8-2009

# A clustering approach for optimizing beam angles in IMRT planning

Gino J. Lim
*University of Houston - Main*

Allen Holder
*Rose-Hulman Institute of Technology*, holder@rose-hulman.edu

Josh Reese
*PROS Revenue Management*

### Recommended Citation

# A clustering approach for optimizing beam angles in IMRT planning

**Gino J. Lim, Allen Holder, Josh Reese**

**Mathematical Sciences Technical Report Series**
**MSTR 09-03**

**August, 2009**

**Department of Mathematics**
**Rose-Hulman Institute of Technology**
**http://www.rose-hulman.edu/math**

**Fax (812)-877-8333**                              **Phone (812)-877-8193**

# A clustering approach for optimizing beam angles in IMRT planning

**Gino J. Lim**
**Department of Industrial Engineering**
**University of Houston, Houston, TX 77204, USA**

**Allen Holder**
**Department of Mathematics**
**Rose-Hulman Institute of Technology, Terre Haute, IN, USA**

**Josh Reese**
**PROS Revenue Management, Houston, TX, USA**

## Abstract

In this paper we introduce a p-median problem based clustering heuristic for selecting efficient beam angles for intensity modulated radiation therapy. The essence of the method described here is the clustering of beam angles according to probability that an angle will be observed in the final solution and similarities among different angles and the selection of a representative angle from each of the $p$ resulting cluster cells. We conduct experiments using several combinations of modeling parameters to find the conditions where the heuristic best performs. We found a combination of such parameters that outperformed all other parameters on three of the four tested instances.

**Keywords**
IMRT, Beam Angle Optimization, p-Median.

## 1. Introduction

Intensity modulated radiation therapy (IMRT) is an advanced form of three-dimensional conformal radiation therapy (3DCRT) [7] for cancer patients. In IMRT, radiation is delivered via a linear accelerator (LINAC) [10]. The LINAC is mounted on a movable arm, called the *gantry*, that is able to make a complete 360 degree rotation around the patient. The gantry rotates around a point called the *isocenter*. The LINAC is equipped with a *multileaf collimator* (MLC), a device with pneumatic "leaves" that move back and forth to block portions of the radiation beam. The multileaf collimator allows the beam to be shaped, thereby partitioning each of the 360 beams into *sub-beams*, also known as *pencil beams* or *beamlets* [4].

IMRT planning is an "inverse problem," where an ideal dose deposition is specified *a priori* as a mathematical objective function. Treatment parameters are then iteratively adjusted and the results are mathematically simulated. The goal of IMRT is to deliver a lethal dose of radiation to the cancer cells while limiting the amount of radiation that deposits in nearby critical and normal healthy tissues in order to reduce damage [3, 8].

There are several different volumes to be considered when planning to treat a tumor. The first is the planning target volume (PTV) that includes the actual tumor volume and further margins for uncertainties such as patient motion or anatomy changes during the course of the treatments. Physicians delineate on each CT slice the PTV as well as organs that may be incident to the radiation beam. These organs are often called critical structures or organs-at-risk (OAR). Any remaining tissue is delineated as normal tissue (NOR).

Physicians specify the *prescription*, characterized by a lower and upper preference bound on absorbed dose in the tumor. The lower bound on the tumor ensures that there are no *cold spots*, defined as dosepoints in the anatomy that are receiving significantly lower than the prescribed dose. Furthermore, certain hot spot control parameters may be specified for OAR and NOR tissue. These specify how to penalize dosepoints in OAR and NOR tissue that receive unusually high dosage. These parameters may vary with the structure type and are determined by the susceptibility and severity of radiation damage to that structure. For instance, a significant amount of damage to the spinal cord may result in paralysis or loss of critical body functions. Therefore, the tolerance assigned to the spinal cord is comparatively lower than most other structures [4].

## 2. Optimization Model

Assuming a finite number of angles *a* and sub-beams *s*, we can calculate the dose contribution of sub-beam $(a,s)$ to the dose point *p*. In three dimensions the dose points result from discretizing the irradiated area into small cubes

called voxels. In two dimensions the dose points may be visualized as squares or pixels. We let $\omega_{(p,a,s)}$ be the rate in Grays per second (Gy/t) at which radiation along sub-beam $s$ in angle $a$ is deposited into dose point $p$, where $\omega_{(p,a,s)} \geq 0, \forall (p,a,s)$. The calculation of this value is beyond the scope of this paper, but the value captures the effects of scattering, attenuation, and several other physical processes on the deposited dose. The total amount of dose deposited in a particular $p$ is

$$D_p = \sum_{(a,s)} \omega_{(p,a,s)} x_{(a,s)}.$$

There are three optimization problems within the realm of IMRT treatment planning – beam angle optimization (BAO) problem [1, 2, 8], fluence map optimization (FMO) problem [8] and beam segmentation problem [5, 6]. Finding the optimal beam angles can be computationally challenging, researchers often focus on heuristic approaches for selecting efficient beam angles, and it is referred to a beam angle selection problem (BASP). BASP is the primary focus of this paper. It involves choosing the most effective beams to use in an intensity modulated radiation therapy (IMRT) treatment. The fluence map optimization (FMO) problem seeks an optimal fluence vector $x$ given certain constraints on the dose deposited in the anatomy. In this paper we focus on BASP but we use FMO in the process of developing a heuristic for BASP. Therefore we develop an optimization model that simultaneously optimizes both BASP and FMO below. The objective function is to capture the nature of how a treatment planner decides between good and bad treatments. The function we use can be found in [8] and we present here involves several parameters, outlined in Table 1.

| Notation | Definition | Prostate | Pancreas |
|----------|------------|----------|----------|
| $\Theta_L$ | Cold spot control parameter on PTV | 0.97 | 0.97 |
| $\Theta_H$ | Hot spot control parameter on PTV | 1.05 | 1.05 |
| $\phi_i$ | Hot spot control parameter on the $i$-th OAR | 0.3 | 0.2 |
| $L_T$ | Lower reference bound on PTV | 0.94 | 0.95 |
| $U_T$ | Upper reference bound on PTV | 1.15 | 1.07 |
| $U_N$ | Upper reference bound on normal structure | 0.78 | 0.83 |
| $\lambda_T^+$ | Penalty term for hot spots on PTV | 1.0 | 1.0 |
| $\lambda_T^-$ | Penalty term for cold spots on PTV | 1.0 | 1.0 |
| $\lambda_N$ | Penalty term for normal structure | 1.0 | 1.0 |

Table 1: Optimization model parameters

Delivering a treatment from all possible 360 gantry angles is not clinically feasible. Current clinical constraints require that IMRT treatments consist of less than 9 gantry angles (the actual number depends on the problem). Experiments have verified that treatment quality does not increase substantially with more beam angles per treatment. Thus, the beam angle selection problem is how to select $k$ angles so that the subsequent $k$-angle treatment is effective. We can model BASP as an integer program with the following, given that $\mathcal{A}$ is a *candidate* set of beam angles to choose from:

$$\min f(\mathcal{A}', x) \tag{1}$$

subject to:

$$\begin{aligned}
\mathcal{A}' &\subset \mathcal{A}, \\
|\mathcal{A}'| &\leq k, \\
D_p &= \sum_{a \in \mathcal{A}'} \sum_{s=1}^{n} x_{(a,s)} \omega_{(p,a,s)}, \quad \forall p \in \{T \cup C \cup N\} \\
L_T &\leq A_T x \leq U_T, \\
0 &\leq x_{(a,s)} \leq M_{(a,s)}, \qquad \forall a \in \mathcal{A}', s \in \{1,\ldots,n\}.
\end{aligned}$$

Here we are deciding which $k$ beams from a candidate set (and an associated fluence vector $x$ for the chosen beams) will satisfy the constraints and best minimize the objective. This problem is combinatorial in nature, exceeding current calculation capabilities. Therefore there is a need for heuristic methods in solving the IMRT beam angle selection problem.

## 3. Set Clustering Heuristic

## *p*-**Median Problem**

The essence of the method described here is the clustering of beam angles according to probability and similarity information and the selection of a representative angle from each of the $k$ resulting cluster cells. The $p$-median problem is commonly stated as a binary optimization problem, originally found in [9]. Let $\Omega(a_i)$ be a probability function designed to capture the likelihood of observing angle $a_i$ in a final treatment plan. Furthermore, let $h(a_i, a_j)$ be a function that captures the similarity between beam angles $a_i$ and $a_j$. We define the following binary allocation variable:

$$\xi_{ij} = \begin{cases} 1 & \text{if angle } a_j \text{ is allocated with } a_i \\ 0 & \text{otherwise,} \end{cases}$$

and with these definitions the binary integer program is stated as:

$$\min \quad \sum_{ij} \Omega(a_j) h(a_i, a_j) \xi_{ij}$$

$$\text{subject to} \quad \sum_i \xi_{ij} = 1, \text{ for } j = 1, \dots, n, \tag{2}$$

$$\sum_i \xi_{ii} = k, \tag{3}$$

$$\xi_{ii} \geq \xi_{ij}, \quad \forall i, j = 1, \dots, n, \tag{4}$$

$$\xi_{ij} \in \{0, 1\}. \tag{5}$$

Here we have simply inserted the BASP probability and similarity information into the $p$-median model. The $k$ resulting beam angles represent the original angles better than any other set of $k$ beams in the candidate set $\mathcal{A}$ (according to the similarity and probability information).

## **Modeling Probability**

In order to apply the $p$-median model to the beam angle selection problem, we need to decide a measure of probability $\Omega(a)$ that captures the likelihood of using any given candidate angle $a \in \mathcal{A}$ in a treatment plan using $k < |\mathcal{A}|$ beams. Calculating the probabilities for each angle is a special case of sensitivity analysis, where we are trying to determine the effect or contribution of each beam angle in the optimal LP fluence solution.

There are only a small number of probability distributions suggested in the literature [2]. We discuss three approaches for modeling probability. The first suggestion is to generate a uniform probability distribution where each angle is equally likely to be used in a final plan, i.e., $\Omega(a) = \frac{1}{|\mathcal{A}|}, \forall a \in \mathcal{A}$. The second approach is called "balanced" probability model. The balanced probability function was introduced in [2]. The goal of the method is to remove dependence on a specific solver, by generating a balanced probability distribution and attempting to make it as uniform as possible. This distribution is obtained by using a special lexicographic optimization technique:

$$\text{lexmin}\,(z(x), \text{SORT}(x)), \tag{6}$$

where SORT is a mapping that reorders the components of $x$ in a non-increasing order. This optimization formulation is solved in an iterative manner.

Finally, we introduce a new approach "Total Weight Probability". The total weight probability function (TW) is obtained after solving for the optimal fluence $x_{(a,s)}$. The total weight probability is simply a sum of the fluency values over the sub-beams for each angle, calculated as

$$\Omega(a) = \sum_s x_{(a,s)}.$$

The essence of the TW method is the assumption that if this value is high for some angle in the fluence optimization problem, it indicates that the selected angle is important to maintaining the constraints of the fluence optimization problem, and hence is more likely to be used in a final $k$ beam plan.

## **Modeling Similarity (Difference)**

We measure similarity between beam angles based upon different characteristics that describe each beam angle. Angle location is perhaps the most obvious characteristic of any given angle and angle distance is a measurement of similarity

or difference between any two given beam angles. Letting $\alpha_{max} = \max\{\alpha_1, \alpha_2\}$ and $\alpha_{min} = \min\{\alpha_1, \alpha_2\}$, the angle distance function is as follows:

$$A(\alpha_1, \alpha_2) = \begin{cases} (\frac{\pi}{180°})[(360° - \alpha_{max}) + \alpha_{min}], & \text{if } (\alpha_{max} - \alpha_{min}) > 180° \\ (\frac{\pi}{180°})(\alpha_{max} - \alpha_{min}), & \text{otherwise.} \end{cases}$$

To derive more descriptive characteristics about each angle, we analyze how the angles are used in two situations: the way they are used in an optimal FMO solution and the way they are used in a uniform (unit) dose solution. The optimal FMO solution is the solution to (1) when we open all candidate beam angles $\mathcal{A}$. The uniform (unit) dose solution is obtained by assigning each sub-beam a fluence value of 1. Hence, the fluence vector $x$ in this case is simply a unit vector. Because of this, dose information for the unit solution may be obtained directly from the dose deposition matrix $A$. From these two different solutions we derive several statistics describing each beam angle. Each statistic defines a quality characteristic for each angle. There are several dose-based statistics for both unit and FMO solutions, and fluence-based statistics for the FMO solution only.

**Characteristic Vectors:** Once each statistic has been calculated we create vectors that contain the statistic values for each angle in observation. Allowing the notation $g_{ij}$ to represent the value of the $j$-th characteristic of angle $i$, we define characteristic vectors for each angle $a$ as follows:

$$c_a = (g_{a1}, g_{a2}, \ldots, g_{am}),$$

where $m$ is the total number of statistics we are considering. We measure $c_a$ by a norm $||c_a||$, and this norm induces the metric:

$$h(a_i, a_j) = ||c_{a_i} - c_{a_j}||.$$

Using the characteristic vectors allows us to measure an angle's effectiveness with respect to a wide variety of user-defined performance measurements, including measurements of the apparent effect of each angle on the quality of treatment.

## 4. Proposed Methodology

In order to apply the $p$-median model to the beam angle selection problem, we first must collect data about how each candidate angle is used in the unit (uniform) and optimal LP solutions, so that we may generate the characteristic vectors. We then apply preprocessing techniques to the data sets as well as the probability information. After this we are ready to use the $p$-median model with the chosen characteristic vectors and probability weights to obtain a pruned collection of angles. Finally we obtain an optimal fluence map using only the pruned angles, and evaluate the resulting plan using various measures of performance.

## Data Collection

In order to collect data about each beam angle in a candidate set $\mathcal{A}$, we analyze two different solutions to the fluency map optimization problem (FMO). In FMO, the objective function in [8] is modified so that $x$, the fluency vector, is the decision variable for a fixed candidate set of beam angles $\mathcal{A}$. We consider two different solutions to the FMO problem. The first solution is a unit solution: $x(a, s) = 1, \forall (a, s)$. Since the mapping $Ax$ determines the amount of dose deposited in the anatomy, dose-based statistics for the unit solution are obtained directly from $A$. We collect 19 attributes from the uniform dose solution: sum of dose deposited to all voxels, maximum dose and minimum dose on the PTV, average dose on the PTV, OAR, and NOR, to name a few. The second solution is an optimal linear programming solution to the FMO problem with an assumption that the FMO solution will provide enough information for finding good angles. We collect 29 statistical values (or attributes) from an LP solution: maximum dose and minimum dose on the PTV, average dose on the OAR, to name a few. Both dose-based and fluence-based statistics may be obtained from this solution. With these definitions of the statistics that we want to consider, we can alter what type of statistics are in the characteristic vector set $\mathbb{X}$ that we pass to the $p$-median model.

## Preprocessing

**Data Standardization:** Recall that in any data set $\mathbb{X}$, the rows correspond to the angles and the columns correspond to the statistics. In any such data matrix, the data in the columns may have different units, and even when they have the same units there may be much variation in the data scaling, sometimes even by several orders of magnitude. For

this reason, we standardize the data in each data set before any further calculation by finding the *z-score* of each data element in each column of $\mathbb{X}$. This generates a *z*-score vector, consisting of a value for each of the $j$ elements of the $i$-th column, given by

$$z_{ij} = \frac{x_{ij} - \hat{\mu}_i}{\hat{\sigma}_i}, \tag{7}$$

where $\hat{\mu}_i$ and $\hat{\sigma}_i$ are estimators of the mean and standard deviation of the $i$-th column, respectively. The original columns of $\mathbb{X}$ are replaced by their respective *z*-score vectors, generating a new standardized data set $\mathbb{X}'$. This helps to avoid the skewing of results by using improperly skewed input data.

**Principal Component Analysis (PCA):** Some statistics (or attributes) in the original sets can be highly correlated, due to the similarity of many of the calculations. We apply a statistical technique known as principal component analysis (PCA) to remedy this issue. PCA removes unnecessary correlation and shrinks the data set so that only the most important indicators are left for further analysis. This method refines the data so that the results are not improperly skewed due to highly correlated input data. This technique is a data reduction method often applied in data mining. In data mining it is used as a feature selection technique to reduce large data sets into the most significant features.

**Probability Scaling** The probability information must be scaled in some way as to balance the influence of the probability and similarity information in finding a set of beam angles. We begin with the modified dataset $\mathbb{X}'$ (after the *z*-score standardization and optional PCA technique have been performed) and find the non-zero distances (using the value of $d$ as used in the $p$-median objective function) between all pairs of the characteristic vectors. Defining $\hat{\mu}_d$ and $\hat{\sigma}_d$ as estimators of the mean and standard deviation of the non-zero distances, respectively, we compute the *z*-score vector of the single column of probability information ($Z^P$). The final probability vector is calculated by

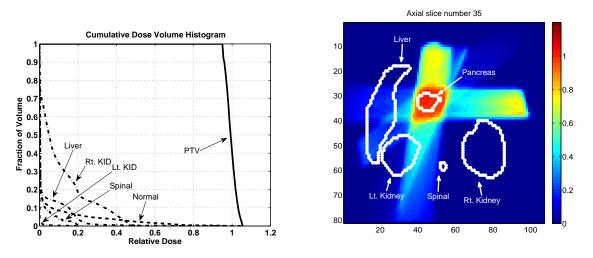$$\Omega_i = \hat{\mu}_d + (\hat{\sigma}_d \times Z_i^P).$$

The original probabilities are scattered using the original *z*-score information but are rescaled to fit within the range of the distances between the characteristic vectors. This ensures that there is a proper balance of influence between the probability and similarity information in the $p$-median model.

## 5. Numerical Results

We test our method on two sets of patient data: prostate cancer and a pancreas tumor. The prostate cancer data contains two main organs of interest: prostate and rectum. Using 4mm precision on the grids, the tumor region has 5,246 voxels, the rectum has 1,936 voxels, and the remaining healthy tissues have over 461,000 voxels on 38 CT slices. The pancreas tumor contains five organs of interest: pancreas, liver, left kidney, right kidney, and spinal cord. The voxel counts are 1,244, 50,391, 9,116, 5,920, and 489, respectively on 90 CT slices. We consider both 12 and 36 candidate beam angle set $\mathcal{A}$, i.e., there are 36 equally spaced beams ($\mathcal{A}_{36} = \{10i\pi/180 : i = 0, 1, 2, \ldots, 36\}$). In the prostate case we aim to select six beam angles from each candidate set. For the pancreatic tumor case, four angles will be selected for the final treatment. Once we obtain the resulting sets of beam angles from the $p$-median model, we evaluate the solutions using two visual performance measures of the final fluence map solutions: dose volume histogram (DVH) and radiation dose distribution plots. Details of these can be found in [8]. We noticed that the best quality plans (according to visual analysis) resulted from the same combination of parameters. This combination was using all statistic values with PCA and the TW probability distribution function. We also tested cases with 12 initial candidate beam angles versus 36 initial beam angles. We observed that increasing the candidate beam angles improved the solution quality. The DVH graph and a dose plot for the pancreatic tumor case with 36 initial beam angles are shown in Figures 1.

## 6. Conclusion

In conclusion, we observed good and consistent results with the proposed heuristic method for solving the IMRT beam angle selection problem under certain modeling conditions. In general, we make the following recommendations. First, we recommend using pre-processing data with PCA and using the TW probability distribution function. This combination obtained good results in many test cases and one of them was presented in this paper (the pancreas case with 36 candidate beams). We have also noticed that increasing the initial candidate beam angles improved the treatment plan quality. Increasing the number of candidate beams increases the time required to solve the fluence map optimization problem, so this recommendation should be followed with that in mind. A balance between an accurate

(a) Dose Volume Histogram on pancreas data with 36 initial candidate beam angles

(b) DAxial dose distribution plot slice on pancreas data with 36 initial candidate beam angles

Figure 1: Pancreas data with 36 initial candidate beam angles The final beams were (90,150,170,300)

representation of problem geometry and solution time is desired. Following these recommendations obtained good results on the tested patient data in our experiments.

# References

[1] D. M. Aleman, A. Kumar, R. K. Ahuja, H. E. Romeijn, and J. F . Dempsey, *Neighborhood search approaches to beam orientation optimization in i ntensity modulated radiation therapy treatment planning*, Journal of Global Optimization **42** (2008), 587–607.

[2] M. Ehrgott, A. Holder, and J. Reese, *Beam selection in radiotherapy design*, Tech. Report 95, Trinity University Mathematics, San Antonio, TX, 2005.

[3] A. Holder, *Radiotherapy treatment design and linear programming*, Operations Research and Health Care: A Handbook of Methods and Applications (M. Brandeau, F. Sainfort, and W. Pierskalla, eds.), Kluwer Academic Publishers, 2004.

[4] A. Holder and B. Salter, *A tutorial on radiation oncology and optimization*, Tutorials on Emerging Methodologies and Applications in Operations Research (H. Greenberg, ed.), Kluwer Academic Publishers, 2004.

[5] T. J. Jordan and P. C. Williams, *The design and performance characteristics of a multileaf collimator*, Physics in Medicine and Biology. **39** (1994), 231–251.

[6] T. Kalinowski, *Realization of intensity modulated radiation fields using multileaf collimators*, General Theory of Information Transfer and Combinatorics, Shannon Foundation, 2004.

[7] G. J. Lim, M. C. Ferris, S. J. Wright, D. M. Shepard, and M. A. Earl, *An optimization framework for conformal radiation treatment planning*, INFORMS Journal on Computing **19** (2007), no. 3, 366–380.

[8] G.J. Lim, J. Choi, and R. Mohan, *Iterative solution methods for beam angle and fluence map optimization in Intensity Modulated Radiation Therapy Planning*, OR Spectrum **30** (2008), no. 2, 289–309.

[9] C. ReVelle and R. Swain, *Central facilities location*, Geographical Analysis **2** (1970), 30–42.

[10] L. J. Verhey, *Principles of radiation physics*, Textbook of Radiation Oncology (S. A. Leibel and T. L. Phillips, eds.), W. B. Saunders Company, Philadelphia, 1998, pp. 91–114.