Optimization Methods and Software
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/goms20

Intensity modulated radiotherapy treatment planning by use of a barrier-penalty multiplier method

M. Alber [a] & R. Reemtsen [b]

[a] Universitätsklinikum Tübingen , Radioonkologische Klinik, Hoppe-Seyler-Strasse 3, D-72076, Tübingen

[b] Brandenburgische Technische Universität Cottbus, Institut für Mathematik , Universitätsplatz 3-4, D-03044, Cottbus, Germany

Published online: 21 Mar 2007.

To cite this article: M. Alber & R. Reemtsen (2007) Intensity modulated radiotherapy treatment planning by use of a barrier-penalty multiplier method, Optimization Methods and Software, 22:3, 391-411, DOI: 10.1080/10556780600604940

To link to this article: http://dx.doi.org/10.1080/10556780600604940

Optimization Methods and Software
Vol. 22, No. 3, June 2007, 391–411

Taylor & Francis
Taylor & Francis Group

Intensity modulated radiotherapy treatment planning by use of a barrier-penalty multiplier method

M. ALBER*† and R. REEMTSEN‡

†Universitätsklinikum Tübingen, Radioonkologische Klinik, Hoppe-Seyler-Strasse 3, D-72076 Tübingen
‡Brandenburgische Technische Universität Cottbus, Institut für Mathematik, Universitätsplatz 3–4, D-03044 Cottbus, Germany

(Received 24 August 2004; revised 22 November 2005; in final form 10 January 2006)

The use of nonlinear functions describing biological effects has recently become a major goal in connection with intensity modulated radiotherapy planning models for cancer treatment. In this article, we present a new biological model for this purpose and discuss the solution of the related large-scale nonlinear optimization problems. The model includes equivalent uniform dose and partial volume constraints and employs tumor control probability as the objective. The resulting optimization problems are convex; there are nonconvex constrained optimization problems with several thousands of variables for which gradients of the involved functions are available, but the computation of Hessians is too costly. It is suggested to solve these problems using the barrier-penalty multiplier method by Polyak ([Polyak, R., 1992, Modified barrier functions (theory and methods). *Mathematical Programming*, **54**, 177–222.], [Polyak, R., 2002, Nonlinear rescaling *vs*. smoothing technique in convex optimization. *Mathematical Programming*, **92**, 197–235.]) and Ben-Tal *et al.* and Ben-Tal and Zibulevsky ([Ben-Tal, A., Yuzefokich, I. and Zibulevsky, M., 1992, Penalty/barrier multiplier methods for minimax and constrained smooth convex problems. Technical Report 9/92, Optimization Laboratory, Faculty of Industrial Engineering and Management, Technion, Haifa, Israel.], [Ben-Tal, A. and Zibulevsky, M., 1997, Penalty/barrier multiplier methods for convex programming problems. *SIAM Journal of Optimization*, **7**, 347–366.]), where this algorithm is modified according to ideas which are motivated by the related Lagrangian barrier algorithm of Conn *et al.* ([Conn, A.R., Gould, N.I.M. and Taint, P., 1992, A globally convergent Lagrangian barrier algorithm for optimization with general inequality constraints and simple bounds. Technical Report 92/07, Department of Maths, FUNDP, Namur, Belgium.], [Conn, A.R., Gould, N.I.M. and Taint, P.L., 1992, A globally convergent Lagrangian barrier algorithm for optimization with general inequality constraints and simple bounds. *Mathematics of Computation*, **66**, 261–288.]). In particular, the subproblems in the algorithm are solved by a conjugate gradient method, as the spectrum of the Hessian of the Lagrangian at a solution of such problem indicates fast (local) convergence of the objective function values to a good approximate (locally) optimal value. Some characteristic numbers showing the average numerical performance of the algorithm are tabulated for various types of tumors and for a set of 127 clinical cases in total. Its capabilities and typical behavior also are illustrated explicitly by a computed therapy plan for a difficult clinical case of a larynx tumor.

Keywords: Radiation therapy treatment planning; Intensity modulated radiation therapy; Barrier algorithms; Multiplier methods; Nonlinear rescaling method; Barrier-penalty functions

AMS Subject Classification: 90C90; 90C06; 90C30

*Corresponding author. Email: msalber@med-tuebingen.de

Optimization Methods and Software
ISSN 1055-6788 print/ISSN 1029-4937 online © 2007 Taylor & Francis
http://www.tandf.co.uk/journals
DOI: 10.1080/10556780600604940

1. Introduction

Each year, about 500,000 people receive radiation therapy for cancer traetment in the USA alone. The fundamental predicament of radiation therapy treatment is that it does not only affect diseased but also healthy tissues. Therefore, for each individual patient, a treatment plan, based on the images of computer tomography (CT), has to be established such that the radiation effects are sufficient for the planning target volumes (PTVs), i.e. the tumor(s) and the possibly involved tissue, and are acceptably small for the organs at risk (OARs).

Conventional techniques in radiation therapy use standardized radiation field arrangements with typically two to five fields. The radiation intensity of these fields is homogeneous or has a constant gradient, where normally (but not always) the fields are shaped according to the projection of the PTVs upon the fields. The latter is achieved by the covering of parts of a rectangular field with blocks of lead or, more recently, by means of a computer-controlled multi-leaf collimator (MLC), which consists of typically 25–60 tungsten slabs that can be shifted from each of two opposite sides.

Treatment planning in the case of such conventional techniques is performed by a forward approach, where, by a trial-and-error procedure, the radiation effects of several different dose distribution arrangements are assessed with respect to the anatomy of the patient. In contrast, inverse dose planning [1–3] in connection with intensity modulated radiation therapy (IMRT), which is the topic of this article, allows an improved tumor control and a possible reduction of side effects [4].

For IMRT, each of, normally, 2–12 radiation fields is partitioned into some 100–2000 field elements. Accordingly, each of the 2–12 beams is divided into the corresponding number of small beamlets. Moreover, the treatment goals are described by an objective function and constraints depending on the dose deposited in the patient's body. The dose is defined by a nonnegative weight for each beamlet (beamlet weight) which, relative to an assumed unit intensity of the radiation source, specifies the radiation intensity of the particular beamlet. Thus, IMRT leads to large-scale optimization problems which have the beamlet weights as variables.

Concerning the optimization goals for IMRT, a variety of different approaches can be distinguished, leading to different types of mathematical optimization problems. At present, in all clinically relevant optimization approaches for inverse dose planning, the angles for the position of the radiation fields are fixed (like in all earlier approaches) and a desired dose distribution is specified by an upper dose bound for a number of normal tissue volumes and lower dose bounds for the PTVs [5–10]. As the set of linear inequalities obtained in this way may be inconsistent, a substitute goal can be to find a beamlet weight profile, which minimizes the sum of all squared violations of the requested dose tolerances subject to nonnegativity constraints for the beamlet weights where, for each treatment goal, an additional weight is used in order to take the presumed significance of its violation into account. The resulting optimization problem is a convex least-squares type problem with nonnegativity constraints for the variables, which has been solved, e.g. by a scaled gradient projection algorithm [11].

This dose-based approach has several disadvantages. The choice of weights specifying the significance of violations of dose tolerances is quite arbitrary, and such a priori choice does not provide insights into the actual violations caused by a solution to the related problem. Therefore, in pursuit of this approach, a physician normally needs to compute several plans for different choices of weights and select the one which is the best by his personal point of view. (A multiple objective linear programming approach in this connection has recently been given in [12].) Moreover, IMRT leads to inhomogeneous dose distributions in organs, which are penalized by quadratic functions while biological side effects may increase in a

nonquadratic way if dose bounds are violated. For these reasons, it has been proposed to include biological considerations in the definition of both dose prescriptions and rules for controlling their violation [13–17]. As a result, alternative biological optimization models, which respect the dose–responses of the different tissues, have been developed [6,18,19]. A general framework for multi-criteria models in this connection has recently been studied [20].

In contrast to dose-based optimization, which constitutes the first attempt to bridge the gap between the desired and the feasible dose distribution, constrained biological optimization aims to incorporate the biological knowledge and clinical evidence of conventional radiotherapy in order to explore the potential of IMRT and simultaneously remain on safe clinical grounds. In particular, it takes both the response of tissues to the dose per treatment fraction and the response to inhomogeneous dose distributions into account. The first purpose of this article is to present a new biological optimization model where the tumor control probability (TCP), i.e. the probability that no tumor cell survives, is maximized under equivalent uniform dose (EUD) and partial volume (PV) constraints for the OARs [see refs. 1–23 for ingredients of the model and notice that another model, showing the interest in such nonlinear approaches, has recently been presented in ref. 24]. The second and main goal of this article is to describe the algorithm which is used to solve the related large-scale nonlinear constrained optimization problems and to demonstrate by an example case, both for the model and the performance of the algorithm, that such nonlinear IMRT optimization problems can be solved in a satisfactory way. The average performance of the algorithm is illustrated by a compilation of some characteristic values for 127 patients, most of which were treated with the dose distributions computed by this algorithm.

The optimization methods which have come to application in connection with IMRT have mainly been the methods for linear optimization and for simple-bound constrained linear least-squares type problems [9,11,25]. The desire of employing a nonlinear biological optimization model for IMRT raises the question of a reliable and efficient algorithm for its solution. Nonlinear programming methods have been used only rarely until now and in special circumstances [26–28]. For the solution of the problems originating from the nonlinear model [24] mentioned earlier, a heuristic algorithm based on a componentwise Newton type algorithm was suggested. Notice that the functions involved in such a model are sufficiently smooth, but that the computation of Hessians typically is too time-consuming [see ref. 29 for the approach discussed here and observe that the authors of ref. [24] also avoid the use of Hessians].

In view of the size of the problems, the available information, the required memory, and the complexity of the subproblems to be solved within an algorithm, the recently developed modified barrier multiplier methods appear particularly attractive [30–41]. The particular choice of update rules for the multipliers and the use of scaling factors, which are both characteristic for this type of multiplier method, have led to a considerable acceleration of the convergence, in comparison to more traditional multiplier methods such as the one used by LANCELOT [42]. In this article, we especially use the nonlinear rescaling method (NR) or barrier (-penalty) multiplier method by Polyak *et al.* [33,34,36], where we employ the modified logarithmic barrier function combined with a quadratic or linear penalty branch [31,32,36,38,40,43,44] and modify this algorithm along the lines of the related Lagrangian barrier algorithm by Conn *et al.* [30,45].

An important reason for the selection of a multiplier type approach in this work was that the resulting subproblems are (nonlinear) unconstrained or simple-bound constrained optimization problems and that the spectrum of the Hessian of the Lagrangian at a solution of the studied problems typically shows only few distinctly different eigenvalues, whereas all other eigenvalues are clustered around these [29]. The latter fact suggests the use of a conjugate gradient method for the solution of the subproblems [46], and indeed, such a method has

turned out to be very efficient to find a satisfactory approximate objective function value for a (local) solution of the considered problem.

The biological optimization model used for IMRT treatment planning is described in section 2. The proposed algorithm for the solution to the resulting optimization problems is motivated in section 3 and discussed and modified in section 4. The article concludes with a clinical example case of a larynx tumor in section 5. This example and a table showing the average performance of the algorithm for 127 clinical cases illustrate the type and possible size of the occurring optimization problems and simultaneously demonstrate the capability and typical behavior of the algorithm.

The presented biological model and optimization algorithm were implemented in the software package HYPERION, which was developed at the university hospital of Tübingen.[†] This software package is being applied in daily clinical routine at some hospitals in Europe and the USA.

2. IMRT treatment model

In our model, as in current clinical routine of radiotherapy, the number p as well as the p angles for the position of the radiation fields, i.e. of the directions of the beams, is determined by experience or by trial and error. Normally, the number p lies between 3 and 6 and is smaller than 12. Optimization with respect to the choice of beam directions would be desirable but is impeded by the computationally expensive dependence of the dose absorbed by the patient's body on the orientation of the radiation fields and the combinatorical nature of the problem. (A heuristic procedure for beam direction optimization, which is based on the availability of an efficient algorithm for the problem with pre-determined field directions, was recently described [47].)

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. The jth field is partitioned into n_j rectangular field elements of equal size and, through that, the jth beam is partitioned into n_j rectangular beamlets. Consequently, the total number of beamlets over all fields equals $n := \sum_{j=1}^{p} n_j$. For each beamlet, a beamlet weight defining its radiation intensity is to be determined. The n beamlet weights $\phi_k \geq 0$ $(k = 1, \ldots, n)$ are the unknowns of an optimization problem for IMRT treatment planning.

We mention in this context that the radiation intensity profile generated by the beamlet weights may be realized for each field either by a lead compensator, which has to be manufactured and inserted into the respective field in a time-consuming manner, or by the MLC, which is part of the treatment machine and can be controlled automatically. The tungsten leaves of an MLC can be placed such that a polygonal geometry is exposed, and several different consecutive exposures from the same angle (typically 10–30) can be used to generate sophisticated intensity patterns (figure 1). Clearly, it would be desirable to include a comprehensive set of constraints in an optimization model which guarantees that the optimal dose obtained by the model is realizable by an MLC. To our knowledge, such a set has not been given yet. Therefore, at present, a beamlet weight profile obtained by some optimization procedure has to be translated properly into a profile which can be realized by a relatively small number of MLC openings. Such an adaptation is a difficult problem in itself, which is not discussed here. Fortunately, for the model used here and its objective function, it can be performed with a negligible loss with respect to optimality [48].

[†]For more information, see http://www.medizin.uni-tuebingen.de/medphys/.

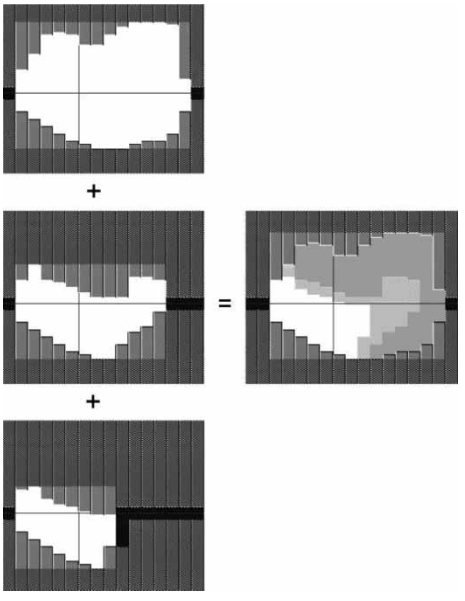

Figure 1. Intensity modulation by superposition of MLC shaped fields. Dark gray bars from top and bottom symbolize the tungsten leaves, white area in the centers symbolizes the exposed area of the field. On the right, total intensity levels are symbolized by gray values.

In regard to these technical realizations of dose distributions, a choice of up to about 1000–2000 elements per field is reasonable, so that the total number n of all beamlet weights ϕ_k typically amounts to about 3000–8000, but may exceed 15,000 in special circumstances (section 5). For obvious physical reasons, the beamlet weights have to be nonnegative. Technical limitations of the MLC may enforce additional constraints on the weights, which have to be included in the optimization problem [see ref. 49 for a minimum area surface smoothing constraint and our remarks in the last part of section 4].

The exposed part of the human body is considered to be divided into q 3D volumes representing the PTV and regions of normal tissue (e.g. OARs) and the jth volume again is divided into m_j 3D cubic (volume) elements of equal size, having a side length of normally ≥ 2.5 mm. Typically, q is a number less than 15, and the total number $m := \sum_{j=1}^{q} m_j$ of volume elements is of order 10^6. We number all volume elements consecutively from 1 to m and let V_j $(j = 1, \ldots, q)$ be the index set of all elements belonging to the jth volume.

If $t_{ik} \geq 0$ denotes the dose deposited in the ith volume element by the kth beamlet at unit intensity, then

$$d_i(\phi) := \sum_{k=1}^{n} t_{ik}\phi_k = T_i^{\mathrm{T}}\phi \geq 0$$

is the total dose in the ith volume element, where $\phi := (\phi_1, \ldots, \phi_n)^{\mathrm{T}}$ is the nonnegative vector of beamlet weights and T_i^{T} is the line vector containing the coefficients of the ith line of the $m \times n$ dose matrix $T := (t_{ik})$, which, for our purposes, is assumed to be given. The matrix T can be determined 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 [cf. ref. 50]. Alternatively, more precise Monte Carlo methods can be used to simulate the radiation transport through the patient, at an increase of the computation time by a factor 100 [51]. As the kth beamlet predominantly affects volume elements only in the proximity of its line of propagation, the dose matrix T is sparse and thus

can be stored in a closed form. Typically, at a reasonable cut-off for the minimal dose, less than 3–8% of the coefficients of T are nonzero.

Recently, nonlinear optimization models for radiotherapy treatment planning have come into the focus of interest (see, e.g., refs. [20,24] for other approaches discussed in section 1). We next describe the (nonlinear) mathematical functions which we use in our model. For that let V denote the index set of all elements belonging to a particular volume and let $|V|$ equal the cardinality of V. For simplicity, we also denote a volume itself as V. Then, the goal of our model is to maximize the TCP for each PTV V, which is given [52] by

$$\prod_{i\in V} \exp\left\{-\frac{1}{|V|}\exp(-\alpha * (T_i^{\mathrm{T}}\phi - \Delta))\right\},$$

where $\alpha > 0$ is a constant related to cell survival and $\Delta > 0$ is the total dose which ideally is administered to V. Now, observe that maximization of this function is equivalent to the maximization of the natural logarithm of this function and hence to the minimization of the following logarithmic tumor control probability (LTCP) which we actually use

$$\mathrm{LTCP}(\phi;\, V,\, \Delta,\, \alpha) := \frac{1}{|V|}\sum_{i\in V}\exp(-\alpha * (T_i^{\mathrm{T}}\phi - \Delta)). \tag{1}$$

In connection with a PTV V, we also use quadratic overdosage penalty (QOP) constraints of the type

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

where $\delta > 0$ is a given bound and

$$[x]_+ := \max\{x, 0\}.$$

The intention of such constraints is to prevent an excessively high dose in V while allowing a mild violation of the acceptable dose Δ that can be controlled by means of the parameter δ. In case some elements of V are overdosed, the dose of Δ Gy should be exceeded by at most δ Gy on average. Such a constraint is also applied in connection with the fact that a sharp dose drop from some PTV to a neighboring volume V is not realizable physically so that a certain overdosing of V needs to be accepted around the target volumes, which would not be acceptable far away from these (section 5). Notice that a function $\mathrm{QOP}(\cdot;\, V,\, \Delta,\, \delta)$ is once but not twice continuously differentiable everywhere. (To compute Hessians in ref. [29], we replaced the power 2 in QOP by 3.)

For each critical serial organ V (e.g. the spinal cord and other nervous structures or, generally, any organ that can be seriously damaged by a high dose in a small spot), we employ an EUD constraint [53]

$$\mathrm{EUD}(\phi;\, V,\, \Delta,\, p,\, \varepsilon) := \frac{1}{|V|}\sum_{i\in V}\left(\frac{1}{\Delta}T_i^{\mathrm{T}}\phi\right)^p - \varepsilon^p \le 0. \tag{3}$$

where $\Delta > 0$ is some critical dose, $\varepsilon > 0$ a given constant, and $p > 1$ some organ specific power. For example, if V is the spinal cord, a choice of $\Delta := 40$, $p := 15$, and $\varepsilon := 1$ effects that only a tiny excess over 40 Gy is allowed for parts of this organ, with the extent of overdosage depending sensitively on the size of the volume in which it occurs (section 5).

Finally, for each critical parallel organ V (lung, parotid gland, kidney, etc.), we apply a PV constraint

$$\mathrm{PV}(\phi;\, V,\, \Delta,\, p,\, \zeta) := \frac{1}{|V|} \sum_{i \in V} \frac{((1/\Delta) T_i^{\mathrm{T}} \phi)^p}{1 + ((1/\Delta) T_i^{\mathrm{T}} \phi)^p} - \zeta \leq 0 \tag{4}$$

with some constants $\zeta \in (0, 1)$ and $p > 1$, where the size of p depends on the organ. Such constraints are related to the fact that parallel organs have functional reserves [see refs. 21,22,54 for details]. For example, the data $\Delta := 26$, $p := 3$, and $\zeta := 0.6$, express that at a dose of 26 Gy, a volume element of V loses 50% of its function (e.g. blood–air gas exchange and production of saliva) and that at most 60% of the total function of the organ may be lost. Notice that if the step of this sigmoidal function is contracted so that the function is either 0 or 1, respectively, outside the step interval, this constraint type becomes similar to a dose–volume (*histogram*) constraint [6].

For more details on the modeling functions, we also refer to refs. [21–23]. LTCP, QOP, and EUD are nonquadratic convex functions, whereas the function PV is formed by the nonconvex sigmoid function $\sigma(x) := x^p/(1 + x^p)$ [9]. Therefore, use of these functions leads either to a convex optimization problem or, if the irradiation of, e.g. parotid glands and lungs is to be controlled, to a nonconvex optimization problem with sufficiently smooth functions in n variables. In this ideal description concerning the beamlet weights (see the last part of section 3 in this regard), the zero vector obviously is a feasible point for all constraints, as no constraint on a minimal dose for any target is used in the model.

3. Algorithm

We now consider the optimization problem

$$\begin{aligned} (P)\text{: Minimize} \quad & f(x) \\ \text{s.t.} \quad & g_i(x) \leq 0 \ (i = 1, \ldots, l), \end{aligned}$$

where f are $g_i \colon \mathbb{R}^n \to \mathbb{R}$ are sufficiently smooth functions and $x := (x_1, \ldots, x_n)^{\mathrm{T}}$. We denote the set of feasible points of (P) by

$$\mathcal{F} := \{x \in \mathbb{R}^n | g_i(x) \leq 0 \ (i = 1, \ldots, l)\}.$$

Problem (P) may include nonnegativity bounds on the variables, i.e. we may have $l > n$ and $g_i(x) := -x_i$ $(i = 1, \ldots, n)$. (Concerning the problems discussed in this article, we will comment on this at the end of section 4.)

In connection with (P), we define the classical Lagrange function or Lagrangian

$$L(x, \lambda) := f(x) + \sum_{i=1}^{l} \lambda_i g_i(x)$$

with (Lagrange) multipliers $\lambda_i \in \mathbb{R}$. Then, for a local minimizer x^* of (P) and for some $\lambda^* \in \mathbb{R}^l$, the Karush–Kuhn–Tucker (KKT) necessary optimality conditions read [55]:

$$x^* \in \mathcal{F}, \quad \nabla_x L(x^*, \lambda^*) = 0, \ \ \lambda^* \geq 0, \quad \lambda_i^* g_i(x^*) = 0 \ \ (i = 1, \ldots, l). \tag{5}$$

The last set of equations in (5) is denoted as complementary slackness condition.

The tremendous success of interior-point methods for linear programming problems within the last two decades has reactivated the research on barrier methods and has led to combinations

of classical Lagrangian multiplier methods and barrier type approaches, which possess the best properties of both and simultaneously avoid most of their main deficiencies (see refs. [46,55,56] for the classical methods). On the basis of fundamental results by Polyak [33] and Polyak and Teboulle [34], several authors have recently reported exciting results for such (modified) barrier multiplier methods or NR methods [31,32,35–41,57].

An NR method starts from the observation that for $s_i > 0$, the constraints in (P) can be equivalently replaced by $-s_i\psi(-g_i(x)/s_i) \le 0$ $(i = 1, \ldots, l)$, where ψ is a function with certain properties, for example, the modified logarithmic barrier function

$$\psi(y) := \log(1 + y), \quad y > -1. \tag{6}$$

The method can be interpreted as a multiplier method for the classical Lagrangian of the problem altered in this way:

$$\mathcal{L}(x, \lambda, s) := f(x) - \sum_{i=1}^{l} \lambda_i s_i \psi\left(-\frac{g_i(x)}{s_i}\right). \tag{7}$$

where variants of the method are obtained by different choices of ψ. Note that for $\lambda \ge 0$ and $s > 0$, the function $\mathcal{L}$ is convex in x on its domain of definition, if ψ is a concave function as ψ in (6) and if f and all g_i's are convex functions.

For $s > 0$, $\lambda \in \mathbb{R}^l$, and e.g. as in (6), the function $\mathcal{L}$ is defined on the extended feasible set $\{x | g_i(x) < s_i, i = 1, \ldots, l\}$ and hence, in contrast to functions used in classical barrier methods, also on the boundary and outside of $\mathcal{F}$. Nevertheless, the barrier multiplier approach for $\mathcal{L}$ with ψ in (6) has similar properties as the classical logarithmic barrier function approach. Especially, the minimization of $\mathcal{L}(\cdot, \lambda^k, s^k)$ for some λ^k and s^k, as required in multiplier methods, may turn out to be a difficult problem, as $\mathcal{L}(\cdot, \lambda^k, s^k)$ is defined on a subset of $\mathbb{R}^n$ only. A breakthrough for this type of multiplier approach was attained by Ben-Tal *et al.* and Ben-Tal and Zibulevsky [31,32] who employed the modified logarithmic-quadratic barrier-penalty function

$$\psi(y) := \begin{cases} -\dfrac{1}{2\sigma^2}(1 + y)^2 + \dfrac{2}{\sigma}(1 + y) + \log(\sigma) - \dfrac{3}{2}, & \text{if } 1 + y \le \sigma, \\ \log(1 + y), & \text{if } \sigma < 1 + y. \end{cases} \tag{8}$$

In this, $\sigma \in (0, 1)$ is some given constant (we usually choose $\sigma = 0.1$) and the coefficients of the quadratic branch of ψ, which is rather a penalty than a barrier function, are determined such that ψ is twice continuously differentiable at $y := \sigma - 1$. With this choice of ψ, the quadratic branch is used for i in (7) when $g_i(x)$ is close to s_i, i.e. when x is close to the boundary of the extended feasible region, and else the modified logarithmic barrier function is employed. Note that the function ψ in (8) is likewise concave, but compared with ψ in (6), has a bounded second derivative. Furthermore, ψ is defined on all of $\mathbb{R}$, which simplifies the solution to the subproblem of minimizing $\mathcal{L}(\cdot, \lambda^k, s^k)$ considerably. The multiplier estimates for ψ as in equation (8) are updated according to the formula

$$\lambda_i^{k+1} := \begin{cases} \left(\dfrac{g_i(x^k) - s_i^k}{\sigma^2 s_i^k} + \dfrac{2}{\sigma}\right)\lambda_i^k, & \text{if } g_i(x^k) - s_i^k \ge -\sigma s_i^k, \\ \dfrac{s_i^k}{s_i^k - g_i(x^k)}\lambda_i^k, & \text{if } s_i^k - g_i(x^k) > \sigma_i^k, \end{cases} \quad (i = 1, \ldots, l), \tag{9}$$

which is motivated by the first-order update formula

$$\lambda_i^{k+1} := \lambda_i^k \psi'\left(-\frac{g_i(x^k)}{s_i^k}\right) \quad (i = 1, \ldots, l) \tag{10}$$

that can be derived from the identity $\nabla_x \mathcal{L}(x^*, \lambda, s) = \nabla_x L(x^*, \lambda^*)$ for some λ and $s > 0$. For $\lambda^k > 0$ and $s^k > 0$, the formula (9) guarantees $\lambda^{k+1} > 0$. Note that the definition of λ_i^{k+1} in (9) depends on the relative distance of $g_i(x^k)$ from s_i^k, which seems to be a reasonable measure.

Lagrange multiplier methods have a long tradition in nonlinear programming [55,58]. Especially, augmented Lagrangian methods, for which the Lagrange function of the problem is augmented by a penalty function for the constraints, have, for a long time, been the only available tool for the solution of large-scale problems (e.g. the package LANCELOT by Conn *et al.* [42]. However, it is known that for these methods, a proper initial choice and update of the penalty parameter may be difficult and that with decreasing penalty parameter, although it can be kept away from zero, numerical difficulties caused by ill-conditioning and hence low accuracy may occur. Also, occasionally, slow convergence has been observed (e.g. see the results for LANCELOT in ref. [59]). Thus, for similar reasons which have led to the development of NR methods and seemingly parallel to them, the authors of ref. [42] have developed an augmented Lagrangian type method with a barrier function for the constraints [30,45]. This method, denoted here as CGT method, employs the Lagrange barrier function

$$\mathcal{L}_{\mathrm{CGT}}(x, \lambda, s) := f(x) - \sum_{i=1}^{l} \lambda_i s_i \log(s_i - g_i(x)) \tag{11}$$

with Lagrange multiplier estimates λ_i and shifts $s_i > 0$. As it turns out, the NR method for $\psi(y) := \log(1 + y)$ and the CGT method are closely related, as

$$\mathcal{L}(x, \lambda, s) = \mathcal{L}_{\mathrm{CGT}}(x, \lambda, s) + \sum_{i=1}^{l} \lambda_i s_i \log(s_i).$$

Hence, $\mathcal{L}$ and $\mathcal{L}_{\mathrm{CGT}}$ have the same local minimizers and gradients with respect to x and lead to the same first-order update formula for the multiplier estimates so that, with these settings, $\mathcal{L}$ and $\mathcal{L}_{\mathrm{CGT}}$ can be used equivalently for both the NR and the CGT methods. Furthermore, the two-parametric barrier function $\log(s + y)$ in $\mathcal{L}_{\mathrm{CGT}}$ could also be combined with a quadratic penalty function, similar as $\psi(y) := \log(1 + y)$ and for the same reasons.

Remark 1 For $\psi(y) := \log(1 + y)$, the function $\mathcal{L}$ is less sensitive than $\mathcal{L}_{\mathrm{CGT}}$ in regard to ill-conditioning in connection with the absence of strict complementary slackness [30, p. 278] for the CGT algorithm in this respect). Namely, if one has $\lambda_i^k \to \lambda_i^* = 0$ and $g_i(x^k) \to g_i(x^*) = 0$ with $k \to \infty$ for some i and a KKT point (x^*, λ^*) of (P), then one has $s_i^k \to 0$ with $k \to \infty$ and henceforth $\log(s_i^k - g_i(x^k)) \to -\infty$, whereas $\log(1 - g_i(x^k)/s_i^k)$ does not necessarily tend to $-\infty$.

NR methods have typically been stated in conceptual form and discussed under convexity assumptions [32,34,36], although many results are true in the nonconvex cases and they have been successfully applied to nonlinear problems also [38,60]. On the other hand, the CGT method with the function in (11) (which should not be used in practice) is stated as a realizable algorithm and has been proven convergence under standard assumptions for general nonlinear problems. Therefore, we amalgamate the ideas of both approaches and give the NR algorithm subsequently in the more detailed form of the CGT method, where we use the barrier-penalty function

$$\mathcal{L}_k(x) := \mathcal{L}(x, \lambda^k, s^k) = f(x) - \sum_{i=1}^{l} \lambda_i^k s_i^k \psi\left(-\frac{g_i(x)}{s_i^k}\right) \tag{12}$$

of the NR method with ψ as in (8) for given $\lambda^k > 0$ and $s^k > 0$. Relating to NR methods, we denote the s_i^k in (12) as scaling constants (rather than the numbers $1/s_i^k$ as, e.g., in ref. [36]).

At initialization, the CGT algorithm requires the specification of 12 parameters. By our numerical experiences, the choice of most of these parameters is not critical and can be quite arbitrary, so that we fixed some of them in the following algorithm. We also used two different vector norms in stopping and branching criteria. In particular, $\|\cdot\|_2$ means the Euclidean norm on $\mathbb{R}^n$.

ALGORITHM

(1) *Initialization. Specify positive constants* α_n, $\alpha_\lambda \le 1$, α_ω, η_s, $\eta_* \ll 1$, $\tau < 1$, ω_s, *and* $\omega_* \ll 1$ *such that* $\alpha_\eta + (1+\alpha_\lambda)^{-1} > 1$ *and set* $\omega_0 := \omega_s$, $\eta_0 := \eta_s$.
Provide initial estimates $x^0 \in \mathbb{R}^n$ *and choose* $\mu_0 \in (0,1)$. *Define* $\lambda_i^0 := 1$ *and* $s_i^0 := \mu_0$ $(i = 1, \dots, l)$ *and set* $k := 0$.

(2) *Inner iteration. By minimizing* $\mathcal{L}_k(x)$ *in* (12) *with* ψ *from* (8) *over* $\mathbb{R}^n$, *find* $x^k \in \mathbb{R}^n$ *such that for* $\vartheta_k := \|\nabla \mathcal{L}_k(x^k)\|_2$,

$$\vartheta_k \le \omega_k. \tag{13}$$

(3) *Test of convergence. Compute multiplier estimates* λ_i^{k+1} $(i = 1, \dots, l)$ *via* (9).
In case

$$\vartheta \le \omega_*, \quad \max_{1 \le i \le l} |\lambda_i^{k+1} g_i(x^k)| \le \eta_*, \tag{14}$$

stop. In case

$$\max_{1 \le i \le l} \left| \frac{\lambda_i^{k+1} g_i(x^k)}{(\lambda_i^k)^{\alpha_\lambda}} \right| \le \eta_k, \tag{15}$$

go to step (3). *Otherwise go to step* (4).

(4) *Update of Lagrange multiplier estimates. Set*

$$\lambda^{k+1} := \lambda^{k+1}, \quad \mu_{k+1} := \mu_k \quad \omega_{k+1} := \omega_k \mu_{k+1}^{\alpha_\omega}, \quad \eta_{k+1} := \eta_k.$$

Go to step (5).

(5) *Update of barrier-penalty parameter. Set*

$$\lambda^{k+1} := \lambda^k, \quad \mu_{k+1} := \tau \mu_k, \quad \omega_{k+1} := \omega_s \mu_{k+1}^{\alpha_\omega}, \quad \eta_{k+1} := \eta_s \mu_{k+1}^{\alpha_\eta}.$$

(6) *Compute scaling constants. Set*

$$s_i^{k+1} := \mu_{k+1} (\lambda_i^{k+1})^{\alpha_\lambda} \quad (i = 1, \dots, l). \tag{16}$$

Set $k := k + 1$ *and go to step* (1).

The NR method as described by Ben-Tal *et al.* [31,32] and Polyak [36] employs the Lagrangian in (7) with $s_i := 1/\nu_i$ and some parameter ν_i, where ψ is a transformation fulfilling certain properties. (One scaling parameter for each constraint seems to have been introduced for the first time in ref. [1].) The update formula for the multipliers coincides with that in (10). The one for the s_i's is identical with that in (16) for $\alpha_\lambda = 1$ in ref. [36] and it is defined in a somewhat more general way in ref. [32], where μ_{k+1} equals some $\nu > 0$ for all k [36] or is allowed to vary [32]. In ref. [34], scaling constants $s_i := 1/\nu$ for all i had been considered, where it was remarked that ν could be decreased at each iteration as long as $\nu \ge \bar{\nu} > 0$ is given for some $\bar{\nu}$. (A modification of the latter approach is found in ref. [35], where primal and dual variables are updated simultaneously by means of Newton's method.)

Although combined barrier-penalty functions seem to have been applied first by engineers [43,44,62], Ben-Tal *et al.* [31,32] seem to have been the first who have proved general convergence of the NR method for convex problems and a class of transformations ψ, which includes the one in (8). Polyak [36] studies the NR algorithm primarily for a log-sigmoid transformation, but his (more detailed) convergence results also apply to ψ in (8), where Polyak can show that the quadratic branch is used by the NR algorithm at most a finite number of times. For the conceptual NR algorithm, where $\nabla \mathcal{L}_k(x^k) = 0$ is true for all k and the (fixed) barrier-penalty parameter ν is sufficiently large, Polyak could furthermore prove under standard assumptions that the algorithm converges with a linear rate to a solution of the problem, that the multipliers for the inactive constraints converge to zero quadratically, and that the scaling constants corresponding to the active constraints decrease linearly. While NR methods seem to have been considered mainly in a convex framework, quite impressive numerical results have been presented for convex as well as nonconvex large-scale optimization problems [32,35–41,57,60].

On the other hand, for the CGT method, which is obtained with $\psi(y) = \log(1 + y)$ in (12), the global convergence results relate to KKT points and a realizable algorithm for general nonlinear problems [30], but only a limited number of numerical experiments has been presented for that [37]. The rate of convergence results proved for this method is similar to those for the NR method mentioned earlier. Moreover, it is shown that under certain assumptions, the generated sequence $\{\mu_k\}$ of barrier parameters is bounded away from zero. In fact, it was proved under stronger assumptions that the barrier parameter remains unchanged for all sufficiently large k [63]. But, note that for the function $\mathcal{L}_{\mathrm{CGT}}$, an auxiliary problem may have to be solved to obtain an appropriate starting point for the solution to the inner problem [30] and all inner iterates have to remain inside the domain of that function. In contrast to that the barrier-penalty function ψ in (8) is defined on all of $\mathbb{R}^n$.

Summarizing we can say that by virtue of the results for NR methods, convergence of this algorithm is guaranteed for ψ in (8) at least if the problem (P) is convex and the barrier-penalty parameter and the scaling constants are chosen as was specified for the NR method mentioned earlier. (Some of the results for the NR methods are also true in the nonconvex case.) On the other hand, the results by Conn *et al.* refer to general nonlinear problems and, in particular, suggest a rule for the reduction of the barrier-penalty parameter and stopping criteria both for the inner and the outer iterations. To our knowledge, however, convergence of this combined algorithm has not been proved yet for nonconvex problems. In fact, convergence would be ensured by the results in ref. [30] if, for $\omega_* = \eta^* = 0$, the quadratic branch in (8) is needed only a finite number of times in the algorithm, as has been verified for the NR method with this function in the convex case [36]. Furthermore, note in this connection that several authors have applied the NR method also to nonconvex problems, although the convergence results do not cover this case yet.

4. Discussion and modification of the algorithm

Our selections of the parameters in the earlier-mentioned algorithm do not agree in all cases with settings used for theoretical results for the CGT algorithm [30, Theorem 5.3]. But, as is said in [30, Appendix, p. 1], the choice in ref. [30] may not be suitable for practical purposes. Workable parameters for the earlier algorithm in case of our problems are

$$\alpha_\eta = 0.5, \quad \alpha_\lambda = \alpha_\omega = 0.25, \quad \eta_s = 1, \quad \eta_* = 0.01,$$
$$\mu_0 = 0.5, \quad \omega_s = 10^{-3}, \quad \omega_* = 10^{-5}, \quad \tau = 0.5.$$

The algorithm may start with an infeasible point x^0 and produce infeasible primal iterates, while $\lambda^k > 0$ is guaranteed for all $k \geq 1$ due to the choice $\lambda^0 > 0$. A big advantage of the use of $\mathcal{L}_k$ in (12) with the barrier-penalty function ψ in (8) is that it is defined for all $x \in \mathbb{R}^n$, and therefore, the inner algorithm needed to determine x^k can be started with the previous iterate x^{k-1}.

The inner iteration in step (1) is stopped when, in addition to condition (13), the relative deviation of the function values of $\mathcal{L}_k$ for two consecutive inner iterates is below some threshold $\tilde{\omega}_k$ in three succeeding iterations, where $\tilde{\omega}_k$ is chosen similarly to ω_k in (13). The algorithm is terminated in step (2) when $\|\nabla \mathcal{L}_k(x^k)\|_2$ is sufficiently small and the complementary slackness condition is satisfied sufficiently well. In addition, in order to prevent false indication of convergence in case some λ_i^k is too small, the maximum constraint violation has to satisfy $\max_{1\leq i\leq l}(g_i(x^k)) \leq \eta_*/\eta_m$, where $\eta_m > 1$ is some suitable constant.

Concerning the criterion in (15), note that for $\psi(y) := \log(1 + y)$ with $1 + y > \sigma$, this is equivalent to [30]

$$\max_{1\leq i\leq l} |\lambda_i^{k+1} - \lambda_i^k| \leq \frac{\eta_k}{\mu_k}$$

This criterion merely serves to detect whether the current iterate x^k is sufficiently close to the feasible region of (P) and hence whether the barrier-penalty parameter is small enough so that it essentially can be kept constant for all further iterations. (It indeed normally remains constant.)

If the criterion in (15) is satisfied, the dual variables are updated, the barrier-penalty parameter remains unaltered, and the termination constant for the inner algorithm is reduced. Following the motivation for the criterion in (15), we do not make the threshold η_k for (15) smaller in this case. (In the CGT algorithm, one has $\eta_{k+1} := c\eta_k$ for some constant $c < 1$ instead, where c may be chosen close to 1.) In the contrary case, when the condition in (15) is not fulfilled, the barrier-penalty parameter μ_k is reduced and the tolerances for the termination of the inner algorithm and the reduction of the barrier-penalty parameter are decreased, whereas the dual variable vector is retained.

Two further modifications described in the following have significantly weakened the sensitivity to the tuning of the parameters and accelerated the convergence of our algorithm for the problems under consideration. First, we have finally replaced the quadratic branch in the definition of ψ by the linear branch

$$\psi(y) := \frac{1+y}{\sigma} + \log(\sigma) - 1, \quad \text{if } 1 + y \leq \sigma,$$

so that we arrive at the modified-logarithmic-linear barrier-penalty function $\psi \in C^1(\mathbb{R})$ and the related multiplier update formula

$$\lambda_i^{k+1} := \begin{cases} \dfrac{1}{\sigma}\lambda_i^k, & \text{if } g_i(x^k) - s_i^k \geq -\sigma s_i^k, \\ \dfrac{s_i^k}{s_i^k - g_i(x^k)}\lambda_i^k, & \text{if } s_i^k - g_i(x^k) > \sigma s_i^k, \end{cases} \quad (i = 1, \ldots, l). \tag{17}$$

(The inverse barrier function had been combined with a linear branch already earlier in refs. [43,44]. The rationale for that is twofold. First, in case x^k lies far outside the feasible set $\mathcal{F}$ of (P), the use of x^k as starting point for the subsequent inner problem may cause problems in the case of a quadratic penalty branch, whereas a linear penalty term grows slower and therefore requires less care with respect to the choice of the starting point (cf. figure 2). Secondly, in order to avoid an overshooting of the multiplier estimates, it is desirable to limit

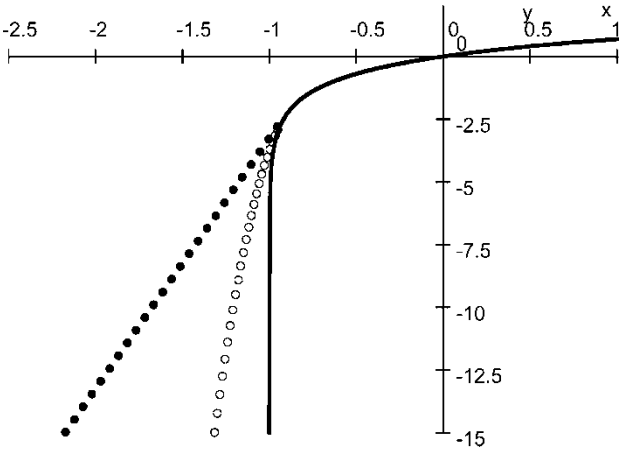

Figure 2. Log linear (dots), log quadratic (circles), and log function (solid) for $\sigma = 0.1$.

the size of the scaling factor in their update formula, which can become considerably larger for (9) than for (17), for which it is bounded. Note that a large change of one multiplier usually leads to a large change of the subsequent iterate x^k and constraint values $g_i(x^k)$ so that the reliability of all multiplier updates suffers if a few are very large.

The combined modified-logarithmic-linear barrier-penalty function does not satisfy the requirements on ψ in refs. [32,36]. However, an alternative to the multiplier estimate and scaling constant updates which are determined by $\psi(y) := \log(1 + y)$ is needed only for a few initial iterations so that any reasonable strategy may serve for that (see the results for the NR method in ref. [36]). Note that finite use of the linear or quadratic branch could also be enforced in the algorithm and then would not destroy the convergence results for the CGT algorithm. In fact, such an enforcement was not necessary, as the update rule related to the linear branch was always applied at most a few times at the beginning of the iteration process.

The fact that $\{\mu_k\}$ does not converge to zero in the CGT algorithm does not necessarily avert numerical ill-conditioning of the inner problems [see ref. 30, p. 278 and Remark 1]. Indeed, numerical experiments with that algorithm have shown that ill-conditioning of inner problems and small values of μ_k can trigger a fatal feedback loop. If the inner problem is ill-conditioned, a value $g_i(x^k)$ for a violated constraint can become quite large relative to s_i^k, especially if λ_i^k is small. (Note that $s_i^k - g_i(x^k) > 0$ is required for the CGT algorithm.) Thus, if μ_k is small, the associated scaling constant s_i^k also is small according to the rule (16), and this in turns leads to a large change in the magnitude of the multiplier (given by the second branch of (9) with $\sigma = 0$). In this case, especially when λ_i^k is small, the criterion (15) can cause a further reduction of μ_k and thereby lead to even worse ill-conditioning of the inner problem. By our observation, typically, some constraint values $g_i(x^k)$ oscillate around 0 until μ_k is reset to μ_0 (which is not planned in the algorithm). The described critical loop requires a meticulous choice of many parameters of the CGT algorithm.

Hence, some kind of reverse step to step (4), so that the scaling constants are enlarged or at least not reduced, is missing in the CGT algorithm. Our experience shows that too small scaling constants caused by a too small barrier-penalty parameter are also not desirable in the final stage of the algorithm. Then, either the termination threshold for the inner problem which is likewise determined by the barrier-penalty parameter may become unreasonably small and hence may entail an unreasonable number of inner iterations or such small scaling constants lead to outer iterates which are not close enough to the exact minimizers of the inner problem. These errors may be amplified in the multiplier update by virtue of the small scaling

constants and may consequently lead to overly large changes of the multipliers, causing them to overshoot.

In our opinion, the purpose of the branching criterion (15), which decides upon a reduction of the barrier-penalty parameter, is twofold. First, the magnitude of the scaling constants has to be adjusted to the present problem. And secondly, by a reduction of the scaling constants, the iterates x^k can be forced to become feasible. Clearly, the scaling constants can be adjusted more sensitively if they are not coupled too tightly via a common scalar μ_k. A large violation of one constraint should not affect the scaling constants of the others unduly.

To achieve the latter goal, our extensive numerical experience with the algorithm led us to exchange the scaling constant update formula (16) for the rule

$$\frac{1}{s_i^k} := \frac{1}{\mu_k(\lambda_i^k)^{\alpha_\lambda}} + \frac{|\lambda_i^{k+1} g_i(x^k)|}{\gamma \lambda_i^k} \quad (i = 1, \ldots, l), \tag{18}$$

where we choose the parameter γ of the same order of magnitude as η_*. Note that at convergence of the algorithm, i.e. if $\lambda^k \to \lambda^*$ and $x^k \to x^*$ with $k \to \infty$ is true for some KKT point (x^*, λ^*) of (P) (possibly only for a subsequence of $\{(x^k, \lambda^k)\}$), the second term in (18) tends to zero so that for equal values of μ_k, λ_i^k, and α_λ, the rules (18) and (16) tend to give similar updates. In particular, provided that the sequence $\{\mu_k\}$ is bounded away from zero, the convergence $\lambda_i^k \to \lambda_i^* = 0$ for $k \to \infty$ implies $s_i^k \to 0$ by (18) as well as by (16) for both $g_i(x^*) = 0$ and $g_i(x^*) < 0$. Thus, in practice, the second term in (18) becomes relevant only when x^k is far away from the feasible set and, in this case, leads to usually much smaller scaling constants for constraints with large value $|g_i(x^k)|$ than the rule (16). On the other hand, seen relatively, it typically yields larger scaling constants in a neighborhood of a solution to the problem, as the second term in (18) loses influence there. With this modification, step (4) of the algorithm needs to be executed only rarely. We also found that the choice of the various parameters in the algorithm becomes much less critical through this change. However, theoretical convergence of the algorithm with the rule in (18) has not been verified yet.

Like most authors, we have assumed nonnegativity for the beamlet weight vector ϕ in our IMRT treatment model (and, for simplicity, we will do so also in the example case of the subsequent section). In clinical practice, however, each beamlet weight $x_i := \phi_i$ $(i = 1, \ldots, n)$ is related to a radiation field generated by an MLC, and therefore, for technical reasons, it either has to equal zero and belong to a field area without radiation or it has to be greater than or equal to some small constant $\delta > 0$. This is often neglected. In the algorithm, we deal with this fact by using constraints $\delta - x_i \le 0$ $(i = 1, \ldots, n)$, where we treat these constraints differently from the other inequality constraints. First, we use a constant scaling constant $s_i := \delta$ for each index i of these, which leads to a term

$$\psi\left(-\frac{\delta - x_i}{\delta}\right) = \psi\left(\frac{x_i}{\delta} - 1\right)$$

in the Lagrangian (7) with ψ as in (8) (see ref. [38] for another treatment of simple bounds). And, secondly, we do not let these constraints enter the branching criterion (15) in step (2) of the algorithm, i.e. the violation of these constraints does not lead to a decrease of the barrier-penalty parameter. Then, at termination of the algorithm, variables with a value $< \delta$ are set to 0 and hence lead to a shrinkage of the respective radiation fields. This procedure has turned out to be satisfactory in our circumstances. A post-optimization on the reduced variable space normally does not bring any sizable progress, as usually only a small fraction of the beamlet weights equals zero. Notice finally that, in general, the point $x_i := \delta$ $(i = 1, \ldots, n)$ is feasible for the studied problems as δ is sufficiently small. We actually start the algorithm with a point

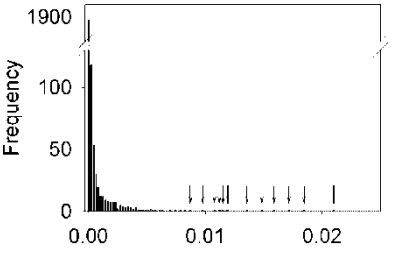

Figure 3. Eigenvalue distribution of a Hessian.

which is motivated by the particular problem, but is feasible usually only with respect to the simple bounds for the variables.

As noted in section 1, for the problems discussed in this article, the spectrum of the Hessian of the Lagrangian at a solution to a problem shows only few clearly distinct eigenvalues, where the majority of the eigenvalues has a positive value close to zero [29]. As a consequence, function values of the Lagrangian, roughly said, increase significantly only into few directions at a solution. A typical spectrum for an example with about 2500 variables is given in figure 3.

This characteristic of the Lagrangian for our problems has motivated us to treat the constraints $x_i \geq \delta$ $(i = 1, \ldots, n)$ as normal constraints and to solve unconstrained optimization problems in step (1) of the algorithm by a conjugate gradient (CG) method, as such an eigenvalue distribution of its Hessian promises rapid convergence of CG methods to a good estimate of the optimal objective function value of the problem, at least in a neighborhood of a solution in which the problem is well approximated through a quadratic model [46, section 5.1]. We employ the Polak-Ribiere method [46] together with exact stepsizes, which we compute via mixed golden section search and quadratic interpolation [64]. In accordance with our general expectations, this method typically needs only $\leq$50 iterations on average for a satisfactory solution of the inner problems, at a run of the earlier algorithm with up to 18,000 variables. We chose the Polak-Ribiere method [55], as other CG methods such as those in refs. [65–67] did not prove to be superior in experiments.

5. Example case of radiotherapy of a larynx tumor

The example case is that of a larynx tumor which had spread to the base of skull and down to the level of the shoulders. Radiotherapy still has a curative intent even for locally advanced tumors such as this, yet side effects can diminish the quality of life for survivors tremendously. Apart from the obvious hazard of paralysis due to damage of the spinal cord, in particular, obliteration of the parotid glands can cause painful chronic impairment. The objective of the treatment therefore has been fourfold: sparing of the spinal cord, conservation of parotid gland functions, total coverage of the macroscopic tumor volume (boost volume), and sterilization of the suspected microscopic spread (adjuvant volume). These goals are duly reflected in the definition of the constraints and objectives for this treatment, given in the following.

Let V_1 and V_2 relate to the disjoint boost and adjuvant volume, respectively. Then, the optimization problem, called (P_1), is specified in the following way, where the functions LTCP, QOP, EUD, and PV are defined in (1)–(4) and the index sets V_j, $j = 3, 4, \ldots, 8$, refer to the volumes denoted together with the constraints:

$$\begin{aligned} \text{Minimize} \quad & \text{LTCP}(\phi; V_1; 70, 0.25) + \text{LTCP}(\phi; V_2, 50; 0.25) \\ \text{s.t.} \quad & \text{QOP}(\phi; V_1, 70, 0.5) \leq 0 \quad \text{(boost volume)}, \end{aligned}$$

$$
\begin{aligned}
&\mathrm{QOP}(\phi;\, V_2, 50, \sqrt{2.5}) \le 0 \quad \text{(adjuvant volume)},\\
&\mathrm{QOP}(\phi;\, V_2, 70, 0.5) \le 0 \quad \text{(adjuvant volume)},\\
&\mathrm{EUD}(\phi;\, V_3, 32, 15, 1) \le 0 \quad \text{(spinal cord)},\\
&\mathrm{PV}(\phi;\, V_4, 26, 3, 0.6) \le 0 \quad \text{(right parotid gland)},\\
&\mathrm{PV}(\phi;\, V_5, 26, 3, 0.15) \le 0 \quad \text{(left parotid gland)},\\
&\mathrm{PV}(\phi;\, V_6, 20, 3, 0.25) \le 0 \quad \text{(right lung)},\\
&\mathrm{PV}(\phi;\, V_7, 20, 3, 0.25) \le 0 \quad \text{(left lung)},\\
&\mathrm{QOP}(\phi;\, V_8, 40, 0.1) \le 0 \quad \text{(remaining volume)},\\
&\phi \ge 0.
\end{aligned}
$$

In this, the remaining volume V_8 refers to the set of volume elements of the irradiated part of the patient's body, which do not belong to V_i, $i = 1, \ldots, 7$, and lie outside an 1 cm wide neighborhood of the adjuvant volume.

Obviously $\phi := 0$ is a feasible point for this problem. We next want to show that this problem (and thus each similar problem) also has a solution. For that observe that the volumes V_i, $j = 3, \ldots, 8$, do not appear in the objective function of the problem and that, therefore, a beamlet which only affects elements of these volumes can obtain a zero weight and thereby reduce the total dose administered. Hence, we can assume that T has no zero columns (else they could be cancelled) and that each beamlet which meets an element of the volumes with index sets V_j, $j \in \{3, \ldots, 8\}$, also meets at least one element of one of the volumes with index sets V_j, $j \in \{1, 2\}$ (else the related beamlet weight could be set equal to zero). Hence, the assumption of the following theorem is reasonable.

THEOREM 1 *Let $T := (t_{ik})$ with $t_{ik} \ge 0$ be the dose matrix and, for each $k \in \{1, \ldots, n\}$, let there exist an $i \in (V_1 \cup V_2)$ such that $t_{ik} > 0$. Then, problem (P_1) has a solution.*

Proof Let $\mathcal{F}_1$ be the feasible set of the problem. Owing to $0 \in \mathcal{F}_1$, this set is nonempty, and by the continuity of the constraint functions of (P_1), it is closed. If we can show that $\mathcal{F}_1$ also is bounded, then the assertion follows from the Weierstrass existence theorem.

Assume that there would be an unbounded sequence $\{\phi^{(s)}\}_{s\in\mathbb{N}}$ in $\mathcal{F}_1$. Then, some $k \in \{1, \ldots, n\}$ would exist such that $\phi_k^{(s)} \to \infty$ is true for $s \to \infty$, and furthermore, by assumption, for k, an $i \in V_j$ with $j = 1$ or $j = 2$ would exist such that $t_{ik} > 0$. Hence, for $s \to \infty$, we would have $T_i^{\mathrm{T}}\phi^{(s)} \ge t_{ik}\phi_k^{(s)} \to \infty$, and therefore, for all sufficiently large s,

$$
\mathrm{QOP}(\phi^{(s)};\, V_j, \Delta, \delta) = \frac{1}{|V|}\sum_{i\in V_j}[T_i^{\mathrm{T}}\phi^{(s)} - \Delta]_+^2 - \delta^2 \ge \frac{1}{|V|}[t_{ik}\phi_k^{(s)} - \Delta]_+^2 - \delta^2 \longrightarrow \infty,
$$

where Δ and δ are specified in the problem. But, as $\phi^{(s)}$ is in $\mathcal{F}_1$, i.e. $\mathrm{QOP}(\phi^{(s)};\, V_j, \Delta, \delta) \le 0$ is true, this is not possible. Consequently, $\mathcal{F}_1$ is bounded. ■

Remark 2 Under the assumption of Theorem 1 and the assumption that $\phi_k := \delta$ $(k = 1, \ldots, n)$ is a feasible point of problem (P_1), also the more realistic problem has a solution where either $\phi_k = 0$ or $\phi_k \ge \delta$ with some constant $\delta > 0$ is required for $k = 1, \ldots, n$. This is true as there exists only a finite number of problems for which none or some of the n beamlet weights ϕ_k are set identical zero and all remaining weights have to satisfy the bounds $\phi_k \ge \delta$. Application of the proof of Theorem 1 to these problems shows that each of them has

a solution. Moreover, each solution of one of these problems, which has the smallest objective function value among all of them, also solves the desired problem.

The total dose administered to the patient was 70 Gy for the boost volume and 50 Gy for the adjuvant volume. Note that a larger part of the left parotid can be saved than of the right one, owing to its greater distance from the tumor. The number of radiation fields in this example was 9, and the total number of beamlet weights, i.e. variables of the problem (P_1), was $n = 14{,}557$. Hence, each of the 9 fields had 1,617 beamlets on average. The size of a beamlet was $10 \times 2\,\text{mm}^2$ so that the average size per field was $330\,\text{cm}^2$. Furthermore, 133 CT slices of the patient with 3 mm distance had been used, and the exposed volume amounted to a cube of $47 \times 27 \times 35\,\text{cm}^3$. This cube was partitioned in about 1,600,000 cubic volume elements with a side length of 3 mm, where naturally only a certain amount of these elements belonged to the patient's body so that the number $m := \cup_{i=1}^{8} |V_i|$ of volume elements entering the optimization problem was approximately 800,000.

The algorithm needed only five outer iterations so that five subproblems had to be solved. The total number of inner iterations amounted to 186 so that the number of iterations needed for the solution of one subproblem was approximately 37 on average. (Typically, in our applications of the algorithm, the number of inner iterations is strongly decreasing with increasing outer iteration index.) For the computation of step lengths, 1731 evaluations of the objective function, i.e. 9.3 function evaluations per inner iteration on average, had to be executed. The total computing time was about 38 min on a Xeon 2.66 GHz processor, and 1.2 GB RAM was occupied. The final dose distribution at the base of the skull and the computed beamlet weight distribution for one of the nine beams are shown in figures 4 and 5, respectively.

We remark in this context that the constants δ, ε, and ζ in the constraint functions QOP, EUD, and PV (see (2)–(4)) depend on the assessment of the physicians and the particular balance between treatment risks and success they are willing to strike. These constants may be altered if the resulting dose distribution is unsatisfactory. In such a case, the solution of

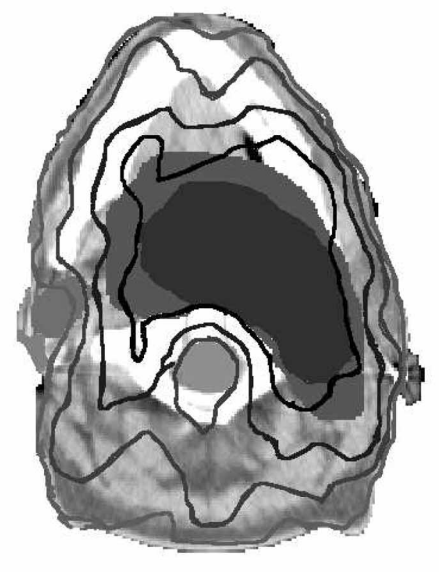

Figure 4. Final dose distribution at the base of skull. The boost volume lies in the center and is surrounded by the adjuvant volume. The right parotid can be seen to the left. Isodose lines in dark from innermost to outermost: 66.5, 50, 35, and 25 Gy.

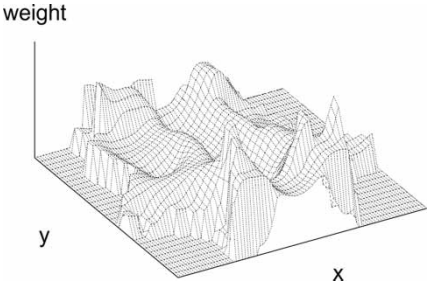

Figure 5. The optimal beamlet weight distribution for the beam impinging frontally onto the patient. The orientation of the profile is similar to the view of figure 6. The central depression of the profile mirrors the task to spare the spinal cord.

the original problem can be used as a starting point for the altered problem so that only few additional iterations of the algorithm are needed. Hence, naturally, also our approach includes some weighting of the experts, but the effect of a change of one of these constants on a solution is deterministic in that only one constraint and the objective function will be affected. Whereas for the weighted least-squares type approach mentioned in section 1, the effects of an alteration of one weight may be various and unpredictable. Also, in our case, adjustments of the constraint parameters can be guided by a sensitivity analysis [68].

In table 1, we give a summary of the average size of the optimization problems and the average number of inner and outer iterations for 127 clinical examples to which our algorithm had been applied. Naturally, these parameters can vary widely in a case class, depending on the individual geometry of the patient which can make certain treatment goals very hard to achieve. We do not provide CPU times here, as these depend strongly on various parameters, for example, the chosen discretization of the exposed volume.

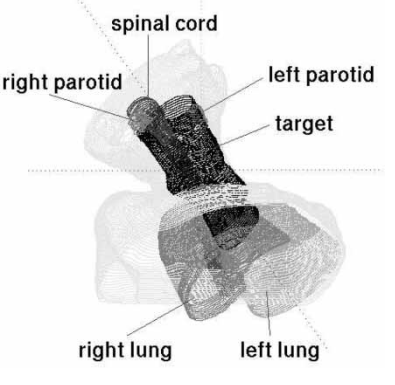

Figure 6. Patient geometry of the example case of a larynx tumor.

Table 1. Typical mean values for the size of clinical optimization problems and the associated number of iterations required to solve them.

Case class	Cases	Constraints	Variables	Outer iteration	Inner iteration
Prostate	43	6	2,600	3.7	140
Prostate/lymph nodes	13	11	9,600	6.3	330
Head and neck + nodes	59	26	10,200	7.4	510
Breast	6	11	13,100	5.1	190
Paranasal sinus	6	21	3,200	7.8	460

Remark 3 Another approach to solve an optimization problem of the given type would be to introduce m additional variables $d_i := T_i^{\mathrm{T}}\phi$, in order to make the problem functions separable. (Note that only V_1 and V_2 occur simultaneously in the objective and a constraint function.) This would add a number of typically $O(10^6)$ linear equality constraints

$$(T\ I_m)\begin{pmatrix}\phi\\ d\end{pmatrix} = 0$$

to the problem, where I_m is the $m \times m$ unit matrix and $d \in \mathbb{R}^m$ is the vector having components d_i. In an augmented Lagrangian type approach similar to ours here, these equality constraints could be treated, e.g. by the addition of the quadratic penalty function

$$\mathcal{P}(\phi, d) := \frac{1}{2\varrho}\sum_{i=1}^{m}(T_i^{\mathrm{T}}\phi - d_i)^2 = \frac{1}{2\varrho}(T\phi - d)^{\mathrm{T}}(T\phi - d)$$
$$= \frac{1}{2\varrho}\begin{pmatrix}\phi\\ d\end{pmatrix}^{\mathrm{T}}\begin{pmatrix}T^{\mathrm{T}}T & -T^{\mathrm{T}}\\ -T & I_m\end{pmatrix}\begin{pmatrix}\phi\\ d\end{pmatrix},$$

to the related Lagrangian, where $\varrho > 0$ is some penalty parameter [cf. 30, section 8.2]. Then, the gradient and Hessian of the function to be minimized in the subproblems of the algorithm have become very simple and Newton's method could be applied to solve them. In case, $(\tilde{\phi}, \tilde{d})$ is the numerical solution to this problem with $T\tilde{\phi} \approx \tilde{d}$, one could use $(\tilde{\phi}, T\tilde{\phi})$ as an approximate solution satisfying the equality constraints exactly. However, at such approach, an $(n + m) \times (n + m)$ linear system with a coefficient matrix, which includes especially the $m \times m$ matrix $T^{\mathrm{T}}T$, has to be solved at each iteration of Newton's method, and the storing and working with $T^{\mathrm{T}}T$ as well as the efficient solution of these systems would require special attention. Nevertheless, the study of such approach seems to be worthwhile to us.

Acknowledgements

This research was sponsored by the Deutsche Forschungsgemeinschaft (DFG) under grant NU 33/7-1 and NU 33/7-2.

References

[1] Brahme, A., Roos, J.E. and Lax, I., 1982, Solution of an integral equation in rotation therapy. *Medical Physics*, **27**, 1221.
[2] Bortfeld, T., Boyer, A.L., Schlegel, W., Kahler, D.L. and Waldron, T.J., 1994, Realization and verification of three-dimensional conformal radiotherapy with modulated fields. *International Journal of Radiation Oncology Biology and Physics*, **30**, 899.
[3] Bortfeld, T., Bürkelbach, J., Boesecke, R. and Schlegel, W., 1990, Methods of image reconstruction from projections applied to conformation radiotherapy. *Physics in Medicine and Biology*, **35**, 1423–1434.
[4] Brahme, A., 1995, Treatment optimization using physical and radiobiological objective functions. In: A.R. Smith (Ed.) *Radiation Therapy Physics*, pp. 209–246 (Berlin: Springer).
[5] Mohan, R., Wang, X., Jackson, A., Bortfeld, T., Boyer, A.L., Kutcher, G.J., Leibel, S.A., Fuks, Z. and Ling, C.C., 1994, The potential and limitations of the inverse radiotherapy technique. *Radiotherapy and Oncology*, **32**, 232–248.
[6] Wang, X.-H., Mohan, R., Jackson, A., Leibel, S.A., Fuchs, Z. and Ling, C.C., 1995, Optimization of intensity-modulated 3d conformal treatment plans based on biological indices. *Radiotherapy and Oncology*, **37**, 140–152.
[7] Bortfeld, T., Stein, J. and Preiser, K., 1997, Clinically relevant intensity modulation optimization using physical criteria. In *Proceeding of 12th International Center for Cancer Research*, pp. 1–4.
[8] Spirou, S.V. and Chui, C.-S., 1998, A gradient inverse planning algorithm with dose–volume constraints. *Medical Physics*, **25**, 321–333.
[9] Shepard, D.M., Ferris, M.C., Olivera, G.H. and Mackie, T.R., 1999, Optimizing the delivery of radiation therapy to cancer patients. *SIAM Review*, **41**, 721–744.

[10] Michalski, D., Xiao, Y., Censor, Y. and Galvin, J.M., 2004, The dose-volume constraint satisfaction problem for inverse treatment planning with field segments. *Physics in Medicine and Biology*, **49**, 601–616.
[11] Bortfeld, T., 1995, Dosiskonfirmation in der Tumortherapie mit externer ionisierender Strahlung: Physikalische Möglichkeiten und Grenzen. Habilitationsschrift, Universität Heidelberg, Heidelberg.
[12] Hamacher, H.W. and Küfer, K.-H., 2002, Inverse radiation therapy planning – a multiple objective optimization approach. *Discrete Applied Mathematics*, **118**, 145–161.
[13] Raphael, C., 1992, Mathematical modelling of objectives in radiation therapy treatment planning. *Physics in Medicine and Biology*, **37**, 1293–1311.
[14] Bortfeld, T., Schlegel, W., Dykstra, C., Levegrün, S. and Preiser, K., 1996, Physical vs biological objectives for treatment plan optimization. *Radiotheraphy and Oncology*, **40**(2), 185, letter, comment.
[15] Goitein, M. and Niemierko, A., 1996, Intensity modulated therapy and inhomogeneous dose to the tumor: A note of caution. *International Journal of Radiation Oncology Biology and Physics*, **36**, 519–522.
[16] Mohan, R. and Wang, X.-H., 1996, Response to Bortfeld *et al.* re physical vs biological objectives for treatment plan optimization. *Radiotherapy and Oncology*, **40**(2), 186–187.
[17] Brahme, A. and Lind, B.K., 1997, The importance of biological modeling in intensity modulated radiotherapy optimization. In *Proceedings of 12th International center for Cancer Research*, pp. 5–8.
[18] Källman, P., Lind, B.K. and Brahme, A., 1992, An algorithm for maximizing the probability of complication free tumor control in radiation therapy. *Physics in Medicine and Biology*, **37**, 871–890.
[19] Gustafsson, A., Lind, B.K. and Brahme, A., 1994, A generalized pencil beam algorithm for optimization of radiation therapy. *Medical Physics*, **21**, 343–356.
[20] Romeijn, H.E., Dempsey, J.F. and Li, J.G., 2004, A unifying framework for multi-criteria fluence map optimization models. *Physics in Medicine and Biology*, **49**, 1991–2013.
[21] Alber, M. and Nüsslin, F., 1999, An objective function for radiation treatment optimization based on local biological measures. *Physics in Medicine and Biology*, **44**(2), 479–493.
[22] Alber, M., 2000, A Concept for the Optimization of Radiotherapy. PhD thesis, Universität Tübingen, Tübingen, Germany.
[23] Alber, M. and Nüsslin, F., 2002, Ein konzept zur optimierung von klinischer IMRT. *Zeitschrift fur Medizinische physics*, **12**, 109–113.
[24] Thieke, C., Bortfeld, T., Niemierko, A. and Nill, S., 2003, From physical dose constraints to equivalent uniform dose constraints in inverse radiotherapy planning. *Medical Physics*, **30**, 2332–2339.
[25] Holder, A. and Salter, B., 2004, A tutorial on radiation oncology and optimization. In: H.J. Greenberg (Ed.) *Tutorials on Emerging Methodologies and Applications in Operations Research* (New York: Springer).
[26] Gustaffson, A., 1996, Development of a versatile algorithm for optimization of radiation therapy. PhD thesis, University of Stockholm, Stockholm, Sweden.
[27] Lim, J., 2002, Optimization in radiation treatment planning. PhD thesis, University of Wisconsin, Madison.
[28] Ferris, M.C., Lim, J. and Shepard, D.M., 2003, Radiosurgery treatment planning via nonlinear programming. *Annals of Operations Research*, **119**, 247–260.
[29] Alber, M., Meedt, G., Nüsslin, F. and Reemtsen, R., 2002, On the degeneracy of the IMRT optimisation problem. *Medical Physics*, **29**, 2584–2589.
[30] Conn, A.R., Gould, N. and Toint, P.L., 1997, A globally convergent Lagrangian barrier algorithm for optimization with general inequality constraints and simple bounds. *Mathematics of Computation*, **66**, 261–288.
[31] Ben-Tal, A., Yuzefovich, I. and Zibulevsky, M., 1992, Penalty/barrier multiplier methods for minimax and constrained smooth convex problems. Technical Report 9/92, Optimization Laboratory, Faculty of Industrial Engineering and Management, Technion, Haifa, Israel.
[32] Ben-Tal, A. and Zibulevsky, M., 1997, Penalty/barrier multiplier methods for convex programming problems. *SIAM Journal on Optimization*, **7**, 347–366.
[33] Polyak, R., 1992, Modified barrier functions (theory and methods). *Mathematical Programming*, **54**, 177–222.
[34] Polyak, R. and Teboulle, M., 1997, Nonlinear rescaling and proximal-like methods in convex optimization. *Mathematical Programming*, **76**, 265–284.
[35] Polyak, R., 2001, Log-sigmoid multipliers methods in constrained optimization. *Annals of Operations Research*, **101**, 427–460.
[36] Polyak, R., 2002, Nonlinear rescaling *vs*. smoothing technique in convex optimization. *Mathematical Programming*, **92**, 197–235.
[37] Breitfeld, M.G. and Shanno, D.F., 1994, Preliminary computational experiences with modified log-barrier functions for large-scale nonlinear programming. In: W.W. Hager, D.W. Hearn and P.M. Pardalos (Eds) *Large Scale Optimization: State of the Art*, pp. 45–67 (Dordrecht: Kluwer Academic Publication).
[38] Breitfeld, M.G. and Shanno, D.F., 1996, Computational experience with penalty-barrier methods for nonlinear programming. *Annals of Operations Research*, **62**, 439–463.
[39] Shanno, D.F., Breitfeld, M.G. and Simantiraki, E.M., 1996, Implementing barrier methods for nonlinear programming. In: T. Terlaky (Ed.) *Interior Point Methods of Mathematical Programming*, pp. 399–414 (Dordrecht/Boston/London: Kluwer Academic Publication).
[40] Kocvara, M. and Stingl, M., 2003, PENNON – a generalized augmented Lagrangian method for semidefinite programming. In: G. DiPillo and A. Murli (Eds) *High Performance Algorithms and Software for Nonlinear Optimization*, pp. 297–315 (Dordrecht: Kluwer Academic Publication).
[41] Kocvara, M. and Stingl, M., 2003, PENNON – a code for convex nonlinear and semidefinite programming. *Optimization Methods and Software*, **18**, 317–333.

[42] Conn, A.R., Gould, N.I.M. and Toint, P.L., 1992, *LANCELOT: A FORTRAN Package for Large-Scale Nonlinear Optimization (Release A)* (New York-Berlin-Heidelberg: Springer).
[43] Kavlie, D. and Moe, J., 1971, Automated design of frame structures. *Journal of Structural Division ASCE*, **ST1**, 33–62.
[44] Cassis, J.H. and Schmit, L.A., 1976, On implementation of the extended interior penalty function. *International Journal for Numerical Methods Engineering*, **10**, 3–23.
[45] Conn, A.R., Gould, N.I.M. and Toint, P.L., 1992, A globally convergent Lagrangian barrier algorithm for optimization with general inequality constraints and simple bounds. Technical Report 92/07, Department of Mathematics, FUNDP, Namur, Belgium.
[46] Nocedal, J. and Wright, S.J., 1999, *Numerical Optimization* (New York-Berlin-Heidelberg: Springer).
[47] Meedt, G., Alber, M. and Nüsslin, F., 2003, Non-coplanar beam direction optimization for intensity modulated radiotherapy. *Physics in Medicine and Biology*, **48**(18), 2999–3019.
[48] Alber, M. and Nüsslin, F., 2001, Optimization of intensity modulated radiotherapy under constraints for static and dynamic MLC delivery. *Physics in Medicine and Biology*, **46**, 3229–3239.
[49] Alber, M. and Nüsslin, F., 2000, Intensity modulated photon beams subject to a minimal surface smoothing constraint. *Physics in Medicine and Biology*, **45**, N49–N52.
[50] Ahnesjö, A. and Aspradakis, M.M., 1999, Dose calculations for external photon beams in radiotherapy. *Physics in Medicine and Biology*, **44**, R99–R156.
[51] Laub, W., Alber, M., Birkner, M. and Nüsslin, F., 2000, Monte Carlo dose computation for IMRT optimization. *Physics in Medicine and Biology*, **45**, 1741–1754.
[52] Munro, T.R. and Gilbert, C.W., 1961, The relation between tumor lethal doses and the radiosensitivity of tumor cells. *The British Journal of Radiology*, **34**, 246–251.
[53] Niemierko, A., 1999, A generalized concept of equivalent uniform dose (EUD) (abstract). *Medical Physics*, **26**, 1100.
[54] Jackson, A., Kutcher, G.J. and Yorke, E.D., 1993, Probability of radiation-induced complications for normal tissues with parallel architecture subject to non-uniform irradiation. *Medical Physics*, **20**, 613–625.
[55] Fletcher, R., 1991, *Practical Methods of Optimization* (2nd edn) (Chichester: John Wiley & Sons).
[56] Fiacco, A.V. and McCormick, G.P., 1968, *Nonlinear Programming: Sequential Unconstrained Minimization Techniques* (New York-London-Sydney-Toronto: John Wiley & Sons).
[57] Nash, S.G., Polyak, R. and Sofer, A., 1994, A numerical comparison of barrier and modified barrier methods for large-scale bound-constrained optimization. In: W.W. Hager, D.W. Hearn, and P.M. Pardalos (Eds) *Large Scale Optimization: State of the Art*, pp. 319–338 (Dordrecht: Kluwer Academic Publishers).
[58] Bertsekas, D.P., 1996, *Constrained Optimization And Lagrange Multiplier Methods* (Belmont, Massachusetts: Athena Scientific).
[59] Mittelmann, H.D., 1999, Benchmarks for optimization software, Available online at: http://plato.la.asu.edu/bench.html.
[60] Polyak, R.A., Griva, I. and Sobieszczanski-Sobieski, J., 1998, The Newton log-sigmoid method in constrained optimization. In Proceedings of the AIAA/USAF/NASA/ISSMO Symposium on Multidisciplinary Analysis and Optimization, St. Louis, pp. 2193–2201.
[61] Tseng, P. and Bertsekas, D.P., 1993, On the convergence of the exponential multiplier method for convex programming. *Mathematical Programming*, **60**, 1–19.
[62] Haftka, R.T. and Starnes, J.H., 1976, Applications of a quadratic extended interior penalty function for structural optimization. *American Institute of Aeronautics and Astronautics Journal*, **14**, 718–724.
[63] Conn, A.R., Gould, N.I.M. and Toint, P.L., 1997, On the number of inner iterations per outer iteration of a globally convergent algorithm for optimization with general nonlinear inequality constraints and simple bounds. *Computational Optimization and Applications*, **7**, 41–69.
[64] Press, W.H., Teukolsky, S.A., Vetterling, W.T. and Flannery, B.P., 1994, *Numerical Recipes in C*. Cambridge University Press.
[65] Touati-Ahmed, D. and Storey, C., 1990, Efficient hybrid conjugate gradient techniques. *Journal of Optimization Theory and Applications*, **64**, 379–397.
[66] Dai, Y.H. and Yuan, Y., 1999, A nonlinear conjugate gradient method with a strong global convergence property. *SIAM Journal on Optimization*, **10**, 177–182.
[67] Dai, Y.-H., 2001, New properties of a nonlinear conjugate gradient method. *Journal of Numerical Mathematics*, **89**, 83–98.
[68] Alber, M., Birkner, M. and Nüsslin, F., 2002, Tools for the analysis of dose optimisation II: Sensitivity analysis. *Physics in Medicine and Biology*, **47**, N265–N270.