CONTINUOUS OPTIMIZATION OF BEAMLET INTENSITIES FOR PHOTON AND PROTON RADIOTHERAPY

R. REEMTSEN* AND M. ALBER†

Abstract. Inverse approaches and, in particular, intensity modulated radiotherapy (IMRT), in combination with the development of new technologies such as multi-leaf collimators (MLCs), have enabled new potentialities of radiotherapy for cancer treatment. The main mathematical tool needed in this connection is numerical optimization. In this article, the variety of continuous optimization approaches, which have been proposed for the computation of optimal beam and beamlet intensities respectively, is surveyed and discussed. The discussion includes a nonlinear optimization model for IMRT with biologically motivated goals, which has recently been presented in [9] and is accompanied by a sensitivity analysis proposed in [4]. At last, new developments in intensity modulated proton therapy (IMPT) are considered. It is shown by a clinical case example that the algorithm introduced in [9] is also capable to solve, within a few minutes of computation times, the much larger problems of treatment planning for the IMPT spot-scanning technique.

Key words. radiotherapy treatment planning, intensity modulated radiation therapy, IMRT, intensity modulated proton therapy, IMPT, spot-scanning technique, optimization, linear programming, nonlinear programming

AMS subject classification. 90C90, 90C06, 90C30

1. Introduction.

Radiation therapy is an essential medical tool for cancer treatment. About $500,000$ patients in the USA and $150,000$ patients in Germany are treated yearly by *radiation therapy*, where about every second patient has undergone surgery before. The hazard with radiotherapy, however, is that it does not only destroy tumor cells, but similarly also affects healthy tissue. Therefore, based on the images of *computer tomography*, for each patient a compromise has to be found between the two conflicting goals, to deposit a sufficiently high dose into the *planning target volume(s)* (PTVs), i.e. the tumor(s) and/or the possibly involved tissue, and to simultaneously spare, as much as possible, the *organs at risk* (OARs) and the other healthy tissue. As a consequence, radiotherapy treatment planning involves the selection of several suitable directions for the incident beams and the fixation of beam intensities or, if these are modulated, beamlet intensities so that, through superposition of the doses delivered by the single beams or beamlets respectively, a desired dose is deposited in the PTVs and simultaneously no critical doses are administered to the normal-tissue volumes lying on their paths of propagation. (Introductions into the field are found e.g. in [13], [21], [41], [49], [87], [101], [102].)

Conventionally in radiotherapy, the radiation is produced by beams of highly energetic photons delivered by a *linear accelerator*. The treatment itself is standardized in most hospitals. Depending on the position and the type of the tumor(s), the number of *radiation fields* or *beams* respectively is prescribed (typically between 2 an 5), the field or beam angels are essentially predetermined, and the beam intensities are homogenous or have a constant gradient. The radiation fields are rectangular, where often custom-made apertures or a collimator like a *multi-leaf collimator* (MLC) are used to cover parts of the fields and thereby protect portions of the patient's body. (A MLC consists of typically $25-60$ tungsten slabs which can be shifted from each

*Brandenburgische Technische Universität Cottbus, Institut für Mathematik, Universitätsplatz 3–4, D-03044 Cottbus, Germany; email: reemtsen@math.tu-cottbus.de

†Universitätsklinikum Tübingen, Radioonkologische Klinik, Hoppe-Seyler-Strasse 3, D-72076 Tübingen; email: msalber@med.uni-tuebingen.de

of two opposite sides by computer control.)

In case of such conventional approach, an individual treatment plan is normally obtained by a trial-and-error procedure, where the radiation effects of a few differing arrangements are considered with respect to their dose distributions. In contrast to such *forward approach*, an *inverse approach* starts from the definition of treatment goals, defined by requirements on the doses for the PTVs and the OARs, and it results in the problem of finding beam or beamlet intensities for a certain number of well positioned radiation fields such that the delivered doses meet these requirements or are close to them (e.g. [15], [16], [24]). Hence, at an inverse approach, restrictions on doses are often established in form of inequalities or equalities, and goals described by one or, as in case of a multicriteria approach, several objective functions are to be optimized. Thus an inverse approach naturally is connected with numerical optimization.

A large number of articles over the last 20 years has dealt with the improvement of conventional arrangements by inverse approaches and the work in this direction still continues. Parallel and starting with the seminal works of Brahme ([20], [23]), the field of *intensity modulated radiation therapy* (IMRT) emerged and has attracted a rapidly growing interest over the last ten years. This inverse approach, which for the first time was employed clinically around 1994, "is regarded by many in the field as a quantum leap forward in treatment delivery capability ([49])". By splitting the beams into thousands of *beamlets* or *pencil beams,* IMRT enables the creation of much more sophisticated and precise dose distributions and, through that, renders possible the treatment of cancer patients by radiation therapy who could not be treated adequately before. Mathematically, IMRT leads to large-scale optimization problems.

For the optimization of an IMRT treatment plan a variety of parameters may be considered. Besides the *beamlet weights* determining the beamlet intensities, the main degrees of freedom are the number of beams used, the beam angles, and parameters connected with the realization of an intensity profile by a MLC. Ideally, all of these parameters should enter an optimization model, and experiments in this direction also have been performed. However, for the time being, such approaches set strong limits in regard to the complexity of the resulting problem as well as the maximum number of variables, where the latter typically does not suffice for clinical routine. Especially, if integer variables are included in a model to find, for example, an optimal set of r from a given set of $s \geq r$ beams with prescribed angles (e.g. [34], [36], [63]), the size of the resulting problems and the state-of-the-art of *mixed-integer programming* exclude nonlinear functions in the model. Therefore, presently, the optimization of IMRT treatment plans requires the a priori decision whether integer variables are allowed in the model, in which case only linear functions should be used for the beamlet weight optimization, or whether certain parameters as beam angles and beam directions are fixed so that also nonlinear constraints can be exploited.

The relevance of biological treatment goals for radiotherapy, leading to nonlinear constraints as *equivalent uniform dose* (EUD) or *partial volume* (PV) *constraints,* has been generally acknowledged during the last years (see Sect. 3.5). On the other hand, nonlinear programming is associated with the risk that local minimizers are computed, having an objective function value far away from the global minimum value. For this reason, some researchers have developed substitutes or approximations of intrinsically nonlinear conditions by linear or convex constraints. For example, it has been suggested to replace the nonlinear convex EUD function of [76] by an expression, which results in a large number of linear constraints ([99]), or to approximate a convex objective function, defined for each volume element of the irradiated volume,

by a finite number of linear constraints ([82]), which likewise increases the size of the problem enormously. Several authors have implemented *dose-volume constraints* (see Sect. 3.5) by means of binary variables and have used *mixed-integer linear programming* (MILP) in order to take partial volume effects, which are naturally described by nonconvex functions, into account (e.g. [11], [60], [63], [80]). However, such treatment of dose-volume effects can lead to tens or hundreds of thousands of additional binary variables, which increases the complexity of the problem and the computing time for its solution enormously. Also, in case of an aforementioned optimal selection of, say, $r = 5$ from $s \geq r$ prescribed beam angles, eventually the number s of preselected angles needs to be relatively small in order to reasonably limit the total number of choices given by the binomial coefficient $\binom{s}{r}$.

At this point it should be pointed out that the process of finding an optimal IMRT treatment plan cannot be fully automated, but in any case requires the participation of an expert, who has to set up the treatment goals, to evaluate the computed treatment plan, and to modify, quite commonly, the original goals, assessing simultaneously the related risks for the patient. By their experience, experts often have a good feeling for reasonable beam numbers and beam directions in a particular case. Also the avoidance of selecting beam angles in an optimal way and hence of binary variables, as described above, may be compensated by the use of a slightly increased number of beams ([93]). Therefore we prescribe the number of beams and beam angles (as this also needs to be done for most clinical software packages) and give preference to the improvement of the model for beamlet weights optimization in the framework of continuous optimization by respecting biological considerations ([9]). Our model may be accompanied by some heuristic procedure to 'optimize' the beam angles ([68]). Also, in order to translate an obtained intensity profile into a sequence of MLC openings, an iterative procedure can be executed which normally leads only to little loss concerning the optimality of the goals ([7]). Other authors have recently suggested MILP procedures ([11], [35], [45], [57], [61]) or the solution of a nonlinear global optimization problem ([95]) for this purpose.

The problem of finding an IMRT treatment plan necessitates compromises between competing goals that may be rated differently. Accordingly, some authors recently have investigated this problem in the framework of *multi-criteria optimization*, with the aim to produce a set of treatment plans which relate to a different weighting of the objectives ([19], [32], [44], [59], [83]). However, compared to ordinary optimization, also multi-criteria optimization increases the complexity of the problem considerably so that practicability forces the treatable number of objectives to be relatively small and the number of beamlet weights to be not too large. In addition, the large numerical effort connected with a multi-objective approach suggests that the constraints in such model should be linear and the objectives linear or convex quadratic.

We therefore handle the IMRT treatment planning problem as an ordinary optimization problem and combine the solution of the problem with a sensitivity analysis, to foretell the approximate change of the optimal objective function value and, at least in some situations, of the EUD in the target which would be caused by a relaxation of optimization goals. In this way, if necessary, normally one or few constraints of the problem have to be changed only and most constraints can be left unaltered. The problems themselves, resulting from our approach, are convex or nonconvex optimization problems, which have the beamlet weights as variables and typically contain only $10 - 25$ constraints, apart from the simple bounds for the beamlet weights. This

is distinguished, for example, from linear models which include at least one inequality constraint and, by that, one additional slack variable for each volume element. Moreover, it is shown in this paper that our optimization model and algorithm for its solution, both presented in [9], can also be extended to the much larger problems of *intensity modulated proton therapy* (IMPT) treatment planning and can yield optimal solutions for these within a few minutes of computing times.

The general tools and our notations for the description of IMRT treatment planning problems are given in Sect. 2. In Sect. 3 we review the most prominent approaches to continuous beamlet weights optimization for inverse treatment planning, where we distinguish between linear, linear approximation type and multicriteria models and models which include or attempt to simulate nonlinear conditions on the doses. In particular the description of the latter models comprises that of our approach from [9]. The sensitivity analysis used in combination with this is presented in Sect. 4. Finally, in Sect. 5, we discuss treatment planning in connection with the $3D$ *spot scanning technique* of IMPT. The paper concludes in Sect. 6 with a case example for both IMRT and IMPT, where the optimization had been performed with the barrier-penalty multiplier method from [9].

2. Preliminaries. Radiotherapy and IMRT in particular require the selection of a number p of *radiation fields*, also called *incident beams*, and associated with that, p *beam angles* where, for practical reasons, normally p is a number between 3 and 6 and is smaller than 12. As we have argued in the introduction, we assume here that the fields and beam angles are predetermined in dependence on the kind and position of the tumor, by experience or by trial-and-error. Clearly, given a number of beams, the continuous optimization of both doses and beam directions would be desirable, but this is impeded by the computationally expensive dependence of the dose absorbed in the patient's body on the orientation of the radiation fields and the combinatorial nature of the problem.

Each of the p *radiation fields* is a $2D$ region with a polygonal boundary, normally originating from a projection of the PTVs onto a plane at the position of the collimator. Each field is partitioned into n_j rectangular *field elements* of equal size, also denoted as *bixels*, where typically the number n_j varies between 100 and $2,000$. Accordingly, each of the p beams is divided into n_j *beamlets* or *pencil beams* respectively so that the total number of beamlets over all fields amounts to $n := \sum_{j=1}^{p} n_j$.

The portion of the human body to be irradiated is considered to be divided into q not necessarily disjoint $3D$ *volumes* which represent the PTVs and the regions of normal tissue, as e.g. OARs. Furthermore, each of these q volumes is partitioned into m_ℓ $3D$ cubic *volume elements* or *'voxels'* of equal size, having a side length of normally ≥ 2 mm. Typically q is smaller than 15, and the total number $m := \sum_{\ell=1}^{q} m_\ell$ of voxels is of order 10^5 or 10^6. We number all volume elements consecutively from 1 to m and let V_ℓ $(\ell = 1, ..., q)$ be the index set of all elements belonging to the ℓ-th volume, having a cardinality $|V_\ell|$. For convenience, we also identify V_ℓ with the ℓ-th volume itself.

Let now $d_{jk} \geq 0$ be the dose deposited in the j-th volume element by the k-th beamlet at unit beam intensity and let $D := (d_{jk})$ be the resulting $m \times n$ *dose matrix*. This matrix D needs to be determined for each individual patient, which can be done by a Monte Carlo simulation of the radiation transport through the patient (cf. [62]) or with sufficient accuracy by a method which adapts a dose distribution computed for a homogeneous medium, so-called *pencil beam kernels*, to the geometry and density distribution of the patient ([2]). The dose matrix D is sparse since the

k-th beamlet predominantly affects volume elements only in proximity of its line of propagation. Typically, at a reasonable cut-off for the minimal dose, less than $3-8\,\%$ of the coefficients of D are nonzero so that D can be stored in a closed form.

For the optimization process the matrix D is assumed to be known. Then the goal of IMRT is to find, for each beamlet and according to the optimization goals of the respective model, a suitable nonnegative *beamlet weight* defining its radiation intensity. The total dose absorbed by the j-th volume element is linearly dependent on the vector $\phi \geq 0$ of beamlet weights, $\phi := (\phi_1, ..., \phi_n)^T$, and is given by

$$D_j^T \phi = \sum_{k=1}^{n} d_{jk}\phi_k \geq 0, \tag{2.1}$$

where D_j^T contains the entries of the j-th line of the dose matrix D. The n beamlet weights $\phi_k \geq 0$ are unknowns of an optimization model for IMRT treatment planning.

The technical realization of a *beamlet weight* or *intensity profile,* which nowadays is typically performed by a MLC, is a difficult problem in itself which is not discussed here. A MLC is part of the treatment machine and can expose a polygonal geometry formed by automatically shifted tungsten leaves. Hence, following the dose optimization, an intensity pattern has to be found for each field, which is close to the optimal profile determined by the optimization process and which can be generated by a relatively small number of MLC openings from the same position in space (typically $10-30$). Clearly, the a priori inclusion of a comprehensive set of constraints into an optimization model, which would guarantee that the optimal dose obtained by the model is realizable by a MLC, would be desirable (see [30], [86], [94], [96] for first models in this direction). For the optimization model used by the authors and for its objective function, however, the translation of an optimal intensity profile into one which can be realized by a MLC, can be performed with a negligible loss concerning the optimality of the doses ([7]).

3. Optimization models for IMRT treatment planning.

3.1. Introduction. In this section we discuss the main continuous optimization models related to inverse approaches for radiotherapy treatment planning. Considering the huge number of papers existing in this connection, we do not intend here to provide a complete review on the topic, but rather to survey the most prevalent ideas and problem types and point out their differences in terms of gains and drawbacks. So in general we give only more recent references, in which also discussions of earlier developments, including stochastic approaches, can be found.

In our review we do not distinguish between inverse radiotherapy treatment planning with and without intensity modulation, since the models used for intensity optimization of unmodulated beams have likewise or similarly been applied to IMRT or could in principle be applied to that. Often, and naturally before IMRT had been invented, the number of incident beams and hence continuous variables in the optimization problem were less than 12 and rarely more than 36. In contrast to that, the number of beamlets and hence continuous variables for IMRT typically amounts to $3,-8,000$, while an optimization problem resulting from IMPT treatment planning, having the same mathematical appearance as a model for IMRT, may possess $40,000$ variables and more (see Sect. 5). Also, for IMRT and IMPT, the resolution in regard to volume elements has to be increased considerably so that the responses of tissues to the inhomogeneous intensities caused by modulation of the beams can be traced appropriately.

Two special treatment techniques of radiotherapy are *tomotherapy* and *radiosurgery* (cf. [13], [87] and [38], [49] respectively for descriptions). While tomotherapy refers to a particular way of delivering IMRT, radiosurgery is a quite specialized treatment technique, which has been primarily designed to destroy malignancies in the brain. The basic mathematical ideas used for treatment planning models in both cases are similar to those for IMRT and are therefore included in our discussion (see e.g. [37], [38], [64], [88] for some recent developments concerning these topics).

For $x \in \mathbb{R}^r$, we employ the l_p-norm

$$\|x\|_p = (\sum_{i=1}^{r} |x_i|^p)^{1/p} \quad (1 \le p < \infty), \qquad \|x\|_\infty = \max_{i=1,\dots,r} |x_i|\,,$$

where the dimension r of the space is assumed to be clear from the circumstances. Especially $e \in \mathbb{R}^r$ is the vector with all elements being 1 and, for given $x \in \mathbb{R}^r$, the nonnegative vector $[x]_+$ is defined by

$$[x]_+ := (\max\{0,\, x_i\})_{i=1,\dots,r}\,.$$

The $|V_\ell| \times n$ matrix with lines D_j^T, $j \in V_\ell$, for some ℓ is denoted by $D_{(\ell)}$. Concerning standard concepts and algorithms of optimization used in our presentation, we refer to textbooks on optimization as e.g. [12], [39], and [78].

3.2. Linear programming models. Surveys on inverse approaches are found e.g. in [41], [49], [87], [101], and [102]. In the early approaches, almost always one treatment goal for each volume V_ℓ was to not exceed an upper dose bound of Δ_ℓ^u Gy, i.e. to satisfy the linear constraints

$$D_j^T \phi \le \Delta_\ell^u, \quad j \in V_\ell. \tag{3.1}$$

Typically, for each PTV V_ℓ, this was combined with the requirement to not fall short of a lower dose bound of Δ_ℓ^l Gy and hence, for $\Delta_\ell^l < \Delta_\ell^u$, to fulfill the constraints

$$D_j^T \phi \ge \Delta_\ell^l, \quad j \in V_\ell. \tag{3.2}$$

The purpose of such lower bound constraints is to guarantee a specified dose and, in combination with upper bounds as in (3.1), a nearly homogeneous dose in the targets. Sometimes, for the sake of a uniform description for all volumes, a lower dose bound is added also for each normal-tissue volume, where this can be set identical zero (see (2.1)). In this way, a large system of linear inequalities

$$A_\ell \phi \le b_\ell \ \ (\ell = 1, ..., q), \qquad \phi \ge 0, \tag{3.3}$$

is obtained, with $A_\ell \in \mathbb{R}^{s_\ell \times n}$, $b_\ell \in \mathbb{R}^{s_\ell}$, $\phi \in \mathbb{R}^n$, and $s_1 + \ldots + s_q \ge m \gg n$, where different actions described in the following have been taken to deal with such system.

Some authors have been of the opinion that each vector ϕ of the (often relatively small) feasible set of the system in (3.3) would be of equal clinical value and have proposed algorithms to find such vector, where special measures have to be considered in case the feasible set is empty (see e.g. [27] and, for a more recent development, [107]). A feasible point of a linear system of inequalities can be computed especially by phase 1 of the *Simplex algorithm*. Moreover, the inequalities satisfied with equality or almost equality for a solution of phase 1 give information about the constraints which should be relaxed in case of infeasibility.

Most authors, however, proposed to determine a feasible vector for the system in (3.3) which minimizes or maximizes some objective function, where different views have been taken concerning a suitable goal to be reached. If we let

$$f_{\mathcal{P}}(\phi) := \frac{1}{\Pi} \sum_{\ell \in \mathcal{P}} \sum_{j \in V_\ell} D_j^T \phi = \frac{1}{\Pi} \sum_{\ell \in \mathcal{P}} \left\| D_{(\ell)} \phi \right\|_1 , \tag{3.4}$$

with $\Pi := \sum_{\ell \in \mathcal{P}} |V_\ell|$ be the *integral dose* for some index set $\mathcal{P}$ of volumes and if we let especially $\mathcal{Q} := \{1, ..., q\}$ be the index set of all volumes, $\mathcal{N}$ that of all normal-tissue volumes including OARs, and $\mathcal{T}$ that of all PTVs, then typical goals have been the minimization of $f_{\mathcal{Q}}(\phi)$ or $f_{\mathcal{N}}(\phi)$ and the maximization of $f_{\mathcal{T}}(\phi)$ or $f_{\mathcal{T}}(\phi) - f_{\mathcal{N}}(\phi)$ (see especially [84] for the latter). In these functions, the factor related to $1/\Pi$ can be ignored for minimization and the sum of the doses for each volume may be weighted differently, according to its presumed importance. In addition, maximization of the minimal dose in the PTVs was suggested (e.g. [65], [69], [80]), which is equivalent to maximization of the variable τ over all vectors (τ, ϕ) under the additional constraints $\tau \leq D_j^T \phi$ $(j \in V_\ell,\ \ell \in \mathcal{T})$. Finally, also the minimization of a linear combination of some linear functions has been suggested, including the integral dose over all volumes and the maximum beamlet weight ([46]). The latter goal can be expressed by a new variable $\phi_{\max}$ and the inclusion of the additional constraints

$$\phi_k \leq \phi_{\max} \quad (k = 1, ..., n). \tag{3.5}$$

Eventually, in [65], an objective function, including linear penalties on the non-preferred beamlet weights, has been investigated.

The resulting *linear programming* (LP) problems include at least one inequality constraint for each voxel. Therefore, in case of IMRT, these problems comprise tens or hundreds of thousands of inequality constraints and as many slack variables in addition to the n unknown beamlet weights, since most codes start from the standard form of a LP problem which requires the introduction of such variables. In case, for the doses, new variables d_j with

$$d_j = D_j^T \phi \quad (j = 1, ..., m), \tag{3.6}$$

are introduced in the problem and the inequalities are written in terms of these d_j's as several authors do, the problem is even enlarged by m variables and equality constraints. Thus LP treatment planning problems typically are large-scale problems in regard to the number of variables and constraints, even for conventional radiotherapy with unmodulated beams. Such problems usually have been solved by the Simplex algorithm and, more recently, also by software packages as CPLEX (e.g. [69], [82], [87]), which includes a LP barrier *interior-point method*.

In clinical routine, initially set up treatment goals often turn out to be too restrictive. Consequently, a natural shortcoming of such LP and any other optimization problem including both upper and lower dose bounds is that the related inequalities may be inconsistent. For this reason, *elastic constraints* have been introduced, which include parameters that allow some over- and underdosage of volume elements, and thereby avoid the possible infeasibility of the inequality system. In the quite general framework of [48], [49], a system $A_\ell \phi \leq b_\ell$ can be replaced, for example, by

$$A_\ell \phi \leq b_\ell + \theta_\ell u_\ell, \tag{3.7}$$

with some given vector $u_\ell > 0$ from $\mathbb{R}^{s_\ell}$ and some parameter $\theta_\ell \geq 0$, where the related sum $\sum_\ell w_\ell \theta_\ell$ with *importance weights* $w_\ell > 0$ defines the objective function. (More generally $u_\ell \theta_\ell$ can be a matrix-vector product with an unknown vector $\theta_\ell \geq 0$.) In this case, each vector $(\phi, \theta_\ell) \geq 0$ with a sufficiently large θ_ℓ satisfies the inequality system. Elastic constraints have similarly been employed for a LP problem in [46] and for a multicriteria weighted sum approach in [44], in which the weights w_ℓ of the objective function $\sum_\ell w_\ell \theta_\ell$ are varied (see Sect. 3.4).

Note that, for $u_\ell := e$ in (3.7), the problem of minimizing the term $w_\ell \theta_\ell$ alone over all vectors $(\phi, \theta_\ell) \geq 0$ under the constraints in (3.7) is equivalent with the problem of minimizing, over all $\phi \geq 0$, the function

$$F_\ell(\phi) := \left\| [A_\ell \phi - b_\ell]_+ \right\|_\infty , \tag{3.8}$$

i.e. the maximum violation of the system $A_\ell \phi \leq b_\ell$. Moreover, if V_ℓ is a target volume and the system in (3.7) stands for

$$\Delta_\ell - \theta_\ell \leq D_j^T \phi \leq \Delta_\ell + \theta_\ell, \quad j \in V_\ell,$$

with some dose $\Delta_\ell > 0$, it is equivalent with the *linear Chebyshev approximation problem* of minimizing

$$F_\ell(\phi) := \|A_\ell \phi - b_\ell\|_\infty \tag{3.9}$$

with respect to $\phi \geq 0$ ([31]). Thus, problems with elastic constraints are closely related to minimum norm (type) problems discussed in Sect. 3.3.

An attempt to overcome some of its limitations and to still remain in the framework of LP is provided in [82], where it is suggested to approximate a convex voxel-based objective function, as given, for instance, in Sects. 3.3 and 3.5, by a piecewise linear function. The price for such action is that, in this way, at least $K \cdot m$ inequality constraints and hence slack variables are added to the problem, where, for the numerical results, K was a number between 2 and 4. Also several authors including those of [82] suggest LP approaches to deal with *partial-volume constraints,* which either are continuous nonlinear or include binary variables. The latter approaches are discussed in Sect. 3.5.

3.3. Linear approximation type models. The possible inconsistency of the constraints in a LP approach to the treatment planning problem has stimulated the study of various constrained linear approximation problems, with the aim of finding an intensity weight vector which is nearest to the desired goals in some sense. Some authors have considered the (squared) simple-bound constrained *linear least-squares approximation problem*

$$\min_{\phi \geq 0} \sum_{\ell=1}^{q} w_\ell \frac{1}{|V_\ell|} \|A_\ell \phi - b_\ell\|_2^2 \tag{3.10}$$

(e.g. [50], [108], [109]). Alternatively the simple-bound constrained *Chebyshev approximation problem*

$$\min_{\phi \geq 0} \max_{1 \leq \ell \leq q} \left\{ w_\ell \frac{1}{|V_\ell|} \|A_\ell \phi - b_\ell\|_\infty \right\} \tag{3.11}$$

has been investigated ([47]). Both problems always have a solution (cf. Remark 3.1). However, minimum norm problems of this type can be interpreted as an attempt

to find an approximate solution of an overdetermined system of equations and hence force all normal tissue volumes to receive doses closely below or above the given upper bounds, which usually is not desirable.

The latter drawback is remedied if, for all normal-tissue volumes, one approximates zero doses with respect to the (squared) weighted l_2-norms, under homogeneity constraints on the targets. This corresponds to a constrained linear least-squares approximation problem of the type

$$\begin{array}{lrcll} \text{Minimize} & \sum_{\ell\in\mathcal{N}} w_\ell \dfrac{1}{|V_\ell|} \left\|D_{(\ell)}\phi\right\|_2^2 & & & \\ \text{s.t.} & A_\ell\phi & \leq & b_\ell & (\ell\in\mathcal{T}), \\ & \phi & \geq & 0, & \end{array} \tag{3.12}$$

where the inequality system stands for lower and upper dose bounds ([54]). The problem in (3.12) much resembles the aforementioned simpler LP problem for the objective function (3.4) with $\mathcal{P} := \mathcal{N}$ and additional importance weights, which is obtained if the (squared) l_2-norm in (3.12) is exchanged for the l_1-norm.

Essentially interchange of the roles of $\mathcal{T}$ and $\mathcal{N}$ in (3.12) yields the alternative problem

$$\begin{array}{lrcll} \text{Minimize} & \sum_{\ell\in\mathcal{T}} \|A_\ell\phi - b_\ell\|_2^2 & & & \\ \text{s.t.} & A_\ell\phi & \leq & b_\ell & (\ell\in\mathcal{N}), \\ & \phi & \geq & 0, & \end{array} \tag{3.13}$$

which has been investigated e.g. in [65] (see also the references in [49]). The matrix inequality constraints in (3.13) typically result from upper dose bounds for healthy volumes so that $\phi := 0$ is feasible for the problem. Instead of the squared l_2-norm in (3.12) and (3.13) one may also exploit the maximum norm, which for problem (3.13) was done in [25]. Other meaningful variations of linear minimum norm problems can be found, for example, in [49] and [87]. Especially the linear least-squares problems with linear constraints can be written as ordinary *quadratic programming* (QP) problems, while (linearly constrained) problems involving the l_1- or l_∞-norm typically can be transformed straightforwardly into LP problems ([31]), where the latter is true for all l_∞-problems given here. Thus the l_2-problems can be solved by an algorithm for QP or some *nonlinear programming* (NLP) method like a *penalty type method* ([54]) or a *gradient projection method* ([13], [16]), while linearly constrained linear Chebyshev approximation problems can be solved by the Simplex algorithm or an interior-point method.

A very popular modification of the least-squares approach in (3.10), which avoids its drawbacks and, in the same context, likewise does include only simple-bound constraints, is to let merely those constraints of the system in (3.3) enter the linear approximation problem, at least for the normal-tissue volumes, which are violated for ϕ (e.g. [13], [14], [16], [36], [53], [92], [105]). The resulting convex simple-bound constrained *linear least-squares type problem* has the form

$$\min_{\phi\geq 0} \sum_{\ell=1}^{q} w_\ell \frac{1}{|V_\ell|} F_\ell(\phi) \tag{3.14}$$

where F_ℓ equals either the quadratic function

$$F_\ell(\phi) := \|A_\ell\phi - b_\ell\|_2^2 \tag{3.15}$$

or the piecewise quadratic function

$$F_\ell(\phi) := \left\|[A_\ell\phi - b_\ell]_+\right\|_2^2 \tag{3.16}$$

and the *importance weights* $w_\ell \geq 0$, not all being zero, may be normalized such that

$$w := (w_1, ..., w_\ell)^T, \quad \|w\|_1 = \sum_{\ell=1}^{q} w_\ell = 1. \tag{3.17}$$

In medical papers a function F_ℓ of type (3.15) or (3.16) entering an objective function often is denoted as a *penalty (function)* for the volume V_ℓ (e.g. [16]), which may irritate optimizers, who rather associate a penalty method with this term. Typically, the quadratic function in (3.15) is used for a PTV, like in problem (3.13), and the piecewise quadratic function in (3.16) for each other volume (e.g. [13], [16], [36], [53], [92]). This approach has been realized in most clinical software, e.g. in the package KonRad of the German Cancer Research Center in Heidelberg ([18], [79]).

The least-squares type problem in (3.14) has been solved, for example, by a scaled gradient projection algorithm ([13]), a variant of a Newton projection method ([36]), and by an active set method ([53]). Some authors also heuristically adapt gradient type methods for unconstrained problems, like conjugate gradient methods, to problems with constraints. Others consider a least-squares type problem including functions of type (3.16) as an ordinary QP problem, which, however, can lead to errors. In particular note that the function in (3.16) possesses a first continuous derivative on $\mathbb{R}^n$, but typically is not twice continuously differentiable everywhere. (If existence of second derivatives is required for an algorithm, it has to be replaced by $F_\ell(\phi) := \left\|[A_\ell\phi - b_\ell]_+\right\|_p^p$ with some $p > 2$.)

Note that, if $F_\ell(\phi) := \left\|[A_\ell\phi - b_\ell]_+\right\|_2^2$ is used in (3.14) for all $\ell \in \{1, ..., q\}$ as in [13], each feasible point of the related linear inequality system is a minimizer of (3.14) with objective function value zero. Hence, in this case, the least-squares type approach in (3.14) is distinguished from the LP feasible-point approach mentioned in Sect. 3.2 only insofar as the types of problems motivate the use of different algorithms and different measures in case the system is inconsistent.

For the LP feasible-point approach, also in [107] an algorithm is discussed which always finds the unique feasible point for which $\|\phi\|_2$ becomes minimal. This latter approach may be viewed as an attempt to find a feasible point which produces a small integral dose over the irradiated volume. Evidently, the feasible point of a linear system having minimal Euclidean norm could also be found by solution of a linearly constrained QP problem with objective function $\|\phi\|_2^2$ ([26]). If the squared Euclidean norm $\|\phi\|_2^2$ in this problem would be exchanged for the maximum norm $\phi_{\max} := \|\phi\|_\infty$, serving the same goal, the problem even could be solved as a LP problem with the additional constraints from (3.5). Observe in this connection also that, if F_ℓ in (3.14) is alternatively defined by (3.9) and (3.8) respectively, i.e. if the the maximum norm is employed rather than the squared l_2-norm, then problem (3.14) is equivalent to the LP problem

$$\begin{array}{lrcl} \text{Minimize} & \sum_{\ell=1}^{q} w_\ell \dfrac{1}{|V_\ell|}\theta_\ell & & \\ \text{s.t.} & A_\ell\phi & \leq & b_\ell + \theta_\ell e \quad (\ell = 1, ..., q), \\ & (\phi, \theta) & \geq & 0, \end{array}$$

with $\theta := (\theta_1, ..., \theta_\ell)^T$, which just is a prominent case of the LP elastic-constraints approach from [48], [49].

Our discussion reveals that the LP and quadratic (type) approaches to the IMRT treatment planning problem discussed up to now are closely related to each other. By their nature all of these problems are linear in some sense and, from the computational point of view, it may also be desirable to achieve linearity for an approach. On the other hand, it is well-known that the response of a complex organ to radiation does depend on the absorbed dose in a nonlinear way and therefore cannot be modelled by linear functions straightforwardly (see Sect. 3.5). Furthermore, the linearity of an approach normally is paid by the presence of at least one constraint for each voxel and it is by no means clear that a LP problem with a very large number of inequality constraints has to be preferred, for example, to a nonlinear convex problem with no or only a small number of inequality constraints, in addition to the simple bounds for the beamlet weights. Especially large numbers of quite similar linear constraints, as they are generated by some discretization process (concerning the volumes here), typically lead to very ill-conditioned constraint matrices and hence may be liable to numerical difficulties at solution of such problems.

A first natural extension of especially the model in (3.14) - (3.17) would be to consider, for each volume V_ℓ, some constraint

$$G_\ell(\phi) \leq 0$$

with some sufficiently smooth goal function G_ℓ defined on a proper subset of $\mathbb{R}^n$, where for simplicity we assume here the presence of only one goal for each volume. Then, in generalization of problem (3.14), we arrive at the problem

$$\min_{\phi \geq 0} \sum_{\ell=1}^{q} w_\ell \frac{1}{|V_\ell|} \left[G_\ell(\phi)\right]_{(+)}^2 . \tag{3.18}$$

where $[\,\cdot\,]_{(+)}^2$ stands for either $[\,\cdot\,]^2$ or $[\,\cdot\,]_+^2$. This is a convex optimization problem if, for example, G_ℓ is convex in each term $[G_\ell(\phi)]_+^2$ and linear else.

REMARK 3.1. *There is some confusion in medical papers concerning local minimizers. Especially there are some inaccuracies, concerning the following, in the much cited paper [33], which are repeated in other papers. (See also Remark 3.2). Note that all minimization problems, studied up to this point and preceding (3.18), are convex problems.*

Consider a convex minimization problem '(CP)' in $\mathbb{R}^n$ with a nonempty feasible set. Then each local minimizer of (CP) also is a global minimizer or, shortly, a 'solution' of the problem. However, (CP) does not necessarily possess a solution. (Confer the problem $\min_{x\in\mathbb{R}} e^x$.) But, for the LP and linearly constrained QP problems of this and the previous section, existence of a solution is guaranteed if the set of feasible points is nonempty since their objective function is bounded below by zero on the respective feasible set (e.g. [103, p. 130]). For problem (3.14) - (3.17), the existence of a solution can be proved along the lines of the proof given for the example case of [9].

If the solution set of (CP) is nonempty, this is a convex set and hence either consists of a single point or is infinite. A sufficient condition for the existence of only one solution is that the objective function is strictly convex (e.g. [12], [39]). Especially the objective function related to a linear least-squares problem is strictly convex if the matrix A, associated with such problem, has full column rank or, equivalently, if $A^T A$

is nonsingular (e.g. [78]). In case a problem has more than one solution, algorithms may converge to different solutions from different starting points.

3.4. Multi-criteria optimization models. The choice of the weights w_ℓ in (3.14) and (3.18) respectively is quite arbitrary. Especially for a prescribed selection of these weights, the maximum amount of a possible constraint violation for a particular volume at a solution of the problem is not predictable and may turn out to be not acceptable clinically. In fact, it has been reported that computed doses are extremely sensitive to the selection of weights (e.g. [44], [69]). Therefore, by trying different settings of weights, one may end up in a very time-consuming trial-and-error process.

From its nature, the problem of finding a radiotherapy treatment plan is a *multi-criteria optimization problem* (e.g. [56]) with a finite number of well-defined objective functions. Such a problem is associated with a manifold of solutions, the *(Edgeworth-) Pareto minimal points*, which refer to the differing importance that may be given to the single objectives. These Pareto minimizers are closely related to minimizers of the *scalar optimization problem* in (3.14) which are obtained for different weights $w \geq 0$. Especially, if the F_ℓ ($\ell = 1, ..., q$) are any convex functions, a solution of problem (3.14) for a given weight vector $w \geq 0$ with $\|w\|_1 = 1$ is a Pareto minimizer of the problem associated with the q objectives F_ℓ, and, conversely, each *properly (Edgeworth-) Pareto minimal point* of that problem solves problem (3.14) for some weights $w > 0$ with $\|w\|_1 = 1$ ([56, p. 299]). The determination of Pareto minimizers via such scalar optimization problem is known in the framework of multicriteria optimization as the *weighted sum approach*. In practice, it requires the solution of a finite set of optimization problems as in (3.14), generated for a proper discrete set of weight vectors $w \geq 0$ with $\|w\|_1 = 1$. Typically either all solutions for a uniform grid of weights in $[0,1]^q$ are offered to the decision maker or the solution of the scalar problems is accompanied by some decision process, according to which irrelevant solutions are ignored and a suitable solution is extracted for use.

Several authors have recently studied multicriteria weighted sum approaches for radiotherapy treatment planning. In [110] the problem in (3.10) is studied in a multicriteria setting and the obtained plans are evaluated by a *dose volume histogram* function. Another approach of this type is discussed in [59] for

$$\begin{aligned}\mathcal{F}(\phi) := w_1 \sum_{\ell \in \mathcal{T}} \frac{1}{|V_\ell|} \left\| D_{(\ell)}\phi - \Delta_\ell e \right\|_2^2 + w_2 \sum_{\ell \in \mathcal{NT}} \frac{1}{|V_\ell|} \left\| D_{(\ell)}\phi \right\|_2^2 \\ + \sum_{\ell \in \mathcal{O}} w_{3,\ell} \frac{1}{|V_\ell|} \left\| \left[D_{(\ell)}\phi - \Delta_\ell e \right]_+ \right\|_2^2, \end{aligned} \tag{3.19}$$

where the Δ_ℓ are reference doses and $\mathcal{T}$, $\mathcal{O}$, and $\mathcal{NT}$ are the index sets of all volumes representing PTVs, OARs, and the remaining normal tissues respectively. (It is not clear to us whether, in [59], one common objective function is taken for all OARs or whether one function is chosen for each single OAR.) Solutions are computed for all weights $w \geq 0$ on a uniform mesh in $[0,1]^{2+|\mathcal{O}|}$ so that the number of mesh points determines the number of problems to be solved. Before, in [32], each OAR had been weighted separately, but the computation time needed for more than 3 goals in total had been found too large. Thus, if there exist significantly more than 3 goals in such type of approach, some of the goals for OARs have to be combined in one objective function. The latter, however, has to be seen critical since, for example, differing distances of OARs from PTVs may require a more subtly differentiated action. (Compare e.g. the example case of Sect. 6 and those in [9], which include up

to 25 goals for head-and-neck cancer cases).

The scalar problem of minimizing the convex function in (3.19) subject to the simple bounds $\phi \geq 0$ could be solved, for example, by some gradient projection method. In order to arrive at an unconstrained optimization problem, the authors of [32] and [59] recommend instead, to replace the weights ϕ_k by weights $\psi_k^2 := \phi_k$. However, this transformation may have consequences concerning the convergence of the used algorithm ([39, p. 147]) and, what is not mentioned, transforms the convex problem into a nonconvex one so that nonglobal local minimizers have to be discussed. (The function $f(x) := (x-a)^2$ with some $a > 0$ is convex, but $g(y) := (y^2-a)^2$ is not.) Then the nonconvex problems are solved by a conjugate gradient method ([32]) and the limited memory BFGS method (e.g. [78]) respectively. Note at this point that several authors use (quasi-) Newton type methods which directly or indirectly need second derivatives, though these do not exist in all points, for example, when functions including expressions of type $\|[\,\cdot\,]_+\|_2^2$ are used. But such action is known to possibly lead to very slow convergence ([91]), as the iteration numbers reported in [59] also seem to indicate.

In [44] a linear multicriteria weighted sum approach is studied using elastic constraints (see Sect. 3.2), where the allowed maximum violations of the constraints for the various structures form the objectives. The approach is combined with a strategy to find a certain representative subset of Pareto solutions rather than to compute solutions for all weights vectors in a grid of $[0,1]^q$, and some numerical experiments are presented.

A more sophisticated multicriteria optimization approach, which requires the solution of optimization problems including constraints on the various goals, is developed in [19] and [97]. The aim again is to find suitable representatives of the set of Pareto minimizers, where the total number of problems to be solved does depend only on the number q of goals (which, as is said, should not exceed about 6) and not on the chosen fineness of a mesh. This approach also makes use of the EUD model (see Sect. 3.5) where, for the numerical realization, the l_p-norm, $1 < p < \infty$, on the dose in the EUD function ([76]) is replaced by a suitable convex combination of the l_1- and l_∞-norm ([99]). This replacement has the advantage of leading to LP problems, but for the price that a single nonlinear convex constraint for a volume V_ℓ is exchanged by $|V_\ell|$ linear constraints. Strategies to reduce the large number of linear constraints are implemented and numerical experiences with the total approach are reported.

The authors of [83] discuss a unifying framework providing conditions under which multicriteria optimization problems including well-known nonconvex treatment planning criteria can be transformed into problems with convex criteria, having the same set of Pareto minimizers.

3.5. Nonlinear conditions. The LP and similarly the linear approximation type models for IMRT considered up to this point are merely based on *physical criteria*, i.e. on measurable physical quantities as volumes and doses. It has been observed by a number of authors that such approaches have serious limitations (see e.g. the discussions and references in [21], [71], [104], [106]). They take the biology of radiation into account only insofar as they try to avoid critical structures, but they do not adequately model the responses of healthy and tumorous tissues to radiation, which behave neither linearly nor quadratically. The sensitivity of a healthy organ to radiation does not simply depend on the maximum dose absorbed by some of its volume elements, but rather on the total dose distribution in the organ. Moreover, for example, a *cold spot*, i.e. a small underdosed volume, in a target may not much

influence a quadratic objective formed by the differences of desired and actual doses, but may significantly reduce the tumor control probability.

Therefore the insertion of biological considerations for both dose prescriptions and the rules for control of their violation has been proposed (e.g. [17], [22], [40], [71], [72], [81]), and alternative biological optimization models, which respect the dose responses of the different tissues and the response to inhomogeneous dose distributions, have been developed (e.g. [3], [5], [8], [9], [42], [58], [100]). Biological conditions are inherently nonlinear so that their direct implementation necessarily leads to large-scale nonlinear convex or nonconvex optimization problems. Naturally, these may have multiple local minimizers, but almost always useful ones seem to have been found (see [14], [33], and [85] for studies in this connection).

Several authors have studied objective functions in an optimization model, representing *normal tissue control probability* (NTCP) and *tumor control probability* (TCP). The authors of [100] optimize, by some gradient technique, an objective function including both probabilities and dose-volume criteria in addition. In [41], which integrates earlier results from [42], [43], [89], and [90], various formulations of optimization problems with biologically motivated linear constraints and a nonlinear objective function have been studied, including the *probability of uncomplicated tumor control* $P_+ := P_B - P_{B\cap I}$ as an objective function, where P_B is the probability of tumor control and $P_{B\cap I}$ is the probability of simultaneous tumor control and severe normal-tissue complications. For the solution of these (by today's standards relatively small) problems, several algorithms based on an augmented Lagrangian approach have been compared with a *Sequential Quadratic Programming* (SQP) method, where it was found that the augmented Lagrangian approach, combined with a limited memory BFGS method, was the most favorable method. In [53] a probability function $P_- := 1 - P_{++}$ was minimized under the constraints $\phi \geq 0$ by an active set method, where P_{++} is taken from [1] and similar to P_+ (see the discussion in [41]). However, it has been remarked that these types of probability functions "are simplistic, and the data they rely on are sparse and of questionable quality ([100])".

The authors themselves favor use of the *logarithmic tumor control probability* (LTCP)

$$\mathrm{LTCP}(\phi; V, \Delta, \alpha) := \frac{1}{|V|} \sum_{j \in V} \exp(-\alpha\,(D_j^T \phi - \Delta)) \tag{3.20}$$

for each PTV V as objective function ([9]), where $\Delta > 0$ is the total dose requested for V and $\alpha > 0$, related to cell survival, is the only biological constant needed . Minimization of this convex function is equivalent to the maximization of the TCP function (see [74])

$$\prod_{j \in V} \exp\left\{ -\frac{1}{|V|} \exp(-\alpha\,(D_j^T \phi - \Delta)) \right\}.$$

Without sufficient prevention, e.g. adequate dose bounds on the normal tissue volumes, minimization of the LTCP could result in a prohibitively high dose in the targets (as is criticized in [49, p. 4-25]). Therefore, for each PTV V, the authors use a *quadratic overdosage penalty* (QOP) constraint of the type

$$\mathrm{QOP}(\phi; V, \Delta, \delta) := \frac{1}{|V|} \sum_{j \in V} [D_j^T \phi - \Delta]_+^2 - \delta^2 \leq 0, \tag{3.21}$$

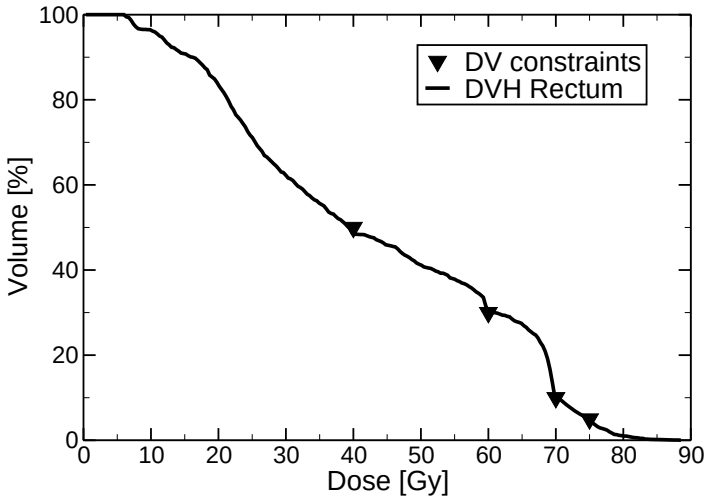

FIG. 3.1. *Cumulative DVH of a rectum in a prostate example case. Four DV constraints were set for the optimzation: (40 Gy/50%), (60 Gy/30 %), (70 Gy/10 %), (75 Gy/ 5%). Each constraint ensures that no more than y % of the organ volume receive more than x Gy dose. The treatment dose of the prostate was 84 Gy.*

where $\delta > 0$ is a given bound. Such a constraint prevents an excessively high dose in V and simultaneously allows a mild mean violation of the acceptable dose Δ in V by some δ. A constraint of this type is also applied to permit a certain overdosing of some volume V neighboring a PTV, since a sharp dose drop from the PTV to V is not realizable physically. Note that the function $\mathrm{QOP}(\,\cdot\,; V, \Delta, \delta)$ is once but not twice continuously differentiable everywhere so that the power 2 in QOP has to be increased by at least one if second derivatives of functions are needed in an algorithm.

In this connection observe that an underdosage in some voxels of V for a solution of some optimization problem, involving a term as in (3.20) in its objective function, would lead to positive powers in the exponential function and hence tends to affect the objective function value considerably more than in case of a quadratic (type) function. This observation implies intuitively, though not rigorously, that using an additional minimum dose constraint for V as in (3.2) would not significantly increase the actual minimum dose attained in such a program. Therefore, like other authors, we avoid the implementation of lower dose bounds since they may cause infeasibility of the program and hence difficulties for algorithms. In either case, when cold spots are detected in a target or if a system of constraints turns out to be inconsistent, the original treatment goals need to be reconsidered and modified.

It has been generally accepted that, for each involved critical *parallel organ* (lung, parotid gland, kidney, etc.), an optimization model should reflect the property that a certain percentage of such organ can be sacrificed without too serious consequences for the patient, if this is of advantage for the overall treatment. Thus, instead of merely pursuing the goal for a particular parallel OAR to stay below an upper dose bound, the model should provide a solution exhibiting an acceptable dose distribution for this organ in regard to the dose versus the percentage-of-volume. Such a relationship can be depicted in a *cumulative dose-volume histogram* (DVH) and is typically considered, in combination with other criteria, to evaluate the quality of a treatment plan.

In this connection many authors start from an ideal clinical DVH curve and intend

to produce dose distributions which match these inherently nonlinear curves at least in one or multiple points. Constraints in a model which are designed for this purpose are often denoted as *dose-volume* (DV) *constraints* (see Fig. 3.1 and e.g. [41, p. 32] for a summary of the application of such constraints).

Some authors suggest the inclusion of certain linear DV constraints to remain within a LP framework. In [73] it is proposed to form several 'collars' around a target and to fix an upper dose-bound constraint of type (3.1) for each such neighborhood, where the dose bound is decreasing with increasing distance from the target and the thickness of the collars is determined by the percentage of volume elements which shall be below the given bound. This procedure is modified for IMRT in [46] where, in addition to the distance from the target, a heuristics concerning the expected number of beamlets meeting a structure is utilized in order to select the voxels related to a certain dose bound. A new type of linear constraints which is derived from a technique used in finance and bounds the tail averages of DVHs is proposed in [82], but entails a number of artificial variables proportional to the number of voxels of the respective structures. While in these approaches the LP program remains unchanged during the iteration process, the authors of [69] employ dose bounds as in (3.1) for subvolumes of a particular structure and make use of sensitivity information to adapt the respective voxel sets in each iteration, in case DVH requirements are not satisfied for the current solution. A dynamic adaptation of such linear bounds is also applied in [52] in combination with a least-squares objective function for the PTVs. However, it is not clear whether these latter procedures always converge to a desired solution.

Another technique of respecting DV conditions, which is applied by some authors and was developed in [18] in connection with the least-squares type problem (3.14) - (3.17) discussed above, is to check at each iteration whether the current solution meets a particular DVH specification, and to else add a 'penalty' $w_\ell[D_j^T\phi - \Delta]_+^2$ for certain $j \in V_\ell$ to the objective function, like, for example, for all or some of those voxels which exceed a desired dose Δ, assuming that the number of these voxels is greater than a permitted number (e.g. [88], [92]). Differing from that, in [29], a continuous, though not everywhere differentiable, linear-quadratic 'penalty' function defined on the total respective volume is added to a least-squares function for the PTVs, where this penalty is multiplied by a factor depending on the current fraction of the structure surpassing a required dose. Similarly, in [28], a least-squares error function is adapted properly in each iteration so that a sequence of least-squares approximation problems is solved. Hence, in these approaches the objective functions of the optimization models are redefined during the iteration process, and it is not clear to what point such process converges, in case it converges at all.

The authors of ([70]) extend the feasible-point idea (see Sect. 3.2) to include a new type of (nonconvex) quasi-convex DV constraints and report satisfying results for an algorithm, which has originally been designed for solution of the convex feasibility problem only.

Though we intend to concentrate on continuous optimization models, we would like to point out that a mathematically rigorous description of pointwise DV constraints is to assign a binary variable to each element of the respective volume, depending on its dose level, and to combine a dose-bound constraint, including these new variables, with an additional constraint that only allows a desired percentage of these binary variables to be 1 (e.g. [11], [60], [63], [69]). In this way, tens or hundreds of thousands of binary variables are added to the continuous variables so that, by the increased complexity, the total treatment model needs to remain linear and much

energy has to be spent in regard to a suitable dealing with such very large-scale MILP problems.

Objections to the discussed pointwise DV conditions are that it cannot be anticipated how many of such conditions are needed so that an obtained solution produces a desired DVH curve. Also the prescription of an ideal DVH curve may entail a significant loss of freedom in the search space, since usually several DVHs have to be considered and curves with a modified shape may still lie inside the tolerance and result in an overall improvement for the patient. On the other hand, trial-and-error procedures in this respect are very time consuming. In addition, some of the approaches, at which definitions of objective functions or constraint sets are altered during the performance of an algorithm, lack a rigorous mathematical convergence analysis and are therefore uncertain concerning their outcomes.

The direct translation of a dose-versus-percentage constraint into a continuous mathematical condition, however, is known to lead to a nonlinear constraint. We apply especially the *partial volume* (PV) constraint

$$\mathrm{PV}(\phi; V, \Delta, p, \zeta) := \frac{1}{|V|} \sum_{j \in V} \frac{(\frac{1}{\Delta} D_j^T \phi)^p}{1 + (\frac{1}{\Delta} D_j^T \phi)^p} - \zeta \leq 0 \tag{3.22}$$

for a parallel OAR V with some constant $\zeta \in (0, 1)$ (for details see [3], [5], [55]). For example, in relation to our example case in Sect. 6, the data $\Delta := 20$, $p := 3$, and $\zeta := 0.1$ for the right parotid gland express that, at a dose of 20 Gy, a volume element of this organ loses 50 % of its function (e.g. production of saliva) and that at most 10 % of the total function of the organ may be lost. The constraint in (3.22) is formed by the sigmoidal function $\sigma(x) := x^p/(1+x^p)$, which has a relatively smoothly increasing step and hence offers some freedom concerning the dose distribution in V. (Alternative experiments with the sigmoidal error function can be found in [81], [87].) Note in this connection that constraints of type (3.22) can also be utilized to obtain continuous pointwise DV constraints as discussed above, when the step of the sigmoidal function is contracted and the function equals 0 left and 1 right from some properly chosen small step interval ([100]).

In contrast to conventional radiation therapy, IMRT normally leads to nonuniform dose distributions in organs. Niemierko ([76]) has introduced the *(generalized) equivalent uniform dose* (EUD)

$$\left\{ \frac{1}{|V|} \sum_{j \in V} (D_j^T \phi)^p \right\}^{1/p} \tag{3.23}$$

as a model for a biologically permissible nonuniform dose distribution in a volume V which, in regard to the irradiation response, is comparable to an uniform distribution of dose Δ. In this function, $p \in \mathbb{Z}$ is some tissue-specific power, which is negative for PTVs and positive for OARs. Note that for $p = 1$ the function in (3.23) becomes the mean dose and for $p = \infty$ the maximum dose for V, both used above. The EUD concept has by now been widely accepted, especially for *serial organs*, i.e. the spinal cord, nerves, and all other structures that can be seriously damaged by a high dose in a small spot. It has been observed that the use of the EUD model can lead to greater normal tissue sparing, compared to merely dose-based optimization ([106]). "Inverse planning based on the probabilities of tumor control and normal tissue complication remains the ultimate goal, and the uniform dose is a step in this direction ([98])".

The EUD concept has been applied for optimization especially in [9], [19], [99], [82], [97], [98], [104], and [106]. Especially the authors of [106] investigate an objective function which makes use of the EUD model for tumors as well as normal tissues. The resulting nonconvex function is minimized by an (unspecified) gradient technique. In [104], the EUD based model is combined with a dose-volume approach to further improve the treatment plans. The recent convex approach from [97], [98] employs an upper bound on the EUD as an optimization constraint for all OARs and PTVs and a lower bound on the EUD of the PTVs. In this way some convex constraint set is obtained and a quadratic least-squares error function, which is adapted in each iteration similarly as it has been suggested in [18] for DV constraints (see above), is minimized over this set by a componentwise Newton method which is combined with a projection technique.

We ourselves employ an EUD constraint of the type

$$\mathrm{EUD}(\phi; V, \Delta, p, \varepsilon) := \frac{1}{|V|} \sum_{j \in V} (\frac{1}{\Delta} D_j^T \phi)^p - \varepsilon^p \leq 0 \tag{3.24}$$

for each serial OAR V only, where $\Delta > 0$ is some given dose, $\varepsilon > 0$ a given constant, and $p \geq 1$ some tissue-dependent power ([9]). For instance, in the optimization problem of the example case in Sect. 6, we include an EUD constraint for the spinal cord with the settings $\Delta := 28$, $p := 12$, and $\varepsilon := 1$. This constraint effects that only a tiny excess of 28 Gy is allowed for fractions of this organ, with the extent of overdosage depending sensitively on the size of the volume in which it occurs. Note that a single convex constraint as in (3.24) normally replaces the $|V|$ constraints which enter a program if an upper dose bound as in (3.1) is set for V.

The functions $LTCP$, QOP, and EUD are nonquadratic convex and the function PV is nonconvex. Moreover, in this ideal description concerning the beamlet weights (see [9] for this), the zero vector is feasible for the respective constraints. Therefore use of these functions leads to a feasible convex or, if the irradiation of e.g. parotid glands and lungs is to be controlled, nonconvex optimization problem with sufficiently smooth functions in n variables. Technical limitations of a MLC may enforce additional constraints on the weights which have to be included in the program (see [6], [7] for examples). However, in contrast to, for example, LP models, the total model involves rarely more than $15 - 20$ constraints. Finally, in view of the possible nonconvexity of a problem, we would like to mention that we have never found useless 'solutions'.

The resulting nonlinear optimization problems are solved by a modification of a recent barrier-penalty multiplier method ([9]) and combined with a suitable sensitivity analysis (see [4] and the following section). Our algorithm has shown to need relatively small execution times also for high dimensional problems in, by now, several hundred clinical examples. Both, the algorithm and the sensitivity analysis, are also implemented in the software package HYPERION, which has been developed at the University Hospital in Tübingen and is already used in daily clinical routine in several hospitals in Germany and the USA.

REMARK 3.2. *In medical papers, it is sometimes not recognized that an optimization problem is convex and that, for a convex problem, each local minimizer also is a global minimizer. Instead, the existence of nonglobal local minimizers seems sometimes to be regarded as a property of certain algorithms like gradient type methods, and local solutions with differing objective values, which have been obtained by different algorithms or by the same algorithm but for different starting points, are debated also for convex problems. The observation of multiple 'solutions' with distinct objective*

values, however, seems to be rather caused by the heuristic nature of the respective algorithm(s) so that convergence to a global minimizer is not guaranteed, and/or by the fact that the initial optimization problem is altered during the iteration process, since an objective function or the constraint set is adapted at each iteration.

It is also often not recognized that, for standard descent algorithms, convergence is proven, under suitable assumptions, only to a point which satisfies the first-order necessary optimality conditions of the problem. Thus in the worst case, applied to a nonconvex problem, a descent algorithm may get trapped in a point that is not a local minimizer, e.g. in a saddle point in case of an unconstrained problem. This argument especially applies to a strongly quasiconvex objective function, as discussed in [33]. If such function possesses a local minimizer, this also is the unique global minimizer, but a strongly quasiconvex function can also have saddle points ([10, p. 113]) and no local minimizer at all, which both is not respected in [33]. Consider, for example, the strongly quasiconvex functions $f(x) := x^3$ *and* $f(x) := x^3(x+1)$.

4. Sensitivity analysis. In clinical routine, the initially provided dose distribution framework, which is needed for the development of a treatment plan, often is too optimistic so that an obtained solution may turn out to be inacceptable. The relaxation of bounds on the other hand can result in serious consequences for the patient and therefore has to include the considerations of physicians. For that, as we show in this section, the physicians can be supported by a standard sensitivity analysis for a solution of our optimization model for IMRT treatment planning.

Let $f : \mathbb{R}^n \to \mathbb{R}$ be the objective function and $g_i(\phi) \leq \delta c$ be some constraint of the problem. Furthermore let $\phi^*(0)$ be the solution of the problem for $\delta c = 0$ and let $\lambda_i^* \geq 0$ be the related Lagrange multiplier. Next consider f as a function depending on δc, i.e. as $f(\phi(\delta c))$. Then, under suitable assumptions and for $|\delta c|$ sufficiently small, the optimization problem has a local minimizer $\phi^*(\delta c)$ and

$$\frac{\partial f}{\partial c}(\phi^*(\delta c))_{\delta c=0} = -\lambda_i^*$$

(see [12, p. 315]). Thus, for some small perturbations δc one arrives at

$$\frac{f(\phi^*(\delta c)) - f(\phi^*(0))}{\delta c} \approx -\lambda_i^*, \tag{4.1}$$

saying that a relaxation of the inequality constraints with the largest multipliers cause the largest local changes of the optimal objective function value.

In case the objective function f equals the LTCP of a single target V, i.e. if $f(\phi) :=$LTCP$(\phi; V, \Delta, \alpha)$ for some prescribed dose Δ, the change in the optimal value of the problem by a small relaxation of an inequality constraint can also be translated into a change of the EUD in V, which is a more significant number for the physicians. The EUD for the dose distribution in V, differing from the function in (3.23) which sometimes is denoted as the generalized EUD, is defined by

$$\mathcal{E}(\phi; V, \alpha) := -\frac{1}{\alpha} \log \left\{ \frac{1}{|V|} \sum_{j \in V} \exp(-\alpha\, D_j^T \phi) \right\}$$

([75]) and, in the assumed case, can be written as

$$\mathcal{E}(\phi; V, \alpha) = -\frac{1}{\alpha} \log \left\{ f(\phi) \exp(-\alpha\Delta) \right\}.$$

Thus, in this case, the change of the EUD in the target

$$\delta_{EUD} := \mathcal{E}(\phi^*(\delta c); V, \alpha) - \mathcal{E}(\phi^*(0); V, \alpha)$$

affected by a constraint perturbation δc is given, with (4.1), approximately by

$$\begin{aligned}\delta_{EUD} &\approx -\frac{1}{\alpha}\left[\log\left\{f(\phi^*(0)) - \lambda_i^*\delta c\right\} - \log\left\{f(\phi^*(0))\right\}\right] \\ &= -\frac{1}{\alpha}\log\left\{1 - \frac{\lambda_i^*\delta c}{f(\phi^*(0))}\right\}.\end{aligned}$$

Note in this connection that, ideally, $f(\phi^*(0))$ would equal 1.

The advantage of the application of a sensitivity analysis as described is that it guides the decision making on the part of the expert in regard to the bounds which need to be relaxed to achieve the desired target dose. This leads to both an elimination of a fruitless trial-and-error process and, typically, to the relaxation of only one or few bounds in a program, while all others can be kept. Also, for the treatment planning model of the authors, the biological interpretation of such change of bounds is immediate, for example, if the percentage of a volume to be sacrificed in the worst case is raised from 30 % to 40 %. This is quite different for least-squares type or multicriteria approaches where the effect of a change of desired dose bounds, percentages of a volume, or importance weights for one volume cannot be foreseen and may effect the outcomes for all volumes.

5. Intensity modulated proton therapy. It has been proposed about 60 years ago, that irradiation with beams of *protons* or *heavy ions*, would often be a better tool for cancer treatment than conventional irradiation with photons. In contrast to photons, which deposit the maximum dose at the beginning of their path through the body, protons deliver the maximum dose briefly before they stop and only relatively little before and almost none behind this point (see Fig. 5.1).

The depth of the *Bragg peak*, i.e. the position of maximum dose deposition, is directly correlated to the energy of the incident particles and can be tuned precisely. Hence, by modulating the kinetic energy of the particles and the beam intensities, i.e. the exposure times of the beams, one can generate a nearly homogeneous *Spread-Out Bragg Peak* (SOBP) in the direction of the beam. This can be performed with *passive scattering techniques*, where the proton beam passes through rotating devices of angularly variable thickness which reduce the particle energy for an appropriate, fixed fraction of the rotation time. In contrast, in *intensity modulated proton therapy* (IMPT), the exposure time of the proton beam for every scanning position and every beam energy is a free variable. This technique is also called *spot-scanning* (SC) which highlights the fact that the irradiated volume is covered by Bragg peaks of narrow beams that are scanned in 3 dimensions (two lateral deflections and the depth via the particle energy).

Experience and small-scale case studies show that, compared to conventional photon radiotherapy, proton therapy normally leads to similar results in terms of the targets, but may yield some or much improvement concerning the OARs. Considerable improvement, however, may be reached in regard to the total dose administered to the patient. (In the case study of [77] it has been reduced with the spot scanning technique by about 46 %.) The latter fact is relevant especially when children have to be irradiated.

For a long time, however, the technical problems and costs to perform irradiation with protons and other heavy charged particles have been prohibitive, at least for

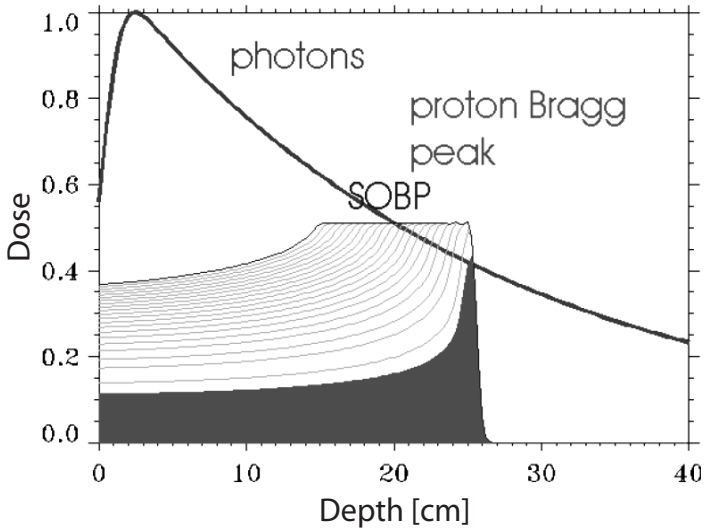

FIG. 5.1. *Schematic depth dependence of 6MV photons (solid line), a proton beam at fixed energy (solid filled curve), and a superposition of proton beams of various energies (thin lines) yielding a SOBP, where the dose is represented in arbitrary units.*

an application on a large scale. During the last 10 years, however, especially the Paul Scherrer Institute (PSI) in Villigen, Switzerland, has reached much progress in this respect and developed the SC technique to clinical applicability ([66], [67]). The recent success with proton therapy also has led to the development of cyclotrons exclusively for proton therapy, while, in the past, the cyclotrons needed for the proton acceleration had been constructed primarily for research in atomic physics and not for medical applications. Several dedicated proton sites will go into operation in the near future.

Treatment planning tools for proton therapy are not that far developed yet as in case of conventional therapy with photons. However, the optimization models discussed earlier in this paper can be straightforwardly transferred to the SC technique. In contrast to IMRT, where the intensities of beamlets that depend on two parameters (the position in the fields) are optimized, for IMPT with the SC technique the intensities of beamlets that depend on 3 parameters (position in the fields and particle energy) need to be determined. On the other hand, due to the favorable properties of protons, irradiation of a patient is performed normally only from 2 or 3 directions. In total, the optimization model has the same appearance as for IMRT. Merely the dose matrix $D := (d_{jk})$ is computed differently and, has a considerably larger number of columns because of the third dimension of beamlet variability. In this context it is remarkable that an obtained proton beam intensity profile can be realized directly up to some negligible deviations so that, in contrast to IMRT, no translation into MLC openings is needed.

Optimization problems for IMPT with the SC technique can have $40,000$ variables and more. This fact and the newness of the method explain why only very few references can be found commenting on an optimization model and algorithms for IMPT treatment planning. In [66] application of a least-squares approach is reported, which is said to be similar to the one used in [51] and [16] (which is not purely least-squares, see Sect. 3.3) and is known as a method for image reconstruction e.g. in

computer tomography. The authors of [77] employ the least-squares type approach from [14] and [79] which has also been implemented in the software package KonRad (see Sect. 3.3). In the following we show that our approach and algorithm from [9] can also be successfully applied to IMPT.

6. Example case. The patient of our example case was an 11 years old boy having a rhabdomyosarcoma, which reached from the interior of the lower jaw to the base of the skull. Irradiation of the patient by conventional radiotherapy was impossible since the tumor had infiltrated the second vertebra and because of the proximity of the tumor to the optic chiasm, optical nerves, and the brain stem. The decision was made to irradiate the vertebra with 36 Gy to avoid unilateral growth inhibition and simultaneously spare the spinal cord. The volume of gross tumor was treated to 57.6 Gy, the volume of suspected microscopic expansion to 48.6 Gy. The organs at risk (chiasm, optical nerves, eyes, spinal cord, brain stem) were defined with a 3 mm margin for setup errors, and a dose reduction in the overlap of the optical chiasm and the PTV was accepted.

The constraints of the optimization model are of the type introduced and explained in Sect. 3.5. In total, constraints for 12 volumes, denoted together with them, entered the optimization model. Especially V_1 is the *gross tumor volume* (GTV), i.e. the solid tumor, $V_2 \supseteq V_1$ is the *clinical target volume* (CTV), which is the GTV together with a margin in which tumor cells are suspected, and $V_3 \supseteq V_2$ is the *planning target volume* (PTV) which adds a safety margin to the CTV, in order to respect small movements of the patient and other inaccuracies. The optimization model had the following form, where "$V_\ell \pm$ 5 mm" means that an area of 5 mm width was added or subtracted respectively from V_ℓ. In particular, the remaining volume "$V_{12} - 5\,mm$" consists of the entire head and neck not otherwise classified as organ at risk or target volume, with an additional margin of 5 mm around all targets.

$$
\begin{aligned}
\text{Min.}\quad & \mathrm{LTCP}(\phi; V_1, 57.6, 0.25) + \mathrm{LTCP}(\phi; V_2, 48.6, 0.25) \\
& \qquad + \mathrm{LTCP}(\phi; V_3, 36, 0.25) \\
\text{s.t.}\quad & \mathrm{QOP}(\phi; V_1, 57.6, 1) \le 0 \quad \textit{(GTV)}, \\
& \mathrm{QOP}(\phi; V_2 \setminus V_1, 57.6, 0.2) \le 0 \quad \textit{(CTV)}, \\
& \mathrm{QOP}(\phi; V_2 \setminus (V_1 + 5\,mm), 48.6, 1) \le 0 \quad \textit{(CTV)}, \\
& \mathrm{QOP}(\phi; V_3 \setminus V_2, 48.6, 0.3) \le 0 \quad \textit{(PTV)}, \\
& \mathrm{QOP}(\phi; V_3 \setminus (V_2 + 5\,mm), 36, 1) \le 0 \quad \textit{(PTV)}, \\
& \mathrm{EUD}(\phi; V_4, 8, 12, 1) \le 0 \quad \textit{(right eye)}, \\
& \mathrm{EUD}(\phi; V_5, 14, 12, 1) \le 0 \quad \textit{(left eye)}, \\
& \mathrm{EUD}(\phi; V_6, 40, 12, 1) \le 0 \quad \textit{(optic chiasm)}, \\
& \mathrm{EUD}(\phi; V_7, 28, 12, 1) \le 0 \quad \textit{(r. optical nerve)}, \\
& \mathrm{EUD}(\phi; V_8, 40, 12, 1) \le 0 \quad \textit{(l. optical nerve)}, \\
& \mathrm{EUD}(\phi; V_9, 28, 12, 1) \le 0 \quad \textit{(spinal cord)}, \\
& \mathrm{EUD}(\phi; V_{10}, 28, 12, 1) \le 0 \quad \textit{(brain stem)}, \\
& \mathrm{PV}(\phi; V_{11}, 3, 20, 0.1) \le 0 \quad \textit{(right parotid)}, \\
& \mathrm{QOP}(\phi; V_{12} - 5\,mm, 30, 0.1) \le 0 \quad \textit{(remaining vol.)}, \\
& \phi \ge 0.
\end{aligned}
$$

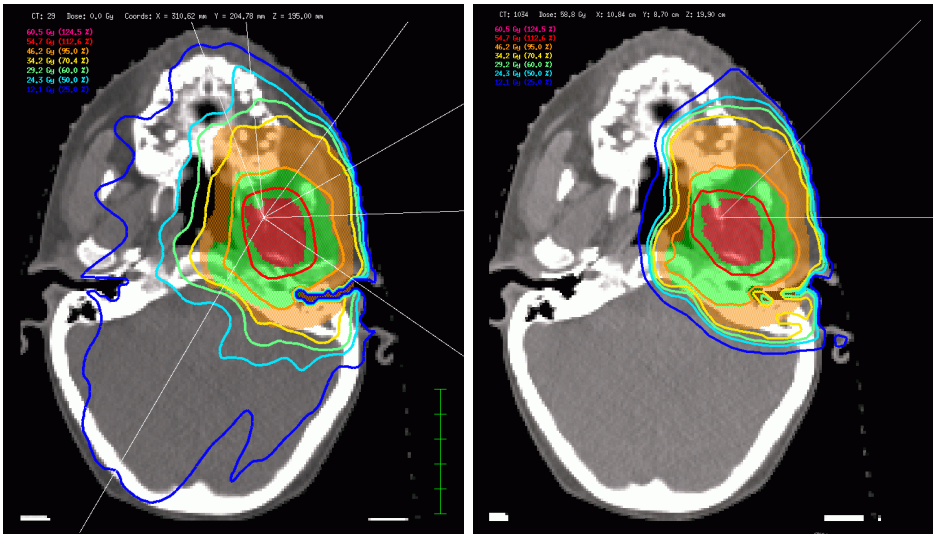

FIG. 6.1. *Transversal section close to the base of skull of the example case. Left, IMRT; right, IMPT. The isodoselines correspond to 25%, 50%, 60%, 70%, 95%, 112.5% of the prescription dose to the CTV of 48.6 Gy.*

For the image processing, 112 computer tomographic slices with 3 mm spacing had been generated. A $18 \times 21 \times 33$ cm^3 box was irradiated and the size of a volume element was $2 \times 2 \times 2$ mm^3 so that the total number of volume elements amounted to about $m = 1,600,000$, of which about 50 % belonged to the patients body. For the application of IMRT, 7 radiation fields were used, partitioned in field elements of 10×2 mm^2 size. The number of field elements and beamlets respectively totalled to $n = 3,727$ so that each field had about 532 elements on average. In contrast to that, for IMPT, only 2 beam directions were chosen. The beams were scanned over a $3 \times 3 \times 2.4$ mm^3 grid ($x \times y \times$ energy), resulting in $n = 47,043$ proton spots, where the number of the spots equals the product of the number of beam directions and of those grid points of the *scanning grid* which belong to the PTV. The maximum proton energy needed to cover the PTV was 138 MeV.

Results	IMRT	IMPT
# outer iterations:	3	6
# inner iterations:	148	274
average # inner iterations per outer iteration:	49	46
# obj. function evaluations for step size:	1,413	2,595
average # obj. function evaluations:	10	9
CPU time (minutes:seconds):	5 : 41	10 : 51

Table 1

The optimization problem was solved with the algorithm introduced in [9]. Some characteristic numbers for its performance are listed in Table 1. The CPU times refer to a XEON 2.66 processor. The given results show the typical behavior of the algorithm. The average sizes of a set of 127 clinical problems and the average iteration numbers for their solution in case of IMRT can be found in [9].

[1] A. Agren, A. Brahme, and I. Turesson, *Optimization of uncomplicated control for head and neck tumors*, Int. J. Radiat. Biol., 19 (1990), pp. 1077–1085.

[2] A. Ahnesjö and M. M. Aspradakis, *Dose calculations for external photon beams in radiotherapy*, Phys. Med. Biol., 44 (1999), pp. R99–R156.

[3] M. Alber, *A Concept for the Optimization of Radiotherapy*, PhD thesis, Universität Tübingen, Tübingen, Germany, 2000.

[4] M. Alber, M. Birkner, and F. Nüsslin, *Tools for the analysis of dose optimization: II. sensitivity analysis*, Phys. Med. Biol., 47 (2002), pp. 1–6.

[5] M. Alber and F. Nüsslin, *An objective function for radiation treatment optimization based on local biological measures*, Phys. Med. Biol., 44 (1999), pp. 479–493.

[6] M. Alber and F. Nüsslin, *Intensity modulated photon beams subject to a minimal surface smoothing constraint*, Phys. Med. Biol., 45 (2000), pp. N49–N52.

[7] ———, *Optimization of intensity modulated radiotherapy under constraints for static and dynamic MLC delivery*, Phys. Med. Biol., 46 (2001), pp. 3229–3239.

[8] ———, *Ein Konzept zur Optimierung von klinischer IMRT*, Z. Med. Phys., 12 (2002), pp. 109–113.

[9] M. Alber and R. Reemtsen, *Intensity modulated radiotherapy treatment planning by use of a barrier-penalty multiplier method*, Submitted (download from http://www.math.tu-cottbus.de/INSTITUT/lsing1/publications/reemtsen.html), (2004).

[10] M. S. Bazaraa, H. D. Sherali, and C. M. Shetty, *Nonlinear Programming - Theory and Algorithms*, Wiley, New York, 1993.

[11] G. Bednarz, D. Michalski, C. Houser, M. S. Huq, Y. Xiao, P. R. Anne, and J. M. Galvin, *The use of mixed-integer programming for inverse treatment planning with predefined field segments*, Phys. Med. Biol., 47 (2002), pp. 2235–2245.

[12] D. P. Bertsekas, *Nonlinear Programming*, Athena Scientific, Belmont, Massachusetts, 2nd ed., 1999.

[13] T. Bortfeld, *Dosiskonfirmation in der Tumortherapie mit externer ionisierender Strahlung: Physikalische Möglichkeiten und Grenzen*. Habilitationsschrift, Universität Heidelberg, Heidelberg, 1995.

[14] ———, *Optimized planning using physical objectives and constraints*, Semin. Radiat. Oncol., 9 (1999), pp. 20–34.

[15] T. Bortfeld, A. L. Boyer, W. Schlegel, D. L. Kahler, and T. J. Waldron, *Realization and verification of three-dimensional conformal radiotherapy with modulated fields*, Int. J. Radiat. Oncol. Biol. Phys., 30 (1994), p. 899.

[16] T. Bortfeld, J. Bürkelbach, R. Boesecke, and W. Schlegel, *Methods of image reconstruction from projections applied to conformation radiotherapy*, Phys. Med. Biol., 35 (1990), pp. 1423–1434.

[17] T. Bortfeld, W. Schlegel, C. Dykstra, S. Levegrün, and K. Preiser, *Physical vs biological objectives for treatment plan optimization*, Radiother. Oncol., 40 (1996), p. 185. letter, comment.

[18] T. Bortfeld, J. Stein, and K. Preiser, *Clinically relevant intensity modulation optimization using physical criteria*, in XIIth International Conference on the Use of Computers in Radiation Therapy, D. D. Leavitt and G. Starkschall, eds., Medical Physics Publ., Madison, WI, 1997, pp. 1–4.

[19] T. Bortfeld, C. Thieke, K.-H. Küfer, M. Monz, A. Scherrer, and H. Trinkhaus, *Intensity-modulated radiotherapy - a large scale multi-criteria programming problem*, Tech. Report ITWM, Nr. 43, Fraunhofer Inst. Techno- und Wirtschaftsmath., Kaiserslautern, Germany, 2003.

[20] A. Brahme, *Optimisation of stationary and moving beam radiation therapy techniques*, Radiother. Oncol., 12 (1988), pp. 129–140.

[21] ———, *Treatment optimization using physical and radiobiological objective functions*, in Radiation Therapy Physics, A. R. Smith, ed., Springer, Berlin, 1995, pp. 209–246.

[22] A. Brahme and B. K. Lind, *The importance of biological modeling in intensity modulated radiotherapy optimization*, in XIIth International Conference on the Use of Computers in Radiation Therapy, D. D. Leavitt and G. Starkschall, eds., Medical Physics Publ., Madison, WI, 1997, pp. 5–8.

[23] A. Brahme, J. E. Roos, and I. Lax, *Solution of an integral equation encountered in rotation therapy*, Phys. Med. Biol., 27 (1982), pp. 1221–1229.

[24] ———, *Solution of an integral equation in rotation therapy*, Med. Phys., 27 (1982), p. 1221.

[25] R. E. Burkard, H. Leitner, R. Rudolf, T. Siegl, and E. Tabbert, *Discrete optimization models for treatment planning in radiation therapy*, in Science and Technology for Medicine: Biomedical Engineering in Graz, H. Hutten, ed., Pabst Science Publ., Lengerich, 1995, pp. 237–249.

[26] Y. Censor, *Mathematical optimization for the inverse problem of intensity-modulated radiation therapy*, in Intensity-Modulated Radiation Therapy: The State of the Art, J. R. Palta and T. R. Mackie, eds., Medical Physics Publ., Madison, WI, 2003, pp. 25–49.

[27] Y. Censor and S. A. Zenios, *Parallel Optimization: Theory, Algorithms, and Applications*, Oxford University Press, Oxford, 1997.

[28] Y. Chen, D. Michalski, C. Houser, and J. M. Galvin, *A deterministic iterative least-squares algorithm for beam weight optimization in conformal radiotherapy*, Phys. Med. Biol., 47 (2002), pp. 1647–1658.

[29] P. S. Cho, S. Lee, R. J. Marks II, S. Oh, S. G. Sutlief, and M. H. Phillips, *Optimization of intensity modulated beams with volume constraints using two methods: Cost function minimization and projections onto convex sets*, Med. Phys., 25 (1998), pp. 435–443.

[30] P. S. Cho and R. J. Marks, *Hardware-sensitive optimization for intensity modulated radiotherapy*, Phys. Med. Biol., 45 (2000), pp. 429–440.

[31] L. Collatz and W. Wetterling, *Optimization Problems*, Springer, New York, 1975.

[32] C. Cotrutz, M. Lahanas, C. Kappas, and D. Baltas, *A multiobjective gradient based dose optimization algorithm for conformal radiotherapy*, Phys. Med. Biol., 46 (2001), pp. 2161–2175.

[33] J. O. Deasy, *Multiple local minima in radiotherapy optimization problems with dose-volume constraints*, Med. Phys., 24 (1997), pp. 1157–1161.

[34] M. Ehrgott and R. Johnston, *Optimisation of beam directions in intensity modulated radiation therapy planning*, OR Spectrum, 25 (2003), pp. 251–264.

[35] K. Engel, *A new algorithm for optimal multileaf collimator field segmentation*, Tech. Report 03/5, Universität Rostock, Fachbereich Mathematik, 2003.

[36] K. Engel and E. Tabbert, *Fast simultaneous angle, wedge, and beam intensity optimization in inverse radiotherapy planning*, Optimization and Engineering, to appear, (2004).

[37] M. C. Ferris, J. Lim, and D. M. Shepard, *An optimization approach for radiosurgery treatment planning*, SIAM J. Optimiz., 13 (2003), pp. 921–937.

[38] ———, *Radiosurgery treatment planning via nonlinear programming*, Ann. Oper. Res., 119 (2003), pp. 247–260.

[39] R. Fletcher, *Practical Methods of Optimization*, John Wiley & Sons, Chichester, 2nd ed., 1991.

[40] M. Goitein and A. Niemierko, *Intensity modulated therapy and inhomogeneous dose to the tumor: A note of caution*, Int. J. Rad. Oncol. Biol. Phys., 36 (1996), pp. 519–522.

[41] A. Gustaffson, *Development of a Versatile Algorithm for Optimization of Radiation Therapy*, PhD thesis, University of Stockholm, Stockholm, Sweden, 1996.

[42] A. Gustafsson, B. K. Lind, and A. Brahme, *A generalized pencil beam algorithm for optimization of radiation therapy*, Med. Phys., 21 (1994), pp. 343–356.

[43] A. Gustafsson, B. K. Lind, R. Svensson, and A. Brahme, *Simultaneous optimization of dynamic multileaf collimation and scanning patterns or compensation filters using a generalized pencil beam algorithm*, Med. Phys., 22 (1995), pp. 1141–1156.

[44] H. W. Hamacher and K.-H. Küfer, *Inverse radiation therapy planning - a multiple objective optimization approach*, Discr. Appl. Math., 118 (2002), pp. 145–161.

[45] H. W. Hamacher and F. Lenzen, *A mixed-integer programming approach to the multileaf collimator problem*, in The Use of Computers in Radiation Therapy, W. Schlegel and T. Bortfeld, eds., Springer, Berlin-Heidelberg-New York, 2000, pp. 210–212.

[46] M. Hilbig, *Inverse Bestrahlungsplanung für intensitätsmodulierte Strahlenfelder mit Linearer Programmierung als Optimierungsmethode*, PhD thesis, Technische Universität München, München, Germany, 2003.

[47] J. Höffner, *New Methods for Solving the Inverse Problem in Radiation Therapy Planning*, PhD thesis, Universität Kaiserslautern, Kaiserslautern, Germany, 1996.

[48] A. Holder, *Designing radiotherapy plans with elastic constraints and interior point methods*, Health Care and Management Science, 6 (2003), pp. 5–16.

[49] A. Holder and B. Salter, *A tutorial on radiation oncology and optimization*, in Tutorials on Emerging Methodologies and Applications in Operations Research, 2004, in print.

[50] T. Holmes and T. R. Mackie, *A comparison of three inverse treatment planning algorithms*, Phys. Med. Biol., 39 (1994), pp. 91–106.

[51] ———, *A filtered backprojection dose calculation method for inverse treatment planning*, Med. Phys., 21 (1994), pp. 303–313.

[52] Q. Hou, J. Wang, Y. Chen, and J. M. Galvin, *An optimization algorithm for intensity modulated radiotherapy - the simulated dynamics with dose-volume constraints*, Med. Phys., 30 (2003), pp. 61–68.

[53] D. H. Hristov and B. G. Fallone, *An active set algorithm for treatment planning optimization*, Med. Phys., 24 (1997), pp. 1455–1464.

[54] ———, *A continuous penalty function method for inverse treatment planning*, Med. Phys., 25 (1998), pp. 208–223.

[55] A. Jackson, G. J. Kutcher, and E. D. Yorke, *Probability of radiation-induced complications for normal tissues with parallel architecture subject to non-uniform irradiation*, Med. Phys., 20 (1993), pp. 613–625.

[56] J. Jahn, *Vector Optimization*, Springer, Berlin-Heidelberg, 2004.

[57] T. Kalinowski, *An algorithm for optimal multileaf collimator field segmentation with interleaf collision constraint*, Tech. Report 03/8, Universität Rostock, Fachbereich Mathematik, 2003.

[58] P. Källman, B. K. Lind, and A. Brahme, *An algorithm for maximizing the probability of complication free tumor control in radiation therapy*, Phys. Med. Biol., 37 (1992), pp. 871–890.

[59] M. Lahanas, E. Schreibmann, and D. Baltas, *Multiobjective inverse planning for intensity modulated radiotherapy with constraint-free gradient-based optimization algorithms*, Phys. Med. Biol., 48 (2003), pp. 2843–2871.

[60] M. Langer, R. Brown, M. Urie, J. Leong, M. Stracher, and J. Shapiro, *Large scale optimization of beam weights under dose-volume restrictions*, Int. J. Radiation Oncology Biol. Phys., 18 (1990), pp. 887–893.

[61] M. Langer, V. Thai, and L. Papiez, *Improved leaf sequencing reduces segments or monitor units needed to deliver IMRT using multileaf collimators*, Med. Phys., 28 (2001), pp. 2450–2458.

[62] W. Laub, M. Alber, M. Birkner, and F. Nüsslin, *Monte Carlo dose computation for IMRT optimization*, Phys. Med. Biol., 45 (2000), pp. 1741–1754.

[63] E. K. Lee, T. Fox, and I. Crocker, *Integer programming applied to intensity-modulated radiation therapy treatment planning*, Ann. Oper. Res., 119 (2003), pp. 165–181.

[64] J. Lim, *Optimization in Radiation Treatment Planning*, PhD thesis, University of Wisconsin, Madison, 2002.

[65] W. A. Lodwick, S. McCourt, F. Newman, and S. Humphries, *Optimization methods for radiation therapy plans*, in Computational Radiology and Imaging: Therapy and Diagnostics, C. Börgers and F. Natterer, eds., Springer, Berlin, 1999, pp. 229–249.

[66] A. Lomax, *Intensity modulated methods for proton radiotherapy*, Phys. Med. Biol., 44 (1999), pp. 185–205.

[67] A. J. Lomax, T. Boehringer, A. Coray, E. Egger, G.Goitein, M. Grossmann, P. Juelke, S. Lin, E. Pedroni, B. Rohrer, W. Roser, B. Rossi, B. Siegenthaler, O. Stadelmann, H. Stauble, C. Vetter, and L. Wisser, *Intensity modulated proton therapy: A clinical example*, Med. Phys., 28 (2001), pp. 317–324.

[68] G. Meedt, M. Alber, and F. Nüsslin, *Non-coplanar beam direction optimization for intensity modulated radiotherapy*, Phys. Med. Biol., 48 (2003), pp. 2999–3019.

[69] M. Merritt, Y. Zhang, H. Liu, and R. Mohan, *A successive linear programming approach to IMRT optimization problem*, Tech. Report TR02-16, Dept. Comp. Appl. Math., Rice Univ., Houston, 2002.

[70] D. Michalski, Y. Xiao, Y. Censor, and J. M. Galvin, *The dose-volume constraint satisfaction problem for inverse treatment planning with field segments*, Phys. Med. Biol., 49 (2004), pp. 601–616.

[71] R. Mohan, X. Wang, A. Jackson, T. Bortfeld, A. L. Boyer, G. J. Kutcher, S. A. Leibel, Z. Fuks, and C. C. Ling, *The potential and limitations of the inverse radiotherapy technique*, Radiother. Oncol., 32 (1994), pp. 232–248.

[72] R. Mohan and X.-H. Wang, *Response to Bortfeld et al. re physical vs biological objectives for treatment plan optimization*, Radiother. Oncol., 40 (1996), pp. 186–187. letter, comment.

[73] S. M. Morill, R. G. Lane, J. A. Wong, and I. I. Rosen, *Dose-volume considerations with linear programming optimization*, Med. Phys., 18 (1991), pp. 1201–1210.

[74] T. R. Munro and C. W. Gilbert, *The relation between tumor lethal doses and the radiosensitivity of tumor cells*, Br. J. Radiol., 34 (1961), pp. 246–251.

[75] A. Niemierko, *Reporting and analyzing dose distributions: A concept of equivalent uniform dose*, Med. Phys., 24 (1997), pp. 103–110.

[76] ———, *A generalized concept of equivalent uniform dose (EUD) (abstract)*, Med. Phys., 26 (1999), p. 1100.

[77] S. Nill, T. Bortfeld, and U. Oelfke, *Inverse planning of intensity modulated proton therapy*, Z. Med. Phys., 14 (2004), pp. 35–40.

[78] J. Nocedal and S. J. Wright, *Numerical Optimization*, Springer, New York-Berlin-Heidelberg, 1999.

[79] U. Oelfke and T. Bortfeld, *Inverse planning for photon and proton beams*, Med. Dosimetry, 26 (2001), pp. 113–124.

[80] F. Preciado-Walters, R. Rardin, M. Langer, and V. Thai, *A coupled column generation, mixed integer approach to optimal planning of intensity modulated radiation therapy for cancer*, Math. Progr., Ser. B 101 (2004), pp. 319–338.

[81] C. Raphael, *Mathematical modelling of objectives in radiation therapy treatment planning*, Phys. Med. Biol., 37 (1992), pp. 1293–1311.

[82] H. E. Romeijn, R. K. Ahuja, J. F. Dempsey, A. Kumar, and J. G. Li, *A novel linear programming approach to fluence map optimization for intensity modulated radiation therapy treatment planning*, Phys. Med. Biol., 48 (2003), pp. 3521–3542.

[83] H. E. Romeijn, J. F. Dempsey, and J. G. Li, *A unifying framework for multi-criteria fluence map optimization models*, Phys. Med. Biol., 49 (2004), pp. 1991–2013.

[84] I. I. Rosen, R. G. Lane, S. M. Morrill, and J. A. Belli, *Treatment plan optimization using linear programming*, Med. Phys., 18 (1991), pp. 141–152.

[85] G. R. Rowbottom and S. Webb, *Configuration space analysis of common cost functions in radiotherapy beam-weight optimization algorithms*, Phys. Med. Biol., 47 (2002), pp. 65–77.

[86] J. Seco, P. M. Evans, and S. Webb, *Modelling the effects of IMRT delivery: constraints and incorporation of beam smoothing into inverse planning*, in Proceedings of the XIII International Conference on the Use of Computers in Radiation Therapy, W. Schlegel and T. Bortfeld, eds., Heidelberg, 2000, Springer, pp. 542–544.

[87] D. M. Shepard, M. C. Ferris, G. H. Olivera, and T. R. Mackie, *Optimizing the delivery of radiation therapy to cancer patients*, SIAM Review, 41 (1999), pp. 721–744.

[88] D. M. Shepard, G. H. Olivera, P. J. Reckwerdt, and T. R. Mackie, *Iterative approaches to dose optimization in tomotherapy*, Phys. Med. Biol., 45 (2000), pp. 69–90.

[89] S. Söderström, A. Gustafsson, and A. Brahme, *The clinical value of different treatment objectives and degrees of freedom in radiation therapy optimization*, Radiother. Oncol., 29 (1993), pp. 148–163.

[90] ———, *Few-field radiation therapy optimization in the phase space of complication-free tumor control*, Int. J. Imaging Syst. Technol., 6 (1995), pp. 91–103.

[91] P. Spellucci, *Numerische Verfahren der nichtlinearen Optimierung*, Birkhäuser, Basel-Boston-Berlin, 1993.

[92] S. V. Spirou and C.-S. Chui, *A gradient inverse planning algorithm with dose-volume constraints*, Med. Phys., 25 (1998), pp. 321–333.

[93] J. Stein, R. Mohan, X.-H. Wang, T. Bortfeld, Q. Wu, K. Preiser, C. C. Ling, and W. Schlegel, *Number and orientations of beams in intensity-modulated radiation treatments*, Med. Phys., 24 (1997), pp. 149–160.

[94] J. Tervo and P. Kolmonen, *A model for the control of a multileaf collimator in radiation therapy treatment planning*, Inv. Probl., 16 (2000), pp. 1875–1895.

[95] J. Tervo, P. Kolmonen, T. Lyyra-Laitinen, J. D. Pintér, and T. Lahtinen, *An optimization-based approach to the multiple static delivery technique in radiation therapy*, Ann. Oper. Res., 119 (2003), pp. 205–227.

[96] J. Tervo, T. Lyyra-Laitinen, P. Kolmonen, and E. Boman, *An inverse treatment planning model for intensity modulated radiation therapy with dynamic MLC*, Appl. Math. Comput., 135 (2003), pp. 227–250.

[97] C. Thieke, *Multicriteria Optimization in Inverse Radiotherapy Planning*, PhD thesis, Univ. of Heidelberg, Heidelberg, Germany, 2003.

[98] C. Thieke, T. Bortfeld, A. Niemierko, and S. Nill, *From physical dose constraints to equivalent uniform dose constraints in inverse radiotherapy planning*, Med. Phys., 30 (2003), pp. 2332–2339.

[99] C. Thieke, T. R. Bortfeld, and K.-H. Küfer, *Characterization of dose distributions through the max and mean dose concept*, Acta Oncologica, 41 (2002), pp. 158–161.

[100] X.-H. Wang, R. Mohan, A. Jackson, S. A. Leibel, Z. Fuchs, and C. C. Ling, *Optimization of intensity-modulated 3d conformal treatment plans based on biological indices*, Radiother. Oncol., 37 (1995), pp. 140–152.

[101] S. Webb, *The Physics of Conformal Radiotherapy: Advances in Technology*, Medical Science Series, IOP Publishing, Bristol, 1997.

[102] S. .Webb, *Intensity-Modulated Radiation Therapy*, Medical Science Series, IOP Publishing, Bristol, 2000.

[103] J. Werner, *Optimization Theory and Applications*, Vieweg, Braunschweig, Germany, 1984.

[104] Q. Wu, D. Djajaputra, Y. Wu, J. Zhou, H. H. Liu, and R. Mohan, *Intensity-modulated radiotherapy optimization with gEUD-guided dose-volume objectives*, Phys. Med. Biol., 48 (2003), pp. 279–291.

[105] Q. Wu and R. Mohan, *Algorithms and functionality of an intensity modulated radiotherapy optimization system*, Med. Phys., 27 (2000), pp. 701–711.

[106] Q. Wu, R. Mohan, A. Niemierko, and R. Schmidt-Ullrich, *Optimization of intensity-modulated radiotherapy plans based on the equivalent uniform dose*, Int. J. Radiation Oncology Biol. Phys., 52 (2002), pp. 224–235.

[107] Y. Xiao, D. Michalski, J. M. Galvin, and Y. Censor, *The least-intensity feasible solution for aperture-based inverse planning in radiation therapy*, Ann. Oper. Res., 119 (2003), pp. 183–203.

[108] L. Xing and G. T. Y. Chen, *Iterative methods for inverse treatment planning*, Phys. Med. Biol., 41 (1996), pp. 2107–2123.

[109] L. Xing, R. J. Hamilton, D. Spelbring, C. A. Pelizzari, G. T. Y. Chen, and A. L. Boyer, *Fast iterative algorithms for three-dimensional inverse treatment planning*, Med. Phys., 25 (1998), pp. 1845–1849.

[110] L. Xing, J. G. Li, S. Donaldson, Q. T. Le, and A. L. Boyer, *Optimization of importance factors in inverse planning*, Phys. Med. Biol., 44 (1999), pp. 2525–2536.