Radiosurgery Treatment Planning via Nonlinear Programming *

Michael C. Ferris† Jinho Lim ‡ David M. Shepard §

January 2001

Abstract

The Gamma Knife is a highly specialized treatment unit that provides an advanced stereotactic approach to the treatment of tumors, vascular malformations, and pain disorders within the head. Inside a shielded treatment unit, multiple beams of radiation are focussed into an approximately spherical volume, generating a high dose shot of radiation. The treatment planning process determines where to center the shots, how long to expose them for, and what size focussing helmets should be used, in order to cover the target with sufficient dosage without overdosing normal tissue or surrounding sensitive structures. We outline a new approach that models the dose distribution nonlinearly, and uses a smoothing approach to treat discrete problem choices. The resulting nonlinear program is not convex and several heuristic approaches are used to improve solution time and quality. The overall approach is fast and reliable; we give several results obtained from use in a clinical setting.

*This material is based on research supported in part by National Science Foundation Grant CCR-9972372, the Air Force Office of Scientific Research Grant F49620-01-1-0040 and Microsoft Corporation.

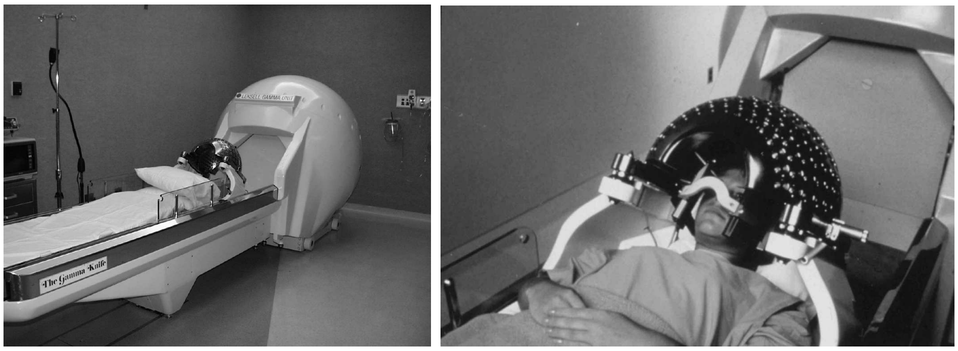
†Computer Sciences Department, University of Wisconsin, 1210 West Dayton Street, Madison, Wisconsin 53706. email: ferris@cs.wisc.edu

‡Department of Industrial Engineering, University of Wisconsin, Madison, Wisconsin 53706. email: jin-ho@cs.wisc.edu

§Department of Radiation Oncology, University of Maryland School of Medicine, 22 South Green Street, Baltimore, MD 21201 email: dshep001@umaryland.edu

1 Introduction

The Gamma Knife (see Figure 1(a)) is a highly specialized treatment unit that provides an advanced stereotactic approach to the treatment of tumor and vascular malformations within the head [6]. The Gamma Knife delivers a single, high dose of radiation emanating from 201 Cobalt-60 unit sources. All 201 beams simultaneously intersect at the same location in space to form an approximately spherical region that is typically termed a shot of radiation. A typical treatment consists of a number of shots, of possibly different sizes and different durations, centered at different locations in the tumor, whose cumulative effect is to deliver a certain dose to the treatment volume while minimizing the effect on surrounding tissue.



(a) The patient lies on the couch and is moved back into the shielded treatment area

(b) A focussing helmet is attached to the frame on the patients head

Figure 1: The Gamma Knife Treatment Unit

Gamma Knife radiosurgery begins (after administering local anesthesia) by fixing a stereotactic coordinate head frame to the patient's head using adjustable posts and fixation screws. This frame establishes a coordinate frame within which the target location is known precisely and also serves to immobilize the patients head within an attached focussing helmet during the treatment (see Figure 1(b)). An MRI or CT scan is used to determine the position of the treatment volume in relation to the coordinates determined by the head frame. Once the location and the volume of the tumor are identified, the neurosurgeon, the radiation oncologist, and the physicist work together in order to develop the patient's treatment plan. Multiple shots are often used in a treatment using a Gamma Knife due to the irregularity and size of tumor shapes and the fact that the focussing helmets are only available in four sizes (4, 8, 14 and 18mm).

The determination of plans varies substantially in difficulty. For example, some tumors are small enough to apply one shot of radiation. On the other hand, when the shape of the tumor is large or has an irregular shape or is close to a sensitive structure, many shots of different sizes could be needed to achieve appropriate coverage of the tumor while sparing the surrounding tissue. The treatment planning process can be very tedious and time consuming and due to the variety of conflicting objectives, the quality of treatment plan produced depends heavily on the experience of the user. Therefore, a unified and automated Gamma Knife treatment process is desired. Further description of the treatment process, along with some more explanatory figures can be found in [5].

The plan aims to deliver a high dose of radiation to the intracranial target volume with minimum damage to the surrounding normal tissue. The treatment goals can vary from one neurosurgeon to the next, so a planning tool must be able to accommodate several different requirements. Among these requirements, the following are typical, although the level of treatment and importance of each may vary.

1. A complete 50% isodose line coverage of the target volume. This means that the complete target must be covered by a dose that has intensity at least 50% of the maximum delivered dosage. This can be thought of as a "homogeneity" requirement.

2. To minimize the non-target volume that is covered by a shot or the series of delivered shots. This requirement is clear and can be thought of as a "conformity" requirement.

3. To limit the amount of dosage that is delivered to certain sensitive structures close to the target. Such requirements can be thought of as "avoidance" requirements.

There are standard rules established by various professional and advisory groups that specify acceptable homogeneity and conformity requirements. In addition to these requirements, it is also preferable to use a small number of shots to limit the treatment times and thus increase the number of patients that can be treated.

The approach for treatment planning that will be used here is based on an optimization model of the physical system. Three characteristics are important in the optimization technique for Gamma Knife treatment planning:

speed, flexibility, and *robustness*. A fast treatment plan is desired primarily for patient comfort. The system must be flexible because the treatment goals vary from patient to patient and neurosurgeon to neurosurgeon. The system also must be robust so that it produces a high quality solution regardless of the size and the shape of the target volume. The solution produced by the optimization must also be practical and implementable.

We assume throughout this work that the number of shots that will be delivered is specified to the optimization tool. While other approaches may try to minimize this number, it is typically straightforward to estimate this number and then develop a plan to optimize other important features for the treatment. In the model we propose, there are three types of decision variables:

1. *A set of coordinates* (x_s, y_s, z_s): for each shot the position of the shot centers is a continuous variable to be chosen.

2. *A discrete set of collimator sizes*: currently four different sizes of focussing helmets are available (4mm, 8mm, 14mm, 18mm).

3. *Radiation exposure time*: the dose delivered is a linear function of the exposure time.

A number of researchers have studied techniques for automating the Gamma Knife treatment planning process [15, 9]. One approach incorporates the assumption that each shot of radiation can be modeled as a sphere. The problem is then reduced to one of geometric coverage, and a ball packing approach [13, 12, 15] can be used to determine the shot locations and sizes. The use of a modified Powell's method in conjunction with simulated annealing has also been proposed [9, 16]. A mixed integer programming and a nonlinear programming approach for the problem is presented in [5, 11]. A mixed integer programming approach for linear accelerator(LINAC) radiosurgery treatment is presented in [8], and a nonlinear approach in this case is given in [10].

This paper is based on the approach outlined in [5], whereby the actual dose distribution is modeled and a formal constrained optimization model is solved to determine the treatment plan. The remainder of this paper is organized as follows. Section 2 reviews the nonlinear programming formulation of the problem used in [5, 11] and details some of the changes that were necessary for implementation of the optimization scheme in the clinic. In particular, a new dose model is described that allows the shots to be modeled as ellipsoids, along with a new conformity estimation problem and a continuation approach to solve the nonlinear program. Section 3 presents the application of the new techniques to several three dimensional real patient cases. Finally, Section 4 concludes the paper with some remarks concerning future directions.

2 Nonlinear Programming Model

2.1 Original formulation

A nonlinear programming approach for solving the treatment planning problem is described in [5]. The input to the nonlinear program consists of several pieces of information, namely the number of shots that are to be used and the widths of shots that are considered appropriate for the target volume, the required isodose level and the target volume itself. The initial locations of the shots are placed randomly within the target, and the initial levels for the exposure time are fixed appropriately.

Given the shot locations and exposure time, the dose distribution for each shot at a given voxel (volume element) on a three dimensional grid is calculated based on a spherical algebraic model. A standard least squares model is used to determine the weights on particular basis functions that are used in the dose model, based on suggestions from the literature. It is assumed that the dose model does not change due to movement of the shot center.

Once a description of the dose is determined, the optimization model can be formulated. The basic variables of the optimization we consider include the coordinates of the center location of the shot (x_s, y_s, z_s), the width of the shot w, and the time $t_{s,w}$ that each shot is exposed. In practice, we consider a grid $\mathcal{G}$ of voxels. There are two types of voxels: $\mathcal{T}$ represents the subset of voxels that are within the target and $\mathcal{N}$ represents the subset of voxels that are out of the target. Since the number of voxels out of the target is vast, we typically use just a small subset of them, generated close to the target volume or in a sensitive structure.

Isodose line coverage. Neurosurgeons commonly use isodose curves as a means of judging the homogeneity of a treatment plan. The 50% isodose curve is a curve that encompasses all of the voxels that receive at least 50% of that maximum dose that is delivered to any voxel in the patient. A treatment plan is normally considered acceptable if a certain percentage isodose curve (typically 50%) encompasses the tumor. We model such a constraint by imposing strict lower and upper bounds on the dose allowed in the target, namely for all $(i,j,k) \in \mathcal{T}$

$$\theta \le Dose(i,j,k) \le 1 \tag{1}$$

In this way, the $100\theta\%$ isodose curve is guaranteed to cover the target. Other isodose curves can be generated by simply modifying the numerical value θ.

Choosing shot widths. The number of shots to be used is given to the optimization model, and the location of the shot center is chosen by a continuous optimization process. Choosing the particular shot width at each shot location is a discrete optimization problem that is treated by approximating the step function

$$H(t) = \begin{cases} 1 & \text{if } t > 0 \\ 0 & \text{if } t = 0 \end{cases}$$

by a nonlinear function,

$$H(t) \approx H_\alpha(t) := \tfrac{2\arctan(\alpha t)}{\pi}$$

For increasing values of α, H_α becomes a closer approximation to the step function H. This process is typically called smoothing.

The set of shot widths for a given number of shots n is chosen by imposing the constraint:

$$n = \sum_{(s,w)\in\{1,\ldots,n\}\times\mathcal{W}} H_\alpha(t_{s,w}). \tag{2}$$

This states that the total number of shot/width combinations that are to be used is n. In practice, we solve a sequence of models, each time increasing the value of α to improve the approximation. Note that the optimization may place two shots of different widths at the same location, and hence none at another location. Typically, we relax the requirement for exactly n shot/widths, and instead impose a range constraint forcing lower and upper bounds on the number of shot/width combinations.

We have tested several optimization formulations. The most obvious model is to minimize the dose outside of the target subject to a constraint on the minimum isodose line that must surround the target:

$$
\begin{array}{lll}
\min & \displaystyle\sum_{(i,j,k)\in\mathcal{N}} Dose(i,j,k) & \\
\text{subject to} & Dose(i,j,k) = \displaystyle\sum_{(s,w)\in\mathcal{S}\times\mathcal{W}} t_{s,w} D_w(x_s, y_s, z_s, i, j, k) & \\
& \theta \le Dose(i,j,k) \le 1, \quad \forall (i,j,k) \in \mathcal{T} & \\
& n = \displaystyle\sum_{(s,w)\in\{1,\ldots,n\}\times\mathcal{W}} H_\alpha(t_{s,w}) & \\
& t_{s,w} \ge 0. &
\end{array}
\tag{3}
$$

The most critical problem is that due to the large number of voxels that are needed when dealing with large irregular tumors (both within and outside of the target) the computational time to complete this treatment plan is too long. To make the solution process faster, we can remove a large number of the non-target voxels from the model. While this improves computational time, this typically weakens the conformity of the dose to the target. How can we maintain conformity without vastly increasing computational time? The second formulation uses a constraint to control the conformity of the plan. The constraint specifies that at least $P\%$ of the total dose must be deposited in the target.

It is known how to simulate the delivery of a shot of width $w \in \mathcal{W}$ centered at the middle of the head of a previously scanned patient on the Gamma Knife. For each shot width we use this to estimate the total dose delivered (at unit intensity) to the complete volume and term this constant $\bar{D}_w$. This is then used to determine an estimate of the total dose delivered to the complete volume by the collection of shots as

$$
\sum_{(s,w)\in\mathcal{S}\times\mathcal{W}} \bar{D}_w t_{s,w}, \tag{4}
$$

without having to calculate the dose at any voxel external to the target. This expression can be used as the denominator of an of the conformity of a given plan without evaluating dose at voxels outside of the target. The numerator would obviously just be the total dose delivered to the target.

We update the basic model to force more conformity at the expense of relaxing homogeneity. Instead of enforcing the strict lower bound of θ on the dose in the target, we instead calculate the amount of dose under this value at every voxel in the target, and sum the “underdose” to form our objective. More formally, a voxel is considered to be underdosed if it receives less than the prescribed isodose, which for the example formulation is assumed to be θ. We actually use the optimization process to model UnderDose. UnderDose is constrained to be greater than or equal to $\max(0, \theta - Dose)$ at every voxel in the target. Since we minimize UnderDose, it will take on the maximum of these two values at optimality. An upper bound is still placed on the dose in the target, and the lower bound on dose is relaxed. The complete formulation is:

$$
\begin{array}{ll}
\min & \displaystyle\sum_{(i,j,k)\in\mathcal{T}} UnderDose(i,j,k) \\
\text{subject to} & Dose(i,j,k) = \displaystyle\sum_{(s,w)\in\mathcal{S}\times\mathcal{W}} t_{s,w} D_w(x_s, y_s, z_s, i, j, k) \\
& 0 \le Dose(i,j,k) \le 1 \\
& 0 \le UnderDose(i,j,k) \\
& \theta - UnderDose(i,j,k) \le Dose(i,j,k), \quad \forall (i,j,k) \in \mathcal{T} \\
& P \le \dfrac{\displaystyle\sum_{(i,j,k)\in\mathcal{T}} Dose(i,j,k)}{\displaystyle\sum_{(s,w)\in\mathcal{S}\times\mathcal{W}} \bar{D}_w t_{s,w}} \\
& n = \displaystyle\sum_{(s,w)\in\{1,\ldots,n\}\times\mathcal{W}} H_\alpha(t_{s,w}) \\
& t_{s,w} \ge 0.
\end{array}
\tag{5}
$$

In this formulation, P is the fraction of the total dose that must be deposited in the target. For solution purposes, we rearrange the equation involving P so that it is a purely linear equation (i.e. multiply both sides by the term in the denominator). While this approach is much more feasible, it is unclear in many patient cases what is an appropriate value to choose for P.

While the results generated using the above approach are typically very good, there are a number of difficulties that are manifested when implementing this approach in the clinic.

1. In reality, the dose delivered by the Gamma Knife is not symmetrical and is quite different when the patient is prone or supine.

2. How do we estimate the conformity value P automatically?

3. The treatment plan can only be input on the Gamma Knife at a particular specificity. How can we translate the optimization output onto the Gamma Knife?

4. Given the nonconvexity of our approach, how do we guarantee that the solutions we generate are both reasonable and close to the optimal values possible? Furthermore, how do we carry out the complete approach quickly?

In the following sections we will treat each of these issues in turn.

2.2 A new dose distribution model

The complete dose distribution can be calculated as a sum of contributions from each shot delivered, once the location of the center of that shot (x_s, y_s, z_s) is known, and the length of time of delivery $t_{s,w}$ is known. In practice this means that for all (i, j, k)

$$Dose(i, j, k) = \sum_{(s,w) \in \mathcal{S} \times \mathcal{W}} t_{s,w} D_w(x_s, y_s, z_s, i, j, k), \tag{6}$$

where $D_w(x_s, y_s, z_s, i, j, k)$ is the dose delivered to the voxel (i, j, k) by the shot of width w centered at (x_s, y_s, z_s).

In previous work [5], we simulated the delivery of a shot of width $w \in \mathcal{W}$, centered at the middle of the head of a previously scanned patient on the Gamma Knife. For each shot width, we determined the dose delivered in the x, y and z directions at given distances from the center of the shot from the simulation. The three values were then averaged to give a value of dose (for each width of shot) at a particular distance from the center. However, we observed that the actual dose delivered was ellipsoidal in nature rather than spherical, so we determined the principal axes and measured the values of dose $\bar{D}_w$ along them. In practice, the axis location depended on whether the patient was lying prone or supine, and thus we rotate the target so its coordinate axes lie along the ellipsoid's principal axes in either case.

The problem is thus reduced to determining a functional form for the dose delivered at a voxel (i, j, k) from the shot centered at (x_s, y_s, z_s). A sum of error functions has been noted in the literature to approximate this dose distribution [2, 7, 14]. We therefore used the following functional form

$$D_w(x_s, y_s, z_s, i, j, k) = \sum_{p=1}^{2} \lambda_p \left(1 - \operatorname{erf}\left(\frac{\sqrt{(i - x_s)^2 + \mu_p^y (j - y_s)^2 + \mu_p^z (k - z_s)^2} - r_p}{\sigma_p}\right)\right) \tag{7}$$

and fit the ten parameters λ_p, μ_p^y, μ_p^z, r_p and σ_p to the data described above via least-squares, with different values for each shot width. The notation erf(x) represents the integral of the standard normal distribution from $-\infty$ to x. The resulting nonlinear optimization problem

$$\min_{\lambda,\mu,r,\sigma} \left\| \begin{pmatrix} \bar{D}^x_w(i) - \sum_{p=1}^{2} \lambda_p \left(1 - \mathrm{erf}\left(\frac{\sqrt{(i-x_s)^2} - r_p}{\sigma_p}\right)\right) \\ \bar{D}^y_w(j) - \sum_{p=1}^{2} \lambda_p \left(1 - \mathrm{erf}\left(\frac{\sqrt{\mu^y_p(j-y_s)^2} - r_p}{\sigma_p}\right)\right) \\ \bar{D}^z_w(k) - \sum_{p=1}^{2} \lambda_p \left(1 - \mathrm{erf}\left(\frac{\sqrt{\mu^z_p(k-z_s)^2} - r_p}{\sigma_p}\right)\right) \end{pmatrix} \right\|^2$$

was solved using CONOPT. These values were then fixed in the nonlinear models used throughout the remainder of this paper.

2.3 Conformity estimation

How do we estimate a good value for the conformity P required in a given plan. Note that if the target is close in size to one of our given shots and ellipsoidal in nature, we can expect good conformity for our plan. For odd shaped targets with a limited number of avaliable shots, the conformity we can expect is likely to be much lower. In fact, for larger treatment volumes it is very hard to estimate a reasonable value for P, so we use optimization to accomplish this.

We propose to solve an additional optimization problem to obtain a good conformity estimate for each particular patient. As outlined in (4), we can use the simulated data $\bar{D}_w$ to derive an accurate estimate for the total dose delivered to the complete volume. We choose to minimize this quantity, subject to the standard constraints of maintaining an appropriate isodose line around the target, and a limit on the number of shots of different widths and locations.

$$\begin{aligned} \min \quad & \sum_{(s,w)\in\mathcal{S}\times\mathcal{W}} \bar{D}_w t_{s,w} \\ \text{subject to} \quad & \theta \le Dose(i,j,k) \le 1, \quad \forall (i,j,k) \in \mathcal{T} \\ & n = \sum_{(s,w)\in\mathcal{S}\times\mathcal{W}} H_\alpha(t_{s,w}) \\ & t_{s,w} \ge 0 \end{aligned} \tag{8}$$

Note that this model uses the data $\bar{D}_w$ instead of calculating the dose outside the target and thus is a much smaller optimization model even if the number of voxels in the complete volume is large. We have to be careful to ensure that the value calculated for P is fairly insensitive to changes in the starting point given to the model; this is shown in Section 3. Some care is taken to choose the value of α appropriately. For large treatment volumes we typically only evaluate the bound constraints in (8) on a small but representative subset of the voxels in the target.

Given this estimate of achievable conformity under the strict homogeneity constraint, we allow the modeler to specify a scaling parameter to increase the conformity P required in the final solution.

2.4 A continuation solution process

After generating a value for P, we then solve (5) four times to produce the treatment plan. We first solve (5) with a coarse grid $\mathcal{G}$ and $\alpha = 7$. That is, instead of evaluating the constraints at all voxels in $\mathcal{T}$, we use a larger grid spacing and enforce the constraints only at $\mathcal{G} \bigcap \mathcal{T}$. We then calculate which constraints are violated on a finer grid and add these grid points into our optimization model. The solution we obtain from the finer grid usually uses more shots than are available. Therefore, we impose a higher value of $\alpha = 100$ on the refined grid problem and resolve to further limit the shot/width combinations used.

The computed solution may not be implementable on the Gamma Knife since the coordinate locations cannot be keyed into the machine. One approach to refine the optimization solution to generate implementable coordinates for the shot locations is to round the solution. Typically, simple rounding degrades the solution and can severely detract from the perceived quality of our results. To overcome this, we round and fix the shot center coordinate values from our process, and then reoptimize the intensity values t to generate a feasible solution. This is always better than simple rounding and recovers a similar quality but precisely implementable solution.

The optimization models are written in the GAMS [1] modeling language and solved using CONOPT [3]. Note that our solution technique does not guarantee that the shots are centered at locations within the target. In the previous work we generated random starting values for (x_s, y_s, z_s) within the target to encourage the final shot locations to also lie within the target.

3 Computational Results on Patient Data

The tool that implements the process described in this paper is in use at the University of Maryland Medical School. We have tested our techniques on

Table 1: Average run time for different sizes of tumors

Average Run Time	Size of tumor		
	Small	Medium	Large
Random (std.dev)	2 min 33 s (40 s)	17 min 20 s (3 min 48 s)	373 min 2 s (90 min 8 s)
SLSD (std.dev)	1 min 2 s (17 s)	15 min 57 s (3 min 12 s)	23 min 54 s (4 min 54 s)

three targets arising from real patient cases. The three targets are radically different in size and complexity. The first patient has a small tumor, the second is medium sized, whereas the third is considered very large for this type of treatment. The small tumor contains 4006 voxels, 10061 voxels for the medium size, and 36088 voxels for the large size tumor. Since our problems are not convex, the choice of parameters in their solution can also have dramatic effects. In this section, we demonstrate how to choose good parameters for the NLP models. Some further description of the medical implications of these results are given in [11].

We generate good initial shot center locations and widths by running a shot location and size determination (SLSD) technique that is described elsewhere [4].

In Figure 2, we show how the conformity parameter P affects the final solution in the small patient case. As we increase P, the solution becomes more conformal, but at a cost in homogeneity, that we measure via the objective function in (8). The conformity estimation problem generates an average value for P of 0.248, with a standard deviation of 0.012 when we run the process 50 times with slightly perturbed starting values. Recall that the planner can specify a scale parameter increase of this value to achieve higher conformity if desired.

We run the optimization multiple times on all three patient cases to test the robustness of our approach. Table 1 shows average run time of the entire model for three different sized tumors, based on runs where we perturbed the initial solution (30% of the voxel locations and 60% of the weights were perturbed by small amounts). We made 50 runs for the small tumor and 25 runs for the medium and large tumors. The solutions used

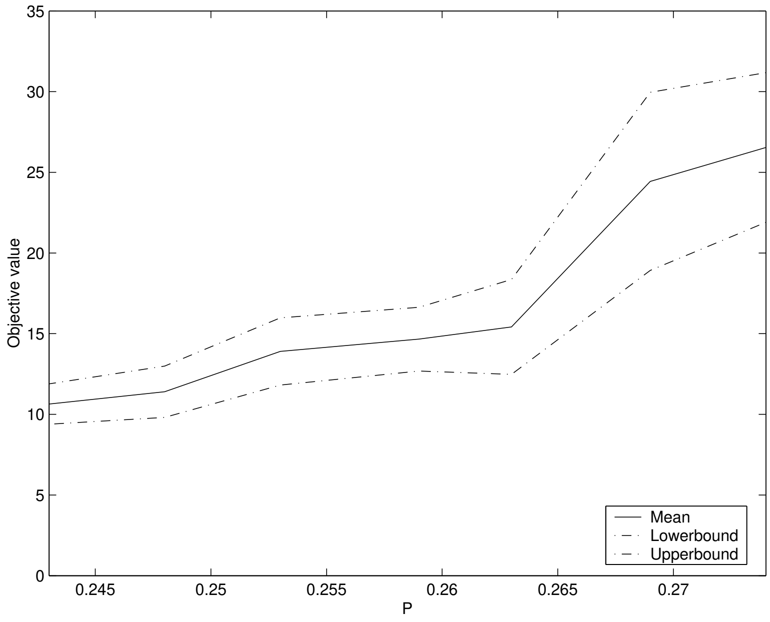


Figure 2: 90% confidence interval for the objective value of (8) as a function of conformity value P

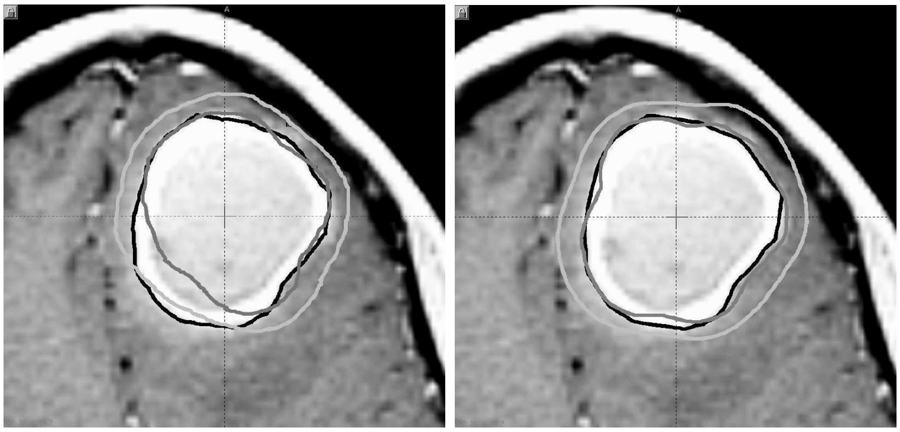
Table 2: A comparison of two different schemes

	18 mm only			14 and 18 mm		
	P	Obj.Val	Run Time	P	Obj.Val	Run Time
Mean	0.205	33.98	15 min 57 s	0.232	17.62	5 min 43 s
Std.dev.	0.002	3.42	3 min 12 s	0.005	6.55	1 min 15 s

five 8mm and two 14mm shots for the small tumor, seven 18mm shots for the medium, and twelve 18mm shots for the large tumor. As we can see, we solve the small size problem in about 1 minute, the medium size problem within 16 minutes with 3 minutes of standard deviation, and 24 minutes with a standard deviation of 5 minutes for the large tumor. Although the gain in speed using SLSD depends on the shape and size of the tumor, the table shows that the SLSD solutions outperform the random starting solutions regardless of the size of tumor. Note that random starting solutions often fail to solve the large problem in a clinically acceptable time limit (typically 20 - 40 minutes).

It is interesting to observe that in the medium size case that SLSD does not provide a distinct advantage over random starting points. In practice, the planners limit the number of different helmets that can be used in planning as a mechanism to improve solution times. In Table 1 we used these prescribed numbers in all cases. However, if more helmet sizes are allowed, sometimes better solutions can be found more quickly. Table 2 shows the performance when two different helmet sizes are allowed, 14mm and 18mm for the medium sized tumor. First of all, we get more conformal solution without losing the target coverage. In fact, we get better coverage, i.e. smaller objective value. We also get a substantial gain from its speed when two different helmet sizes are allowed. The run time was reduced from 16 minutes to 6 minutes on average. Clearly, the advantage of using SLSD is its ability to choose a set of reasonable shot center locations and their appropriate helmet sizes.

Finally, Figure 3 shows several pictures of the large tumor solutions that are used by the planners to understand the quality of the solutions. While these figures show the SLSD solution is much more conformal in this slice, and seems much better in quality, it is hard to make a definitive judgement from these figures. Radiation oncologists often use a cumulative dose volume



(a) Random starting point solution. Note that the target and the 50% isodose curve do not match closely.

(b) SLSD starting point solution. Note that the target and the 50% isodose curve match closely.

Figure 3: Large patient example. Three contours drawn represent target, 50% and 30% isodose curves respectively in decreasing greyscale

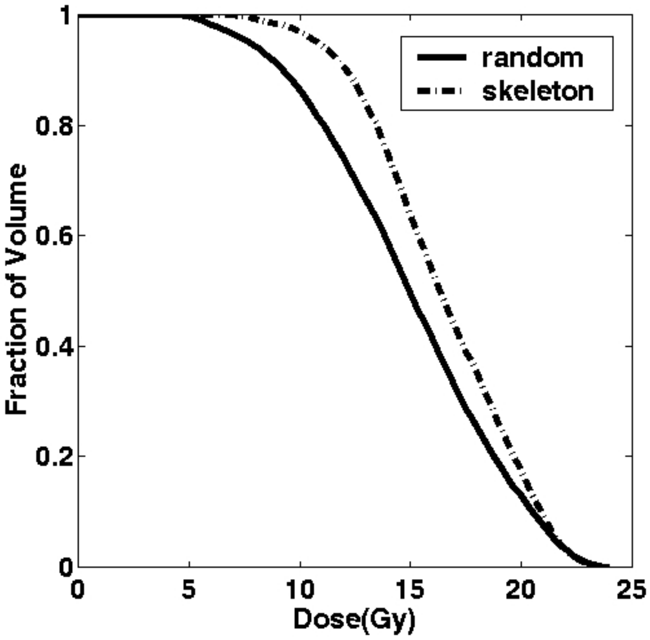


Figure 4: A dose volume histogram depicting the percentage of dose delivered to the target volume by the random and SLSD (skeleton) solutions.

histogram as a means of determining the quality of a treatment plan (see Figure 4). A cumulative dose volume histogram displays the fraction of the patient that receives at least a specified dose level. Since the SLSD solution lies entirely to the right of the random solution, the SLSD solution is uniformly better.

4 Conclusion and Future Directions

Automation of the planning process for Gamma Knife radiosurgery is highly desired since it has the potential to produce more uniform and better quality plans. The key to the success of an automated approach depends on how quickly and accurately the solution can be generated, and how reliable the method is in various problem settings. Our approach is fast to generate good quality solutions. It is flexible because it provides problem-dependent initial solutions for the NLP model which can then be solved using standard optimization tools. It takes no more than 7 seconds to generate the starting solution for any size (the biggest tumor size tried has length 7.2 cm) and regardless of shape of the tumor.

Note that a user of this procedure has only to provide the following information: the target volume, an estimate of the number of shots to use, the isodose prescription level, the conformity relaxation parameter, whether the patient is supine or prone and the helmet sizes that should be used (optional).

There are a variety of research questions whose solution would make the current tool even more useful. Many of these are related to the speed and robustness of the solution process. Is there a better continuation process for solving the nonlinear program? Are there other methods to generate a good starting solutions for NLP model? Can we improve the quality of the starting solution by taking into account the exposure times? The solution of these and related questions are the topic of future research.

References

[1] A. Brooke, D. Kendrick, and A. Meeraus. *GAMS: A User's Guide.* The Scientific Press, South San Francisco, CA, 1988.

[2] P. S. Cho, H. G. Kuterdem, and R. J. Marks. A spherical dose model for radiosurgery treatment planning. *Physics in Medicine and Biology*, 43:3145–3148, 1998.

[3] A. Drud. CONOPT: A GRG code for large sparse dynamic nonlinear optimization problems. *Mathematical Programming*, 31:153–191, 1985.

[4] M. C. Ferris, J.-H. Lim, and D. M. Shepard. Optimization approaches for treatment planning on a gamma knife. In preparation.

[5] M. C. Ferris and D. M. Shepard. Optimization of gamma knife radiosurgery. In D.-Z. Du, P. Pardolas, and J. Wang, editors, *Discrete Mathematical Problems with Medical Applications*, volume 55 of *DIMACS Series in Discrete Mathematics and Theoretical Computer Science*, pages 27–44. American Mathematical Society, 2000.

[6] J. C. Ganz. *Gamma Knife Surgery.* Springer-Verlag Wien, Austria, 1997.

[7] H. M. Kooy, L. A. Nedzi, J. S. Loeffler, E. Alexander, C. Cheng, E. Mannarino, E. Holupka, and R. Siddon. Treatment planning for streotactic

radiosurgery of intra-cranial lesions. *International Journal of Radiation Oncology, Biology and Physics*, 21:683–693, 1991.

[8] E. K. Lee, T. Fox, and I. Crocker. Optimization of radiosurgery treatment planning via mixed integer programming. *Forthcoming in Medical Physics*, 2000.

[9] L. Luo, H. Shu, W. Yu, Y. Yan, X. Bao, and Y. Fu. Optimizing computerized treatment planning for the gamma knife by source culling. *International Journal of Radiation Oncology, Biology and Physics*, 45(5):1339–1346, 1999.

[10] D. M. Shepard, M. C. Ferris, G. Olivera, and T. R. Mackie. Optimizing the delivery of radiation to cancer patients. *SIAM Review*, 41:721–744, 1999.

[11] D. M. Shepard, M. C. Ferris, R. Ove, and L. Ma. Inverse treatment planning for gamma knife radiosurgery. *Medical Physics*, 27:12, 2000.

[12] R. A. Stone, V. Smith, and L. Verhey. Inverse planning for the Gamma Knife. *Medical Physics*, 20:865, 1993.

[13] J. Wang. Packing of unequal spheres and automated radiosurgical treatment planning. *Journal of Combinatorial Optimization*, 3:453–463, 1999.

[14] S Webb. Optimisation of conformal radiotherapy dose distributions by simulated annealing. *Physics in Medicine and Bilology*, 34(10):1349–1370, 1989.

[15] Q. J. Wu and J. D. Bourland. Morphology-guided radiosurgery treatment planning and optimization for multiple isocenters. *Medical Physics*, 26(10):2151–2160, 1999.

[16] Y. Yan, H. Shu, and X. Bao. Clinical treatment planning optimization by Powell's method for Gamma unit treatmen system. *International Journal of Radiation Oncology, Biology and Physics*, 39:247–254, 1997.