of the brain case for the PCA and the isomap method.

kidney and the stomach are balanced against the left kidney and the liver. Variations in the third PC are less pronounced. The most prominent trade-off is observed between the left kidney and the liver together with the right kidney.

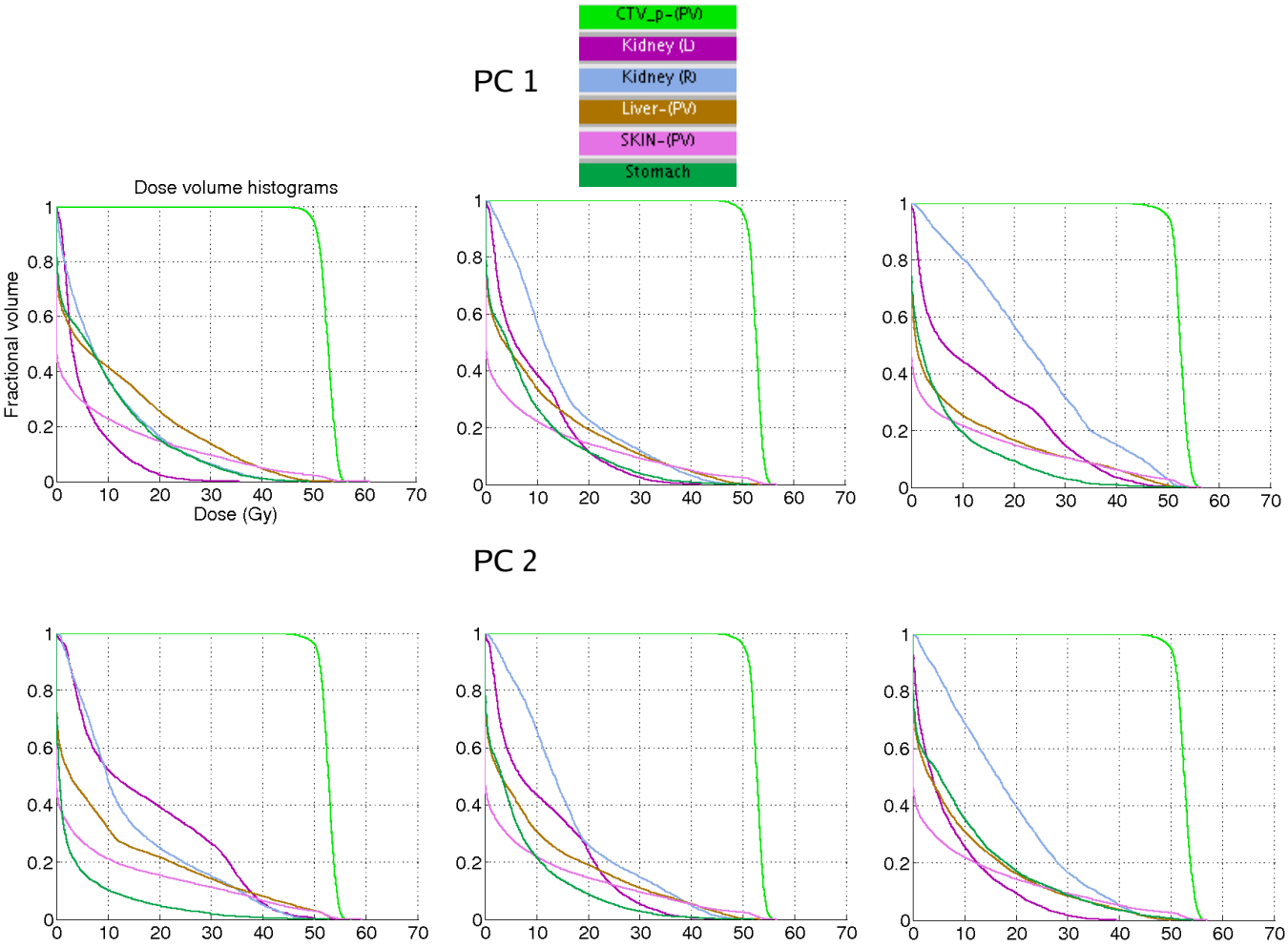

Figure 5.20: DVHs of the pancreas case for the first and second PC.

The analysis of the pancreas case results in an effective dimension of three (cf. Figure 5.23). So two dimension are not sufficient to describe the database, but rather three dimensions are needed for this case. Nevertheless we can get an insight by looking at the 2d-embedding shown in Figure 5.24. Dimension one distinguishes between vertical (right) and horizontal (left) dose entry channels. Due to the geometry of the organs the corresponding tradeoff is observed between the liver and the kidneys. The second dimension improves the stomach when moving downwards, either by distributing dose horizontally (left side) or vertically (right side). So good stomach solutions can involve good kidney sparing or good liver sparing.

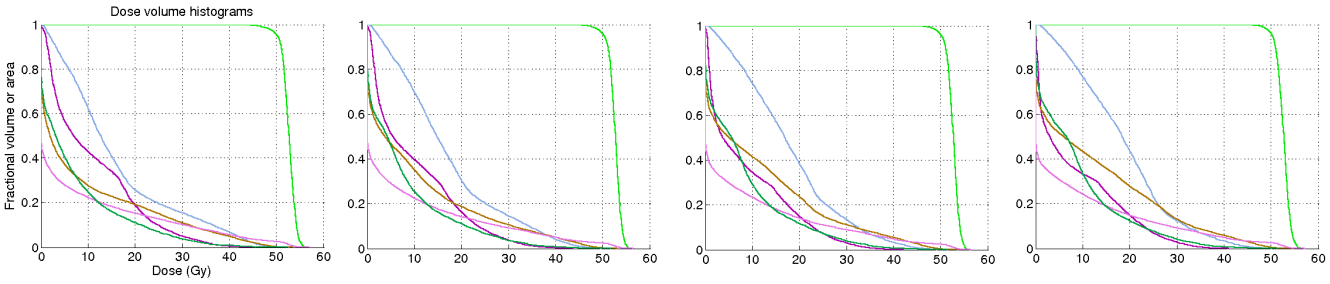

Figure 5.21: The DVHs for the third PC.

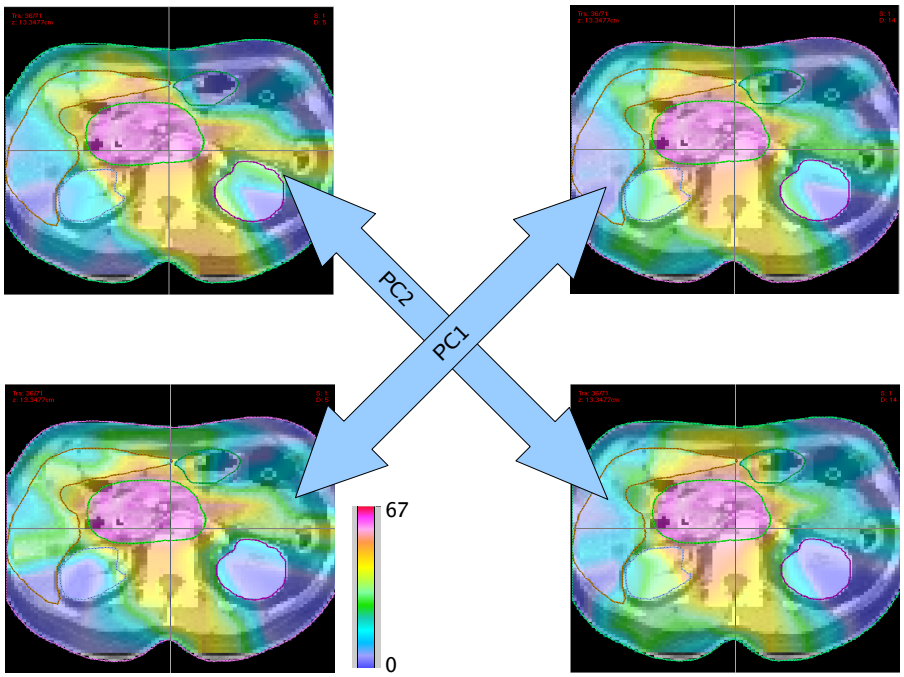

Figure 5.22: Dose distributions of the pancreas case along the first PC (lower left to upper right) and the second PC (upper left to lower right).

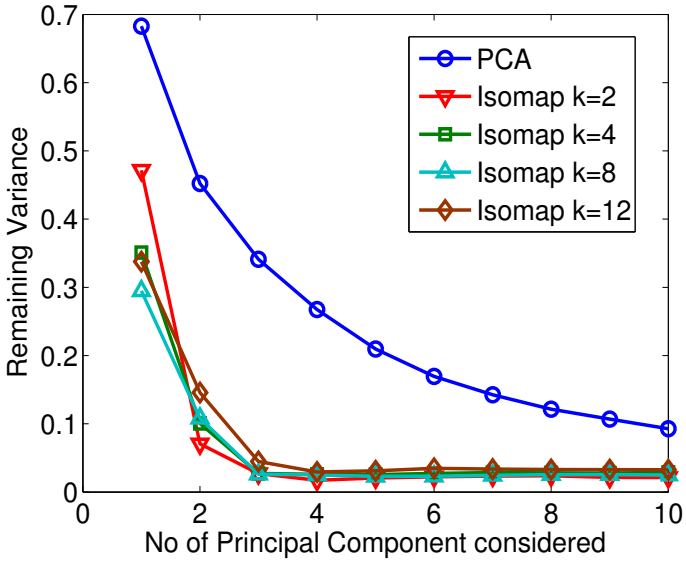

Figure 5.23: The remaining variance of the pancreas case for the PCA and the isomap method.

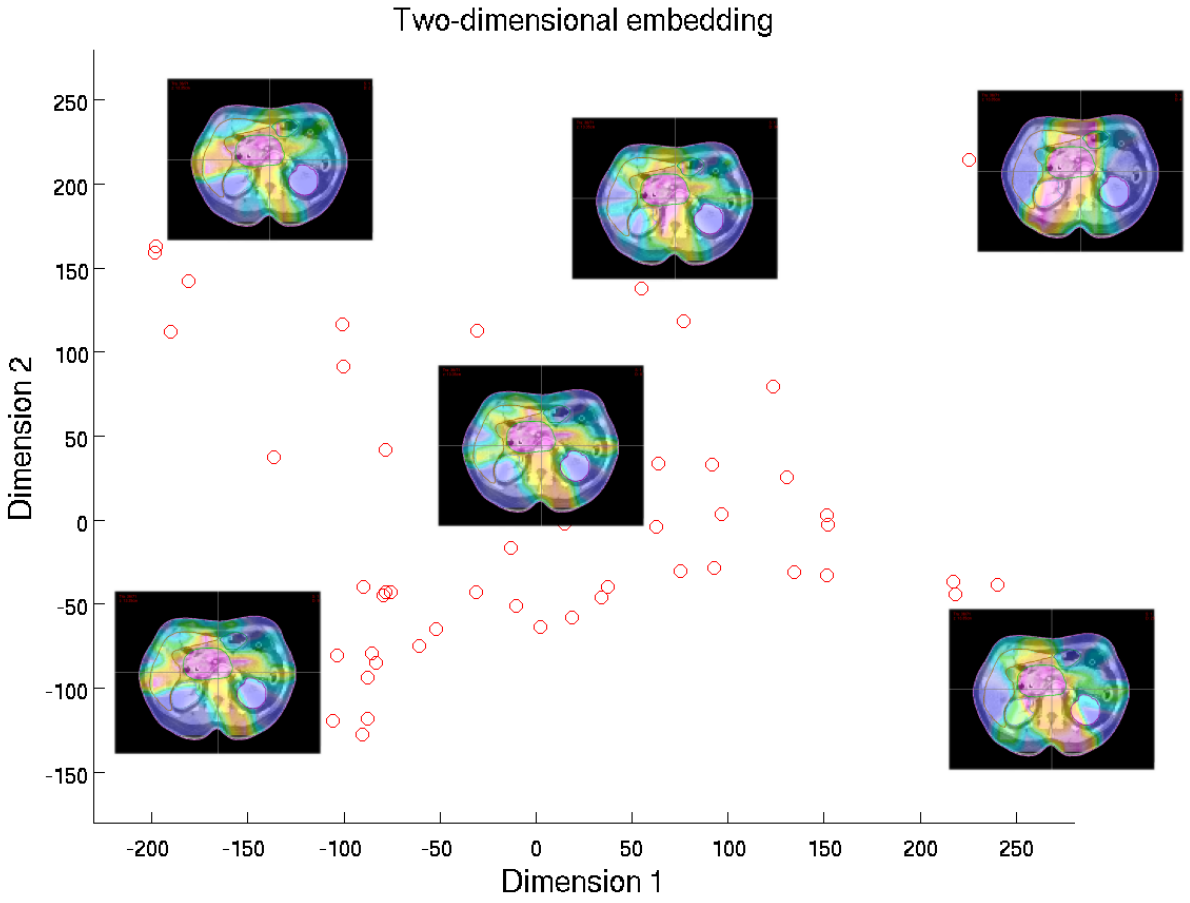

Figure 5.24: The two dimensional embedding of the pancreas case.

Effective Dimension

Can we understand why the cases differ in the effective dimension, or even predict it? The four phantoms containing between one and four OARs are considered. We study how the effective dimension changes when new objectives are added. As shown in table 5.1 the effective dimension can increase with additional OARs. The corresponding Spearman rank-correlation matrices are shown in figure 5.25.

Table 5.1: Effective dimension and corresponding indicator values for all cases. Another case (p7b) is included in which OAR3 and OAR4 are moved towards OAR2.

case	objectives	negative entries	neg. entries / objectives	effective dimension
p4	4	3	0.75	2
p5	5	6	1.2	2
p6	6	8	1.33	3
p7	7	11	1.57	4
p7b	7	9	1.28	3
brain	5	5	1	2
pancreas	5	6	1.2	3

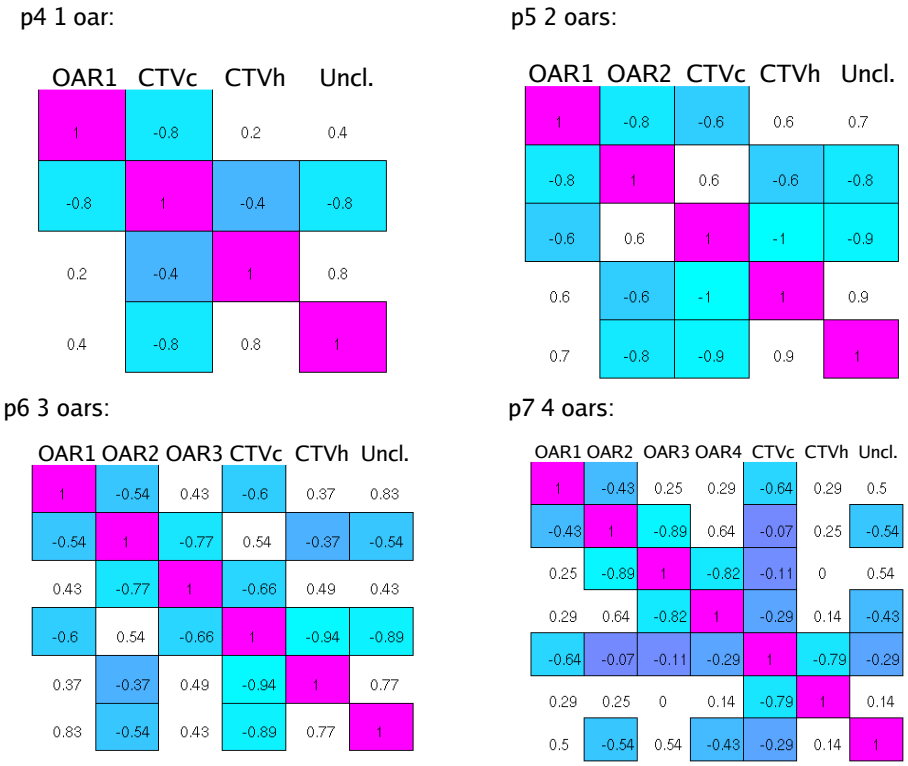

Figure 5.25: Spearman rank correlation coefficients of the phantom cases.

A simple way to approach the effective dimension is to look at Spearman correlation matrices

in the objective space. The complexity of a database should also be reflected in the entries of the correlation matrices. If two VOIs have a high positive correlation in their objectives then this decreases the degrees of freedom of a database: these VOIs cannot be traded independently, this compromise is not accessible. Therefore the number of negative entries in the correlation matrix is expected to be related to the effective dimension of the database, which we indeed see in Table 5.1 and in Figure 5.26.

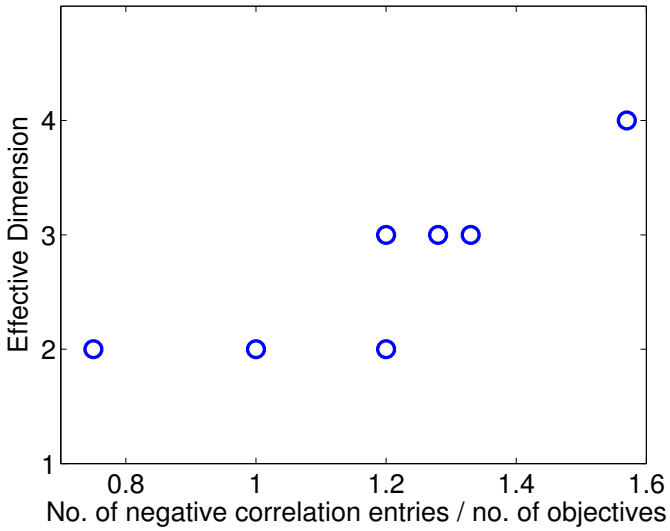

Figure 5.26: The effective dimension as a function of the indicator value.

5.2.4 Discussion

With PCA the main trade-off directions can be identified and relations between VOIs become comprehensible. Few (2-3) PCs are appropriate to describe the main variations in a database.

The isomap method, better suited for data that inherently exists on a nonlinear manifold, reveals a low dimensional structure of the efficient set in beamlet space for all cases. It should be noted that the effective dimension is only an estimate for the true dimension of the manifold Tenenbaum et al. (2000). This insight into the topology of the efficient set is not possible by looking at the PCA alone, which tends to overestimate the dimensionality for nonlinear data.

Both methods allow for a systematic analysis of a database. The isomap method achieves better estimates on the true dimensionality and can produce meaningful visualizations. The direction of a single isomap dimension is not constant in beamlet space, but will depend on the current value of the isomap projection. With regard to further practical proceedings

the PCA is advantageous. As each PC defines a fixed direction in beamlet space one could e.g. easily navigate along the PCs and thereby explore the whole range of the database. If the nonlinearity of the efficient set is less pronounced the differences in both methods will become smaller.

The effective dimension can change with added objectives. An increase in dimension stands for an increase in complexity. If the additional VOI does not change the efficient set much the effective dimension will stay constant. The presented indicator derived from the correlation matrices should give a reasonable guess for the effective dimension for a wide range of cases.

The pancreas case has a higher effective dimension than the brain case and the two phantom cases. This can be understood by geometry of the organs. If the organs are large and separated from each other the dose distributions can largely differ in shape. Hence the dimension of the efficient set in beamlet space is likely to be higher. Similarly, we see a decrease in the effective dimension in the phantom case p7b in Table 5.1 when we move the critical structures closer together.

The low effective dimension revealed by the isomap method also supports the findings that the number of plans needed for an accurate covering of the Pareto front is small and does not necessary increase exponentially Craft and Bortfeld (2008). The correlation between different VOIs and the imposed constraints in the optimization problem restricts the feasible set and therefore also the efficient set.

The answers to our questions from the introduction are as follows:

- Main trade-offs: PCA allows to identify the main trade-offs in a database. The distribution of the data is captured by the main PCs and each PC exposes one key trade-off. The 2d visualization of the database provided by the isomap method enables one to see the compromises and understand the relation between different VOIs.
- Complexity: The effective dimension can serve as a measure of complexity. If it is small, e.g. two, then it is possible to visualize the whole database, by laying out the plans in a 2D figure, see e.g. Figure 5.16. If the effective dimension stays unchanged after adding an objective (or other actions that potentially increase complexity) this indicates that the additional plans can still be understood with the same number of degrees of freedom. The newly added plans can be inserted in the existing embedding of plans and a smooth transition between new and old plans is possible. If on the other hand the effective dimension increases after adding a new OAR this means that the only way to achieve an satisfactory embedding is to extend the dimension and move some points in this direction.

- Variety: The variety of available planning options can be judged by considering the database visualizations and by looking at the effective dimension. The 2d visualization allows one to grasp the whole range of the database and also the transitions between the different plans. The higher the effective dimension the more degrees of freedom are contained in the database. The different planning options are also reflected in the first PCs. Their effect on the dose distribution reveals the extent of the database and thus the main planning options. Although local fine tunings are certainly possible the first PCs describe the primary characteristics of the data for a specific case.

5.2.5 Conclusion

We presented two methods of dimensionality reduction for the analysis of multiobjective databases. They provide the basis to extract the main trade-offs from a database. For each individual case opposing structures and the available planning options contained in a database can be revealed. It is possible to visualize the efficient set in a meaningful way even for high dimensional databases with many objectives (providing literally an insight into the Pareto surface). The gained information can be helpful in the planning process and improve understanding of a case.

6 Summary and Outlook

Summary

In this thesis we have presented and evaluated a new prototypical treatment planning system which uses a multiobjective approach. We further on developed methods that allow to systematically extract information from an efficient set, increasing the understanding of a Pareto optimal database as a whole.

We evaluated different modelling frameworks and studied the influence of the corresponding modelling parameters on the resulting solutions. Furthermore we introduced a quantitative measure which allows to determine the change of the efficient set due to a change in the model framework, so different efficient sets become comparable.

To determine the quality of a finite approximation of a high dimensional Pareto surface we developed a set of indicators. These indicators yield additional information on the approximation quality and can assist the quality assurance for the MIRA system. The indicators are of general nature and are not restricted to the models presented here but can be applied to characterize any efficient set.

In retrospective planning studies we examined the practical feasibility of the new multiobjective approach. We could demonstrate that the system is able to produce clinically meaningful plans. For a first clinical application a certified quality assurance has to be carried out. Furthermore, the normalization problem has to be dealt with.

In the last part we have introduced several methods to obtain a better understanding of the Pareto optimal databases and the underlying cases. These methods use successive hyperplane restrictions of the efficient set and linear as well as non-linear dimensionality reduction techniques. The revealed low dimensional structure of the efficient set in the beamlet-space allows for meaningful visualizations of high dimensional Pareto fronts and an insight into the main trade-offs of a case.

As a consequence of this thesis the developer as well as the practial planner will benefit from the achievement of this work. The developing side can directly use the presented methods to assist the quality assurance in the further development process. The planner is provided

with additional systematic information on the case that supports him in choosing the best suited treatment plan for the patient. Alltogheter we are confident that a multiobjective treatment planning environment will soon be established in clinical practice to allow for an efficent and flexible decision support in inverse radiotherapy planning.

Outlook

For future developments there are still a number of open questions. The problem of the beam angle optimization is at present separated from the subsequent planning process. The efficient set of the whole planning problem though, would include plans with different beam angle setups. Besides the computational challenges in generating the databases there have to be found ways how to practically navigate these databases which are not convex anymore and the navigation conceptually relies on this property. It remains to be seen if the potential non-convex effects play a role in a practical context.
To extend the control of the dose distribution in a particular regions of a structure one would have to incorporate local objectives or one could modify the search of the databases to navigate DVH-points instead of values of a whole objective.

It is possible to relate the multiobjective framework with some other current developments in the field of radiotherapy planning.
The inclusion of several biological models is already implicitly achieved as the efficient sets are equivalent if the biological model is expressible as a monotone function of the gEUD. Many other biological models can potentially be included via the convex reformulation of objectives.
Some research has focused on the inclusion of uncertainties in treatment planning leading to robust optimization. As the resulting formulations are also convex optimization problems, the extension to a robust multiobjective framework is also possible. The databases could either contain robust plans or the robustness could even be included as an independent controllable meta objective.
A challenging topic might be the connection with adaptive planning where the treatment plan is adapted on the basis of new information like additional imaging information before a treatment. If effective ways can be found to adapt a whole plan database due to the new gained information, using (some of) the previous knowledge on the efficient set, has to be seen.

7 Acknowledgment

Einer Dissertation geht immer eine lange Lebensphase des Lernens voraus. Obwohl nicht viel Physik in dieser Arbeit durchscheint, so empfinde ich sie doch als mein eigentliches wissenschaftliches zu Hause. Ich bin dankbar die Möglichkeit gehabt zu haben, dieses faszinierende und grundlegende Feld zu studieren.
Viel des Gelernten erfährt erst mit der Zeit die eigene Wertschätzung und gelangt in eine gedankliche Form, die eine Erfassung der Dinge in einem größeren Zusammenhang ermöglicht (und ich glaube, dieser Prozess ist - zum Glück - noch nicht am Ende). Ich möchte all denen danken, die darauf Bedacht waren, meinen physikalischen und kritischen Verstand zu schärfen und mich auf meinem wissenschaftlichen Weg bis hier her begleitet haben. Dies gilt in gleichem Maße für Lehrende und Mentoren wie auch für Weggefährten und Freunde.

Im Bezug auf diese Arbeit möchte ich an erster Stelle Uwe Oelfke sehr herzlich danken, für die Betreuung dieser Arbeit und das Vertrauen, welches er mir stets entgegengebracht hat. Er hat mich immer in meinem Tun unterstützt, sich für mein fachliches Weiterkommen eingesetzt und mich durch seine Betonung der physikalischen Hintergründe für das Gebiet der Medizinischen Physik begeistern können. Herrn Wolfgang Schlegel danke ich für die Bereitschaft, die Arbeit als Zweit-Gutachter zu bewerten und für seine angenehme Leitung der Medizinischen Physik Abteilung des DKFZs.

Desweiteren bin ich all denen zu Dank verpflichtet, die sich ebenfalls für das Gelingen dieser Arbeit eingesetzt haben. Dazu gehört Christian Thieke, der sich in vielen Gesprächen stets motivierend und mit hilfreichen Ratschlägen einbracht hat. Herzlich möchte ich mich auch bei den Kooperationspartnern vom ITWM in Kaiserslautern Alexander Scherrer, Michael Monz, Phil Süss und Karl-Heinz Küfer bedanken, die durch Ihre Entwicklungsarbeit dieses Projekt erst ermöglichten, sich immer mit großer Hilfsbereitschaft und Unterstützung einbrachten und mir viele Geheimnisse der multikriteriellen Optimierung vermitteln konnten.

Eine sehr schöne und anregende Zeit in Boston konnte ich in der Gruppe von Thomas Bortfeld verbringen. Ihm gilt mein aufrichtiger Dank für seine weitsichtige Betreuung, das Teilen vieler interessanter Ideen und seine Großzügigkeit, mit der er mir stets begegnet ist. Eine hervorragende und bereichernde Zusammenarbeit verdanke ich David Craft, aus der viele Ergebnisse und neue Ideen entstanden.

Ich habe die Arbeitsbedingugen und die Atmosphäre am DKFZ immer als sehr gut und kollegial erlebt. Bei meinen Zimmerkollegen, meiner Arbeitsgruppe und vielen weiteren Kollegen bedanke ich mich herzlich.

Ohne eine Finanzierung ist das Verfassen einer Dissertation kaum möglich. Für die breite finanzielle Unterstützung durch die Deutsche Krebshilfe, den DAAD und das DKFZ möchte ich mich anerkennend bedanken.

Letzlich befinden sich die tragenden Säulen immer jenseits des Fachlichen. Ich bin all meinen Freunden dankbar für die gemeinsame Zeit, die vielen heiteren Momente und Ihre ehrliche Freundschaft! US, DPR, AJK, MA, BKK, DWB, AK, UG, KO, ...
Mein abschließender und tief empfundener Dank gilt Euch, lieber Vater, liebe Mutter und liebe S.

Bibliography

D. Bertsekas. *Nonlinear Programming*. Athena Scientific, 2004.

G. Birkhoff. On drawings composed of uniform straight lines. *Journ. de Math.*, 19(3): 221–236, 1940.

T. Bortfeld. Imrt: a review and preview. *Physics in Medicine and Biology*, 51(13):R363–R379, 2006. URL `http://stacks.iop.org/0031-9155/51/R363`.

T. Bortfeld and W. Schlegel. Optimization of beam orientations in radiation therapy: some theoretical considerations. *Physics in Medicine and Biology*, 38:291–304, 1993.

S. Boyd and L. Vandenberghe. *Convex Optimization*. Cambridge University Press, March 2004. ISBN 0521833787.

A. Brahme, J. Roos, and I. Lax. Solution of an integral equation in rotation therapy. *Physics in Medicine and Biology*, 27:1221–1229, 1982.

O. Chernykh. Approximating the pareto hull of a convex set by polyhydral sets. *Computational Mathematics and Mathematical Physics*, 35:1033–1039, 1995.

C. Cotrutz, M. Lahanas, K. Kappas, and D. Baltas. A multiobjective gradient based dose optimization algorithm for external beam conformal radiotherapy. *Physics in Medicine and Biology*, 2001.

D. Craft and T. Bortfeld. How many plans are needed in an imrt multi-objective plan database? *Physics in Medicine and Biology*, 53(11):2785–2796, 2008. URL `http://stacks.iop.org/0031-9155/53/2785`.

D. Craft, T. Halabi, H. Shih, and T. Bortfeld. Approximating convex Pareto surfaces in multi-objective radiotherapy planning. *Medical Physics*, 33(9):3399–3407, 2006.

D. Craft, T. Halabi, H. Shih, and T. Bortfeld. An approach for practical multi-objective IMRT treatment planning. *Int. J. Radiation Oncology Biol. Phys.*, 69(5):127–135, 2007.

I. Das. An improved technique for choosing parameters for pareto surface generation using normal-boundary intersection. *WCSMO-3 proceedings*, 1999.

B. Fruhwirth, R. E. Bukkard, and G. Rote. Approximation of convex curves with application to the bicriterial minimum cost flow problem. *European Journal of Operational Research*, 42(3):326–338, October 1989.

A. Gustafsson, B. Lind, and A. Brahme. A generalized pencil beam algorithm for optimization of radiation therapy. *Medical Physics*, 21(3):343–356, 1994.

W. Haerdle and L. Simar. *Applied multivariate statistical analysis*. Springer, 2003.

T. Halabi, D. Craft, and T. Bortfeld. Dose-volume objectives in multi-criteria optimization. *Physics in Medicine and Biology*, 51(15):3809–3818, 2006. URL http://stacks.iop.org/0031-9155/51/3809.

A. Höss, J. Debus, R. Bendl, R. Engenhart-Cabillic, and W. Schlegel. Computerized procedures in 3-dimensional radiotherapy planning. *Radiologe*, 35:583–6, 1995.

W. S. J. Bille. *Medizinische Strahlenphysik*. Springer, Berlin ; Heidelberg, 2002. ISBN 3-540-65254-X, 978-3-540-65254-0.

M. Krause, A. Scherrer, and C. Thieke. On the role of modeling parameters in imrt plan optimization. *Physics in Medicine and Biology*, 53(18):4907–4926, 2008. URL http://stacks.iop.org/0031-9155/53/4907.

K. Kuefer, M. Monz, A. Scherrer, P. Süss, F. Alonso, A. Sultan, T. Bortfeld, D. Craft, and C. Thieke. Multicriteria optimization in intensity modulated radiotherapy planning. Technical report, Fraunhofer-Institut für Techno- und Wirtschaftsmathematik, Kaiserslautern, 2005.

K. Küfer, H. Hamacher, and T. Bortfeld. A multicriteria optimization approach for inverse radiotherapy planning. In T. Bortfeld and W. Schlegel, editors, *Proceedings of the XIIIth ICCR, Heidelberg 2000*, pages 26–29. Springer, 2000.

K. Küfer, A. Scherrer, M. Monz, F. Alonso, H. Trinkaus, T. Bortfeld, and C. Thieke. Intensity modulated radiotherapy - a large scale multi-criteria programming problem. *OR Spectrum*, 25:223–249, 2003.

K. Küfer, M. Monz, A. Scherrer, P. Süss, F. Alonso, A. Sultan, T. Bortfeld, and C. Thieke. *Multicriteria optimization in intensity modulated radiotherapy planning*. Kluwer, Ed. H. Romeijn, 2006.

A. Makhorin. *GNU Linear Programming Kit, Version 4.36*. GNU Software Foundation, http://www.gnu.org/software/glpk/glpk.html, 2009. URL http://www.gnu.org/software/glpk/glpk.html.

A. Messac, A. Ismail-Yahaya, and C. A. Mattson. The normalized normal constraint method for generating the Pareto frontier. *Journal of the International Society of Structural and Multidisciplinary Optimization (ISSMO)*, 25:86–98, 2003.

K. Miettinen. *Nonlinear Multiobjective Optimization*, volume 12 of *International Series in Operations Research and Management Science*. Kluwer Academic Publishers, Dordrecht, 1999.

M. Monz. *Pareto Navigation - Interactive Multiobjective Optimisation and its Application in Radiotherapy Planning*. PhD thesis, Technische Universität Kaiserslautern, 2006.

M. Monz, K. H. Küfer, T. R. Bortfeld, and C. Thieke. Pareto navigation, algorithmic foundation of interactive multi-criteria imrt planning. *Physics in Medicine and Biology*, 53(4):985–998, 2008. URL http://stacks.iop.org/0031-9155/53/985.

A. Niemierko. Reporting and analyzing dose distributions: a concept of equivalent uniform dose. *Med Phys*, 24(1):103–10, 1997. Journal Article.

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

U. Oelfke and T. Bortfeld. Inverse planning for photon and proton beams. *Med. Dosim.*, 26: 113–124, 2001.

J. R. Palta and T. R. Mackie. *Intensity Modulated Radiation Therapy - The State of the Art.* Medical Physics Publishing, Madison, WI, 2003.

V. Pareto. Cours déconomie politique. *Rouge*, 1896.

H. Romeijn, J. Dempsey, and J. Li. A unifying framework for multi-criteria fluence map optimization models. *Physics in Medicine and Biology*, 49:1991–2013, 2004.

S. Ruzika and M. Wiecek. A survey of approximation methods in multiobjective programming. Technical report, Department of Mathematics, Univeristy of Kaiserslautern, 2003.

S. Sayin. Measuring the quality of discrete representations of efficient sets in multiple objective mathematical programming. *Mathematical Programming*, 87(3):543–560, 2000.

A. Scherrer, K.-H. Küfer, T. Bortfeld, M. Monz, and F. Alonso. Imrt planning on adaptive volume structures – a decisive reduction in computational complexity. *Physics in Medicine and Biology*, 50(9):2033–2053, 2005.

D. Sommerville. *An Introduction to the Geometry of N Dimensions.* Dover, New York, 1958.

J. B. Tenenbaum, V. d. Silva, and J. C. Langford. A Global Geometric Framework for Nonlinear Dimensionality Reduction. *Science*, 290(5500):2319–2323, 2000. doi: 10.1126/science.290.5500.2319.

K. Tesarczyk. Entwurf und realisierung eines plugins für das bestrahlungsplanungssystem virtuos zur navigation in einer datenbank mit multikriteriell optimierten bestrahlungsplänen. Master's thesis, University Heidelberg / FH Heilbronn, 2006.

C. Thieke, T. Bortfeld, and K. Kuefer. Characterization of dose distributions through the max and mean dose concept. *Acta Oncol.*, 41:158–161, 2002.

N. S. U. Oelfke, S. Nill. Inverse treatment planning based on eud: Analyzing first order biological effects. In *Proceedings 44th AAPM Annual Meeting*, 2002.

S. Webb. *Intensity-modulated radiation therapy.* CRC Press, 2001.

Y. Yu. Multiobjective decision theory for computational optimization in radiation therapy. *Med Phys*, 24(9):1445–54, 1997.

M. Zeleny. *Multiple Criteria Decision Making.* McGraw Hill, New York, 1982.

List of Figures

2.1 Efficient solution (green) on the Pareto front (blue) for a bicriterial problem. The solution in red is dominated by the green one, because it lies in the red cone. 6
2.2 The example by Brahme with a rotating source and a centrally blocked beam. Only a mudulated beam profile (dashed lines) leads to a homogeneous dose distribution in the target. Figure from (Brahme et al., 1982). 7
2.3 The tumor (red) is irradiated with high energetic photons from seven directions. The irradiation fields with spatially non-uniform intensity maps are superimposed to achieve the prescribed dose in the tumor and a sufficient sparing of crticial structures. 8
2.4 All feasible solutions in $\mathcal{X}$ from the parameter space (left) are mapped to the design space (middle). The set of efficient solutions is symbolized in red. The solutions in the design space are evaluated by a vector valued function leading to a representation in the objective space. Here the efficient solutions are located at the boundary of the set $\mathcal{F}$ of feasible objective vectors. 9
2.5 Graphical user interface of the navigator. Each objective is represented by a slider (left) and can be manipulated interactively. The corresponding dose distribution and DVH information is updated in real time. 12

3.1 my short caption 16
3.2 The geometries of three cases. A horseshoe phantom case (left) with one OAR and a horseshoe shaped target volume, a paraspinal tumor (paraspinal I) with one boost volume (middle) and a small paraspinal case (paraspinal II) (right). 16
3.3 Dose volume histograms of the horseshoe phantom. The parameters p_T and q_T are the only parameters which are modified. On the left $p_T = 1$ and $q_T = 8$, setting a high priority on the target overdosage. On the right we have $p_T = 8$ and $q_T = 1$. The aspired objective value for the OAR was $\mathrm{gEUD}_8^{asp} = 30\,\mathrm{Gy}$. 17
3.4 DVHs of two solutions from two different databases that only differ in the setting of the parameters p_T and q_T. On the right $p_T = 8$ and $q_T = 1$, setting a high priority on the target underdosage. On the left we have $p_T = 1$ and $q_T = 8$. Here the aspired objective value for the OAR was set to $\mathrm{gEUD}_8^{asp} = 10\,\mathrm{Gy}$. 17
3.5 Influence of the organ parameter p on the efficient solutions. All three solutions are optimized with the pq-model and the organ parameter is varied from $p = 2$ to $p = 8$ ($p = 2, 4, 8$ from left to right). The maximum OAR dose decreases in total by 3.7% ($d_{max}^{OAR} = 79.5, 77.6, 75.8$). 18

3.6 Effect of changing the aspired function value for the OAR. The aspired function value for the OAR is varied between $\mathrm{gEUD}_8^{asp} = 10\,\mathrm{Gy}$ (left) 30 Gy (middle) and 40 Gy (right). The maximum OAR dose changes by more than 40%. 19
3.7 The horseshoe phantom is optimized with the pt-model. The first three solutions correspond to different values of p_t ($p_t = 2, 4, 8$ from left to right), while in the solution on the right d_{up} was changed from 30 Gy to 10 Gy. 19
3.8 Influence of the parameter p for the spine. Two solutions with $p = 3$ (dashed lines) and two solutions with $p = 8$ (solid lines) are shown. While the maximum dose of the spine changes by 6 Gy on the left hand side there is nearly no difference on the right. 20
3.9 Influence of the parameter p for the stomach. All other OARs are virtually the same. 20
3.10 A paraspinal case is optimized with three different models. One solution from each database is shown, the pq-model (left), the pt-model (middle) and the std-model (right). 21
3.11 Dose distributions for the three optimized with the pq-model (left), the pt-model (middle) and the std-model (right). 21
3.12 Maximal distances for the solutions from $\mathcal{D}_2$ to the convex hull of $\mathcal{D}_1$ (blue) and from the convex hull of $\mathcal{D}_1$ to the solutions from $\mathcal{D}_2$ (red). 22
3.13 Schematic picture of the hypervolume indicator. 27
3.14 Example of two dimensional efficient sets. On the left successive solutions (blue) have the same arc length. The inner approximation (red) is close to the Pareto front (black). On the right hand side the random placement of the same number of solutions can result in large gaps between the approximation and the efficient set. The corresponding indicator values for H_s are $H_s = 1.09$ (left) and $H_s = 0.33$ (right). 29
3.15 For the solutions above the largest hyperspheres are calculated. The gap in the distribution on the right results in a large hypersphere. For the hypersphere indicators we have $H_l = 0.00$ (left) and $H_l = 1.38$ (right). 29
3.16 Example of a two dimensional Pareto front in three dimensions. Hyperspheres for three arbitrary solutions are shown. The even distribution results in very similar spheres (left) with $H_s = 1.15$ and $H_l = 0.03$ while for the stochastically placed solutions the size of the radii can differ ($H_s = 0.65$ and $H_l = 0.53$). . . 30
3.17 The simplex (bluish) formed by a facet and the utopia point for that facet is restricted by the hyperplane (red) defined by a neighbouring facet. 33
3.18 The hypervolume indicator for the phantom with two OARs and a database of 615 solutions (left). The upper bound of the coverage error for the database is shown on the right. 33
3.19 Distributions of the nearest neighbour distances (left) and the largest hyperspheres (right) for the phantom case. The corresponding indicator values are $H_s = 0.69$ and $H_l = 1.25$. 34
3.20 Hypervolume indicator of the phantom case with a database of 800 solutions. The upper bound of the coverage error for the database is shown on the right. 34
3.21 Hypervolume indicator for two different databases with 800 and 1200 solutions for the same phantom case. 34

3.22 Hypervolume indicator (left) and the upper bounds on the coverage error (right) for a prostate database. 35
3.23 For a prostate case with a 110 solutions the nearest neighbour distribution of the solutions (left, $H_s = 0.59$) and the distribution of the largest hyperspheres (right, $H_l = 1.07$) are shown. 35

4.1 Geometry of the vois for the paraspinal case (left) and the DVHs from the Pareto optimal database. 41
4.2 The KonRad reference plan compared to the the convex combination of five solution from the database (dashed lines). 42
4.3 Incorporation of the navigation principle into the routine planning system VIRTUOS at the DKFZ. 44
4.4 Three different trade-offs from a database compared to a KonRad plan. On the left the sparing of both OARs is achieved by a worse target coverage. In the middle the coverage is improved and both OARs are inferior in return. On the right both OARs are decoupled and can be traded independently. . . 44
4.5 Dosedistributions of a KonRad plan (left) and a plan from the database (right) corresponding to the left DVH in figure 4.4. 45
4.6 The plan from a database optimized with the pt-model is similar to the KonRad plan (left). In a database optimized with the std-model a clinically slightly superior plan compared to KonRad can be found. 45
4.7 DVHs for a prostate case. Two solutions from a database are compared to the KonRad plan. 47
4.8 For a paraspinal case a clinical inferior plan can be found in the database (left). Both OARs exibit larger maximum doses and the coverage of the boost volume is inferior. For a prostate case there can occur hot spots (right) in the normal tissue. 47

5.1 Scatter plots of the projected solution for a phantom case with 500 solutions. The figure (i, j) in the matrix corresponds to the projection of all solutions onto the j, i-plane. The diagonal (i, i) shows the histograms of the solution when projected onto the axis of objective i. 52
5.2 Scatter plots of the projected solution for a prostate case. 52
5.3 For the efficient set $\mathcal{P}$ two restricting hyperplanes are shown. 53
5.4 Geometries of two clinical cases of a prostate case (left) and a head and neck case (right) with the corresponding CTs. 54
5.5 For the horseshoe phantom with three objectives the sensitivities for the objectives is shown when the OAR (red) is restricted. 55
5.6 The sensitivities of the objectives for a prostate case where the objective values for the bladder are restricted. 55
5.7 Restriction of the objective values of the rectum and the corresponding sensitivities for the other objectives. 56
5.8 A head and neck case with eight independent objectives. The sensitivities of the objectives is shown when the spine (red curve) is restricted. 56

5.9 Example of a one dimensional curve embedded in three dimensions and the directions of the PCs (a). The PCA yields three nonzero eigenvalues (b) and is not able to discover the true dimensionality due to the non-linear structure. The isomap method unambiguously reveals the true dimension (c). 60
5.10 Geometry of the phantoms. OAR1 and OAR2 are used for the isomap and PCA analysis, whereas up to four OARs are added to study the effective dimension. 62
5.11 Solutions along the first PC for the phantom with one OAR. The prescribed dose of 70 Gy is marked. 63
5.12 Solutions along the second PC for the phantom with one OAR. 63
5.13 Solutions along the first PC for the phantom with two OARs. 64
5.14 Solutions along the second PC for the phantom with two OARs. 64
5.15 The remaining variance for the PCA and the isomap method is shown for the phantom with one OAR. 65
5.16 Two dimensional embedding of all solutions of the phantom case with two OARs. The anchor point solutions for both OARs, the target under- and overdosage (ctv-cold, ctv-hot) and the unclassified tissue (ut) are labeled. . . 66
5.17 DVHs of the brain case for solutions along the first and the second PC. . . . 67
5.18 Dose distributions of the brain case along the first PC (lower left to upper right) and the second PC (upper left to lower right). 68
5.19 The remaining variance of the brain case for the PCA and the isomap method. 68
5.20 DVHs of the pancreas case for the first and second PC. 69
5.21 The DVHs for the third PC. 70
5.22 Dose distributions of the pancreas case along the first PC (lower left to upper right) and the second PC (upper left to lower right). 70
5.23 The remaining variance of the pancreas case for the PCA and the isomap method. 71
5.24 The two dimensional embedding of the pancreas case. 71
5.25 Spearman rank correlation coefficients of the phantom cases. 72
5.26 The effective dimension as a function of the indicator value. 73

List of Tables

3.1 Hausdorff distances $d_H = (\text{conv}(\mathcal{D}_1), \mathcal{D}_2)$ based on the first order approximation of the efficient sets for the three models. The values in brackets correspond to the values of $\max_{y \in \mathcal{D}_2} \min_{z \in \text{conv}(\mathcal{D}_1)} d(y, z)$. 22

4.1 Values of the interpolation coefficients and the corresponding solution numbers after solving the quadratic program. 41
4.2 Results for the prostate planning study. The range covered by the database is shown and the values of the corresponding clinical reference plan are reported in brackets. 46
4.3 Results for the prostate planning study. 46

5.1 Effective dimension and corresponding indicator values for all cases. Another case (p7b) is included in which OAR3 and OAR4 are moved towards OAR2. 72