Elimination of importance factors for clinically accurate selection of beam orientations, beam weights and wedge angles in conformal radiation therapy

James L. Bedford and Steve Webb

Joint Department of Physics, The Institute of Cancer Research and The Royal Marsden NHS Trust, Downs Road, Sutton, Surrey SM2 5PT, United Kingdom

(Received 29 October 2002; accepted for publication 18 April 2003; published 25 June 2003)

A method of simultaneously optimizing beam orientations, beam weights, and wedge angles for conformal radiotherapy is presented. This method removes the need for importance factors by optimizing one objective only, subject to a set of rigid constraints. This facilitates the production of inverse solutions which, without trial-and-error modification of importance factors, precisely satisfy the specified constraints. The algorithm minimizes an objective function which is based upon the single objective to be optimized, but which is forced to an artificially high value when the constraints are not met, so that only satisfactory solutions are allowed. Due to the complex nature of the objective function space, including multiple local minima separated by large regions of plateau, a random search technique equivalent to fast simulated annealing is used for producing inverse plans. To illustrate the novel features of the new algorithm, a simulation is first presented, for the case of a cylindrical phantom. The morphology of the objective function space is shown to be significantly different for the new algorithm, compared to that for a conventional quadratic objective function. Clinical cases for prostate and craniopharyngioma are then presented. For the prostate case, the objective is to reduce irradiated rectal volume. Three-field, four-field, and six-field optimizations, with or without orientation optimization, are shown to provide solutions which are consistent with previously reported plans and class solutions. For the craniopharyngioma case, which involves the use of a high-precision stereotactic conformal technique, the objective is to reduce the irradiated volume of normal brain. Practically feasible beam angles are produced which, compared to a standard plan, provide a small but worthwhile sparing of normal brain. The algorithm is thereby shown to be robust and suitable for clinical application. © 2003 American Association of Physicists in Medicine. [DOI: 10.1118/1.1582471]

Key words: conformal radiation therapy, inverse planning, optimization, importance factors, beam orientations

I. INTRODUCTION

Although optimization and inverse treatment planning have become firmly established as part of the conformal radiation therapy process, the important issue of the selection of beam orientations has received relatively little attention. Moreover, beam orientation optimization presents a particular challenge in view of the computational burden of searching through many combinations of beam angle, each of which is dependent upon the beam weights and wedge angles used in conjunction with it.¹⁻⁵

Various methods have been proposed to overcome this computational demand. These methods are comprehensively reviewed by Webb.⁶ Principally, they consist of (1) limiting the geometry to a single plane, so as to reduce the number of possible beams and the complexity of the dose calculation, (2) considering a single beam in isolation and finding which are the most suitable angles for irradiation, then basing the treatment plan on these angles, and (3) simplifying the dose calculation used, so that the computation required for each beam orientation combination is minimal. All of these methods are successful in achieving solutions within a reasonable length of time. However, they inevitably introduce some aspect of compromise into the problem. For example, limiting the geometry to a single plane excludes the more challenging non-coplanar cases where orientation selection is of greatest benefit. Considering a single beam in isolation neglects the effects of dose superposition, which is the essence of providing adequate dose to the planning target volume (PTV) whilst avoiding the critical structures. A classic example of this is the four-field box plan for prostate radiotherapy, where all of the beams pass through critical structures, yet the resulting plan spares the rectum and bladder well because the high dose region (treated volume) is cuboidal in shape.⁷ Simplifying the dose calculation is probably the most promising approach, although this must be negotiated carefully, otherwise a so-called “optimal” solution may be obtained, which, upon application of a more accurate dose computation, may turn out to be less than optimal.

Recently, Rowbottom et al.³ studied the beam orientation optimization problem in detail. They devised a scheme in which the proximity of the PTV to the surface of the patient, and the position of the organs at risk (OARs) were evaluated using a single-beam objective function. This provided a measure of optimality for each coplanar beam orientation. The
overall treatment plan was then derived by choosing several beams according to the single-beam objective function. This had the advantage that it avoided the need to search intensively through many combinations of beam angle, but neglected beam overlap effects. This approach was then extended to non-coplanar geometry and a multiple-beam objective function was introduced to take into account the superposition of the various beams. A more sophisticated method performed a simulated annealing search over beam orientations, with a downhill simplex search over beam weights at each step. This work has also been recently extended to intensity-modulated radiation therapy (IMRT), where fast simulated annealing was used for optimization of beam orientations, and a filtered projection was used to optimize the beam weights for each angle combination.

A similar approach has been taken by Pugachev et al. for the case of IMRT. For the relatively fast case of coplanar optimization, they used a scheme in which the beam orientations were chosen by simulated annealing, and at each iteration, the beam weights were chosen by filtered backprojection. For the more computationally demanding case of three-dimensional optimization, they applied a beam’s eye view method in which the beam was divided into beamlets. For each of these beamlets, the maximum dose deliverable to the planning target volume without exceeding the tolerance doses of the critical structures, was evaluated. The sum of these individual scores then formed the total score for that beam. This provided a measure of the goodness of a particular beam orientation; consideration of all orientations then allowed an informed selection of appropriate beams. However, as discussed above, this method was not able to take into account the true dose distribution resulting from the interaction of the separate fields. The effects of overlapping fields were taken into account by performing a simulated annealing search which was weighted towards the promising angles indicated by the beam’s eye view method.

As well as limitations on accuracy due to dosimetric simplifications needed to overcome the computational complexity of the problem, all of the methods so far proposed are limited by the inherent imprecision of importance factors. These relative weights govern the compromise between the competing goals in the treatment plan, and they are universally used in optimization of radiation therapy, whether for beam weights or beam orientations. The difficulty, however, is that they are unintuitive, and they must often be set to irrelevant values in order to simply provide a sensible dose distribution. Most significantly, it is almost impossible to
know in advance which importance factors to use. The dosimetrist must therefore apply the optimization algorithm, adjust the importance factors, and then repeat the procedure. This is cumbersome and impractical. To overcome this, it has been suggested that the importance factors should be themselves optimized, but this requires an alternative means of assessing the goodness of the plan.

This paper describes a new optimization strategy which addresses these issues to provide a robust tool for clinical treatment plan optimization. The technique is addressed to conformal radiotherapy without intensity-modulation, since this type of technique still constitutes the majority of clinical treatments. However, the concepts are also applicable to intensity-modulated radiation therapy. The principal features of the method are that (1) it avoids the use of importance factors, thereby allowing the dosimetrist to produce a solution without unnecessary iterations of the scheme; (2) it is accurate, creating conformally blocked fields for a variety of beam orientation combinations, and calculating dose using a fast convolution algorithm including heterogeneity correction; and (3) it is robust, using dose-volume statistics and biological models to produce feasible clinical solutions.

The paper is divided into three sections. In the first, the algorithm is described in detail, and in the second, it is illustrated with reference to a simple simulation, which shows clearly the behavior of the algorithm and its difference to established methods. In the final section, the algorithm is applied to two case studies illustrating its performance and utility. In particular, the results are compared with previous optimizations for prostate radiotherapy, as a means of benchmarking the technique.

II. METHOD

A. Clinical objectives and constraints

The algorithm is initiated by supplying a clinical objective and a set of clinical constraints. The user constructs a statement of the form: “Maximize (minimize) X, such that Y is less than (greater than) Z,” where X and Y are dose statistics for selected anatomical structures and Z is the constraint on Y. The top panel of Fig. 1 illustrates the software implementation of this concept. The first row of the grid gives the clinical objective of the optimization. The objective is specified by an anatomical structure and a dose-volume or biological statistic for this structure. Note that there can only be one objective. The subsequent rows give a set of constraints which must be met. These constraints are also specified by an anatomical structure, a dose-volume or biological statistic and the constraint level for this statistic. Table I lists the complete set of items from which the objective and constraints may be constructed. In the software implementation of the algorithm, Table I, column 5 offers dose, volume or neither, depending upon the statistic selected from column 4.

The use of one objective and a set of constraints is a key feature of the algorithm, which obviates the need for importance factors: because there is only one objective, the need for a series of competing objectives combined together into a single mathematical function is not necessary. This allows the algorithm to be guided by the user much more precisely. The reader may consider the use of only one objective to be a limitation, but in practice, this is usually the method by which treatment plans are optimized. For example, if “dose to hottest” is selected from column 4, a list of percentage volumes appears in column 5. If “volume irradiated to” is selected from column 4, a list of doses appears in column 5, and if “mean dose” is selected, nothing appears in column 5. Similarly, the constraint levels in column 7 are offered as doses in Gray or percentage volumes, depending upon what is selected in column 4.

The use of one objective and a set of constraints is a key feature of the algorithm, which obviates the need for importance factors: because there is only one objective, the need for a series of competing objectives combined together into a single mathematical function is not necessary. This allows the algorithm to be guided by the user much more precisely. The reader may consider the use of only one objective to be a limitation, but in practice, this is usually the method by which treatment plans are optimized. For example, there is often one important dose-limiting structure, which, when selecting beam orientations, a dosimetrist may wish to avoid more than all others. In this case, the dose-limiting structure would form the basis for the objective statement, and the

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Maximize</td>
<td>structure A</td>
<td>RMS dose around</td>
<td>1, 2, 3 Gy, etc.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Minimize</td>
<td>structure B</td>
<td>dose to hottest</td>
<td>1, 2, 3%, etc.</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>structure C</td>
<td>vol. irradiated to</td>
<td>1, 2, 3 Gy, etc.</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>structure D</td>
<td>mean</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>structure E</td>
<td>maximum</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>structure F</td>
<td>minimum</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>structure G</td>
<td>TCP</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>structure H</td>
<td>NTCP</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>structure A</td>
<td>RMS dose around</td>
<td>1, 2, 3 Gy, etc.</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>structure B</td>
<td>dose to hottest</td>
<td>1, 2, 3%, etc.</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>structure C</td>
<td>vol. irradiated to</td>
<td>1, 2, 3 Gy, etc.</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>structure D</td>
<td>mean</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>structure E</td>
<td>maximum</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>structure F</td>
<td>minimum</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>structure G</td>
<td>TCP</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>structure H</td>
<td>NTCP</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
other organs at risk, and perhaps the planning target volume, would then become constraints. Alternatively, having selected orientations, a dosimetrist may concentrate upon achieving PTV dose homogeneity, without unduly considering the organs at risk. In this situation, the PTV is the structure for the objective statement and all of the organs at risk form constraints. Hence, the use of a single objective, together with a variety of constraints is precise, and works well in practice, as will be demonstrated in the results section.

The drawback with this approach is that in some cases, no combination of beam parameters is able to satisfy the set of constraints. When this occurs, the program returns a message to state that no solution has been found. The user must then manually modify the constraints. This could give rise to a situation in which the constraints are being manipulated instead of importance factors, but in practice, suitable constraints can be determined for particular treatment sites.

The converse situation, in which the dose statistics satisfy the constraints but are not necessarily optimal, may also exist. For example the user may specify a maximum dose of 48 Gy to the spinal cord. The program may return a treatment plan which gives 40 Gy to the spinal cord, when another similar plan exists which gives only 35 Gy to the spinal cord. This is because the algorithm is designed simply to test that the constraints are satisfied, rather than to perform optimization within the constraint boundaries. In this instance, it may be appropriate to run the optimization several times, with progressively tighter constraints, to prove that the current set of beam parameters cannot be improved upon. However, the experience of the authors is that optimizing the single objective cannot be improved upon. However, the progress of the treatment plan which gives 40 Gy to the spinal cord, when another similar plan exists which gives only 35 Gy to the spinal cord. This is because the algorithm is designed simply to test that the constraints are satisfied, rather than to perform optimization within the constraint boundaries. In this instance, it may be appropriate to run the optimization several times, with progressively tighter constraints, to prove that the current set of beam parameters cannot be improved upon. However, the experience of the authors is that optimizing the single objective is sufficient to force most of the constrained dose statistics to the constraint boundaries.

B. Physical constraints

The second panel of Fig. 1 contains a list of physical constraints on beam characteristics. There can be up to six fields in the treatment plan, each field being specified by completing a row of the grid. There is nothing fundamental about the choice of six fields as the maximum number of fields. This number of fields has been found to be sufficient so far; more fields could be allowed by simple changes to the software. Working from left to right across a row of the grid, the field parameters are (1) the beam, including machine and energy, (2) the blocking device, (3) the wedge angle, and (4) a range of gantry, couch and collimator angles which the algorithm is to search.

The blocking device is a conformal block, shaped accurately to the beam’s eye view of the PTV. The margin between the PTV and the block edge can be selected as required. The methods used for this are similar to those described by Bedford and Shentall.14 The conformal block is sufficient to provide a conformal field, without introducing issues relating to leaf fitting and collimator orientation. Note that this method of shaping gives rise to dose distributions which are always convex; additional segments covering only part of the PTV would be required to introduce concavities.

The wedge angle can be set from $0^\circ$ (i.e., open field) to $60^\circ$ in $15^\circ$ steps, or left variable for the optimization to determine the most appropriate angle. In the latter case, the wedge angle is varied from $0^\circ$ to $60^\circ$ in $10^\circ$ steps. The wedge is universal, with the dose distribution being constructed from the beam data for an open field and a $60^\circ$ wedged field.15 If the wedge angle is set specifically, then the appropriate open and wedged fields are combined with the appropriate weight. If the angle is to be optimized, the relative weights of the open and wedged fields are adjusted during the optimization process, as proposed by Oldham et al.16 and Xing et al.17

A range of gantry and couch angles can be searched, in $10^\circ$ steps. Specifying a range of gantry and couch angles is a practical solution to the problem of avoiding collisions on the treatment unit. By empirical means, it is possible to define a range of feasible couch angles for each gantry angle. Geometrical18 and model-based solutions19 to this problem have also been described in the literature. The geometrical approaches have the disadvantage that they tend to be applicable only to a limited range of treatment sites, with different techniques having different constraints, according to the devices used for that treatment and the precise position of the patient. The model-based solutions attempt to provide a true model of the linear accelerator gantry and couch, and the patient, to give an accurate indication of treatment possibilities. However, in order to describe the treatment geometry with sufficient accuracy, the model is required to be complex, and therefore difficult to generate. The method used in the present work provides the user with a simple means of reliably specifying what range of couch and gantry angles is to be searched, based upon their own knowledge of what is feasible for the particular linear accelerator and patient. This prevents beams from entering the patient from regions where no CT information is available and from exiting through the entire patient length. It also reduces the search space and therefore facilitates the production of a practical solution in a shorter time. It does have limitations: for example, over a given range of gantry angles, the feasible couch angles vary, and this cannot be specified using the current method. However, in practice, by selecting the beams so that each one can be given a different range of couch angles, this is not found to be a problem.

The collimator can also be varied over a range of angles, in steps of $10^\circ$. This allows wedged fields to be positioned with the fluence gradient optimally oriented. Note that the collimator orientation does not affect the field shaping, as the conformal block is fitted to the beam’s eye view of the PTV. However, this would no longer be the case if a multileaf collimator were used to shape the beam.

C. Beam weights

The options at the lower left of Fig. 1 specify which of the anatomical structures is to be taken as the PTV, and what the isocentric dose is to be. The beam weights can be selected to be adjusted in fine intervals, coarse intervals, or not to be optimized at all. If fine weights are selected, the algorithm assigns to each field a relative isocentric weight of between
0.1 and 1, in steps of 0.1. If coarse weights are selected, the algorithm increments the beam weights in steps of 0.25, from 0.25 to 1. If unit weights are selected, each beam contributes equally to the dose at the isocentre. After calculation of the dose distribution using the arbitrary relative weights, the dose distribution is normalized to the prescribed dose at the isocenter.

**D. Treatment plans**

The optimization proceeds by constructing a series of treatment plans with the number of fields given in the list of physical constraints. As described above, these plans are isocentric, conformally blocked plans with universally wedged beams. The treatment plans are constructed in conjunction with a model of the patient. The patient model is a three-dimensional grid of voxels which cover a computed tomography (CT) scan of the patient. In the longitudinal (z-) direction, the centers of the voxels are aligned with the CT slices. The superior and inferior edges of the voxels lie midway between CT slices. The lateral (x) and anteroposterior (y) dimensions of the voxels are equal for all voxels, and are chosen to be 5 mm, although this can be changed if required.

Each voxel contains all essential information relating to a particular region of the patient. It includes a position coordinate, the width, height and length of the voxel, the CT number of that voxel, the primary fluence at the center of the voxel, the absorbed dose due to each field of a treatment plan, and the total dose. Storage of the dose due to each treatment field allows for the total dose to be simply re-summed if only the beam weights of the plan change. The voxel also contains a list of the volumes of interest of which that voxel forms part, for calculation of dose-volume histograms (DVHs) and dose statistics.

For each treatment plan, dose is calculated using a fast convolution algorithm.\(^{20}\) This scheme uses a restricted scatter kernel to provide a full three-dimensional scatter convolution without the time requirement of a traditional convolution dose calculation algorithm. The calculation has been shown to provide accuracy in isocenter dose of better than 5%, and root-mean-square accuracy in dose-volume histograms of better than 5%. The calculation time is around 4 s per beam.\(^{20}\) This therefore facilitates the evaluation of a wide range of treatment plans in a feasible length of time.

**E. Optimization scheme**

The purpose of the optimization algorithm is to determine the combination of beam orientations, collimator rotations, beam weights and wedge angles which minimizes or maximizes the clinical objective, subject to the clinical and physical constraints. Accordingly, an objective function is constructed to reflect the suitability of each plan. The objective function is simply equal to the statistic to be optimized, as given by the top row of the grid in Fig. 1. If the goal is to maximize a statistic, the negative of the statistic is used, so that optimization always consists of minimizing the objective function. If the clinical constraints are met, the objective function is not further adjusted. However, if the clinical constraints are not met, the objective function is set to an artificially high value (120), larger than any value of the statistic being minimized, to signify an unacceptable plan. Thus, treatment plans which do not satisfy the constraints are effectively excluded from the optimization, and the algorithm returns the optimal value of the statistic in the first row of the grid, subject to the given constraints. If no treatment plan is found to satisfy the constraints, the algorithm concludes by stating that this is the case, and no treatment plan is offered.

Mathematically, the problem can be considered as follows. The goal is to minimize an objective function \(f(d(x))\), where \(d(x)\) is a distribution of dose among the voxels of the patient model, and \(x\) is a vector of parameters to be selected, such as beam orientations and weights. The objective function may be divided into various terms, \(f_1, f_2, f_3, \ldots\), involving different dosimetric attributes such as mean-square dose, integral dose, dose-volume functions or biological indices. These may relate to the same or different anatomical structures. The traditional approach has been to choose \(f\) such that it is quadratic with respect to \(d\), in which case it is also quadratic with respect to \(x\), since \(d\) is a linear function of \(x\). However, in the present approach, this is not required. Instead, the objective function takes the form

\[
 f(d(x)) = A - (A-f_1(d)) \prod_{i=2}^{C+1} H(c_i-f_i(d)),
\]

where \(A\) is a constant which is larger than any of \(f_1, f_2, f_3, \ldots\), the \(c_i\) are limits on the \(C\) constrained values of \(f_i\), and \(H\) is the Heaviside step function. The effect of this expression is that when all of the constraints are satisfied, the product becomes unity, and the objective function takes the value of \(f_1\). If any one of the constraints is not met, the product vanishes and the objective function takes the value of the constant \(A\). The above form of the expression is for constraints specifying that the \(f_i\) should be less than the tolerances \(c_i\); for the corresponding case where they should be greater than the tolerances, the argument of the Heaviside step function is negated. There can be a mixture of these two types of term within one objective function. The complete objective function is then:

\[
 f(d(x)) = A - (A-f_1(d)) \prod_{i=2}^{C_L+1} H(c_i-f_i(d)) \times \prod_{i=C_L+2}^{C_L+G+1} H(f_i(d)-c_i)
\]

for \(C_L\) constraints where the \(f_i\) must be less than the \(c_i\) and \(C_G\) constraints where the \(f_i\) must be greater than the \(c_i\). It is well known that the objective function for beam orientation optimization contains local minima.\(^{21}\) The situation with regard to beam weights is less clear, Deasy\(^{22}\) having shown that local minima exist, but Wu and Mohan\(^{23}\) having shown that they are little problem in practice. In the present study, there are clearly local minima when considering beam orientations (see results), and although it is less clear whether they are present when considering beam weights, the objective function does have a very irregular form, with disconti-
nuities at constraint boundaries. Furthermore, the objective function can include biological indices, with tumor control probability being calculated by the Webb and Nahum model, and normal tissue complication probability being calculated by the Lyman model in conjunction with the Kutcher–Burman histogram reduction scheme. The objective function may therefore contain local minima when considering beam weights. Hence, a gradient-based technique is not used. Instead, either an exhaustive search, classical simulated annealing or a random search is used to optimize the beam orientations and beam weights simultaneously. Each of these searches is discrete, with the starting, stopping and incremental values as described above. The exhaustive search is provided for use when beams with predetermined angles are to be optimized. In this case, the parameters to be optimized are $N$ beam weights and $N$ wedge angles, where $N$ is the number of fields in the plan. This number of parameters is just about small enough to permit the entire set of parameters to be evaluated in turn.

Classical simulated annealing reduces the time required when orientations are to be searched as well as beam weights and wedge angles. In this case, the parameters to be optimized are $N$ couch angles, $N$ gantry angles, $N$ collimator angles, $N$ beam weights, and $N$ wedge angles, so that an exhaustive search is no longer feasible. The annealing process consists of an exponential temperature reduction in 100 steps, with a temperature reduction factor of 0.9 between each step, from a value equal to the highest possible value of the objective function (120, which is also used to signify that a constraint has not been satisfied). At each temperature, 10 iterations are carried out per parameter to be optimized. In other words, for a four-field plan, with couch angles, gantry angles, collimator angles, beam weights and wedge angles to be optimized, there are $10 \times 5 \times 4 = 200$ iterations per temperature. At each iteration, each parameter is adjusted by $-1$, 0 or $+1$ steps and a Metropolis criterion is used to determine whether this combination is accepted. This criterion states that if the objective function has reduced, the new combination is accepted unconditionally, and if the objective function has increased, the new combination is accepted with probability $\exp(-d f / t)$, where $d f$ is the change in the objective function, and $t$ is the temperature. If one successful adjustment per parameter is made at a given temperature, before the total number of iterations is reached for that temperature, the temperature is prematurely reduced. In the above illustration for four fields, this would occur after $1 \times 5 \times 4 = 20$ successful adjustments.

The number of iterations used is a compromise between an accurate solution and a practical optimization time. Given the complexity of the search space, it is desirable to search the space comprehensively. However, this would take a prohibitive length of time. On the other hand, a fast optimization would be practical but would lead to suboptimal solutions. The present optimization scheme is therefore arranged so as to provide a satisfactory solution after a reasonable length of time.

In the random search, the algorithm begins at the center of the space of parameter combinations and then searches randomly throughout the space. The first evaluation of the objective function takes place after one random step has been taken away from the center of parameter space. The number of iterations is set to be equal to 100 times the number of parameters to be optimized. Given a particular current combination of parameters, the next combination is found by sampling each parameter from a uniform distribution centered at the current point in parameter space. The width of this distribution is reduced linearly from the entire width of the parameter space at the beginning of the optimization, to just $\pm 1$ parameter increment at the end of the optimization. This method has sometimes been referred to as fast simulated annealing. As with the classical simulated annealing scheme, the number of iterations is chosen to be a pragmatic compromise between speed and optimality of solution. Note that neither the classical simulated annealing schedule nor the random search require gradient information from the objective function, and they are therefore well suited to the extensive “plateau” regions and deep “valleys” of the objective function. As is usually the case, the random search (or fast simulated annealing) is faster than classical simulated annealing, because the former is able to explore the whole extent of the search space initially, whereas the latter “wanders” around on the plateau regions using small steps for rather longer before finding the appropriate valley.

**F. Software design**

The scheme described above has been implemented using Java technology (Sun Microsystems, Palo Alto, CA) for maximum cross-platform portability. At present, it is resident on a 440-MHz Sun Ultra 10 workstation. A summary of the modular design is shown in Fig. 2. The program is multi-threaded, with one parallel process controlling the graphical user interface, and a second parallel process controlling the application, including optimization and dose computation. The thread controlling the user interface runs at high priority.
so that the interface remains responsive to the user, while the application thread runs at lower priority, so that calculations are performed whenever the system performance is not required for graphical purposes.

Separate objects store and handle the data associated with the graphical user interface and application. The application object stores the patient and plan data, together with the beam data for dose calculation. For optimization, an optimization engine is specified, which suggests various treatment plan parameters according to which optimization method has been selected. These treatment plan parameters are then implemented into the plan object, dose is calculated, and the objective function updated.

G. Simulations

The functionality of the algorithm was demonstrated by means of two simulated clinical cases. The purpose of these simulations was to provide a simple illustration of the concepts used by the new algorithm. In particular, the algorithm was contrasted with the traditional approach which used a quadratic objective function. For both simulations, the patient was taken to be a cylinder of diameter 250 mm, with a central cylindrical PTV of diameter 50 mm and length 50 mm. Three OARs, also of diameter 50 mm and length 50 mm, surrounded the PTV (Fig. 3). Both simulations used two fields, so that the results could be conveniently represented in two dimensions. A dose of 10 Gy was prescribed to the isocenter, which lay at the center of the PTV. In the first simulation [Fig. 3(a)], an anterior field and a lateral field were used, and the wedge angles and weights of the fields were optimized such that the RMS dose variation in the PTV was minimized, subject to the mean dose in the OARs being less than 5 Gy. The number of variables in the optimization was sufficiently small for all combinations of beam weights to be searched. For each combination of beam weights, the effective wedge angle was also optimized for each field, and the minimum objective value taken as the objective function value for that beam weight combination.

In the second simulation [Fig. 3(b)], the dose prescription was the same as for the first simulation. However, a fixed anterior field was used, and the second field varied in angle from 0° to 180° in 10° steps. Since the problem was laterally symmetric, this covered all possible angles for the second field. Both fields were specified to contribute equally to the isocentric dose, so as to limit the number of variables to be optimized in this demonstration. For each combination of beam angles, the wedge angles were optimized and the lowest objective function taken to be the objective function for that combination of beam angles.

H. Clinical cases

The clinical application of the optimization scheme was illustrated by means of two clinical cases: a prostate tumor and a cranioopharyngeal tumor, the former illustrating coplanar optimization and the latter noncoplanar optimization.

The prostate patient had previously been CT scanned and treated in a supine position. The clinical target volume (CTV) for the prostate tumor consisted of prostate plus seminal vesicles. A margin of 10 mm (7 mm posteriorly/superiorly/inferiorly) was added to the CTV to account for patient repositioning and internal organ motion. The prescription dose was 74 Gy to the isocenter, and the clinical objective (Table II) was to reduce the volume of rectum irradiated to 60 Gy ($V_{60}$). To control the occurrence of hotspots, all tissue further than 30 mm from the PTV was designated normal tissue and contoured accordingly, the region of 30 mm around the PTV being unconstrained to allow for the projecting corners of the polygonal treated volume. (The normal tissue volume was created by expanding the PTV 30 mm, then expanding the outer contour of the patient 0 mm, while excluding the expanded PTV.) Conformal blocks, with a margin of 6 mm between the block and the...
PTV, were used. Three-field, four-field, and six-field 6-MV coplanar plans were considered (Table III). For each type of plan, a standard plan was used as a reference. This consisted of standard beam orientations, beam weights and wedge angles. A weight-optimized plan was then created, in which the beam weights and wedge angles were optimized, but the orientations were unchanged. Finally, the beam orientations, beam weights and wedge angles were all simultaneously optimized. Since this type of optimization had previously been carried out by optimizing the beam angles manually and the beam weights and wedge angles using simulated annealing,7,31–33 this study had the advantage that it could be compared against previous results.

The brain tumor was a suprasellar craniopharyngioma,

### Table II. Clinical objective and clinical constraints used for optimization of the prostate case. Prescription dose is 74 Gy to the isocenter. (* represents the normal tissue maximum dose is 74 Gy for the three-field orientation-optimized plan in view of the higher maximum doses occurring with a three-field plan.)

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Minimize PTV</td>
<td>Rectum</td>
<td>volume irradiated to 60 Gy</td>
<td>volume irradiated to 70 Gy</td>
<td>greater than 90%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Left femoral head</td>
<td>volume irradiated to 52 Gy</td>
<td>less than 10%</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Right femoral head</td>
<td>volume irradiated to 52 Gy</td>
<td>less than 10%</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Normal tissue</td>
<td>maximum dose</td>
<td>less than 70 Gy *</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Table III. Allowed beam parameters for optimization of the prostate case using three-field (3F), four-field (4F), and six-field (6F) plans. Beam weights are isocentric weights, and are expressed as a percentage of the isocentric dose.

<table>
<thead>
<tr>
<th>Plan</th>
<th>Gantry angle (deg)</th>
<th>Couch angle (deg)</th>
<th>Collimator angle (deg)</th>
<th>Wedge angle (deg)</th>
<th>Beam weight (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>3F standard</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>33.3</td>
</tr>
<tr>
<td>3F weight-optimized</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>33.3</td>
</tr>
<tr>
<td>3F orientation optimized</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>33.3</td>
</tr>
<tr>
<td>4F standard</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>25</td>
</tr>
<tr>
<td>4F weight optimized</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>25</td>
</tr>
<tr>
<td>4F orientation optimized</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>25</td>
</tr>
<tr>
<td>6F standard</td>
<td>50</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>16.7</td>
</tr>
<tr>
<td>6F weight optimized</td>
<td>50</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>16.7</td>
</tr>
<tr>
<td>6F orientation optimized</td>
<td>50</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>16.7</td>
</tr>
</tbody>
</table>

Medical Physics, Vol. 30, No. 7, July 2003
situated medially. The patient had previously been treated in a supine position. The gross tumor volume (GTV) was expanded by 5 mm (8 mm superiorly/inferiorly) to allow for microscopic spread and patient motion, the latter being minimized by use of a stereotactic frame.34 A dose of 50 Gy was prescribed to the isocenter. For this case, the beam orientations, beam weights and wedge angles were to be optimized such that the volume of normal brain irradiated to 25 Gy (V25) was minimal, subject to satisfactory PTV coverage (Table IV). In accord with the local protocol for stereotactically-guided conformal radiotherapy, the treatment used four noncoplanar 6-MV fields, an anterior superior oblique, a posterior superior oblique and two lateral inferior oblique fields. The fields were shaped by conformal blocks with a margin of 4 mm between the block edge and the PTV. The possible geometries specified to the optimization algorithm are shown in Table V. As with the prostate case, a standard plan was compared to a weight-optimized plan and an orientation-optimized plan. The range of beam orientations allowed for the orientation-optimized plan was chosen so that the beams did not enter or exit the head through the stereotactic frame. The thyroid and mouth were also to be avoided. In addition, by outlining several slices of thorax and constraining the dose in this volume to be less than 3 Gy (see Table IV), the beams were discouraged from exiting through the thorax.

### III. RESULTS

#### A. Simulations

For the first simulation, considering the optimization of beam weights and wedge angles for fixed beam orientations, the form of the objective function is shown in Fig. 4(a). It is seen that there are a large range of beam weights where the objective function has a uniform value of 20, signifying that the solutions are unacceptable, that is, that the constraint on the OAR mean dose is not met. Within this plateau is a narrow “valley” region where the solutions are satisfactory, and the minimum value of objective function within this region marks the solution to the problem. The valley region is radial with respect to the beam weights, since multiplying each of the beam weights by a constant factor simply alters the absolute dose but not the distribution of dose. After normalization to 10 Gy at the isocenter, the dose statistics are unaffected by the relative change in beam weights.

Figure 4(b) shows the variation of the objective function with the weight of Beam 1. (The weight is expressed as a percentage contribution of Beam 1 to the isocentric dose of 10 Gy.) It is clear from Fig. 4(b) that the global solution to the optimization problem occurs when W1 = 64%; i.e., 6.4 Gy should be delivered using Beam 1 and the remaining dose using Beam 2. The dose statistics for this solution are

<table>
<thead>
<tr>
<th>Plan</th>
<th>Gantry angle (deg)</th>
<th>Couch angle (deg)</th>
<th>Collimator angle (deg)</th>
<th>Wedge angle (deg)</th>
<th>Beam weight (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>4F standard</td>
<td>320</td>
<td>270</td>
<td>270</td>
<td>0</td>
<td>25</td>
</tr>
<tr>
<td></td>
<td>230</td>
<td>270</td>
<td>90</td>
<td>0</td>
<td>25</td>
</tr>
<tr>
<td></td>
<td>250</td>
<td>10</td>
<td>180</td>
<td>45</td>
<td>25</td>
</tr>
<tr>
<td></td>
<td>110</td>
<td>350</td>
<td>180</td>
<td>15</td>
<td>25</td>
</tr>
<tr>
<td>4F weight-optimized</td>
<td>320</td>
<td>270</td>
<td>270</td>
<td>0–60</td>
<td>10–90</td>
</tr>
<tr>
<td></td>
<td>230</td>
<td>270</td>
<td>90</td>
<td>0–60</td>
<td>10–90</td>
</tr>
<tr>
<td></td>
<td>250</td>
<td>10</td>
<td>180</td>
<td>0–60</td>
<td>10–90</td>
</tr>
<tr>
<td></td>
<td>110</td>
<td>350</td>
<td>180</td>
<td>0–60</td>
<td>10–90</td>
</tr>
<tr>
<td>4F orientation-optimized</td>
<td>290–0</td>
<td>270</td>
<td>0–350</td>
<td>0–60</td>
<td>10–90</td>
</tr>
<tr>
<td></td>
<td>180–250</td>
<td>270</td>
<td>0–350</td>
<td>0–60</td>
<td>10–90</td>
</tr>
<tr>
<td></td>
<td>220–320</td>
<td>340–20</td>
<td>0–350</td>
<td>0–60</td>
<td>10–90</td>
</tr>
<tr>
<td></td>
<td>40–140</td>
<td>340–20</td>
<td>0–350</td>
<td>0–60</td>
<td>10–90</td>
</tr>
</tbody>
</table>
OAR2 are only irradiated by Beam 2, and therefore as a compensate for any imbalance in the beam weights. OAR1 and since the wedge angles are adjusted by the algorithm to equally weighted. The curve is fairly flat around this region at approximately 50%, i.e., Beam 1 and Beam 2 are approximately equally weighted. The curve is fairly flat around this region since the wedge angles are adjusted by the algorithm to compensate for any imbalance in the beam weights. OAR1 and OAR2 are only irradiated by Beam 2, and therefore as $W_1$ increases, the weight of Beam 2 decreases, and the mean doses to OAR1 and OAR2 therefore decrease accordingly. The reduction in the objective function for these OARs is linear, as the mean dose is a linear function of Beam 2 weight. The mean dose for OAR2 is higher than for OAR1 as Beam 2 enters through OAR2 but exits through OAR1, with a corresponding reduction in depth dose on exit. OAR3 is only irradiated by Beam 1 and therefore its mean dose increases linearly as $W_1$ increases. The magnitude of mean dose for this OAR is similar to that for OAR1, since it receives exit dose.

The optimization scheme selects the lowest value of PTV RMS dose around 10 Gy, subject to OAR mean dose being less than 5 Gy. From Figs. 4(d)–4(f), it can be seen that this OAR constraint is only met for values of $W_1$ between 64% and 73%. Outside of this region, the optimization algorithm sets the objective function to an artificially high value to signify that solutions in this range are invalid. Between $W_1$ values of 64% and 73%, the constraints are met, and the algorithm therefore uses the value of PTV RMS dose as the objective value. The lowest value of the objective function, at $W_1 = 64\%$, provides the solution.

This method stands in stark contrast to the traditional method of using a quadratic objective function. In the quadratic approach, the curves in Figs. 4(c)–4(f) are linearly combined to provide the objective function. This is illustrated in Fig. 4(g). In this figure, the PTV and the OARs all have importance factor equal to unity, which is a reasonable starting point for this method. However, the curve is heavily biased towards high $W_1$, which is not surprising, as Beam 1 only irradiates one OAR instead of two. An optimization scheme using this set of importance factors would return a solution consisting of Beam 1 only, with correspondingly unacceptable PTV homogeneity and dose to OAR3 (Table VI). Increasing the PTV importance factor to four times that of the OARs [Fig. 4(h)] causes the shape of the PTV objec-

Table VI. Dose statistics for optimization of beam weights and wedge angles in the simulation study. The constrained objective function is compared with a quadratic objective function with various importance factors (IF).

<table>
<thead>
<tr>
<th>Optimization</th>
<th>PTV RMS dose (Gy)</th>
<th>OAR1 mean dose (Gy)</th>
<th>OAR2 mean dose (Gy)</th>
<th>OAR3 mean dose (Gy)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constrained</td>
<td>6.48</td>
<td>2.58</td>
<td>4.96</td>
<td>4.40</td>
</tr>
<tr>
<td>Quadratic (PTV IF 1)</td>
<td>8.77</td>
<td>0.65</td>
<td>1.24</td>
<td>6.19</td>
</tr>
<tr>
<td>Quadratic (PTV IF 4)</td>
<td>6.50</td>
<td>2.53</td>
<td>4.87</td>
<td>4.42</td>
</tr>
<tr>
<td>Quadratic (PTV IF 16)</td>
<td>6.37</td>
<td>2.92</td>
<td>5.61</td>
<td>4.07</td>
</tr>
</tbody>
</table>
of the required solution, but in routine use, this would not be the case. It then becomes more difficult to determine precisely where the solution lies, especially when the OARs have different tolerances and different importance factors. Even more importantly, the form of the objective function for the quadratic method, shown in Figs. 4(g)–4(i), is very different from that for the constrained algorithm [Fig. 4(b)], which is taken to represent the true clinical situation of PTV and OARs with specific tolerance doses. This is generally the case, with the quadratic methods yielding a simplified objective function which may approximate, but not equate to, the real clinical decision process.

The results for beam orientation optimization are shown in Fig. 5(a). The global solution to the problem occurs when Beam 2 is at gantry angle 110°. Table VII shows the corresponding dose statistics. Figures 5(b)–5(e) show the values of the dose statistics for each of the individual structures as a function of Beam 2 gantry angle. The plot for the PTV shows the RMS dose around 10 Gy to be reasonably constant for gantry angles of greater than about 60°. A gantry angle for Beam 2 of less than this means that both Beam 1 and Beam 2 are entering at a similar angle, so that the wedge angle is unable to compensate. OAR1 and OAR2 show peaks at around 90° due to the passage of Beam 2 through them, OAR2 showing greater magnitude due to the entry of Beam 2 rather than its exit. Note that this would be reversed if the symmetrical case of gantry angle 270° were considered. OAR3 shows similar behavior, with a peak at around 180°, due to the entry of Beam 2, and a smaller peak at around 0°, due to its exit. For OAR1, the tolerance of 5 Gy is not exceeded at any of the gantry angles considered, for OAR2, the tolerance is exceeded between 80° and 100°, and for OAR3, the tolerance is exceeded between 0° and 30° and 140° and 160°.

### Table VII. Dose statistics for optimization of beam orientations and wedge angles in the simulation study. The constrained objective function is compared with a quadratic objective function with various importance factors (IF).

<table>
<thead>
<tr>
<th>Optimization</th>
<th>PTV RMS dose (Gy)</th>
<th>OAR1 mean dose (Gy)</th>
<th>OAR2 mean dose (Gy)</th>
<th>OAR3 mean dose (Gy)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constrained</td>
<td>5.96</td>
<td>2.80</td>
<td>4.25</td>
<td>3.47</td>
</tr>
<tr>
<td>Quadratic (PTV IF 1)</td>
<td>6.37</td>
<td>0.60</td>
<td>0.56</td>
<td>5.50</td>
</tr>
<tr>
<td>Quadratic (PTV IF 4)</td>
<td>5.74</td>
<td>0.01</td>
<td>0.00</td>
<td>8.72</td>
</tr>
<tr>
<td>Quadratic (PTV IF 16)</td>
<td>5.34</td>
<td>3.00</td>
<td>6.05</td>
<td>3.48</td>
</tr>
</tbody>
</table>
180°, giving the plateau regions to the constrained optimization objective function in Fig. 5a. Between these plateau regions, Fig. 5a approximately follows the form of the PTV objective function, with the minimum at 110°. There is actually a small difference between the PTV RMS dose in Fig. 5b and the troughs of the constrained objective function in Fig. 5a, because each curve represents the minimum value obtained with a series of wedge angle combinations. The constrained algorithm may reject certain combinations due to the constraints not being met, in contrast to the dose statistics for the individual structures, where the minimum is always taken from the entire set of wedge angles.]

The corresponding quadratic case is shown in Fig. 5(f) for PTV and OAR importance factors all equal to unity. The overall objective function shows a peak around 90° due to the passage of the beams through OAR1 and OAR2, with lesser peaks at 0° and 180° due to the passage of beams through OAR3. Increasing the PTV importance factor to 4 or 16 [Figs. 5(g) and 5(h), respectively] changes the curve to more closely represent the PTV dose homogeneity. The resulting minima are at 140°, 160° then 100°. Unfortunately, none of these solutions satisfy the OAR constraints (Table VII). To produce a valid solution, the importance factors for PTV, OAR1, and OAR2 would need to be set equal, but the importance factor for OAR3 would need to be raised. Thus, as with the case of beam weights discussed earlier, it is dif-
difficult to choose the importance factors, and even when they have been chosen, the result may not be acceptable in terms of dose to critical structures and/or target homogeneity. Furthermore, the form of the quadratic objective function does not necessarily bear any resemblance to the mathematical form of the true clinical decision-making process.

B. Clinical cases

Figure 6 summarizes the results of the prostate optimization. For the three-field plans, the effect of optimizing the weights and wedge angles is to increase the weight of the anterior field and to introduce wedges on the posterior fields, as might be expected. This reduces the rectal $V_{60}$ from 65.8% to 60.7%. Optimizing beam orientations, beam weights and wedge angles simultaneously causes one of the fields to become more lateral and for the wedging to be further adjusted. This makes a substantial reduction in $V_{60}$ from 60.7% to 36.3%.

For the four-field plans, optimizing beam weights and wedge angles causes the weight of the posterior beam to be reduced, and for wedges to be introduced, allowing a small reduction in rectal $V_{60}$ from 45.9% to 43.5%. Optimization of the beam orientations as well as the beam weights and wedge angles produces an asymmetric plan, with an accompanying reduction in $V_{60}$ from 43.5% to 35.9%. In the case of the six-field plans, similar effects are observed, with the rectal $V_{60}$ reducing from 49.7% to 45.2% with optimization of beam weights and wedge angles, and from 45.2% to 32.1% with beam orientation optimization. The dose-volume histogram for the six-field plan with optimized beam orientations, beam weights and wedge angles is shown in Fig. 7. Note that all of the clinical constraints have been met in both the standard and optimized plans, without exceeding the limits.

The results for the craniopharyngioma are shown in Fig. 8. The standard plan with equally weighted fields, and wedges chosen to provide satisfactory PTV dose homogeneity, satisfies the clinical constraints (Table IV) and produces a brain $V_{25}$ of 6.4%. This is reduced to 5.5% by optimization of beam weights and wedge angles. The optimized beam weights are slightly more asymmetric than the standard plan, but not substantially different. When beam orientation optimization is used simultaneously with optimization of beam weights and wedge angles, the superior anterior oblique becomes more anterior, the superior posterior oblique becomes more superior, and the gantry angles of the lateral fields are increased, thereby effecting a mutual rotation of the lateral fields. The $V_{25}$ for the normal brain is in this case reduced to 4.7%. The reduction in $V_{25}$ of 1.7% corresponds to 25 cm$^3$ of healthy neural tissue out of the normal brain volume of 1478 cm$^3$. However, the overlapping dose-volume histograms for the standard and orientation-optimized plans (Fig. 9) show that this is not very significant. As with the prostate case, it can be seen that the constraints are all met.
IV. DISCUSSION

In general, an optimization problem consists of a function to be minimized, subject to a set of constraints. However, it is the way in which this general framework is applied to radiotherapy optimization which accounts for the variety of approaches described in the literature. In particular, radiotherapy treatment involves a number of diverse, and often competing, dosimetric goals. Thus, the function to be minimized contains a number of terms for various anatomical structures. It is this which gives rise to the presence of importance factors in many algorithms. Moreover, the constraints may be handled in a variety of different ways. On the one hand, the function to be optimized may be re-expressed in terms of constraints, so that the entire problem is formulated using constraints. This problem can then be solved by the Cimmino algorithm, which has been used on various occasions. However, when applied to radiation therapy, this method requires importance factors to guide the algorithm towards the solution. Moreover, the method only provides a feasible solution, and not necessarily the optimum solution, unless the constraints are chosen very carefully. Conversely, if the constraints are too restrictive, it may fail to provide a satisfactory solution at all. One advantage of the method is that in this situation, the nearest that the algorithm can provide to a feasible solution can be presented to the user, for them to modify the constraints intelligently.

The method of projections onto convex sets behaves similarly. The optimization problem is presented in the form of a series of convex constraints, and the algorithm searches for a feasible solution. However, as with the Cimmino algorithm, if the constraints are all met, the solution is not necessarily optimal, and if they are not met, the result is only a compromise, which may not be to the user’s satisfaction, and may include negative beam weights.

On the other hand, the most common approach is to neglect the constraints completely and use simple terms in the function instead. The traditional quadratic objective function is an example of this. This approach requires importance factors, and although it always produces a solution, this is not always in accord with the clinical requirements.

Between the two extremes are the linear methods, which use a linear objective function and retain the constraints explicitly. These techniques behave very similarly with respect to the constraints as the method described in this paper. However, the objective function retains the use of importance factors, which are a disadvantage. The objective function and the constraints are required to be linear, so that a significant improvement in solution time can be achieved. However, the objective function in radiation therapy is inherently non-linear. This difficulty can be overcome by assuming that the dose falls off away from the planning target volume, but biological considerations have not yet been satisfactorily incorporated into this framework.

The intermediate category also includes penalty methods, which remove the explicit constraints, but include them implicitly as penalty factors in the function to be minimized. Typically, this type of method is used to incorporate dose-volume constraints. With this method, it is necessary to ensure that the importance factors of the penalty terms are set high enough, or the constraints are violated. More properly, as pointed out by Hristov and Fallone, the mathematical framework for this method requires that the problem is repeated with successively increasing values of the importance factors for the penalty terms. However, this is rarely carried out in practice. Apart from this, the method reduces to that of a quadratic objective function, with the attendant problems of importance factors.

The new algorithm presented in this paper is also in the intermediate category. In some respects, the algorithm can be considered a special case of a dose-threshold algorithm, with the penalty factors set to infinity, so that any failure to satisfy a constraint is deemed unacceptable. However, it differs from previous algorithms used in radiation therapy in that a single dose statistic is used for the optimization function. This obviates the need for importance factors. Constraints are used for the remaining dose statistics. By retaining one dose statistic as an objective in the optimization, the problem remains a true optimization, rather than becoming a feasibility search. Moreover, in most practical cases, the single statistic being optimized has the effect of pulling the constrained statistics to the constraint boundaries. This approach accurately models the rationale behind both clinical practice and the derivation of ‘class solutions’. The resulting objective function space, consisting of smoothly varying regions interspersed by areas where the solution is unacceptable, requires the use of a stochastic algorithm to find the global minimum. This is much slower than a gradient algorithm, but results in solutions which satisfy, without intervention, the clinical goal.

The difference between the present approach and the more deterministic methods has been illustrated by means of a comparison with the minimization of a quadratic objective function. It has been found that the quadratic method with equal importance factors returns a very different solution to the constrained optimization algorithm. Only after variation
of the importance factors does the quadratic method yield a similar solution. Even then, the form of the objective function does not really correspond to the clinical specification.

The results demonstrate that the algorithm behaves rationally for both beam weight and beam orientation optimization. For the latter, allowing the user to specify a range of search angles has been found to be a practical means of allowing the solution to be guided in the absence of an elaborate model of the linear accelerator. However, further work to enhance the algorithm by incorporating collision avoidance is expected. The simulations confirm the presence of local minima in the objective function for beam orientation optimization. The stochastic search over the range of angles and weights is time consuming, but satisfactory solutions are generally found in less than 24 hours. This is assisted by the use of a fast convolution dose calculation for computation of multiple plans in a short time interval, with sufficient accuracy. The large time interval required for optimization is not found to be a problem, as the program is left running overnight. Because the clinical objective and constraints are specified precisely, and importance factors are not required to be adjusted, the program can be left to run unattended. By scheduling the optimization to take place ahead of the time allocated to the dosimetrist for work on the treatment plan, it has been found possible at this center to use the program regularly for selected cases.

The clinical prostate case shows results which are consistent with previous studies. Successive optimization of first beam weights and wedge angles only, and then orientations as well, produces a steady reduction in irradiated rectal volume, all of the other parameters being constrained to be equal. The three-field orientation-optimized plan is similar to that evaluated previously. Moreover, it can be considered a variation of a plan consisting of an anterior field and two lateral fields, which has been proposed as a class solution for prostate radiotherapy. In this case, however, a more individualized solution is possible than when using a repeated manual method in a cohort of patients. This accounts for the precise optimization of wedge angle and colimator angle seen in Fig. 6.

The four-field orientation-optimized plan also has features in common with a plan consisting of two anterior oblique fields and two lateral fields, which was found to be optimal in a previous study of four-field plans, and which has also been proposed as a class solution. For the six-field plan, a previous study found that a laterally symmetric plan with anterior oblique fields angled at 65° to the mid-coronal plane and posterior oblique fields at 30° to the mid-coronal plane was generally optimal for a PTV including the seminal vesicles. The six-field orientation-optimized plan in the present study is similar to this, with the exception of the left posterior oblique field, which is somewhat more posteriorly positioned. It should also be noted that the beam weight of the right anterior oblique field is relatively low. In summary, all of the fully orientation-optimized prostate plans show similar irradiated rectal volume, which is in accord with previous observations that the number of fields is relatively unimportant in prostate planning.

In order to produce satisfactory solutions, it has been found necessary to set the constraint on minimum PTV dose such that 90% of the PTV must receive 95% dose, instead of the usual requirement that all of the PTV must receive this dose. This is because in practice, a dosimetrist would increase the length of the fields to ensure satisfactory superior and inferior coverage of the PTV in these coplanar techniques. However, the automated algorithm does not currently perform this, but instead simply adds a uniform margin around the beam’s eye view of the PTV. This is due to the difficulty of knowing exactly where the superior and inferior edges of the PTV are located in a potentially complex-shaped volume. This points to the need to introduce beam aperture optimization53 to shape the aperture as required for complete PTV coverage, but this has not yet been implemented.

The technique of limiting the dose to the normal tissue has also been found to be of widespread utility. This prevents the algorithm from depositing dose in areas which are otherwise unconstrained. This is particularly important for IMRT inverse planning, where several beam elements can sum to produce a small but significant hotspot. However, it is important not to constrain the dose immediately around the PTV, because then the PTV dose must be reduced in order to satisfy the normal tissue constraint. In the framework of the constrained algorithm, this generally results in a solution not being found at all. The method of considering the volume located more than 30 mm away from the PTV as normal tissue has been found to work well.

The craniopharyngeal case is more challenging than that of the prostate in view of the noncoplanar nature of the fields, and the need to avoid the stereotactic frame as well as the organs at risk. This case principally demonstrates that, by a combination of informed specification of the possible search angles and by provision of appropriate clinical constraints (including the use of an artificial clinical constraint designed to prevent beam exit through the patient’s thorax), the optimization algorithm is able to find a rational solution. This solution delivers a small but worthwhile reduction in the volume of normal brain irradiated. The limited improvement over the standard plan is due to the standard plan being very acceptable to begin with, and also because dose has to be delivered to the PTV through one part of the brain or another, so that a reduction to one part of the DVH must be accompanied by an increase elsewhere. Optimizing a dose-volume objective such as V25 may also improve the volume irradiated to a specific dose (25 Gy in this case), without due regard for other doses. This may be an indication for the use of biological indices, which more closely represent the whole of a DVH curve.

As with many optimization problems, the selection of beam orientations, beam weights and wedge angles is actually a search for a local minimum amongst the global radiation therapy optimization problem. Various parameters, such as the beam energy and radiation type are fixed, and the remaining parameters are optimized. However, if the fixed parameters are also varied, a more global solution may result, which improves the previous “optimal” plan. Thus, it
should be remembered that optimization is invariably a restricted search over part of the global parameter space and that the solutions may differ as the remainder of the parameter space is in time explored.

The algorithm as presented provides a versatile mechanism for optimization in conformal radiation therapy, with possible applications in the planning of individual patient treatments, and the determination of class solutions in patient cohorts. Several of such class solution studies are under way at this center.53 The proposed future development of this algorithm is expected to be in the introduction of multiple segments for each field, together with aperture shape optimization, so that the algorithm can be used for beam orientation optimization of IMRT.11,12 This may be achieved using a similar method to that described by De Gersem et al.57 or Shepard et al.,58 and is expected to be major challenge in terms of obtaining a sufficiently accurate solution in a realistic computation time. However, it is thought that optimization of beam angles for IMRT using such an approach may offer simple, more clinically applicable, solutions.

V. CONCLUSION

By optimizing one variable of a treatment plan, subject to a set of constraints, it is possible to eliminate the need for importance factors, thereby facilitating accurate production of inverse plans. This strategy has been shown to be widely applicable to a number of treatment planning problems. Simulation studies show that the resulting objective function is very different from a quadratic objective function, but models the clinical decision process more closely. By the use of a random search technique, this new constrained method is applied to the optimization of beam orientations, beam weights and wedge angles for conformal radiation therapy. It is shown that the optimization provides solutions which are consistent with previous studies, and that the beam orientations produced are clinically feasible.

ACKNOWLEDGMENTS

The author gratefully acknowledges the financial support of Cancer Research UK (grant reference SP2312/0201) and Philips Radiation Oncology Systems. The author would also like to thank Jim Warrington for his supervision of this work, and Dr. Vincent Khoo (Christie Hospital and Holt Radium Institute, Manchester) for his collaboration, including the provision of the prostate case. The contributions of Dr. Vibeke Nordmark Hansen, Dr. Amin Mosleh-Shirazi, and Dr. Judy Christian in discussing the paper, are also acknowledged.

6 S. Webb, Intensity-Modulated Radiation Therapy (Institute of Physics, Bristol, 2001).


