Optimization of Gamma Knife Radiosurgery

Michael C. Ferris and David M. Shepard

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, beams from 201 radioactive sources are focused so that they intersect at the same location in space, resulting in a spherical region of high dose referred to as a shot of radiation. The location and width of the shots can be adjusted using focussing helmets. By properly combining a set of shots, larger treatment volumes can be successfully treated with the Gamma Knife.

The goal of this project is to automate the treatment planning process. For each patient, an optimization seeks to produce a dose distribution that conforms closely to the treatment volume. The variables in the optimization can include the number of shots of radiation along with the size, the location, and the weight assigned to each. Formulation of such problems using a variety of mathematical programming models is described, and the solution of several test and real-patient examples is demonstrated.

1. Introduction

The Leksell 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 [**Gan97**]. Over 100 Gamma Knife units are installed worldwide, and more than 20,000 patients are treated each year. Inside the shielded treatment unit (see Figure 1), beams from 201 Co-60 radioactive sources are focused so that they intersect at the same location in space. The result is a spherical region of high dose referred to as a shot of radiation. Four different shot sizes are available to the user. By combining multiple shots of radiation, the treatment plan can be customized to treat lesions of varying sizes and shapes.

The goal of this project is to automate the treatment planning process. For each patient, the optimization seeks to produce a dose distribution that conforms closely to the treatment volume. The variables in the optimization can include the number of shots of radiation along with the size, the location, and the weight assigned to each. This paper will address the approaches that we have investigated for solving this problem. The advantages and disadvantages of two approaches will be discussed, and optimized patient plans will be shown.

1991 *Mathematics Subject Classification.* Primary: 92C50; Secondary: 90C30, 90C11.

This material is based on research supported in part by National Science Foundation Grants CDA-9726385 and CCR-9972372, the Air Force Office of Scientific Research Grant F49620-98-1-0417 and Microsoft Corporation.

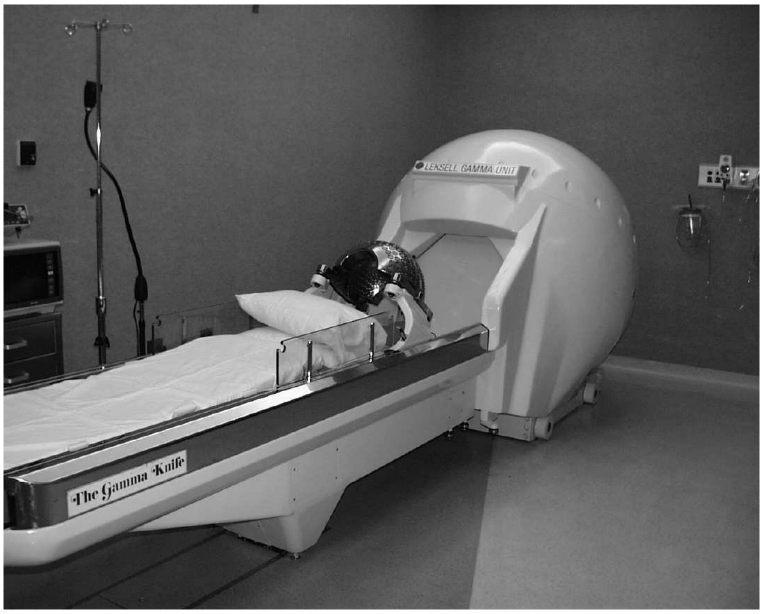


FIGURE 1. Gamma Knife Treatment Unit

The success of radiosurgery, in large part, depends upon the ability to accurately locate the treatment volume. The localization procedure begins with the fitting of a stereotactic head frame onto the patient's head (see Figure 2). This

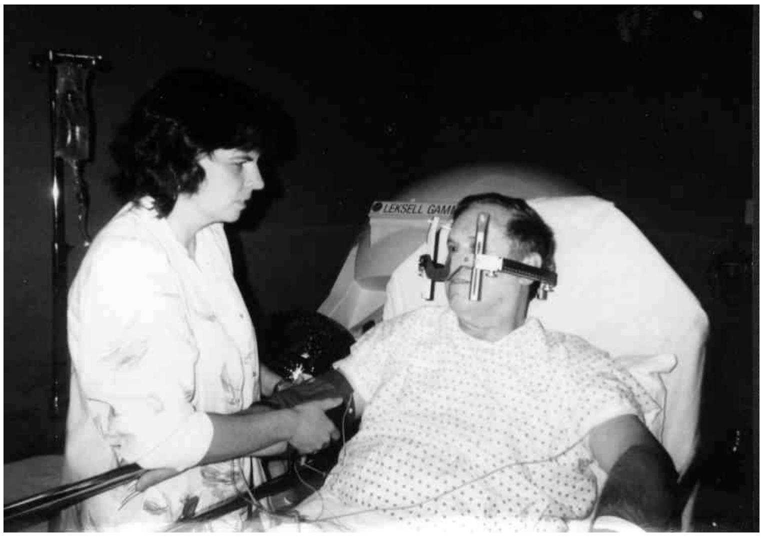


FIGURE 2. Stereotactic Head Frame

head frame is screwed into the patient's skull on the morning of the treatment.

Once the head frame is in place, the patient undergoes MRI or CT imaging. The imaging studies make it possible to precisely determine the position of the treatment volume with respect to the stereotactic head frame.

During the treatment, the patient's head frame attaches to a focusing helmet (see Figure 3). This focusing helmet serves as a collimator that further reduces the

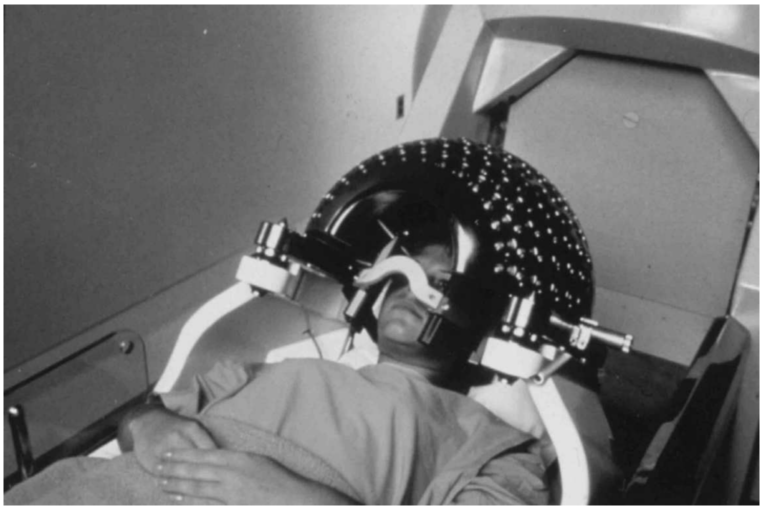


FIGURE 3. Focussing Helmet Attached to Frame

size of each beam. There are a total of four focusing helmets. These helmets can be used produce a shot of radiation that is 4mm, 8mm, 14mm, or 18mm in diameter (see Figure 4).

The neurosurgeon, the radiation oncologist, and the physicist work together in order to develop the patient's treatment plan. For some cases, the treatment planning process is relatively straightforward. For example, some small lesions can be covered with a single shot of radiation. Consequently, the treatment planning process can be completed in a few minutes. Unfortunately, the treatment planning process becomes much more complex when the tumor volume is large or irregularly shaped. These cases typically require several shots of radiation. Through an iterative process, the planners must determine the number of shots of radiation that are required along with the size, the location, and the weight that should be assigned to each. Each time that there is a change in the shot size, the staff must remove the current focusing helmet and replace it with the helmet of the correct size. This is a fairly time consuming process, because each helmet weighs approximately 500 pounds and must be manipulated mechanically. For each shot of radiation, the location of the center of the shot is determined by the connection between the helmet and the patient's head frame. Adjustments in this connection are made between each shot. At the time of delivery, the door to the treatment unit opens and the patient couch is advanced inside the shielded treatment vault (see Figure 5).

For many patients, the treatment planning process becomes both tedious and time consuming, and the quality of the treatment plan that is produced depends upon both the experience and the patience of the planner. Because of these factors, we have sought to develop an automated process for creating Gamma Knife

Figure 4. Stored Focussing Helmets for 4mm, 8mm, 14mm and 18mm shots

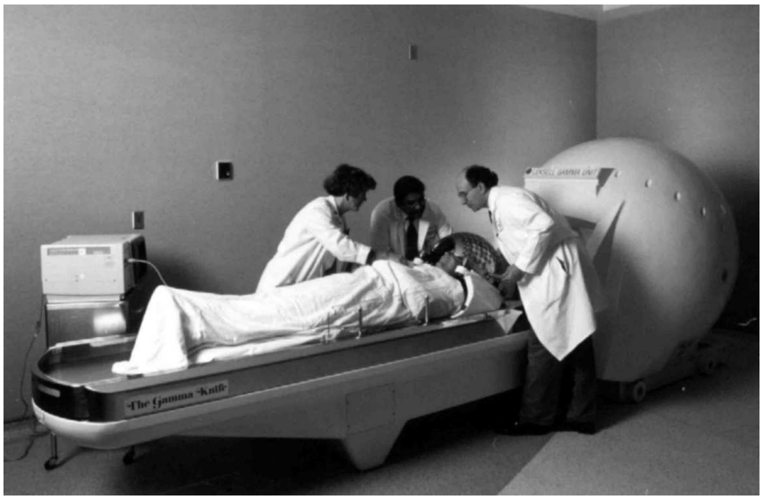


Figure 5. Operation of the Gamma Knife

treatment plans. As a first step in this process, the planner outlines the treatment volume and any sensitive structures. A series of treatment goals are then defined. Based upon these goals, an optimization algorithm determines the best possible treatment plan.

The ideal optimization technique for Gamma Knife treatment planning must be fast, flexible, and robust. The system must be fast, because the treatment planning process cannot begin until after the patient's head frame is in place and MRI or CT images have been obtained. It is therefore unacceptable to have a planning procedure that takes hours to complete. Our goal has been to create an automated planning system that can produce a treatment plan in 20 minutes or less.

The system must also be flexible. This is because the treatment goals will vary from one patient to the next. Finally, the system must be robust enough that it produces a high quality solution for treatment volumes of all sizes and shapes. A robust technique ensures a reasonable result regardless of the quality of the initial guess. An automated treatment planning process can potentially provide several benefits. Most importantly, the quality of patient care could be improved. Due to the large number of variables involved, it is unlikely that a manual approach will produce a truly optimal dose distribution. Improvements in the quality of a dose distribution can lead to an increase in the the probability of uncomplicated tumor control. An automated approach to treatment planning would also greatly reduce the amount of time that the neurosurgeon must commit to treatment planning. Finally, with optimization, it may be possible to reduce the number of shots and the variation in the shot size without sacrificing the quality of a patient's dose distribution. Therefore, the total treatment time would be reduced. Reduced treatment times are appreciated by the patient, and can make it possible to treat more patients in a day.

A number of researchers have studied techniques for automating the Gamma Knife treatment planning process [**WB99, LSY$^+$99**]. One approach incorporates the assumption that each shot of radiation can be modeled as a sphere. The problem is then reduced to one of geometric coverage, and a ball packing approach [**SSV93, WB99**] can be used to determine the shot locations and sizes. The use of a modified Powell's method in conjunction with simulated annealing has also been proposed [**LSY$^+$99, YSB97**].

This paper addresses a different approach, whereby the actual dose distribution is modeled and a formal constrained optimization model is solved to determine the treatment plan. In Section 2, we outline a nonparametric optimization approach for generating a dose model for the Gamma Knife. This dose model is a critical component for developing the ensuing optimization models. In Section 3, we describe two different optimization formulations that can be used for the treatment planning problem, one based on mixed integer programming, the other on nonlinear programming. We demonstrate the effectiveness of these approaches in Section 4 by comparing them on several test examples that we generate in a new testing environment. We also report computational results on several real patient cases using the nonlinear programming approach. The modeling and optimization uses various combinations of GAMS, MATLAB, CPLEX and CONOPT. The paper concludes with several remarks regarding incorporation of the methodology into the real patient planning process, along with some pointers regarding future work.

2. Dose Model

Each patient is currently modeled on a three dimensional grid of pixels. Given such a grid, the first task is to determine the dose that will be delivered by a particular plan at each pixel. We require an algebraic model of the distribution of the dose for use in our optimization formulations. In this model, we let $\mathcal{S}$ represent the set of the shots that we will consider, and $\mathcal{W}$ represent the possible shot sizes (typically 4mm, 8mm, 14mm and 18mm).

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)

$$Dose(i,j,k) = \sum_{(s,w)\in\mathcal{S}\times\mathcal{W}} t_{s,w} D_w(x_s, y_s, z_s, i, j, k), \tag{2.1}$$

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

To determine the form of D_w, the following procedure was followed. First, 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, for example,

$$\bar{D}_w(d) = \frac{D_w(0,0,0,d,0,0) + D_w(0,0,0,0,d,0) + D_w(0,0,0,0,0,d)}{3}.$$

These values were used as data in a nonlinear parameter estimation problem.

The problem is reduced to determining a functional form for the dose delivered at a pixel that is a distance d away from the center of the shot. A sum of error functions has been noted in the literature to approximate this dose distribution [**CKM98**]. We therefore used the following functional form

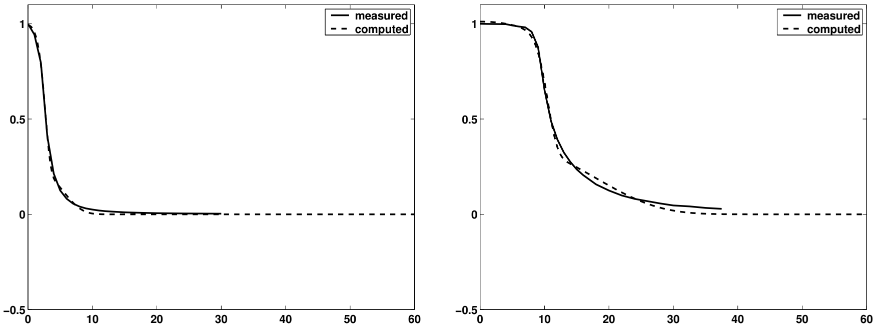
$$\sum_{i=1}^{2} \lambda_i \left(1 - \operatorname{erf}\left(\frac{d - r_i}{\sigma_i}\right)\right) \tag{2.2}$$

(with d representing the distance) and fit the six parameters λ_i, r_i and σ_i to the data described above via least-squares, with different values for each width shot. In our work, we use the notation $\operatorname{erf}(x)$ to represent the integral of the standard normal distribution from $-\infty$ to x. The resulting nonlinear optimization problem

$$\min_{\lambda,r,\sigma} \left\| \bar{D}_w(d) - \sum_{i=1}^{2} \lambda_i \left(1 - \operatorname{erf}\left(\frac{d - r_i}{\sigma_i}\right)\right) \right\|^2$$

was solved using CONOPT. As can be seen by two representative pictorial representations of the resulting fits in Figure 6, the functional fit is very close to the observed data. The fit is best for the small shots (see Figure 6(a) for example), and decreases slightly in accuracy for the larger ones (see Figure 6(b) for example). The particular values of the parameters that we generated are given in Table 1. It is clear that a closer fit can be achieved using more parameters in (2.2). However, this will lead to more computational overhead in the optimization, which we believe is not necessary.

It is possible that the averaging in the x, y and z directions creates a small error for the particular patient at hand. Furthermore, there may be variations in dose across patients. These effects have been ignored in the subsequent formulations. In future work, along with investigations of the above issues, we may fit ellipsoids instead of spheres to the dose data since various researchers have commented that



(a) 4mm shot

(b) 18mm shot

FIGURE 6. Nonlinear Parametric Fitting for Dose Data

Shot	λ_1	r_1	σ_1	λ_2	r_2	σ_2
4mm	0.649200	1.365916	4.413680	0.599844	2.661771	0.668291
8mm	0.401007	7.035785	5.702334	0.648584	4.849365	1.149176
14mm	0.363704	13.97259	7.196694	0.657808	8.199979	1.321161
18mm	0.381801	17.67857	8.194611	0.634696	10.31583	1.441725

TABLE 1. Parameter for Dose Model

the dose is skewed in certain directions. While cubic b-spline approaches may outperform the fit that we achieve here, it is currently unclear how to use such approaches within the GAMS modeling format that we employ.

3. Optimization Formulation

Once a description of the dose is determined, an optimization model can be formulated. The basic variables of the optimization include the number of shots of radiation that will be delivered, along with the width of the shot w, the coordinates of the center location of the shot (x_s, y_s, z_s), and the time $t_{s,w}$ that each shot is exposed. In practice, we consider a grid of pixels, noting two subsets of this grid, namely $\mathcal{T}$ and $\mathcal{N}$, that represent the subset of pixels that are in (out) of the target respectively.

Neurosurgeons commonly use isodose curves as a means of judging the quality of a treatment plan. The 50% isodose curve is a curve that encompasses all of the pixels that receive at least 50% of that maximum dose that is delivered to any pixel in the patient. The neurosurgeon may wish to impose a requirement that the entire target is surrounded by an isodose line of x%. For example, a constraint that the 50% line must surround the target can be modeled by imposing strict lower and upper bounds on the dose allowed in the target, namely for all $(i, j, k) \in \mathcal{T}$

$$1 \leq Dose(i, j, k) \leq 2.$$

In this way, the 50% isodose curve is guaranteed to cover the target. The isodose requirements can be changed by simply modifying the numerical values of the bounds.

We have tested a variety of optimization formulations. For this paper, however, we will simply discuss two. The first formulation imposes a constraint on the minimum isodose line that must surround the target. Given this constraint, the goal is to minimize the dose outside of the target. The formulation is:

$$\begin{array}{lll} \min & & \sum_{(i,j,k)\in\mathcal{N}} \mathit{Dose}(i,j,k) \\ \text{subject to} & \mathit{Dose}(i,j,k) = & \sum_{(s,w)\in\mathcal{S}\times\mathcal{W}} t_{s,w} D_w(x_s,y_s,z_s,i,j,k) \\ & & 1 \leq \mathit{Dose}(i,j,k) \leq 2,\ \forall (i,j,k)\in\mathcal{T} \\ & & t_{s,w} \geq 0 \\ & & \operatorname{card}\left(\{(s,w)\in\mathcal{S}\times\mathcal{W} | t_{s,w} > 0\}\right) \leq n. \end{array} \tag{3.1}$$

The final constraint states that no more than n shots are to be used in the plan.

The second formulation uses a constraint to control the conformity of the plan. The constraint specifies that at least y% of the total dose must be deposited in the target. An upper bound is also placed on the dose to the target. Given these constraints the optimizer seeks to minimize the total underdosage in the target. A pixel is considered to be underdosed if it receives less than the prescribed isodose, which for the example formulation is assumed to be 1. We actually use the optimization process to model UnderDose. UnderDose is constrained to be no less that $\max(0, 1 - \mathit{Dose})$ at every pixel in the target, and since we minimize UnderDose, it will take on the maximium of these two values at optimality. The complete formulation is:

$$\begin{array}{ll} \min & \sum_{(i,j,k)\in\mathcal{T}} \mathit{UnderDose}(i,j,k) \\ \text{subject to} & \mathit{Dose}(i,j,k) = \sum_{(s,w)\in\mathcal{S}\times\mathcal{W}} t_{s,w} D_w(x_s,y_s,z_s,i,j,k) \\ & \dfrac{\sum_{(i,j,k)\in\mathcal{T}} \mathit{Dose}(i,j,k)}{\sum_{(i,j,k)} \mathit{Dose}(i,j,k)} \geq P \\ & 0 \leq \mathit{UnderDose}(i,j,k) \geq 1 - \mathit{Dose}(i,j,k),\ \forall (i,j,k)\in\mathcal{T} \\ & \mathit{Dose}(i,j,k) \leq 2,\ \forall (i,j,k)\in\mathcal{T} \\ & t_{s,w} \geq 0 \\ & \operatorname{card}\left(\{(s,w)\in\mathcal{S}\times\mathcal{W} | t_{s,w} > 0\}\right) \leq n. \end{array} \tag{3.2}$$

In this formulation, P is the fraction of the total dose that must be deposited in the target; typically we choose values for P between 25% and 40%. 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). In practical application, rather than calculate the dose at every pixel, it is easy to accurately estimate the total dose delivered by a plan based solely on the $t_{s,w}$ variables and other precalculated constants.

Both of the formulations are based on the assumption that the neurosurgeon can use the volume of the target and the irregularity of its shape in order to a-priori determine a realistic upper bound n on the number of shots required for treatment. Therefore, the model bounds the number of shots a priori, rather than forcing the model to minimize this number. Since most of the comments we make in the sequel apply equally well to either formulation, we will restrict our exposition mainly to (3.1) in the sequel solely for clarity.

Practical issues must also be considered. For example, a large number of shots can lead to a treatment time that is unacceptably long. This is because after each shot is delivered, the staff must enter the room and perform a series of adjustments. Also, if a change in the shot width is required, the user must exchange focusing helmets.

Several issues need to be resolved to create models that are practical, implementable, and solvable (in a reasonable time frame). Two main approaches are proposed in this paper, namely using mixed integer programming (MIP) and nonlinear programming (NLP).

3.1. MIP approach. Since the optimization technology associated with solving nonlinear mixed integer programming problems is very limited, we restrict attention to linear models and thus need to ensure that the dose calculation constraint (2.1) is linear. To do this, we need to have $D_w(\cdots)$ as data, and thus need to determine all possible (x_s, y_s, z_s) before optimization. We therefore a-priori choose a grid of possible shot locations $\mathcal{G}_\mathcal{S}$, and precalculate $D_{s,w}(i,j,k) := D_w(x_s, y_s, z_s, i, j, k)$ for each grid location $s \in \mathcal{G}_\mathcal{S}$, every pixel (i,j,k) and each width $w \in \mathcal{W}$ and use the optimization algorithm to decide whether or not to use a shot at a particular location. A major drawback of this approach is the enormous amount of data that is required. We will discuss approaches to overcome this in the sequel. It is clear that the only shots that will be considered in this approach are shots that lie within the target. This is because our grid is generated a-priori, and we can easily determine whether or not a shot lies within the target.

However, the mixed integer approach allows us to introduce a new binary variable $\psi_{s,w}$ (heuristically, use the shot (s,w) or not) and impose the constraint

$$t_{s,w} \le \bar{T}\psi_{s,w}$$

to force $\psi_{s,w} = 1$ whenever $t_{s,w} > 0$. Here $\bar{T}$ is an (easy to estimate) upper bound on the length of time that a particular shot can be exposed. Once these variables are introduced, we then write the limiting shot constraint as

$$\sum_{(s,w)\in\mathcal{G}_\mathcal{S}\times\mathcal{W}} \psi_{s,w} \le n.$$

For the problem outlined in (3.1), the resulting mixed integer optimization problem is:

$$\begin{array}{ll} \min & \displaystyle\sum_{(i,j,k)\in\mathcal{R}} Dose(i,j,k) \\ \text{subject to} & Dose(i,j,k) = \displaystyle\sum_{(s,w)\in\mathcal{G_S}\times\mathcal{W}} t_{s,w} D_{s,w}(i,j,k) \\ & 1 \le Dose(i,j,k) \le 2,\ \forall (i,j,k) \in \mathcal{T} \\ & 0 \le t_{s,w} \le \bar{T}\psi_{s,w},\ \psi_{s,w} \in \{0,1\} \\ & \sum_{(s,w)\in\mathcal{G_S}\times\mathcal{W}} \psi_{s,w} \le n. \end{array}$$

Note that we have replaced $\mathcal{N}$ with $\mathcal{R}$ in the above problem. In order to reduce the amount of data required and since the dose drops off rapidly away from its center, we only minimize the dose in a rind $\mathcal{R}$ around the target, instead of everywhere.

The above formulation is a large-scale mixed integer problem [**NW88, Wol98**] that is typically computationally intractable in the time allowed. Several methods are outlined below to further reduce the large amounts of data arising from the realistic dose calculations.

The benefits of reducing the possible shot locations $\mathcal{G_S}$ are two-fold. Firstly, the number of integer variables $\psi_{s,w}$ (typically the limiting computational factor for solution) is reduced. Secondly, the amount of precalculated data required is reduced. The following heuristic was used to attempt to reduce the number of shots as much as possible. We first generate a coarse grid of large shots and for a particular grid spacing and grid offset, we calculate the number of these shots that hit the target. By searching over all grid offsets, we determine the offset that maximizes the number of large shots from this grid that are on target. Having established this, we randomly place small shots near the boundary of the target until a prespecified number of shots is generated. It is an unresolved research issue to determine the proportion of large to small shots, and the number of possible locations that we should use.

There are several interesting features of the MIP approach. As noted above, there are large amounts of data, and many integer variables unless $\mathcal{G_S}$ is severely restricted. By restricting the grid too much, the resulting constrained problem is infeasible. Furthermore, even with the restrictions and a very fast MIP solver such as CPLEX, the problems take enormous amounts of time to solve or determine integer infeasibility. The major benefit of this approach, however, is that it provides a guarantee of global optimality for that particular choice of $\mathcal{G_S}$ and $\mathcal{R}$.

3.2. Nonlinear Approach. A very different approach is to use nonlinear programming and approximate the counting of shots. The basic idea is to use a migrating shot technique. This approach takes a limited number of shots, and the position each shot can change over the course of the optimization. Given we are using n shots, we introduce variables x_s, y_s and z_s and use the nonlinear function formulation (2.2) for the dose calculation. Note that with this approach there are just $3n$ variables to specify shot location, and n is typically small in the application (namely 5 - 20).

This has the key advantage over the MIP formulation that the “data” is much smaller. However, the nonlinear programming problems are still quite difficult to solve (they are not convex, and hence may have local solutions that are not global solutions). Secondly, while it is easy to limit the number of shot locations to n, the actual constraint requires that there are no more than n shots are to be used in the plan. The missing feature is to choose a particular width of shot at each location.

To implement this feature, we use a two phase approach. Note that a shot of width w is used at location s if $t_{s,w} > 0$. The main idea is to approximate the step function $H(t) = 1$ if $t > 0$ and $H(t) = 0$ when $t = 0$ by a nonlinear function,

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

For increasing values of α, H_α becomes a closer approximation to the step function H. Therfore, for the approach outlined with (3.1), we solve the following problem

$$\begin{array}{lll}
\min & & \displaystyle\sum_{(i,j,k)\in\mathcal{R}} Dose(i,j,k) \\
\text{subject to} & Dose(i,j,k) = & \displaystyle\sum_{(s,w)\in\{1,\dots,n\}\times\mathcal{W}} t_{s,w} D_w(x_s, y_s, z_s, i, j, k) \\
& & 1 \leq Dose(i,j,k) \leq 2, \ \forall (i,j,k) \in \mathcal{T} \\
& & t_{s,w} \geq 0 \\
& & \sum_{w\in\mathcal{W}} H_\alpha(t_{s,w}) \leq 1, \ \forall s \in \{1, \dots, n\},
\end{array}$$

with this approximation and $\alpha = 10$, then we fix any shots that have zero width to have $t_{s,w} = 0$, and reoptimize the following problem (with $\alpha = 100$)

$$\begin{array}{lll}
\min & & \displaystyle\sum_{(i,j,k)\in\mathcal{R}} Dose(i,j,k) \\
\text{subject to} & Dose(i,j,k) = & \displaystyle\sum_{(s,w)\in\{1,\dots,n\}\times\mathcal{W}} t_{s,w} D_w(x_s, y_s, z_s, i, j, k) \\
& & 1 \leq Dose(i,j,k) \leq 2, \ \forall (i,j,k) \in \mathcal{T} \\
& & t_{s,w} \geq 0 \\
& & \sum_{(s,w)\in\{1,\dots,n\}\times\mathcal{W}} H_\alpha(t_{s,w}) \leq n \\
& & t_{s,w} = 0, \ \forall (s,w) \in F.
\end{array}$$

Here F represents the subset of $\{1, \dots, n\} \times \mathcal{W}$ of variables that were fixed at 0 based on the solution of the first optimization. Note that both problems are highly nonconvex, so there is no guarantee of global optimality. We experimented with starting points somewhat and use equal values of t for the results of the sequel.

Another disadvantage of the nonlinear approach is that it is difficult to constrain the location of the shots to within the target. In fact, we do not add such a constraint, but choose appropriate starting values for (x_s, y_s, z_s) within the (interior of the) target to encourage the final shot locations to also be within the target. In all our experiments, this was indeed the case, although several shots can be located close to the boundary.

4. Computational Results

4.1. Two-dimensional testing environment. We created a testing environment for our models within MATLAB, using the image processing toolbox. The testing environment provided a means to generate two-dimensional "tumors" of various shapes and sizes and visualize the results of the optimization applied to these targets. The optimization models were written in the GAMS modeling language [**BKM88**], a high-level modeling system for mathematical programming problems. It consists of a language compiler and a variety of integrated high-performance solvers. We used the MATLAB/GAMS interface [**Fer98**] to communicate between these two software packages. GAMS was chosen for the optimizations based on its ability to formulate various types of models and then apply state-of-the-art optimization algorithms for their solution. In particular, for the mixed integer problems we used the CPLEX optimization package [**ILO**], whereas for nonlinear programs, CONOPT [**Dru85**] was used. While the two dimensional testing environment significantly reduces the amount of data present in the model, we found this to be an excellent testing tool to pinpoint the advantages and disadvantages of various formulations. In particular, it is very easy to change between a wide variety of model types, objective functions and constraints.

We present results for two cases. In the first instance, we prescribe 3 shots of radiation. The rind around the target has width 10. The resulting size of the target and rind is respectively 235 and 1092 pixels. The results of applying the mixed integer and nonlinear programming approaches are shown in Figure 7. The tumor outline is shown is cyan, while the scale of the intensities is given on the bar to the right of each figure. The MIP solution, shown in Figure 7(a), results in two 8mm and one 14mm shots, with a total dose to the rind of 510. The optimization took 165 seconds, terminating with a relative optimality tolerance of 1%. The size of the grid of potential shots was 150.

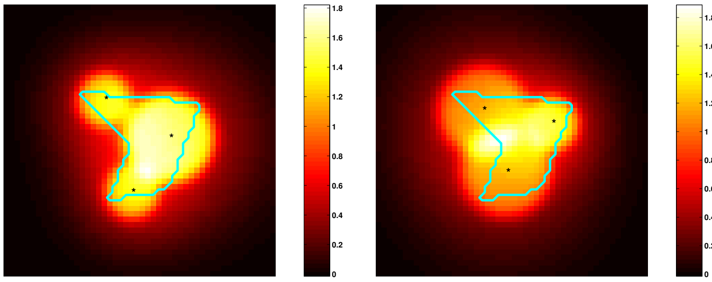
The NLP solution, shown in Figure 7(a), results in one 8mm and two 14mm shots, that combine to give a total dose to the rind of 490. The optimization took 65 seconds, somewhat faster than the MIP algorithm. Although the solution is not guaranteed optimal, it results in a comparable solution to the MIP code. Since the NLP code can choose the locations of the shots (as opposed to selecting one of the shots from the grid) it is possible to find better solutions using the NLP approach than the MIP approach.

To show the effects of more data on our approaches, Figure 8 exhibits the corresponding solutions for a larger case. Again, the target is outlined in cyan, this time containing 1008 pixels. The rind again has radius 10, resulting in 1365 pixels. We attempted to find solutions using seven shots.

The NLP solution, shown as Figure 8(b) took 1354 seconds to generate. It uses two 8mm, three 14mm and two 18mm, with a total dose to the enclosing rind of 1017. Note that the optimization time is slightly over the 20 minute limit, but can be reduced (for example) by reducing the number of pixels in the rind.

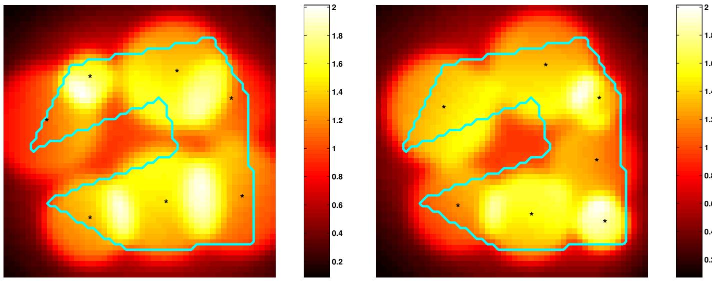
The MIP solution that is shown as Figure 8(a) consists of one 8mm, two 14mm and four 18mm, with a total dose to the disk of 1123, again with 150 potential locations for shots. The optimization was terminated after two hours with relative optimality tolerance 17%, having explored 15,000 nodes of the branch and bound tree. We believe that this demonstrates that more work is needed in this area before the MIP approach becomes a viable alternative to the NLP approach.

4.2. Real patient (three-dimensional) data. After determining which models gave satisfactory results, we switched to three dimensional models, based on real patient data. Currently, the mixed integer programming solvers are unable to process the real patient data in an acceptable time, so we present only nonlinear programming results in this section.



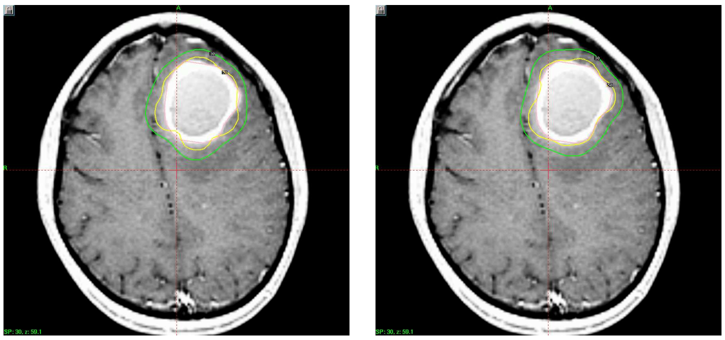
(a) MIP solution (b) NLP solution

FIGURE 7. Smaller Two-Dimensional Test Case



(a) MIP solution (b) NLP solution

FIGURE 8. Larger Two-Dimensional Test Case



(a) Manual 13 shot solution (b) Optimized 13 shot solution

FIGURE 9. Isodose Curves for Patient 1; Axial Slice

The nonlinear programming approach to treatment planning has currently been applied to eight patient cases, and each optimized plan was compared to the original treatment plan produced by the neurosurgeon. These comparisons were based upon analysis of the dose-volume histogram, the minimum target dose and the the conformity index (the volume of the prescription isodose divided by the target volume). Treatment plans using 5-8 shots were typically optimized in less than 15 minutes on a Sparc Ultra-10 with a 330 MHz processor.

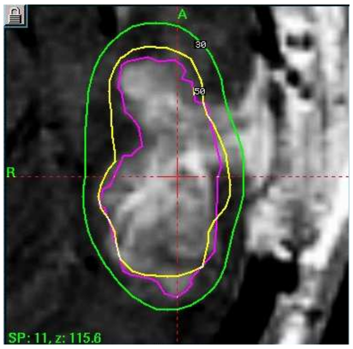
A treatment plan comparison is shown in Figure 9. Each image includes a pink line that indicates the border of the tumor. The 50% isodose line is shown in yellow, and the 30% isodose line is shown in green. Due to the massive size of this patient's tumor (35.3 cubic centimeters), a large number of shots were required in order to achieve adequate tumor coverage. In fact, this tumor is at the very limit of size of tumors that are treated using the Gamma Knife at the University of Maryland. Consequently, the treatment planning process was very tedious and required a great deal of artful manipulation on the part of the neurosurgeon.

The manual treatment plan (Figure 9a) used 13 shots of radiation, and is an excellent plan for such a large tumor. An optimized plan (Figure 9b) was produced using the same number of shots. A constraint guaranteed that the entire tumor volume was surrounded by the 47% isodose line, and the optimizer sought to minimize the total dose delivered outside of the tumor. The optimized plan is very similar in quality to the plan produced by an experienced nuerosugeon. In fact, the optimized plan increased the minimum target dose by 2% and provided a small improvement in the conformity index (1.28 versus 1.27). The purpose of including these results here is to show that an automatic plan can reproduce (or slightly improve) a plan determined by an expert in the field, for even the most difficult cases.

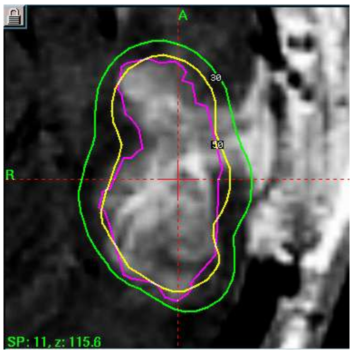
The results for a second patient are shown in Figure 10. In each figure, the purple line denotes the border of the tumor. The 50% isodose line is in yellow, and the 30% is line is in green. In this case, the manual treatement plan used 7 shots of radiation. The optimization was set up to use 6 shots of radiation. A constraint was included that specified that at least 30% of the total dose delivered to the patient had to be deposited in the target. Given this contraint, the optimizer sought to minimize the underdosage to the tumor (as in (3.2)). A voxel was considered to be underdosed if it received less than 50% of the maximum tumor dose. Again, the figures show the plan to be similar in quality, although in this case, the minimum tumor dose was increased from 43% to 44% and the conformity index was simultaneously improved from 1.70 to 1.75. Furthermore, the optimization process used one less shot of radiation than the expert.

The results shown in Figures 9-10 illustrate that this inverse treatment planning can be used to produce high-quality conformal treatment plans. A more detailed analysis of patient data will be provided in a future publication.

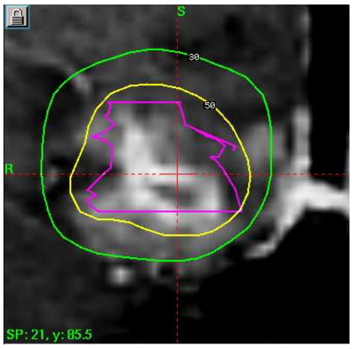
The robustness of the technique is of critical importance. For the eight patient cases optimized thus far, each shot was initially assigned a random location inside of the target. The initial guess incorporated uniform beam weights. Despite this somewhat arbitrary starting point, high-quality dose distributions have been obtained in all cases.



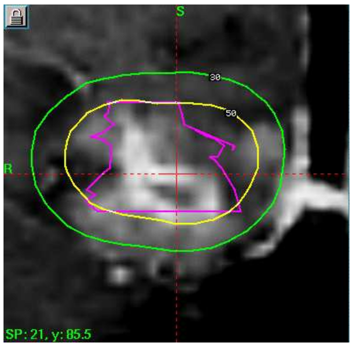
(a) Manual 7 shot solution; Axial Slice



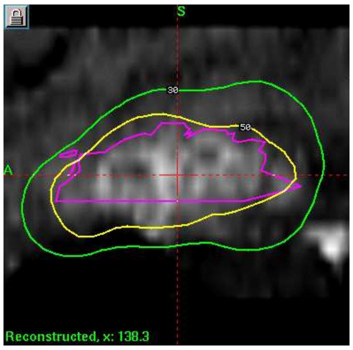
(b) Optimized 6 shot solution; Axial Slice



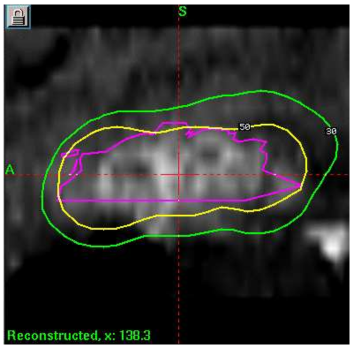
(c) Manual 7 shot solution; Coronal Slice



(d) Optimized 6 shot solution; Coronal Slice



(e) Manual 7 shot solution; Sagittal Slice



(f) Optimized 6 shot solution; Sagittal Slice

FIGURE 10. Isodose Curves for Patient 2

5. Conclusions and Future Work

Currently neurosurgeons produce suboptimal treatment plans by hand. We believe an automated approach based on optimization will generate not only more uniform, better treatment plans, but also produce them in less time than is currently used. The results given in this paper seem to indicate that the approach outlined here is sufficiently robust and quick to be used in a practical setting.

We believe it is important to integrate our procedure into the real system. Currently, we generate the target volume on the Gamma Knife and export it to the optimization program. The optimization program then generates a treatment plan that is imported back onto the Gamma Knife. This is somewhat time consuming, and likely to produce some data errors due to finite precision arithmetic and fixed size output. A full integration into the Gamma Knife planning system would reduce this time, and remove any of these errors.

We believe it is important to reduce the optimization times still further. This would give the ability to try "what if scenarios" at treatment planning time. For example, can conformality be improved if more shots are allowed? Can we treat the tumor at the 60% isodose level?

Apart from the time considerations, the outstanding question from an optimization viewpoint is the issue of global versus local optimality. While the nonlinear programming approach has proven very effective in practice, there is no guarantee that there is not a better treatment plan. Understanding how to improve the robustness of the model and solution techniques will enable more definite bounds to be given on the solution quality compared to the best possible solution.

6. Acknowledgements

We would like to thank Lijun Ma, Ph.D. and Cedric Yu, D.Sc., for their assistance with this project.

References

[BKM88] A. Brooke, D. Kendrick, and A. Meeraus, *GAMS: A user's guide*, The Scientific Press, South San Francisco, CA, 1988.

[CKM98] P. S. Cho, H. G. Kuterdem, and R. J. Marks, *A spherical dose model for radiosurgery treatment planning*, Physics in Medicine and Biology **43** (1998), 3145–3148.

[Dru85] A. Drud, *CONOPT: A GRG code for large sparse dynamic nonlinear optimization problems*, Mathematical Programming **31** (1985), 153–191.

[Fer98] M. C. Ferris, *MATLAB and GAMS: Interfacing optimization and visualization software*, Mathematical Programming Technical Report 98-19, Computer Sciences Department, University of Wisconsin, Madison, Wisconsin, 1998.

[Gan97] J. C. Ganz, *Gamma knife surgery*, Springer-Verlag Wien, Austria, 1997.

[ILO] ILOG CPLEX Division, 889 Alder Avenue, Incline Village, Nevada, *CPLEX optimizer*, http://www.cplex.com/.

[LSY+99] 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** (1999), no. 5, 1339–1346.

[NW88] G. L. Nemhauser and L. A. Wolsey, *Integer and combinatorial optimization*, John Wiley & Sons, New York, NY, 1988.

[SSV93] R. A. Stone, V. Smith, and L. Verhey, *Inverse planning for the Gamma Knife*, Medical Physics **20** (1993), 865.

[WB99] Q. J. Wu and J. D. Bourland, *Morphology-guided radiosurgery treatment planning and optimization for multiple isocenters*, Medical Physics **26** (1999), no. 10, 2151–2160.

[Wol98] L. A. Wolsey, *Integer programming*, John Wiley & Sons, New York, NY, 1998.

[YSB97] 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** (1997), 247–254.

Computer Sciences Department, University of Wisconsin, 1210 West Dayton Street, Madison, Wisconsin 53706
E-mail address: `ferris@cs.wisc.edu`

Department of Radiation Oncology, University of Maryland School of Medicine, 22 South Green Street, Baltimore, MD 21201
E-mail address: `dshep001@umaryland.edu`