

VECTOR QUANTIZATION, GRAPH THEORY, AND IMRT DESIGN

JOSHUA G. REESE

ABSTRACT. Vector quantization is an important concept in the field of information theory and coding theory. The p -median problem is a classic graph theory problem with natural extensions to operations research facility location problems. We investigate some previously unrecognized connections between these problems, and ultimately show that the problem of designing an optimal vector quantizer on a discrete set \mathbb{X} is closely related to solving the p -median problem on a graph $G = (\mathbb{X}, E)$. We also demonstrate how vector quantization and its connection to the p -median problem have important applications in radiotherapy treatment design, specifically in the IMRT *beam selection problem*.

1. INTRODUCTION

Claude E. Shannon laid the groundwork of vector quantization (VQ) by assuming the existence of a distortion measure that would quantify the difference in quality between an input and a coded reproduction. He worked on VQ under the guise of *source coding subject to a fidelity criterion* [3].

The p -median problem is a classic graph theory problem with natural extensions in operations research to problems such as the facility location problem. This problem involves locating p facilities on a graph so that the sum of the weighted distances from the vertices of the graph to their nearest facility is minimized [1].

We begin by defining and describing these concepts in detail, and then we note some previously unrecognized connections between them. We then apply VQ, the p -median problem, and the connections between them to the problem of designing a feasible radiotherapy treatment design. A critical radiotherapy design question is how to choose a collection of angles from which to emit radiation into the tumor. It is important that these angles satisfy clinical restrictions and constraints on the amount of radiation deposited in different areas of the body. We show that vector quantization and the p -median problem are important tools for solving this problem, and conclude by introducing some new avenues of research that our results present.

2. VECTOR QUANTIZATION

2.1. Basic Definitions. The material in this section is treated exhaustively in [3], and interested readers are directed to this text. Vector quantization is a process for coding continuous or discrete vector signals subject to a fidelity criterion. It is often used to compress images or other data.

Date: April 27, 2005.

Key words and phrases. Graph Theory, p -Median, Vector Quantization, Radiotherapy.
Special thanks to Dr. Allen Holder, Department of Mathematics, Trinity University.

A *vector quantizer* \mathcal{Q} maps vectors from an input set of vectors \mathbb{X} into a finite set \mathcal{C} (called the *codebook*) containing N output points (called *codevectors* or *codewords*). Thus,

$$\mathcal{Q} : \mathbb{X} \rightarrow \mathcal{C},$$

where $\mathcal{C} = \{y_1, y_2, \dots, y_N\}$ and $y_i \in \mathbb{X}$ for each $i \in \mathcal{J} = \{1, 2, \dots, N\}$. The vector quantizer separates \mathbb{X} into N distinct regions called *cells*, denoted \mathbb{X}_i for $i \in \mathcal{J}$, and defined as

$$\mathbb{X}_i = \{x \in \mathbb{X} \mid \mathcal{Q}(x) = y_i\}.$$

These cells are disjoint and together include all elements of the input set. Thus we have,

$$\bigcup_i \mathbb{X}_i = \mathbb{X} \text{ and } \mathbb{X}_i \cap \mathbb{X}_j = \emptyset \text{ for } i \neq j.$$

So, these individual cells form a *partition* of \mathbb{X} , defined as

$$\mathbb{P}_{\mathcal{Q}} = \{\mathbb{X}_i \mid i \in \mathcal{J}\}.$$

The basic operation of a vector quantizer is to partition an input set of vectors into partition cells and select a codevector from each cell to serve as a representative for all of the vectors within that cell. Thus, vector quantizers may be separated into two separate processes, known as the *encoder* \mathcal{E} and *decoder* \mathcal{D} . The encoder's job is to examine each input vector and to determine which partition cell it should belong to. Thus, the encoder maps elements from \mathbb{X} into the set \mathcal{J} , which contains indices of the cells of \mathbb{X} . The decoder then chooses a vector from each numbered cell to serve as the codevector for that region. Thus, the decoder maps elements from \mathcal{J} into the codebook \mathcal{C} . We have,

$$\mathcal{E} : \mathbb{X} \rightarrow \mathcal{J} \text{ and } \mathcal{D} : \mathcal{J} \rightarrow \mathcal{C}.$$

As an example, let \mathbb{X} be a finite set of points on a city map. These points could be houses, schools, businesses, or any other location within the city. Consider an encoder that maps children's houses into partition cells, or school districts, each represented by a school district number. These school districts are disjoint subsets of \mathbb{X} . The decoder then maps the index number of a particular school district to a location within that district representing the location of the school. Depending on the design of the decoder, the quantizer might choose a location that already has a house, school or business built on it, or it could choose an unused location.

2.2. Representing VQ With Binary Selector Functions. We represent the operation of the encoder in terms of *selector functions* $S_i(x)$, defined as binary functions that indicate whether a vector x is in a given cell \mathbb{X}_i . Thus,

$$S_i(x) = \begin{cases} 1 & \text{if } x \in \mathbb{X}_i \\ 0 & \text{otherwise.} \end{cases}$$

Since the decoder creates a codebook by selecting a codevector from each cell \mathbb{X}_i , it is possible to create a selector function that describes the entire operation of the quantizer, as follows:

$$\lambda_{ij} = \begin{cases} 1 & \text{if } x_j \text{ is quantized as } x_i \\ 0 & \text{otherwise.} \end{cases}$$

Notice that the codebook is:

$$\mathcal{C} = \{x_i \mid \lambda_{ii} = 1\}.$$

2.3. Vector Quantizer Performance. We quantify the nonnegative cost of quantizing any input vector $x \in \mathbb{X}$ with the codevector $\mathcal{Q}(x)$ by measuring *distortion*, denoted by $d(x, \mathcal{Q}(x))$. The most common measurement of distortion between vectors is the Euclidean distance ,

$$d(x, \mathcal{Q}(x)) = \|x - \mathcal{Q}(x)\|.$$

The methods of quantifying distortion are numerous, however, and often depend on the particular application of VQ. For now we will simply assume that $d(x, \mathcal{Q}(x))$ is a distortion metric meaningful to its application.

We use this distortion measure to determine the overall performance of the vector quantizer. Vector quantization is inherently a statistical problem. One of its original uses was analog to digital signal conversion. For this reason, vector quantizers were, and still are, designed so that even though the input signal is not previously known, a suitable measure of expected distortion may be obtained. In order to work towards this measure of expected distortion, we first state the following definition.

Definition 2.1. A function $f_{\mathbb{X}}$ defined on \mathbb{X} is called a *joint probability density function (pdf)* if

$$f_{\mathbb{X}}(x) \geq 0, x \in \mathbb{X},$$

$$\int_{x \in \mathbb{X}} f_{\mathbb{X}}(x) dx = 1,$$

where the integral is taken to be a multiple integral over the k -dimensional space.

In the school district example, if \mathbb{X} corresponds to the set of points on a map, $f_{\mathbb{X}}(x)$ is the probability that a location x is a house with school-age children. There are many different factors that may determine probability, depending on the particular application of VQ.

If we model the quantizer as a random process, its performance corresponds to its expected distortion, which by definition is

$$D_{\mathcal{Q}} = E d(x, \mathcal{Q}(x)) = \int d(x, \mathcal{Q}(x)) f_{\mathbb{X}}(x) dx,$$

where $f_{\mathbb{X}}(x)$ is a joint probability density function (pdf) of the input vector space \mathbb{X} . Here we are evaluating a statistical average of the distortion measure over the entire input vector space. This measure is called the *average distortion* of the vector quantizer. If the input set of vectors has a discrete distribution (as opposed to a continuous distribution), then we have

$$D_{\mathcal{Q}} = E d(x, \mathcal{Q}(x)) = \sum_i f_{\mathbb{X}}(x_i) d(x_i, \mathcal{Q}(x_i)),$$

where each x_i is a value in \mathbb{X} that has nonzero probability. In terms of the selector function described in Section 2.2,

$$D_{\mathcal{Q}} = \sum_i \sum_j f_{\mathbb{X}}(x_j) d(x_i, x_j) \lambda_{ij}, \quad (2.1)$$

where $\lambda_{ij} = 1$ if and only if $\mathcal{Q}(x_j) = x_i$.

2.4. Optimal Vector Quantizer Design. The vector quantizer design problem is: Find a codebook \mathcal{C} specifying the decoder \mathcal{D} and a partition $\mathbb{P}_{\mathcal{Q}}$ specifying the encoder \mathcal{E} such that the average distortion $D_{\mathcal{Q}}$ is minimized. If we find such a codebook and partition, then the resulting encoder \mathcal{E} and decoder \mathcal{D} produce a *globally optimal* quantizer.

We now state three conditions that lead to algorithms for codebook improvement. These conditions are often called the *necessary conditions* for optimal vector quantizers. See [3] for more details and proofs of the following conditions.

- (1) **Partition Condition.** Given a codebook \mathcal{C} , a partition is optimal if and only if it assigns vectors to the codevectors y_i in a nearest neighbor fashion, i.e.,

$$\mathcal{Q}(x) = y_i \iff d(x, y_i) \leq d(x, y_j) \quad \forall j \in \mathcal{J}.$$

In other words, given a fixed codebook, it is both necessary and sufficient for optimality that the partition satisfy the Partition Condition.

- (2) **Codebook Condition.** Given a partition $\mathbb{P}_{\mathcal{Q}}$, a codebook is optimal if and only if it minimizes the distortion between vectors $x \in \mathbb{X}_i$ and the codevectors y_i , averaged over the probability distribution of x given that x lies in \mathbb{X}_i . Hence,

$$E[d(x, y_i) | x \in \mathbb{X}_i] \leq E[d(x, y) | x \in \mathbb{X}_i]$$

for all $y \in \mathbb{X}$. This condition is fairly intuitive, as it ensures that each codevector minimizes the expected distortion over its respective partition cell. This condition is both necessary and sufficient for producing an optimal codebook, given a fixed partition.

- (3) **Boundary Condition.** Let

$$\mathcal{B} = \{x \mid d(x, y_i) = d(x, y_j) \text{ for some } i \neq j\}$$

be the *boundary points* of the given partition. These points are equidistant from two codevectors and hence do not have a unique nearest neighbor. An optimal codebook for a given source distribution must assign these points to one and only one partition cell. In order to practically ensure that this condition holds, an effective tie-breaking scheme must be implemented when designing the encoder.

Returning to the school district example, the first condition ensures that an optimal partition, or school district plan, is one in which the houses are closer to that district's school than any other school. Likewise, the second condition guarantees that an optimal set of school locations given a set of school districts is one in which the schools minimize distortion, averaged over the probability distribution, between the school and each house in a given district. The third condition ensures that no house lies within two school districts at the same time.

2.5. Lloyd Algorithm. The Lloyd algorithm is useful for designing an effective quantizer. It is based on the fact that the Partition Condition and Codebook Condition allow us to obtain an optimal codebook given a fixed partition, and an optimal partition given a fixed codebook. The algorithm requires an initial, discrete set of vectors called the *training set*. If the set we are trying to quantize is discrete, then we may simply apply the Lloyd Algorithm to this set. If it is continuous, then we select a set of observations from it and use these as the training set for the Lloyd Algorithm. The algorithm begins with an initial codebook, and at every iteration

produces a new codebook that is guaranteed to either reduce or leave unchanged the overall distortion \mathcal{D} . The algorithm is a three step process, as follows:

Lloyd Algorithm

- (1) Set $m = 1$. Begin with an initial codebook \mathcal{C}_1 .
- (2) (*The Lloyd Iteration*)
 - Given the codebook \mathcal{C}_m , create partition cells by using the Partition Condition as follows:

$$R_i = \{v \mid d(v, y_i) < d(v, y_j), \forall j \neq i\}.$$
 If $d(v, y_i) = d(v, y_j)$ for any $j \neq i$, then break the tie by assigning v to the cell R_j for which j is smallest, thus satisfying the Boundary Condition.
 - Using the Codebook Condition, we create a new codebook \mathcal{C}_{m+1} , which is optimal for the partition $\mathbb{P}_m = \{R_i \mid 1 \leq i \leq N\}$.
- (3) Compute \mathcal{D} , the average distortion for \mathcal{C}_{m+1} , and if it has decreased less than δ , stop. Otherwise, set $m + 1 \rightarrow m$ and go to step 2.

Theorem 2.2 (Gersho and Gray [3]). *Each application of the Lloyd Iteration reduces or leaves unchanged the average distortion.*

Proof. From the necessary conditions for optimality, the Partition Condition implies that the first part of the Lloyd Iteration can only improve or leave unchanged the encoder for the given decoder. Similarly, the Codebook Condition implies that the second part can only improve or leave unchanged the decoder for the given encoder. Hence, the average distortion of the vector quantizer cannot increase. \square

Theorem 2.3 (Gersho and Gray [3]). *For a finite training set, the Lloyd Algorithm always produces a sequence of vector quantizers whose average distortions converges in a finite number of iterations.*

Proof. There are a finite number of ways to partition the input set into N subsets. For each partition, the Codebook Condition yields a codebook with a particular value of average distortion. The monotone nonincreasing average distortion implies that the algorithm cannot return in subsequent iterations to a partition yielding a higher average distortion. Hence, the average distortion of the sequence of vector quantizers produced must converge to a fixed value in a finite number of steps. \square

3. GRAPH THEORY AND DESIGNING OPTIMAL VQ CODEBOOKS

3.1. Introduction. The *p*-median problem has long been useful in solving operations research problems dealing with facility location. The problem is to choose a *p*-element subset of a graph such that the weighted sum of the distances between the vertices of the graph and their nearest element in the subset is minimized [1].

3.2. P-Median Problem. Let $G = (\mathbb{X}, E)$ be a complete, weighted and undirected graph, where \mathbb{X} is the set of vertices and E is the set of edges. Let

$$d_{ij} = [d(x_i, x_j)]$$

be the shortest distance matrix where the entry in the *i*th row and *j*th column contains the shortest distance between vertices x_i and x_j , according to some metric

d. Let $w(x_i)$ be the vertex weights. In the facility location problem, these are the costs of satisfying the demand at each vertex x_i . Let ξ_{ij} be an *allocation variable* such that

$$\xi_{ij} = \begin{cases} 1 & \text{if vertex } x_j \text{ is allocated to vertex } x_i \\ 0 & \text{otherwise.} \end{cases}$$

The integer optimization formulation of the p -median problem is:

$$\left. \begin{aligned} & \min \sum_i \sum_j w(x_j) d_{ij} \xi_{ij} \\ & \text{subject to:} \\ & \sum_i \xi_{ij} = 1 \text{ for } j = 1, \dots, n \\ & \sum_i \xi_{ii} = p \\ & \xi_{ij} \leq \xi_{ii} \quad \forall i, j = 1, \dots, n \\ & \xi_{ij} \in \{0, 1\}. \end{aligned} \right\} \quad (3.1)$$

The constraint $\sum_i \xi_{ij} = 1$ for $j = 1, \dots, n$ ensures that each vertex is allocated to one and only one element in the p -element subset. The second constraint, $\sum_i \xi_{ii} = p$, guarantees that there are p vertices allocated to themselves, which forces the cardinality of the p -median subset to be p . The last two constraints ensure that the allocation variables are either 0 or 1.

If ξ_{ij}^* is an optimal answer to the problem defined by (3.1), then the p -medians are the elements of:

$$\mathbb{X}_p = \{x_i \mid \xi_{ii}^* = 1\}.$$

3.3. Using the P-Median Problem to Design Optimal VQ Codebooks.

The following theorem shows an important connection between VQ and the p -median problem in that we may design an optimal VQ codebook by solving a particular p -median problem for a fixed $p > 0$.

Theorem 3.1. *Let \mathbb{X} be a finite set of vectors. Let $f_{\mathbb{X}}(x)$ be a pdf of \mathbb{X} , and $d(x_i, x_j)$ be a metric on \mathbb{X} . An optimal vector quantizer $\mathcal{Q}: \mathbb{X} \rightarrow \mathcal{C}$, $|\mathcal{C}| = p$, may be found by solving the p -median problem on the graph $G = (\mathbb{X}, E)$ with vertex weights $f_{\mathbb{X}}(x)$, obtaining an optimal solution $\mathbb{X}_p = \mathcal{C}$, where \mathcal{C} is the optimal codebook for \mathcal{Q} .*

Proof. Let $w(x_j) = f_{\mathbb{X}}(x_j), \forall j$, and let $d_{ij} = [d(x_i, x_j)]$. Then by solving (3.1), we obtain $[\xi_{ij}]$ such that

$$\sum_i \sum_j f_{\mathbb{X}}(x_j) d(x_i, x_j) \xi_{ij}$$

is minimized. The constraint $\sum_i \xi_{ij} = 1$ for $j = 1, \dots, n$ corresponds to the Boundary Condition for vector quantizers and ensures that each vertex is allocated to one and only one element in the p -element subset (codebook). The constraint $\sum_i \xi_{ii} = p$ ensures that the number of outputs is p , i.e. $|\mathbb{X}_p| = |\mathcal{C}| = p$. The last two constraints guarantee that the allocation variables, which correspond to the selector functions λ_{ij} in VQ, are either 0 or 1. Then, if $\lambda_{ij} = \xi_{ij}$, we have that the average distortion of the quantizer given by (2.1) is minimized. We conclude that $\mathcal{C} = \{x_i \mid \lambda_{ii} = 1\} = \mathbb{X}_p = \{x_i \mid \xi_{ii} = 1\}$ is an optimal codebook for \mathcal{Q} . \square

3.4. Solution Techniques. Kariv and Hakimi [6] showed that the problem of finding a p -median on a graph is NP-hard in p and n , even when the graph is planar, has maximum vertex degree 3, and all of the edges have length 1 and all vertex weights are 1. Polynomial algorithms exist, however, for the case when p is fixed.

Theorem 3.2. *Let $n = |\mathbb{X}|$. The worst-case complexity of the p -median problem on $G = (\mathbb{X}, E)$ for a fixed $p \in \mathbb{N}$ is $O(pn^{p+1})$.*

Proof. Before the algorithm begins, a shortest path distance matrix is created using Dijkstra's algorithm, containing the shortest weighted distances between all vertices of the graph. The complexity of this step is $O(n^2)$ [1]. The search then proceeds through every p -element subset of \mathbb{X} , requiring

$$\binom{n}{p} \leq n^p$$

calculations. For each p -element subset of \mathbb{X} , we must compare every $x \in \mathbb{X}$ to the p elements of the subset, to satisfy the nearest neighbor condition. This requires pn calculations. So, the overall complexity is no greater than

$$O(n^2 + (pn)n^p) = O(pn^{p+1}).$$

□

Given Theorems 3.1 and 3.2, the following corollary becomes apparent:

Corollary 3.3. *The problem of designing an optimal quantizer on a discrete set of data is polynomial.*

4. APPLICATIONS IN RADIOTHERAPY TREATMENT DESIGN

4.1. Introduction. Intensity Modulated Radiotherapy Treatment (IMRT) is a method of cancer treatment that directs ionizing beams of radiation through the anatomy and into a tumorous region. The goals of IMRT are to 1) direct a uniform amount of radiation into the tumor and 2) limit the amount of radiation that deposits in nearby critical and normal healthy tissues. The tumor should receive enough radiation to kill the cancer cells, and the critical structures should receive no more than a previously specified amount of radiation in order to reduce damage [4]. The radiation deposited per unit mass of tissue is called *absorbed dose* and is measured in *Gray (Gy)*, which is equal to one Joule of energy deposited per kilogram [5].

4.2. Therapeutic Advantage. Cancer cells are innately focused on reproducing and lack the effective capability to repair themselves. Consequently, cancer cells are killed when their DNA is damaged, while healthy cells are able to continue reproducing after minor damage by irradiation [2]. The difference between the radiation tolerance of cancer and normal cells is called the *therapeutic advantage*. IMRT treatments take advantage of this therapeutic advantage by delivering the full dose prescribed to the tumor in smaller daily doses [5]. This method of treatment kills the tumor cells, but gives enough time between doses for the normal tissue to heal. It also allows the dead cancer cells to be carried away by the healthy tissue around the tumor, preventing a collection of dead tissue where the tumor was located.

4.3. IMRT Treatment Method. Before the treatment takes place, physicians use a series of sequential CT scans to obtain a 3D representation of the tumor and any nearby critical structures. Physicians then indicate on each CT scan the location of these structures, and the remaining tissue is labeled as *normal tissue*. The *prescription* is then specified by the following variables [4]:

- TLB \rightarrow The lower bound for absorbed dose in the tumor.
- TUB \rightarrow The upper bound for absorbed dose in the tumor.
- CUB \rightarrow The upper bound for absorbed dose in the critical structures.
- NUB \rightarrow The upper bound for absorbed dose in the normal tissue.

Each critical structure is assigned a different upper bound that varies with the structure type and is determined by the susceptibility and severity of 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 [5].

After the physicians determine an IMRT plan that meets the constraints of the prescription, the patient is laid on a couch and immobilized. The device that emits the radiation is called a *linear accelerator*, or LINAC. The LINAC is mounted on a movable arm, called the *gantry*, that is able to make a complete 360 degree rotation around the patient. This allows physicians to deliver varying radiation doses from any particular collection of angles. The gantry rotates around a point called the *isocenter*, and patients are positioned so that the isocenter is located in the middle of the tumor.

LINACs are usually equipped with a *multileaf collimator*, 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 angles into *sub-beams*, also known as *pencil beams* [5].

4.4. Optimal IMRT Plan Design. The following is a brief discussion of the construction of an IMRT plan. See [2], [4] and [5] for further details.

There are three separate problems to solve when designing an IMRT plan. They are 1) determining which angles should be used to focus radiation on the patient, 2) finding a fluency pattern (a collection of exposure times) corresponding to these angles, and 3) choosing a delivery sequence that efficiently administers this treatment [2]. This paper investigates the first problem, which is known as the *beam selection problem*.

Linear programs are often used to determine *optimal* treatments, consisting of fluency values, or radiation exposure times, for each angle and sub-beam. Together these values define the fluency vector x , where $x_{(a,i)}$ is the exposure time for the i th sub-beam of angle a , for $a \in \mathcal{A} = \{i\pi/180^\circ \mid i = 0, 1, 2, \dots, 359\}$. The *dose matrix* A is an approximation of how radiation is deposited into each dose point in the anatomy. The units of A are Gy per unit time. The linear transformation $x \mapsto Ax$ maps the fluency vector into the anatomy and determines the amount of radiation deposited at every dose point in the anatomy [2]. The linearity is not just an assumption of the model—it is experimentally validated that the dose is deposited linearly [5]. We partition A into full-column sub-matrices A_T , A_C and A_N , containing the dose points corresponding to the tumor, critical structures and normal tissue, respectively. There are several linear programs in the literature for producing optimal IMRT plans, and the following is an example:

$$\begin{aligned}
& \min\{\omega \cdot l^T \alpha + u_C^T \beta + u_N^T \gamma\} \text{ subject to:} \\
TLB - L\alpha & \leq A_T x \leq TUB \\
& \qquad \qquad \qquad A_C x \leq CUB + U_C \beta \\
& \qquad \qquad \qquad A_N x \leq NUB + U_N \gamma \\
0 & \leq L\alpha \leq TLB \\
-CUB & \leq U_C \beta \\
0 & \leq U_N \gamma \\
0 & \leq x,
\end{aligned}$$

where ω is a weight determining the importance of meeting the lower bound on the tumor. The first three constraints are *elastic* and are allowed to vary with α , β and γ , respectively. These vectors represent how much the model is allowed to violate the radiation constraints on the tumor, critical structures and normal tissue, respectively. The matrices L , U_C , and U_N define how elasticity is measured, and each corresponds with a respective vector, l , u_C and u_N that decides how discrepancies are penalized. The decision values are x , α , β and γ [4].

This model has been implemented in a prototype treatment package named *Radiotherapy Optimal Design (RAD)*. *RAD* allows users to delineate a tumor, critical structures, and normal tissue on a 64 x 64 grid (similar to how a physician would indicate these structures on a CT scan). One can then assign prescription values to these structures and solve for the optimal plan corresponding to that problem instance [4].

4.5. IMRT Beam Selection Problem. Unfortunately, delivering a treatment from all 360 gantry angles is not clinically feasible. Current clinical constraints require that IMRT treatments consist of 7-9 gantry angles from which to deliver radiation. This is primarily because physicians limit the treatment time in order to reduce patient movement, which may result in treatment error. Thus, the beam selection problem is how to select p angles so that the subsequent p -angle treatment is effective. Solving integer programs designed to choose p angles exceeds calculation capabilities. An alternative strategy is to begin with an optimal treatment like the one described above and strategically prune angles to form an effective and clinically feasible treatment. In the next section we discuss how an optimal vector quantizer addresses this problem.

5. OPTIMAL VQ AND IMRT TREATMENT DESIGN

5.1. Characteristic Vectors. We define a k -dimensional *characteristic vector* c_a for a given angle a as

$$c_a = (g_1(a), g_2(a), \dots, g_k(a)),$$

where $g_i : \mathcal{A} \rightarrow \mathbb{R}$, $\mathcal{A} = \{i\pi/180^\circ \mid i = 0, 1, 2, \dots, 359\}$. The components $g_i(a)$ are different statistics or characteristics of the angle a as it is used in the optimal 360 degree plan. We measure c_a by a norm $\|c_a\|$, and this norm induces the metric:

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

Using the characteristic vectors allows us to measure an angle's effectiveness with respect to a wide variety of user-defined statistics, including statistics on what the apparent effect each angle has on the quality of treatment. Angle location is perhaps the most obvious characteristic of a given angle. Preliminary efforts at incorporating

VQ and IMRT have met with some success in using a pure angle distance metric [2]. This yields a one-dimensional characteristic vector, which contains angle location, i.e. $c_a = (\pi a/180^\circ)$. The method presented here allows us to incorporate numerous other factors, including how each angle contributes to the overall prescription and the absorbed dose within the anatomy.

If we let \mathbb{X} be a set of characteristic vectors, then using VQ on \mathbb{X} allows us to 1) group angles according to how similar their characteristics are, and 2) to select a codebook of p characteristic vectors to represent \mathbb{X} such that the average distortion of doing so is minimized. Therefore, the angles that correspond to these characteristic vectors satisfy the objective of the beam selection problem. These angles may or not be truly optimal however, as their optimality depends on what statistical data is present in the characteristic vectors. Hence, this technique is simply a heuristic to solving the overall problem.

5.2. Vector Quantization and the Beam Selection Problem. We first let $f(a_i)$ be a pdf that describes the statistical characteristics of the angles in the optimal 360 angle plan. Ultimately the process presented here is an attempt to solve a deterministic optimization problem via a statistical technique. This fact places this technique squarely in the realm of probability modelling.

The pdf should contain information regarding the likelihood of using any given angle. It is an open question as to how to create an effective pdf for this problem, although there are several theories. One example is to calculate how much absorbed dose each individual angle can deliver to the tumor before violating the constraints on the critical structures or normal tissue. Here we are measuring each angle's potential for treating the tumor, and the assumption is that the potential for a given angle translates into the probability of that angle being used. Until further results are gathered we will assume that $f(a_i)$ is an appropriate pdf for this particular application.

Once we have an appropriate pdf, we define a quantizer

$$Q : \mathbb{X} \rightarrow \mathcal{C},$$

where $|\mathcal{C}| = p$, and $\mathcal{C} \subseteq \mathbb{X}$. If we let $w(c_{a_i}) = f(a_i), \forall i$, then by Theorem (3.1) we can find the optimal codebook \mathcal{C} by solving the p -median problem on the graph $G = (\mathbb{X}, E)$. Furthermore, we have that this codebook contains the characteristic vectors associated with the angles that are solutions to the beam selection problem. This collection of p angles is a solution to the beam selection problem in the sense that there exists no other collection of p angles that can be selected with lower average distortion (based upon the information in the characteristic vector) of the original 360 angle plan.

6. NEW DIRECTIONS

Our results suggest new areas of research, and the following questions should be investigated.

- What angle statistics or characteristics should be included in the characteristic vector? How do we best select these characteristics to accurately depict angle differences according to what is actually happening in the anatomy?
- What is an appropriate probability density function for an optimal treatment consisting of 360 angles?

- How may we reduce the complexity of solving the p -median problem, and what are intelligent heuristics?
- Can we use the p -median problem to make statements about the Lloyd Algorithm?

REFERENCES

- [1] N. Christofides. *Graph Theory: An Algorithmic Approach*. Academic Press, 1975.
- [2] M. Ehrgott and A. Holder. Beam selection in radiotherapy design. Unpublished Draft, 2005.
- [3] A. Gersho and R. M. Gray. *Vector Quantization and Signal Compression*. Kluwer International Series in Engineering and Computer Science. Kluwer Academic Publishers, 1992.
- [4] A. Holder. Radiotherapy treatment design and linear programming. In M. Brandeau, F. Sainfort, and W. Pierskalla, editors, *Operations Research and Health Care: A Handbook of Methods and Applications*, chapter 29. Kluwer Academic Publishers, 2004.
- [5] A. Holder and B. Salter. A tutorial on radiation oncology and optimization. In H. Greenberg, editor, *Tutorials on Emerging Methodologies and Applications in Operations Research*, chapter 4. Kluwer Academic Publishers, 2004.
- [6] O. Kariv and S. Hakimi. An algorithmic approach to network location problems. ii: The p -medians. *SIAM Journal on Applied Mathematics*, 37(3):539–560, December 1979.
Current address: One Trinity Place 1817, San Antonio, Texas 78212
E-mail address: jreese@trinity.edu