

Inverse radiation therapy planning – a multiple objective optimisation approach

H. W. Hamacher¹, K.-H. Küfer,

Institute for Techno- and Econo-Mathematics (ITWM) and Department of Mathematics,
University of Kaiserslautern, Erwin-Schroedinger-Straße, Postfach 3049, D-67653 Kaiserslautern, Germany
e-mail: hamacher@mathematik.uni-kl.de, kuefer@mathematik.uni-kl.de

Abstract: For some decades radiation therapy has been proved successful in cancer treatment. It is the major task of clinical radiation treatment planning to realise on the one hand a high level dose of radiation in the cancer tissue in order to obtain maximum tumour control. On the other hand it is obvious that it is absolutely necessary to keep in the tissue outside the tumour, particularly in organs at risk, the unavoidable radiation as low as possible.

No doubt, these two objectives of treatment planning – high level dose in the tumour, low radiation outside the tumour – have a basically contradictory nature. Therefore, it is no surprise that inverse mathematical models with dose distribution bounds tend to be infeasible in most cases. Thus, there is need for approximations compromising between overdosing the organs at risk and underdosing the target volume.

Differing from the currently used time consuming iterative approach, which measures deviation from an ideal (non-achievable) treatment plan using recursively trial-and-error weights for the organs of interest, we go a new way trying to avoid a priori weight choices and consider the treatment planning problem as a multiple objective linear programming problem: with each organ of interest, target tissue as well as organs at risk, we associate an objective function measuring the maximal deviation from the prescribed doses.

We build up a data base of relatively few efficient solutions representing and approximating the variety of Pareto solutions of the multiple objective linear programming problem. This data base can be easily scanned by physicians looking for an adequate treatment plan with the aid of an appropriate online tool.

1 The inverse radiation treatment problem – an introduction

Every year, in Germany about 450.000 individuals are diagnosed with life-threatening forms of cancer. About 60% of these patients are treated with radiation; half of them are considered curable because their tumours are localised and susceptible to radiation. Nevertheless, despite the use of the best radiation therapy methods available, one third of these “curable” patients – nearly 40.000 people each year – die with primary tumours still active at the original site.

Why does this occur? Experts in the field have looked at the reasons for these failures and have concluded that radiation therapy planning – in particular in complicated anatomical situations – is often inadequate, providing either too little radiation to the tumour or too much radiation to nearby healthy tissue. Effective radiation therapy planning for treating malignant tumours is always a tightrope walk between ineffective underdose of tumour tissue – the *target volume* – and dangerous overdose of *organs at risk* being relevant for maintaining life quality of the cured patient.

Therefore, it is the challenging task of a radiation therapy planner to realise a certain high dose level conform to the shape of the target volume in order to have a good prognosis for tumour control and to avoid overdose in relevant healthy tissue nearby.

Part of this challenge is the computer aided representation of the relevant parts of the body. Modern scanning methods like computer tomography (CT), magnetic resonance tomography

¹ on sabbatical leave at the Department of Engineering Science, University of Auckland, New Zealand

(MRT), or high resolution ultrasound scans deliver 2-dimensional transaxial slices of the patient's anatomy, which can be digitised and visualised in three dimensions by use of methods from computer aided geometric design.

In this paper we do not deal with this subject (although this is part of the research project we are currently pursuing) and we will assume that this part is satisfactorily resolved in the following.

In most hospitals radiation therapy planning is performed with a *design-reuse-strategy*, i.e., an old treatment set-up proven successfully for a clinically related case is reused for the incoming patient. The experienced planner will modify the old set-up parameters by trial-and-error until a satisfying dose distribution for the anatomy of the current patient is achieved.

It is no surprise that for complicated anatomical situations and in particular in case of non-standard target volumes the described conventional planning strategy might be time consuming and unsatisfying. Moreover, it is very likely – in particular under the pressure of tight time restrictions in daily hospital care – that not the best possible treatment plans are evaluated for lack of time.

For this reason treatment planning problems have been discussed as *inverse problems* for several years as well, cf. the survey paper [2]; i.e., given a desired dose distribution find appropriate parameters for the treatment set-up. This inverse model is much stronger than the classical forward strategy, as it avoids the time consuming trial-and-error search for good parameters, but is mathematically harder to handle.

For a more precise definition of the *inverse treatment planning problem* it is necessary to introduce the common set-up parameters of radiation therapy: besides modality and energy spectrum of radiation and the dose fraction schedule, which are usually fixed by physicians and therefore assumed to be given in our context, we are particularly interested in the technical set-up parameters like number and positions of the radiation beams, widths, shapes and intensity profiles of radiation beams, cf. [1], [2], [5].

Given the direction of irradiation the shape and the width of a particular radiation beam are essentially determined by the projection of the target volume and the technical set-up of the treatment machinery.

Beyond a certain number of radiation beams the precise directions of these are of minor interest as the choice of appropriate intensity profiles gives enough freedom to establish good treatment plans. But, on the other hand one should try to keep the number of radiation beams as small as possible in order to decrease the set-up time of the treatment unit and henceforth indirectly the treatment time for the patient. So, the question for good or in some sense optimal directions of irradiation is important in particular in case of a small number of radiation beams.

Today, the choice of irradiation directions is mostly done by planners who base their decisions on their experience from clinically related situations and adapt the old parameter set-up by trial-and-error to the new situation. To our best knowledge there have not been any efficient automatic search routines for good arrangements of radiation beam directions up to now.

At first glance, it seems natural to choose all mentioned physical set-up parameters – number and direction of radiation beams, widths, shapes and intensity profiles – in a unified process by some optimisation strategy to be fixed. But, unfortunately such a complex optimisation process is far beyond the capacities of today's computer equipment, even if we assume the state-of-the-art in optimisation techniques, cf. the approach in [3].

Therefore, we propose to search for suboptimal solutions in a *two-stage optimisation process*: first we determine a promising arrangement of beam directions – the set-up geometry – and afterwards we calculate optimal intensity profiles for the beams.

2 The choice of the set-up geometry – a problem of locational analysis

Like most researchers in radiation treatment planning we favour an *isocentric choice of irradiation directions*, i. e., all directions of irradiation meet a single point, the *isocentre of irradiation*.

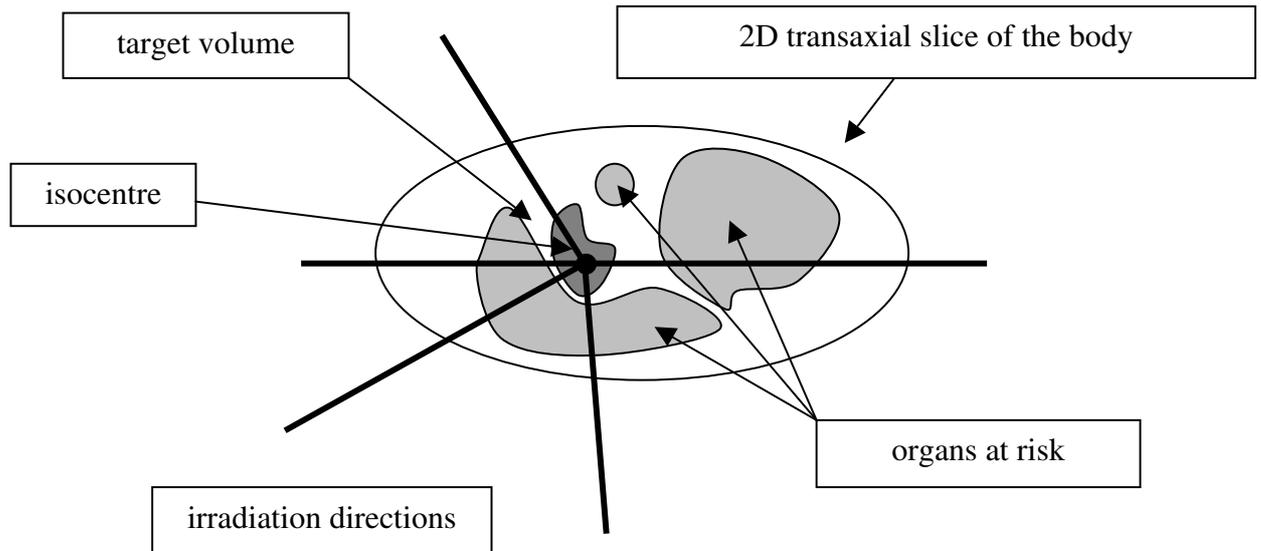


Figure 1: Isocentric geometry of irradiation

For good reasons it is common sense to use the target volume's centre of gravity as the isocentre of irradiation directions, cf. [1], [2] or [3]. This choice is meaningful at least for strictly convex targets; but for a difficult geometry – i.e., in case of tumours with concave indentations around an organ of risk, e.g. prostate carcinomas near the rectum, lymphomas near the spinal cord or meningiomas near the eyes – the isocentre issue needs to be discussed. Besides the choice of the isocentre, number and directions of the irradiation beams are to be chosen in an optimal way within an appropriate setting. While the number of beams has to be determined adaptively the question for optimal directions for a given number n of beams can be interpreted as continuous $(n+1)$ -*facility location problem*: in order to compare a number of possible combinations of beam directions and isocentre we define a scalar *score function* that measures the quality of the arrangement of beams: the greater the value of the score function the better the choice of the arrangement. Having designed such a score function it is the task of mathematical programming to find an arrangement of isocentre and beams whose associated score value maximises the score function; such an arrangement is called optimal with regard to the score function.

Currently we work with score functions that operate on a discrete set of possible locations for isocentre and beam directions. More precisely, we choose a small number of promising places for the isocentre inside or near the target volume and take each of these possible isocentres as a midpoint of a possible beam arrangement. Then, we choose for each such isocentre a number of possible directions of irradiation from a unit sphere around. Typically, we discretise the sphere in such a way that the angles between neighbouring points – i.e., between irradiation directions – are not bigger but nearly 15° . Depending on the particular case, some regions of the sphere might be forbidden for anatomical or technical reasons.

For any possible arrangement of isocentre and beam directions the score function measures the value of the arrangement in terms of the estimated irradiated dose deposition in target and organs at risk. In order to describe the score function precisely we need some more insight concerning the calculation and judgement of dose depositions, which will be provided in the next section. A precise formulation of the score function we work with will be given in Section 4.

3 The calculation of intensity profiles – a multicriteria linear programming problem

If the irradiation geometry – i.e., place of isocentre, number and directions of irradiation beams – is given, the unknown intensity profiles of the radiation beams can be defined as solutions of a linear integral inequality system and after an appropriate discretisation as solutions of a linear inequality system respectively. Shape and width of the irradiation beams are determined by the intensity profiles as well.

The discretisation is done in two respects, cf. Figure 2: the irradiated volume of the human body is dissected in M *volume elements* (voxels) and the union of the beams is dissected in N *beam elements* (bixels). The absorbed doses in the volume elements can be approximated by a matrix vector equation

$$(1) \quad D = P x,$$

where D is an M -dimensional dose vector, whose i th entry represents the absorbed dose in the

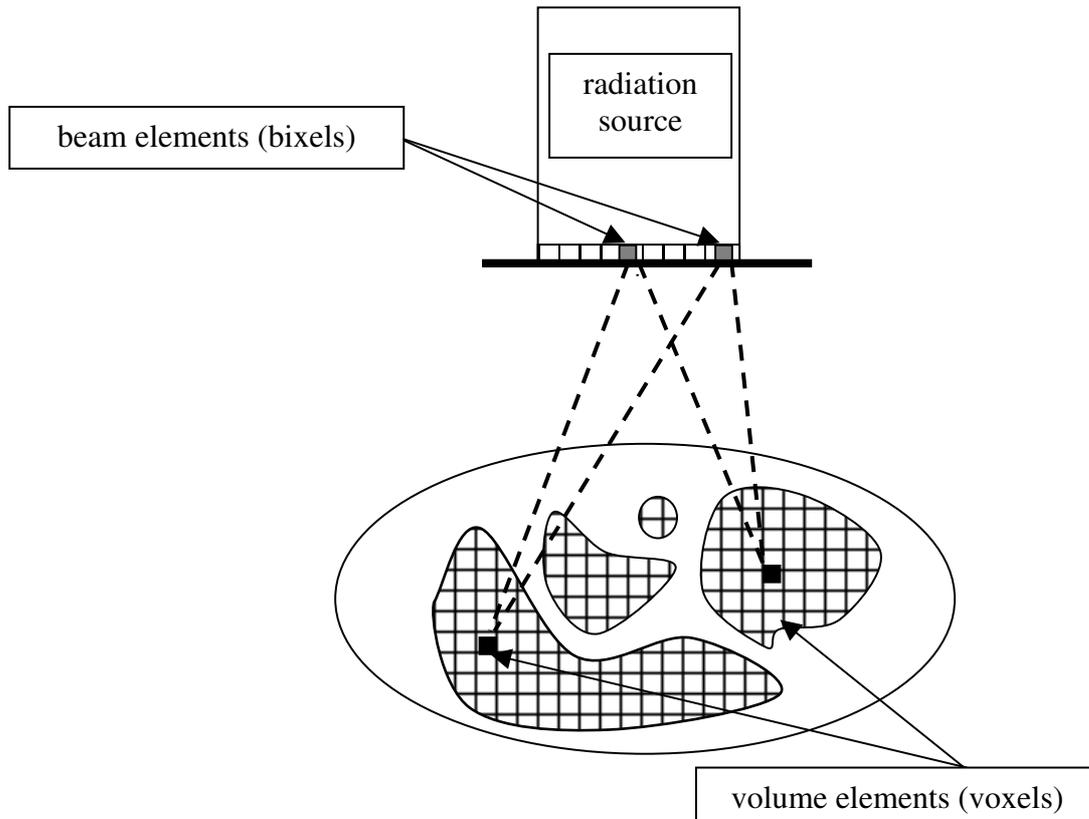


Figure 2: Volume elements and beam elements

i th volume element. x is an N -dimensional vector, which describes the union of all beam element intensities; the j th entry of x equals the radiation intensity in the j th beam element.

Finally, the entry $P(i, j)$ of the matrix P describes the contribution of the j th bixel to the total dose in voxel i under unit intensity. The entries of P depend on modality and energy spectrum of the radiation used. The values $P(i, j)$ are either patient independent and taken from data bases (or are approximated) with regard to the water phenomenon, cf. [9], or are calculated specifically for each patient using Monte-Carlo simulation as it is done e.g. by the PEREGRINE research program of the Lawrence Livermore National Laboratory, Los Alamos, cf. [4]. In this paper we do not discuss this important issue; we assume the $P(i, j)$ to be given in some satisfying way.

For the sake of simplicity we will assume in the following that volume and beams are equidistantly discretised; this is natural with regard to the volume as the 2D-scans of the body are usually taken in equidistant steps. But, for the discretisation of the beam elements it should be discussed whether an adaptive choice of non necessarily equidistant bixels could lead to a simpler form of intensity profiles which are easier to realise with the treatment unit.

The basis of the physical treatment planning are dose constraints: typically, there are lower bounds for the dose in the target volume and upper bounds for the dose levels in organs at risk. Therefore, it is useful to partition the dose calculation formula (1) in fractions that correspond to the different organs of interest. In the following we will assume that there are K organs of interest and throughout the paper we will identify the target with organ number 1 and the organs at risk with numbers 2, 3, ..., K . Dose constraints usually have the form

$$(2) \quad \begin{aligned} D_1 &= P_1 x \geq L_1, & \text{„target (1)“}, \\ D_k &= P_k x \leq U_k, & \text{„organs at risk (} k = 2, \dots, K \text{)“}, \\ x &\geq 0. \end{aligned}$$

Here, D_k , L_k , and U_k are vectors, whose dimension is given by the number N_k of voxels in organ k . L_k and U_k are vectors with constant entries, which represent in case of the target the necessary minimum dose level in order to achieve tumour control and in case of organs at risk the maximum levels which are likely to avoid complications in the patient's future life. P_k is the submatrix of P that consists of those rows of P that correspond to organ k .

Besides the dose constraints given in (2) in most models of radiation treatment planning there is a homogeneity condition for the dose in the target. Homogeneity of dose has been proved to be useful in order to obtain maximal tumour control in many clinical cases. An easy way to formulate dose homogeneity is to compare the dose level in each voxel of the target with the dose in a reference voxel. Let s be a fixed relative tolerance of dose deviations from the absorbed dose in the reference voxel and let P_{ref} be a rowwise constant matrix with row vectors p_{ref} , whose entries $p_{\text{ref}}(j)$ represent the dose contribution of bixel j to the total dose in the reference voxel. Then, the *homogeneity condition* has the form

$$(3) \quad (1-s) P_{\text{ref}} x \leq P_1 x \leq (1+s) P_{\text{ref}} x.$$

In general the inequality system (2) and (3) will be inconsistent such that the system has *no solutions*. This is essentially due to the physical condition $x \geq 0$ in (2), which means that there is no negative radiation intensity.

Thus, mathematics shows that there is in general no choice of the parameter set-up that satisfies the given restrictions (2) and (3) completely. Hence, in order to achieve solutions

satisfying the restrictions at least in some approximate sense the restrictions have to be weakened somehow: either organs at risk have to be (partially) overdosed or the target has to be (partially) underdosed. So, compromises have to be made.

In order to quantify and to value deviations from the constraints (2) and (3) there is need for a definition of approximate solutions.

In the literature, cf. [1], [2], [3], [5], the quality of approximate solutions is characterised with the aid of some scalar quality function F that measures the deviation of an approximate solution from the constraints provided in (2) and (3). Given such a quality function F , it is the planner's task to find approximate solutions x of the inequality system defined in (2) and (3) that minimise the measure $F(x)$ of deviation.

It is worthwhile to discuss the choice of the quality function F very carefully, as the choice of F is somehow arbitrary and often governed by practicability considerations alone.

Typical choices of F result from least squares approximation approaches, as proposed in [1], [2], or from minimising maximal deviations, as proposed in [3] or [5].

In both cases the function F takes the form

$$(4) \quad F(x) = \sum_{k=1}^K \mu_k F_k(x) \longrightarrow \text{Min},$$

s.t. $x \geq 0$ and (3), where $F_k(x)$ measures the least square distance or the minimax distance respectively from the dose restriction for organ k in (2). The parameters μ_k are positive weights, whose particular choice controls the tolerance of deviations from the restrictions for the organs of interest. A small weight makes a deviation more likely, a large weight more unlikely. The a priori choice of the weights is the crucial point here, as it is a hard job to transfer the relative preferences of possible deviations in different organs formulated by physicians into absolute weight values.

A solution x of problem (4) is an optimal approximation of the inequality system (2), (3) with regard to the quality function F . The resulting treatment set-up will be evaluated by the physician with the aid of isodose curves and dose volume histograms, cf. Figure 3.

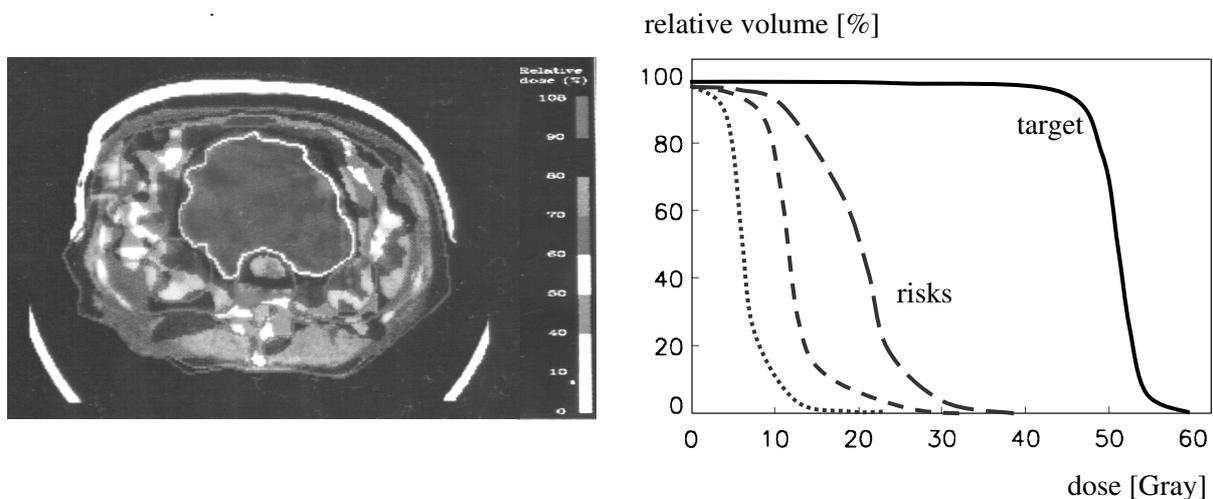


Figure 3: Isodose curves and dose volume histograms

In practice the first choice of weights μ_k is often not appropriate as it does not reflect the requirements of the physician sufficiently. Optimality of a treatment set-up in the setting of

(4) does not guarantee practicability. Therefore, in this approach there is need for a careful trial-and-error change of the weights μ_k in order to improve the outcome until a clinically satisfying weighting of deviations will be achieved.

Unfortunately, problems of type (4) are highly parameter sensitive, i.e., even slight changes may cause tremendous effects on the solutions. This phenomenon makes a search for good weights even harder. The parameter sensitivity of (4) can be reduced by weakening the constraints in (2) making use of so called *dose volume restrictions*, cf. [1] and [7]. A dose volume restriction for an organ at risk k allows a prescribed additional dose \tilde{u}_k in not more than $\lfloor v_k N_k \rfloor$ voxels of organ k , $0 \leq v_k \leq 1$; i.e., a dose volume restriction weakens the upper bound from (2) for an arbitrary but restricted volume part of the organ. Dose volume restrictions apply successfully for those organs that preserve their functionality even if they are partly destroyed, e.g. liver, lung, etc. Mathematically, dose volume restrictions can be formulated by *mixed integer inequalities*

$$(5) \quad \begin{aligned} D_k x &\leq U_k + \tilde{u}_k \Delta_k, \\ \mathbf{1}^T \Delta_k &\leq \lfloor v_k N_k \rfloor, \end{aligned}$$

where Δ_k is a 0-1-vector and $\mathbf{1}$ is a vector with all 1 entries, each of them having appropriate dimension. If the i th entry of Δ_k equals one there is an allowance for an additional dose $\leq \tilde{u}_k$ in the corresponding voxel and if the i th entry of Δ_k equals zero the original tighter bound from (2) applies. The second equation guarantees that the volume part with additional dose does not exceed v_k . Note, that in general the use of dose volume restrictions for certain organs does not lead to consistent linear inequality systems. Henceforth, we have still to deal with an approximation problem for solutions of linear inequality systems of type (2) and (3) partly weakened by inequalities of type (5). For ease of notation in the following, we replace the constraints (2) by constraints of type (5) for all $k=1, \dots, K$: if dose volume restrictions do not apply for a particular k , we set $v_k=0$, $\tilde{u}_k=0$ and $\Delta_k=0$.

Sensitivity problems as discussed before are very familiar for standard approximation approaches for inconsistent inequality systems like (3) and (5).

In the particular context of treatment planning it is an interesting basic question, whether the quality of an approximating solution should be measured by a single function like F in the light of a problem that is a *multicriteria problem*: the question to find simultaneously a high dose in the target and small doses in organs at risk obviously deals with contradictory objectives. If we fix a single quality function F , we have to couple these contradictory objectives a priori in an arbitrary way. This coupling is problematic as it causes at best an a posteriori trial-and-error process for choosing the right weights – a phenomenon which should principally have been avoided by the inverse setting of the treatment planning problem.

Therefore, we propose to model the inverse treatment planning problem in a way that reflects the multicriteria character appropriately: we formulate the problem as a *multicriteria linear optimisation problem*. For every organ of interest – target as well as organs at risk – we establish independent quality functions that measure the deviation from the constraints provided in (5). This system of quality function will be minimised simultaneously.

More precisely with the notation introduced above we consider the *mixed integer, linear inequality system*

$$D_1 = P_1 x + t_1 \mathbf{1} \geq L_1, \quad \text{“target (1)”}$$

$$\begin{aligned}
(6) \quad & D_k = P_k x - t_k \underline{1} \leq U_k + \tilde{u}_k \Delta_k, && \text{“risks (} k = 2, \dots, K \text{)”} \\
& \underline{1}^T \Delta_k \leq [v_k N_k], \\
& (I-s) P_{\text{ref}} x \leq P_I x \leq (I+s) P_{\text{ref}} x, && \text{“homogeneity”} \\
& t_1, \dots, t_K \geq 0, \quad x \geq 0, \\
(6a) \quad & \Delta_k \in \{0,1\}^{N_k},
\end{aligned}$$

with unknown positive thresholds t_k , unknown positive intensity profiles x and unknown 0-1-vectors Δ_k . The tolerance s of inhomogeneity and the volume parts v_k for dose volume restrictions are considered to be given. In contrast to the inequality system (3) and (5), system (6), (6a) is obviously always feasible. The scalar threshold t_k equals the maximal deviation from the corresponding dose bound for organ k stated in (5).

We now consider the linear multicriteria optimisation problem

$$(7) \quad t_1 \longrightarrow \text{Min}, t_2 \longrightarrow \text{Min}, \dots \dots \dots, t_K \longrightarrow \text{Min}$$

subject to the constraints in (6), (6b), i.e., we minimise the maximal deviations from the constraints in (5) simultaneously.

In general situations, there will be no feasible tuples (x^*, t^*) that fulfil $t^* \leq t$ for all feasible solutions (x, t) of (6), (6a), i.e., there will be no solution that minimises all objective functionals from (6), (6a) simultaneously. For this reason Pareto, cf. [8], introduced the setting of *efficient solutions of multicriteria problems*: a solution (x^*, t^*) of the multicriteria problem (7) subject to (6) and (6a) is called *efficient* or *Pareto-optimal* if there is no other feasible solution (x, t) for which there exists a j such that $t_j < t_j^*$ and $t \leq t^*$, i.e., (x^*, t^*) is *efficient if it is not dominated* by some other solution (x, t) .

Unfortunately, general mixed integer multicriteria (monocriterion) linear programs are known to be NP-hard. Thus, the stated optimisation problem (7) subject to (6) and (6a) seems to be intractable in case of clinically typical problem dimensions.

Therefore, we consider a *continuous relaxation* of (6a) and replace (6a) by

$$(6b) \quad \Delta_k \in [0,1]^{N_k},$$

allowing that the entries of Δ_k vary in the interval $[0,1]$. The problem (7) subject to the constraints (6) and (6b) is a continuous linear multicriteria programming problem which is solvable in polynomial time.

At first sight, this relaxation might be justified by the argument that the additional dose for organ k cumulated on all voxels does not exceed the postulated total overdose but can be spread over all voxels of the organ. We are aware that solutions of such relaxations in general lead to unacceptable treatment plans. Therefore we propose to solve problem (7) subject to (6) and (6a) by a *relaxation and roundoff procedure*. First, we solve the continuous relaxation (7) subject to (6) and (6b) and find Pareto-optimal solutions. Given such a solution we look at the Δ_k -vectors and round the entries of these vectors by the following rule: If the entry is one of the $[v_k N_k]$ biggest entries of Δ_k , we round this entry to 1, otherwise we round it to zero. This way, we establish a 0-1-vector $\hat{\Delta}_k$. Now, we solve the continuous linear multicriteria problem(7) subject to the constraints

$$D_I = P_I x + t_I \underline{1} \geq L_I, \quad \text{“target (I)”}$$

$$\begin{aligned}
(8) \quad & D_k = P_k x - t_k \underline{1} \leq U_k + \tilde{u}_k \hat{o}_k, \quad \text{“risks (} k = 2, \dots, K \text{)”} \\
& (1-s) P_{\text{ref}} x \leq P_l x \leq (1+s) P_{\text{ref}} x, \quad \text{“homogeneity”} \\
& t_1, \dots, t_K \geq 0, x \geq 0
\end{aligned}$$

with given *modified upper bounds* $U_k + \tilde{u}_k \hat{o}_k$.

With this relaxation and roundoff procedure we can approximate the set of solutions of the NP-hard mixed integer programming problem (7) subject to (6) and (6a) by solutions of the polynomial continuous problem (7) subject to the constraints (8) that arise from the relaxation and roundoff procedure.

Note, that in general we will obtain suboptimal solutions of the original problem this way. But due to our practical experience, in many cases the number of nonzero entries of Δ_k is smaller than $\lfloor v_k N_k \rfloor$, which means that the additional dose allowance will not be exhausted completely and the relaxation and roundoff procedure delivers optimal solutions.

The set of efficient solutions of continuous linear multicriteria programming problems is a connected and closed subset of the topological boundary of the polyhedron of feasible solutions. In the language of monocriterion linear programming the set of efficient solutions can be characterised as the union of all optimal solutions of linear programs whose objective functions are convex combinations of the objectives in (7). This property will be widely exploited in our algorithms, cf. Section 4. For computational solutions of linear multicriteria programs it is useful to know that the set of efficient solutions of (7) can be described in terms of efficient vertices of the feasible polyhedron. Hence, it is the major algorithmic task to calculate the efficient vertices of the feasible polyhedron. This can principally be done in two stages: first we calculate an arbitrary efficient solution with an interior point method or a variant of the simplex method, in the second stage we enumerate all efficient vertices with a parametric variant of the simplex method.

For the solution of linear multicriteria programs there are well known public domain algorithms at hand. But, for clinically typical dimensions of the problem, i.e., $M \approx 10^6$ and $N \approx 10^3$, $2 \leq K \leq 7$, it will not be possible to apply these algorithms to calculate all efficient vertices, as the expected number of efficient vertices is far too large for the power of today's computers, cf. [6].

Fortunately, a calculation of all efficient vertices is not necessary for two reasons: First, the number of efficient vertices depends strongly on the used bixel/voxel discretisation, since the set of efficient vertices will lie the denser the finer the discretisation is. This means in particular that there will be no significant change between the treatment plans of neighbouring efficient vertices. So, it will be of no use for a planner to know all efficient vertices. It will be sufficient for him/her to know a representative set of efficient vertices the neighbours' of whose t -vectors differ significantly but moderately. Second, it will not be of interest to calculate efficient solutions that completely neglect some of the organs in question.

For these reasons it is absolutely necessary and meaningful to concentrate on an appropriate representative subset of efficient vertices.

We propose to approximate the set of efficient solutions by choosing a set of representative efficient vertices in such a way that for all efficient solutions we know a representative in an appropriate neighbourhood. The width of this neighbourhood must be independent of the width of the underlying discretisation and is chosen adaptively, such that the differences of the t -vectors between neighbouring representatives are not too large.

Having generated a number of efficient solutions of the treatment problem, we store the physical treatment parameters together with the associated isodose curves and dose volume histograms in a data base. This data base can be scanned by the physician in order to find a treatment plan that satisfies his/her requirements. It is not necessary that the physician values and compares all efficient solutions in this data base. The search process is done with an online tool that works as follows. The process starts with a central efficient solution, i.e., an efficient solution corresponding to nearly equiweighted organs. If this proposal is not acceptable, since the physician would like the constraints (5) to be fulfilled more tightly in some organs, an alternative can be found that will satisfy the restrictions for these organs more strongly. The new proposal is available without any delay worth mentioning, because the only thing to be done is to go down in a search tree whose keys consist of the t -vectors of the efficient solutions stored in the data base. If the new proposal is not satisfying, the physician recursively repeats the procedure until a satisfying solution is found. This search process will converge within a few iterations.

If we compare this multicriteria based tool with the conventional inverse planning mechanism, it is clear that radiation therapy planning can be significantly improved, as the physician can – in a couple of minutes – interactively scan a significant number of efficient planning proposals guided by his/her own wishes and preferences. A plan can be chosen without knowledge of the actual physical planning parameters, and decisions are only based on appropriate visualisations of the resulting dose distributions. The physician can “change” the treatment plan without additional interaction with the planning physicists and – even more important – without additional time delay consumed by trial-and-error procedures.

Obviously, the patients will mostly profit from the new tool, since they can be sure that an individually optimised treatment plan will be found and realised, which is not governed by strongly suboptimal compromises resulting from the everyday time pressure in hospitals.

4 Practical realisation

Based on the theoretical considerations presented in Sections 2 and 3 the authors pursue a research project within the department “Optimisation Methods in Mathematics of Economy” of the Fraunhofer Management “Institute of Techno- and Econo-Mathematics (ITWM)” in Kaiserslautern. The department “Medical Physics” of the German Cancer Research Centre at Heidelberg (DKFZ) and the German Research Centre for Artificial Intelligence (DFKI) are the major co-operating partners.

It is the objective of the project to develop and to implement a software package for inverse radiation treatment planning.

As basis for our program serves the digitised and stored anatomy of the patient, which allows the following data manipulations: for every point of the relevant geometry it is known in which organ it lies and for every plane that intersects the relevant geometry it is possible to draw 2-dimensional sections containing surface lines of relevant organs and coloured dose levels.

Moreover, we assume that the voxels are equidistant cubes with side lengths 0.2cm (or 0.25cm) geared to the standard distances of MRC or CT slices. Finally, we work with a maximal bixel resolution of 1 x 1 cm adapted to the resolution of modern available treatment units with multileaf collimators.

Taking the clinical treatment parameters like radiation modality, energy spectrum and dose restrictions and the patient’s anatomy in visualised form as inputs, the program calculates a data base with a number of complete therapy plans each of which is given by the following specifications as output keys:

- high resolution 3D-dose distribution (to be visualised in 3D-form from different point of views or in 2D-form for arbitrary sections)
- dose volume histograms for all relevant organs (target and risks)
- irradiation geometry with geometrical data of isocentre and beam directions
- intensity profiles for all beams
- t -vector of maximal deviations from dose restriction (internal)

The data base can be scanned by an interactive online tool as described in Section 3. The search for a better solution with regard to the physicians specific interaction is internally done alone with the aid of the t -vectors. All dose calculations in the program are currently done with Ulmer's and Harder's approximations of the water phenomenon, cf. [9].

The software currently consists of two main stages: the search for an optimal irradiation geometry, cf. Section 2, and the calculation of efficient intensity profiles, cf. Section 3. We only outline the main features of the algorithm.

Stage 1: (search for an optimal irradiation geometry)

Using the discrete locational analysis framework sketched in Section 2, we choose the following strategy to find a good irradiation geometry: For every possible isocentre, usually less than $I = 10$ possible choices chosen by eye, we formulate a conventional LP-problem with constraints as in (6) and (6b), where the vector x is a discretisation of the intensity profiles of all, say J , possible beam directions.

For the determination of the radiation geometry we do not exploit the maximal degree of freedom for choosing the intensity profiles given by the maximal bixel resolution indicated above nor do we use the finest voxel discretisation possible. Instead, we use an equidistant discretisation of cubic voxels with 0.4 cm or 0.5 cm edge length. The bixels of each beam j are divided into two relevant classes: the first class consists of those bixels whose midpoints lie in the target's shadow under the orthogonal projection onto the radiation source and not in the shadow of any organ at risk, the second class consists of those bixels whose midpoints lie in the shadow of some organ at risk and in the shadow of the target. Bixels that do not meet the shadow of the target at all are provided with intensity 0, width and shape of each beam are determined this way. Next, we associate with the bixels in class one the intensity level $x_1^{(j)}$ and the level $x_2^{(j)}$ with the bixels in class two. Hence, in this stage the intensity profile of each beam has not more then two nonzero intensity levels.

Finally, we consider for each possible choice of the isocentre an objective function c defined as a weighted sum of the maximal deviations t_k from the restrictions for organ k .

$$(9) \quad c = \mu_1/L_1 t_1 + \mu_2/U_2 t_2 + \dots + \mu_K/U_K t_K$$

The weights μ_k are fixed in a standard way with regard to the preference list given by the physician and to the volumes of the organs for all possible choices of the isocentre.

We solve the resulting LPs and afterwards cancel in each particular LP all beam directions whose dose contribution in a time unit is less than $a\%$ of the beam(s) with maximal dose contribution. a ranges between 5 and 15 depending on the complexity of the particular situation. This way, we reduce the number of beams adaptively to a number $J_{\text{adapt}}(i) \leq J$ for any possible isocentre $i = 1, \dots, I$. Afterwards, we calculate the c -values c_1, \dots, c_I for the truncated problems and choose a geometry j that maximises the score function

$$(10) \quad S(j) = I/c_j.$$

Note, that for the determination of the irradiation geometry we only work with the continuous relaxation of problem (7) s.t. (6) and (6a), since the roundoff procedure does not effect the outcome significantly. This is due to the relative stability of the treatment planning problem with regard to the directions of irradiation.

Stage 2: (calculation of efficient intensity profiles)

Based on the fixed irradiation geometry – fixed number of beams, fixed isocentre and fixed directions – calculated in Stage 1 we search recursively for a representative set of efficient vertices of the multicriteria linear optimisation problem (7) s.t. (6) and (6a) with maximal resolution in voxel and bixel discretisation. We solve this mixed integer multicriteria program approximately with the relaxation and roundoff procedure given in Section 3. We start the procedure by solving an ordinary LP with objective function (9) and take the outcome as first representing efficient solution. The weights μ_k for the first representative are taken from a data base with clinically related cases or are fixed corresponding to the preference list of organs in a standard way.

Given such a representative efficient solution (x^*, t^*) , we consider for each $k = 1, \dots, K$ the parametric objective function

$$(11) \quad c_\lambda = \mu_1/L_1 t_1 + \mu_2/U_2 t_2 + \dots + \lambda \mu_k/U_k t_k + \dots + \mu_K/U_K t_K$$

and follow the co-optimal vertex path with increasing λ starting with $\lambda = 1$ and (x^*, t^*) until we reach a solution (x, t) that fulfils the constraints

$$(12a) \quad t_k \leq (1-q) t_k^*$$

$$(12b) \quad t_j \leq (1+q/(K-1)) t_j^*, \quad j \neq k.$$

The constraints (12b) guarantee that the new efficient solution with better t_k -value does not affect the other entries of the t -vector too hard. The algorithm stops if a new efficient solution is obtained that satisfies (12a) or – if such a solution does not exist – delivers a solution (x, t) with the best possible t_k -value s.t. constraints (12b).

The K efficient vertices that result this way from the current representative efficient solution (x^*, t^*) are called the *neighbours* of (x^*, t^*) and are added to the set of representative efficient solutions. Next, we recursively take the new representatives as starting points for determining further neighbouring efficient solutions in an analogous way. This recursion is repeated 3 to 5 times in total depending on the number of organs K . The value q may be different for different k and recursions and varies usually between 0.05 and 0.2.

This method is related to the *reference point method* due to Wierzbicki, cf. [8]. Here, equations (12a) and (12b) serve as *reference points* for the controlled search of neighbouring efficient solutions. In contrast to the latter method our approach is however not interactive to avoid time delays in the design of the radiation plan.

Our method establishes an adaptive controlled search in the objective space excluding typical drawbacks of inherent parameter sensitivity.

Finally, we stress that all calculations of LP-solutions are done with parametric variants of the simplex algorithm. Obviously, commercial LP-solvers like XPRESS-MP or CPLEX can be

used to solve the LPs.. In order to keep the time complexity at a minimum, all LPs are preprocessed and searched for obvious redundant constraints.

5 First experiences and concluding remarks

First practical experiments using a prototype software to deal with some hard clinical and artificial examples like the *horseshoe target* are very promising with regard to time complexity and the requirements on the results we stated above. The detailed discussion of numerical examples is beyond the scope of this paper.

In the future it is planned to improve and to supplement the software in several respects:

New objectives:

- adaptive bixel design
- biological quality functions
- patient individual dose calculations respecting tissue inhomogeneity
- geometric (i.e. dose independent) search for an irradiation geometry
- optional restriction to a fixed or optimised 2D-irradiation geometry

Technical realisation:

- paralleled dose calculations on lean PC-architectures
- interior point methods for the calculation of starting solutions

6 Acknowledgment

We would like to thank our graduate students Frank Lenzen and Thang Nguyen for their help in designing the first prototype version of the software and for doing the first test runs on data.

References:

- [1] T. Bortfeld, J. Stein, K. Preiser. Clinically Relevant Intensity Modulation Optimisation Using Physical Criteria. In: D. D. Leavitt and G. Starkschall (eds.). *Proceedings of the XIIth ICCR*, Salt Lake City, 1997.
- [2] A. Brahme. Treatment Optimisation Using Physical and Radiobiological Objective Functions. In: A. Smith (ed.). *Radiation Therapy Physics*. Springer, Berlin, 1995.
- [3] R. E. Burkard, H. Leitner, R. Rudolf, T. Siegl, E. Tabbert. Discrete Optimisation Models for Treatment Planning in Radiation Therapy. In: H. Hutten (ed.). *Science and Technology for Medicine, Biomedical Engineering in Graz*. Pabst, Lengerich, 1995.
- [4] <http://www-phys.llnl.gov/peregrine>
- [5] J. Höffner. *New Methods for Solving the Inverse Problem in Radiation Therapy Planning*. Shaker Verlag, Aachen, 1997.
- [6] K.-H. Küfer. On the Asymptotic Average Number of Efficient Vertices in Multiple Objective Linear Programming. *J. Complexity* 14: 333-377, 1998.
- [7] M. Langer, R. Brown, M. Urie, J. Leong, M. Stracher, J. Shapiro. Large scale optimisation of beam weights under dose-volume restrictions. *J. Radiation Oncology. Biol. Phys.* 18: 887-893, 1990.
- [8] R. Steuer. *Multicriteria Optimisation: Theory, Computation and Applications*. Wiley, New York, 1985.
- [9] W. Ulmer, D. Harder. A triple Gaussian pencil beam model for photon beam treatment planning. *Z. Med. Phys.* 5: 25-30, 1995.