Project Summary
A New Framework for Discrete Convexity and Its Applications in Radiotherapy
This project consists of two main research components, each from basic and applied research. The basic research component explores row-convexity (or row-concavity), a discrete analogue of a weaker definition of convexity (or concavity), which is useful for certain discrete optimization problems. Incorporating convexity measures developed in row-convexity, the applied research component focuses on algorithms for and comparisons of treatment plans of Intensity Modulated Arc Therapy (IMAT), a recently developed advanced treatment modality in radiation therapy. The compute-intensive tools and algorithms proposed in this research will be implemented with parallel computation in a high-throughput computing (HTC) environment (e.g., via free software provided by the Condor project at the University of Wisconsin-Madison).

Intellectual Merit: The IMAT problem considered in this project requires approximating a real-valued function defined on a subset of the 2-dimensional integer domain ($\mathbb{Z}^2$) with a sum of a fixed number ($K$) of step functions satisfying a convexity property in a given $x$ or $y$ dimension. The proposed row-convexity definition generalizes the above concept to $\mathbb{Z}^n$ and provides a new framework of discrete convexity that can be effectively used to improve the efficiency of a related algorithm. Although previous definitions of discrete convexity (i.e., $L$-convex and $M$-convex functions in matroid theory) were aimed at theoretical frameworks where fundamental theorems of continuous convexity could be extended to a discrete setting, they are too restrictive for the above purpose where convexity properties in a given direction is more of a concern than those on the entire domain. Exploring and identifying sets and functions that satisfy row-convexity, comparing and finding relationships and differences between row-convexity and previous definitions of discrete convexity, and proving optimality and separation theorems on this new platform are major goals in this research. Recent introductions of commercial IMAT systems (Elekta Volumetric Arc Therapy (VMAT) in 2008 and Varian Medical Systems Rapid Arc VMAT in 2008), which have been restricted to plans with $K = 1$ due to optimization challenges, resulted in great interest for deciding relative effectiveness of IMAT plans with different values of $K$. The applied research proposal plans to exploit row-convexity measures and a high throughput computing environment to develop an effective framework to implement this comparative effectiveness study and produce valuable results for the radiation community.

Broader Impacts: The theoretical results in row-convexity in this research will enhance the role of convexity in discrete optimization. The applied results of this research will help improve the quality and efficiency of radiation treatment planning and hence make a positive impact on the quality of life of cancer patients who receive IMRT treatments. This research will yield comparative effectiveness data, allowing clinicians and researchers to decide the better IMAT treatment plan with a high degree of certainty. The proposed research crosses boundaries of several traditional disciplines at the University of Wisconsin-Whitewater (UWW), and hence a wide audience of science and mathematics students at UWW will be impacted directly or indirectly through the implementation of this project. Some of the benefits to UWW are industry-oriented undergraduate research, undergraduate preparation for graduate programs in optimization, curricular impact, contribution of new research tools to solve compute-intensive problems, participation of students from underrepresented groups (ethnic minorities, disabled students, women, first generation college students), and undergraduate employment.
Project Description

1 Introduction

This project consists of two main research components, each from basic and applied research. The basic research component explores row-convexity (or row-concavity), a discrete analogue of a weaker definition of convexity (or concavity), which is useful for certain discrete optimization problems. Incorporating convexity measures developed in row-convexity, the applied research component focuses on algorithms for and comparisons of treatment plans of Intensity Modulated Arc Therapy (IMAT), a recently developed advanced treatment modality in radiation therapy. The compute-intensive tools and algorithms proposed in this research will be implemented with parallel computation in a high-throughput computing (HTC) environment (e.g., via free software provided by the Condor project at the University of Wisconsin-Madison). The specific goals for this proposal are:

1. to explore and establish a new theoretical framework based on row-convexity that can be useful for discrete optimization,
2. to use measures based on row-convexity (row-non-convexity measures) to predict the complexity and quality of an IMAT plan to be generated under a fixed set of treatment plan parameters,
3. to investigate net dose distributions of opposing radiation beams of an IMAT plan and use those results to introduce a new strategy for IMAT plan quality improvement,
4. to generate multiple IMAT plans for each tumor site under a fixed set of treatment plan parameters for given prostate, head and neck and pancreas cases, and check their comparative effectiveness with available plan comparators and study the correlations among clinical results and their predictions based on non-row-convexity measures,
5. to build a high-throughput computing cluster via software provided by the Condor project at the University of Wisconsin-Madison, and to implement and test the compute-intensive tools and algorithms proposed in this research on the Condor platform using parallel computation.

This proposal requires a multi-disciplinary approach that is best conducted within the framework of the Operations Research program. The PI has been actively collaborating on research (Gunawardena et al., 2006; Gunawardena and Meyer, 2007, 2008; Gunawardena et al., 2010) with an interdisciplinary team of investigators who are active in radiotherapy research with expertise in mathematical programming (Professor Robert Meyer, University of Wisconsin-Madison), medical physics and clinical radiation oncology (Professor Warren D. Souza, University of Maryland), and industrial engineering (Dr. Howard Zhang, University of Maryland). The PI plans to work with a student team at the University of Wisconsin-Whitewater and continue collaborations with the above experts for the development of the proposed frameworks.

2 Background and Significance

2.1 Discrete Convexity

Convex analysis (Fenchel, 1953; Valentine, 1964) is the branch of mathematics devoted to the study of properties of convex functions and convex sets, often with applications in convex minimization, a subdomain of continuous optimization theory (Mangasarian, 1969). A discrete analog of convex analysis called discrete convex analysis provides a useful contribution for discrete optimization. Unlike in traditional convex analysis which has a unifying theme, discrete convex analysis has been studied with multiple definitions motivated by various challenges in
different discrete optimization applications and theoretical frameworks (Hochbaum et al., 1992; Ibaraki and Katoh, 1988; Kindler, 1988; Miller, 1971). A detailed account of discrete convex analysis based on matroid theory can be found in Murota (2003).

While designing the IMAT algorithm presented in Gunawardena et al. (2010), the PI and his collaborators (Michael C. Ferris and Robert R. Meyer) introduced a convexity measure for approximating a real-valued function defined on a subset of the 2-dimensional integer domain ($\mathbb{Z}^2$) with a sum of a fixed number ($K$) of step functions satisfying a convexity property in a given $x$ or $y$ dimension. The proposed row-convexity definition in this project generalizes the above concept to $\mathbb{Z}^n$ and provides a framework of discrete convexity that can be used to improve the efficiency of a related algorithm. Although previous definitions of discrete convexity (i.e., $L$-convex and $M$-convex functions in matroid theory) were aimed at theoretical frameworks where fundamental theorems of continuous convexity could be extended to a discrete setting, they are too restrictive for the above purpose where convexity properties in a given direction is more of a concern than those on the entire domain. Exploring and identifying sets and functions that satisfy row-convexity, comparing and finding relationships and differences between row-convexity and previous definitions of discrete convexity, and proving optimality and separation theorems on this new platform are major goals and contributions in this research.

2.2 Radiation Therapy

A detailed historical account of the field of Radiation Therapy (since the discovery of x-rays by Wilhelm Roentgen in 1895 followed by Marie Curie’s discoveries of radium and Polonium) can be found in Mould (1993) and Bentel (1995). An exciting development in the 1960s was the introduction of high energy (megavoltage) treatment machines, known as linear accelerators or linacs. Such machines were capable of producing high energy, deeply penetrating beams, allowing for the very first time treatment of tumors deep inside the body without excessive damage to the overlying skin and other normal tissues. By the late 1980s, computer tomography (CT)-based treatment planning had become commonplace, allowing the development of 3-dimensional conformal radiotherapy (3DCRT). 3DCRT greatly improved the quality and delivery of radiotherapy by reducing dose to surrounding normal tissues.

The rapid advances in computer hardware and software in the 1990s led to the development of Intensity Modulated Radiation Therapy (IMRT) which represents arguably the most important revolution in Radiation Oncology. Like conventional 3DCRT, IMRT links CT scans to treatment planning software that allows the cancerous area to be visualized in three dimensions. However, regular 3DCRT and IMRT differ in how the pattern and volume of radiation delivered to the tumor is determined. In conventional 3DCRT, a uniform intensity is applied across the beam for each orientation. In IMRT, the physician designates specific doses of radiation (constraints) that the tumor and normal surrounding tissues should receive. A sophisticated computer algorithm, which is the main focus of this research, is used to develop an individualized plan to meet the constraints. This process is termed inverse treatment planning. While initially available at only a limited number of academic centers, IMRT has been quickly adopted by the Radiation Oncology community in recent years. The growth of IMRT was remarkable given that surveys found that less than a third of radiotherapists in the USA were using IMRT in 2003 (Mell et al., 2003), 75% in 2005 Mell et al. (2005), and all major clinics at present.

Since IMRT involves delivering radiation of variable or modulated spatial intensity with the help
of a multileaf collimator (MLC) (see Figure 1), it is computationally much more challenging than other forms of radiation delivery. IMRT provides the potential to deliver dose distributions that conform to the irregular shape of the target while ensuring a rapid falloff in dose away from the target (leading to sparing of nearby organs-at-risk (OARs)).

![Figure 1: Beam opening (aperture shape) formed by tungsten leaves of an MLC](image)

Traditionally, IMRT treatment plans are generated with a two-step process (Bortfeld et al. (1994); Chui et al. (1994); Galvin et al. (1993); Webb (1994); Xia and Verhey (1998); Boyer and Yu (1999); Gunawardena et al. (2006); Alber and Reemsten (2007); Boland et al. (2004); Romeijn et al. (2005, 2006); Ahuja and Hamacher (2005)). In the first step the face of the beam at each beam angle is conceptually partitioned into a 2-D grid of small beamlets (depending on the size of the beam aperture, the number of beamlets varies from 200-1000). The matrices of beamlet (also called pencil beams) intensities for the set of specified beam orientations (usually between 4 and 9 directions) are optimized. The resulting optimized intensity matrices or maps are then sent to a leaf-sequencer which determines a set of multi-leaf collimator (MLC) shapes and intensities for each beam angle whose combination approximates the intensity matrix at that beam angle. This set of aperture-intensity pairs is used for treatment delivery. This set of MLC aperture shapes and their corresponding weights (known clinically as monitor units) is referred to as a segmentation. In practice, Intensity Modulated Radiation Therapy can be delivered in two ways, namely, fixed gantry and rotational. Fixed-gantry IMRT (i.e., traditional) is achieved by delivering overlapping fields from a small number of fixed beam directions. Rotational IMRT is achieved by dynamically changing the aperture shapes as the gantry moves around the patient in one or more sweeps along an arc during radiation delivery. Intensity-modulated arc therapy (IMAT) was first proposed by Yu (1995) as an alternative rotational IMRT delivery technique to tomotherapy (Mackie et al. (1995); Carol et al. (1997)). In tomotherapy, a narrow MLC is used to control aperture shapes as the source rotates about the patient while the couch on which the patient rests is moved simultaneously. In contrast, IMAT is performed with a conventional linear accelerator, and the large set of tungsten leaves of the MLC is used to change the “shape” of the aperture as the gantry rotates during delivery (see Figure 1). The treatment is delivered along arcs with a single sweep or multiple sweeps, each with a start and stop position and the patient remains stationary during the delivery process. The MLC field shape changes continuously during gantry rotation from the beginning to the end of each arc. The multiple overlapping arcs provide multiple apertures at each angle which thereby achieve a modulated intensity distribution from each given delivery angle. A key advantage of IMAT over tomotherapy is that the delivery is achieved using a conventional linear accelerator and a conventional MLC. Therefore, IMAT treatments can be delivered using...
existing equipment in most radiation oncology departments. However, it should be noted that to deliver IMAT plans, the linear accelerator must be equipped with the capability for dynamic delivery. Elekta Volumetric Arc Therapy (VMAT) and Varian Medical Systems Rapid Arc VMAT are two commercially available IMAT systems that generally use only a single sweep through one arc for treatment delivery (Otto, 2008; Matuszak et al., 2010; Kuijper et al., 2010). Wang et al. (2008) has proposed Arc-modulated Radiation Therapy (AMAT), which approximates multiple sweeps by a single sweep, as another form of a single sweep IMAT. Previous comparison studies on IMRT and IMAT can be found in Wu et al. (2010); Bortfeld and Webb (2009); Bratengeier (2005). We use the mathematical optimization point of view to develop our Radiation Treatment Planning (RTP) research tools. Related RTP results with similar approaches can be found in ´Olafsson and Wright (2006); Shepard et al. (1999); Ferris et al. (2006b, 2003, 2006a); Hamacher and Küfer (2002).

2.3 High Throughput Computing via Condor

Condor is the product of the Condor Research Project at the University of Wisconsin-Madison (UW-Madison), and it was first installed as a production system in the UW-Madison Department of Computer Sciences nearly 15 years ago. The Condor software and complete documentation is freely available from the Condor project’s website at URL http://www.cs.wisc.edu/condor. Condor is a specialized workload management system for compute-intensive jobs on computer networks. Like other full-featured batch systems, Condor provides a job queuing mechanism, scheduling policy, priority scheme, resource monitoring, and resource management. Users submit their serial or parallel jobs to Condor, Condor places them into a queue, chooses when and where to run the jobs based upon a policy, carefully monitors their progress, and ultimately informs the user upon completion. This project’s use of Condor will demonstrate the potential of parallel platforms in implementing RTP systems to the commercial vendors. The proposal will also provide a capstone project that can showcase benefits of Condor for HTC to other faculty members and students on campus who are involved in compute intensive research. In the long run, the PI plans to coordinate efforts to expand the Condor computer cluster at UWW so that it will be available for faculty and students as a powerful research tool for their research projects. Some of the other suitable applications for Condor are sensitivity analyses, parametric studies or simulations to establish statistical confidence, neural-network training, Monte Carlo statistics, and a very wide variety of simulations, including computer hardware, scheduling policies, and communication protocols, annealing, even combustion-engine simulations, where 100 or even 1000 jobs are submitted to explore the entire parameter space.

3 Definitions and Preliminary Results

3.1 Row-Convex Sets and Functions

First we define a weaker definition of convexity called row-convexity in $\mathbb{R}^n$ using partial row-convexities. Let $x = (x(i) : i = 1, \ldots, n) \in \mathbb{R}^n$ and 1-norm of $x$, $|x|_1 = \sum_{i=1}^{n} |x(i)|$. A set $\Gamma \subseteq \mathbb{R}^n$ is partially row-convex in the $i$th dimension if

$$x, y \in \Gamma, |x - y|_1 = |x(i) - y(i)|, \alpha \in \mathbb{R}, 0 \leq \alpha \leq 1 \Rightarrow (1 - \alpha)x + \alpha y \in \Gamma.$$ 

$\Gamma$ is said to be row-convex if it is partially row-convex in every direction $i = 1, \ldots, n$. Clearly convex subsets in $\mathbb{R}^n$ are also row-convex. Figure 2 (a) shows a row-convex example in $\mathbb{R}^2$ which is not convex.
We restrict our row-convexity definition to $\mathbb{Z}^n$ as follows. A set $\Gamma \subseteq \mathbb{Z}^n$ is partially row-convex in the $i$th dimension if

$$x, y \in \Gamma, |x - y|_1 = |x(i) - y(i)|, \alpha \in R, 0 \leq \alpha \leq 1, (1 - \alpha)x + \alpha y \in \mathbb{Z} \Rightarrow (1 - \alpha)x + \alpha y \in \Gamma.$$

$\Gamma$ is row-convex if it is partially row-convex in every direction $i = 1, \ldots, n$. Figure 2 (b) shows a row-convex example in $\mathbb{Z}^2$.

We use a weaker version of quasi-convexity (Luenberger, 1968) to define row-convex functions as below. Let $\Gamma \subseteq \mathbb{R}^n$ be row-convex. A function $f : \Gamma \rightarrow \mathbb{R}$ is partially row-convex in the $i$th dimension if

$$x, y \in \Gamma, |x - y|_1 = |x(i) - y(i)|, \alpha \in R, 0 \leq \alpha \leq 1 \Rightarrow f((1 - \alpha)x + \alpha y) \leq \max(f(x), f(y)).$$

$f$ is said to be row-convex if it is partially row-convex in every direction $i = 1, \ldots, n$. Similarly we can define row-concave functions as follows. A function $f : \Gamma \rightarrow \mathbb{R}$ is partially row-concave in the $i$th dimension if

$$x, y \in \Gamma, |x - y|_1 = |x(i) - y(i)|, \alpha \in R, 0 \leq \alpha \leq 1 \Rightarrow f((1 - \alpha)x + \alpha y) \geq \min(f(x), f(y)).$$

$f$ is said to be row-concave if it is partially row-concave in every direction $i = 1, \ldots, n$.

Now we state the discrete versions of row-convex and row-concave functions as follows. Let $\Gamma \subseteq \mathbb{Z}^n$ be row-convex. A function $f : \Gamma \rightarrow \mathbb{R}$ is partially row-convex in the $i$th dimension if

$$x, y \in \Gamma, |x - y|_1 = |x(i) - y(i)|, \alpha \in R, 0 \leq \alpha \leq 1 \Rightarrow f((1 - \alpha)x + \alpha y) \leq \max(f(x), f(y)).$$

$f$ is said to be row-convex if it is partially row-convex in every direction $i = 1, \ldots, n$. A function $f : \Gamma \rightarrow \mathbb{R}$ is partially row-concave in the $i$th dimension if

$$x, y \in \Gamma, |x - y|_1 = |x(i) - y(i)|, \alpha \in R, 0 \leq \alpha \leq 1 \Rightarrow f((1 - \alpha)x + \alpha y) \geq \min(f(x), f(y)).$$

$f$ is said to be row-concave if it is partially row-concave in every direction $i = 1, \ldots, n$.

Observe that the role of partial row-convexities plays a similar role to that of partial derivatives in differentiation and row-convexity definitions can be used in both continuous and discrete settings.
3.2 Network Based IMAT Algorithm (NIMAT)

Let \([K]\) denote the set \(\{1, 2, \ldots, K\}\). For a given \(m \times n\) intensity map (a nonnegative real matrix) \(A\), a segmentation of \(K\) shapes for \(A\), is a set

\[
S := \{(\alpha_k, S_k) : k \in [K]\},
\]

where \(\alpha_k\) represents the beam-on-time (intensities) through the aperture \(S_k\) (represented by a binary matrix in which 1s correspond to open beamlets). In the segmentation problem in IMRT, we are given \(A\), a specified number of segments \(K\), and a positive scalar \(\gamma\) that weights the relative importance of the total radiation delivery time relative to the error. The goal is to find a segmentation \(S\) that minimizes

\[
\sum_{k \in [K]} \alpha_k + \gamma \text{err}(A, S)
\]

subject to machine-specific constraints on the achievable binary matrices \(S_k\) (shapes or apertures), where \(\text{err}(A, S)\) is a measure of the difference between the intensity map \(A\) and the delivery \(\sum_{k \in [K]} \alpha_k S_k\) from the given segmentation \(S\) given by

\[
\text{err}(A, S) = |A - \sum_{k \in [K]} \alpha_k S_k|_1*/|A|_1*
\]

where \(|D|_1^*\) denotes the 1-norm of a vector consisting of all the entries of the matrix \(D\). Although different machines may force different constraints on \(S_k\), two common constraints are as follows. Since the set of open beamlets in each row of the collimator is determined by the separation of a pair of opposing leaves of an MLC, any nonzero row of an achievable shape contains a single interval of ones, and this property is called the single-interval constraint. Another constraint needed for forming a deliverable shape is the inter-digitation constraint which does not allow opposing adjacent leaves to cross each other. A segmentation is called “exact” if \(\text{err}()\) is equal to zero. In IMRT, we try to achieve good (i.e., close to exact) segmentations for clinically desirable values of \(K\) (e.g., if the number of beam angles is fixed to 7, limit \(K\) to be between 15 and 20). In practice, for clinically used intensity maps, finding good segmentations is a very hard problem.

In IMAT, we assume that we deliver our plan along an arc using \(P\) equally spaced steps (angles \(\theta\)) along an arc and \(K\) sweeps through the arc. We assume that we are given the optimized intensity maps \(A^{\theta_1}, A^{\theta_2}, \ldots, A^{\theta_P}\) (each of which is an \(m \times n\) nonnegative real matrix) representing total beamlet doses that should be delivered at \(\theta_i\) over the \(K\) sweeps. Then, for each angle \(\theta\) we wish to generate a segmentation of \(K\) shapes for \(A^{\theta}\), (see Figure 3), i.e., the set

\[
S^{\theta} := \{(\alpha_k^{\theta}, S_k^{\theta}) : k = 1, \ldots, K\},
\]

where \(\alpha_k\) represents the beam-on-time (intensities) through the aperture \(S_k\).

Unlike fixed gantry IMRT, IMAT segmentations have to satisfy a maximum leaf movement constraint which forces the segmentations corresponding to two adjacent angles in an arc to have pairs of closely overlapping apertures. This is due to physical constraints on the speed at which a leaf can change its position. Thus in IMAT, we need to produce segmentations with \(K\) shapes for each intensity matrix \(A^{\theta}\) and to enforce leaf movement constraints between shapes on adjacent angles. The approximation problem may be stated as follows;

\[
\text{Minimize } \sum_{i \in [P]} \text{err}(A^{\theta_i}, S^{\theta_i})
\]
subject to machine-specific constraints on the achievable binary matrices $S_k^\theta, k \in [K]$, and the maximum leaf movement constraint for the pairs $S_k^\theta$ and $S_{k+1}^\theta$ for all $k \in [K]$ and $i \in [P-1]$. Such highly constrained segmentations result in large approximation errors $\text{err}(A, S)$ and generating segmentations that yield close to the minimum possible error is a very challenging problem. In practice, relatively few sweeps ($K$ is typically between 1 and 6) will be performed, so we also need to produce segmentations with a relatively small number of shapes $K$ while minimizing the errors of those segmentations.

In this subsection we sketch a network-based IMAT algorithm (presented at the INFORMS meeting, San Diego, 2009, a detailed description can be found in Gunawardena et al. (2010)). NIMAT provides for the combination of segmentations arising from a variety of segmentation procedures. For NIMAT, we develop three network models: the first (Network Model 1) addresses the issue of the ordering (in terms of sweep assignment) of the apertures for segmentations at adjacent angles, the second (Network Model 2) chooses which segmentation $j$ (a collection of $K$ intensities and apertures) to use at each angle $\theta_i$, and the third (Network Model 3) locally optimizes the shapes and intensities to reduce segmentation errors of a given feasible IMAT solution.

### 3.2.1 Network Model 1 (NM1): Assignment Model for Aperture Ordering Problem

For NM1, we consider the problem of morphing shape/shapes as we move along a sweep. To approach this problem in a computationally tractable manner, we consider the total leaf movement required to morph one shape $S_k$ into another $S_{k'}$. (To define this simple intuitive notion formally, we let $\text{leftmove}[i]$ be the amount of the $i$th left leaf movement between shape $S_k$ and $S_{k'}$, and $\text{rightmove}[i]$ be the amount of the $i$th right leaf movement between the two shapes. Then $TLM(S_k, S_{k'}) = \sum_{i \in [m]} (\text{leftmove}[i] + \text{rightmove}[i])$. Provided open beamlets overlap, it is the sum of the absolute differences between the binary matrices $S_k$ and $S_{k'}$.) Since this problem does not depend on intensities $\alpha$ but simply on the apertures, we drop the dependence of segmentations on $\alpha$ for ease of presentation in this subsection.

**Aperture Ordering**: Given two adjacent angles $\theta$ and $\phi$ with given segmentations $S^\theta = \{ S_k^\theta : k \in [K] \}$ and $S^\phi = \{ S_k^\phi : k \in [K] \}$, assign each shape $S_k^\phi$ uniquely to a shape $S_k^\theta$ to minimize the total overall leaf movement between all pairs of shapes.
This ordering problem may be solved as the following assignment problem:

\[
\min_x \sum_{k \in [K]} \sum_{k' \in [K]} x_{kk'} TLM(S_k^\theta, S_{k'}^\phi)
\]

subject to

\[
\sum_{k \in [K]} x_{kk'} = 1 \text{ for } k' \in [K]
\]

\[
\sum_{k' \in [K]} x_{kk'} = 1, \text{ for } k \in [K]
\]

\[
x_{kk'} \in \{0, 1\}
\]

Given the solution of the assignment problem, we know that if \(x_{kk'} = 1\), then for a particular sweep, if aperture \(k\) is used at angle \(\theta\), then aperture \(k'\) should be used at angle \(\phi\). In the discussion below we will refer to an aperture \(k'\) determined in this manner as the optimal successor of aperture \(k\) with respect to the segmentation pair \(S_k^\theta, S_k^\phi\). Typically \(\phi\) is the next angle in a sweep. The TLM minimum value from the above assignment problem to morph from \(S_k^\theta\) to \(S_{k+1}^\theta\) is denoted by \(c_{k,k+1}^\theta\), and will be used in the objective function of our second network model that determines which segmentation to use at each angle.

### 3.2.2 Network Model 2 (NM2): Shortest Path Model for Selection of Segmentations

We now assume that at each angle \(\theta_i\) we have a collection of \(R\) segmentations \(S_{j_i}^\theta\), \(j \in [R]\) generated from different segmentation procedures Gunawardena et al. (2010). The network model chooses the segmentation \(j\) to be used at each angle and uses as data:

1. the TLM minimum values \(c_{jj'}^{\theta_i}\) for each angle \(\theta_i\) and each angle-adjacent pair of segmentations \(j, j'\), and
2. the segmentation error \(e_{ij} = \text{err}(A_{\theta_i}^\theta, S_j^\theta)\) at each angle \(\theta_i\) for each segmentation \(S_j^\theta\), where \(A_{\theta_i}^\theta\) represents the intensity matrix at \(\theta_i\) and \(\text{err}()\) is given by (1).

Now we can build a network with nodes \((i, j)\) corresponding to angles \(\theta_i\) and segmentations \(S_j^\theta\) for \(i \in [P], j \in [R]\). Figure 4 shows the corresponding network.

![Figure 4: A shortest path network model with nodes \((i, j)\) corresponding to angles \(\theta_i\) and segmentation \(S_j^\theta\) for \(i \in [P], j \in [R]\)](image-url)

We assign a cost for each arc as follows. The cost of the arc between \((i, j)\) and \((i + 1, j')\) is equal to \(0.5 \ast (e_{ij} + e_{i+1,j'}) + \gamma c_{ij,j'}^{\theta_i}\), where \(\gamma\) is a nonnegative constant. The cost of the arc between \(\text{source}\) and \((1, j)\) is equal to \(0.5 \ast e_{ij}\) and the cost of the arc between \((P, j)\) and \(\text{sink}\) is \(0.5 \ast e_{ij}\). The
supply at the source is +1 and the demand at the sink is −1, and the problem is thus a shortest path problem. Note that the solutions produced by solving this network have low TLM but may not be feasible with respect to maximum leaf movement constraints. We have developed an efficient optimization procedure for obtaining such a feasible solution by modifying as little as possible the solution obtained from Network Model 2. This feasible solution is input to the network model 3 which optimizes segmentation error.

### 3.2.3 Network Model 3 (NM3): Shortest Path Model for Local Optimization of Error

We suppose we are given a feasible IMAT solution with $K$ sweeps for which the maximum leaf movement is at most $D$ between consecutive angles. The third network model attempts to modify the shape $S^k_{\theta}$ that is found on the $k$th sweep at angle $\theta_i$ to reduce the error measure $\text{err}(A^\theta_i, S^\theta_i)$. The constraints relevant to such a shape are the machine specific constraints $S \in FS$ and the maximum leaf movement constraints between this shape and the previous and next shapes in the sweep. The details can be found in Gunawardena et al. (2010).

### 3.2.4 NIMAT Algorithm

NIMAT is implemented in four phases.

**Phase 1 (Generating Initial Segmentation Pool)**

In Phase 1, we generate a pool of $R$ segmentations, each having $K$ shapes for each angle $\theta_i$, using a collection of segmentation algorithms or an atlas of apertures collected from previous tumor cases (D’Souza and Zhang, 2011) (It is better to use segmentations that produce relatively low errors and which are not similar to each other.).

**Phase 2 (Network models 1 and 2 applied to the relaxed problem)**

In Phase 2, we find an approximate solution to the IMAT optimization problem, relaxing the maximum leaf movement constraint, but finding solutions with minimum total leaf movement. We build NM2 using the ordering of the segments at each angle determined by the solutions of the assignment problems solved using NM1. We solve the shortest path NM2.

**Phase 3 (Finding a locally optimal IMAT solution minimizing the total leaf movement)**

In Phase 3, we use the set of segmentations from Phase 2 and solve a mixed integer program to find a feasible IMAT solution that satisfies the maximum leaf movement constraint by making minimum changes to the set of feasible shapes from Phase 2.

**Phase 4 (Finding a locally optimal IMAT solution minimizing the segmentation errors via NM3)**

In Phase 4, we use the set of feasible segmentations from Phase 3 and find a locally optimal IMAT solution (which minimizes the segmentation errors) by modifying shapes one at a time, with adjacent shapes fixed. We apply NM3 to each angle in an order generated using the row-non-convexity measures of the I-maps.

### 4 Proposed Research

The proposed research is described below under 5 subtopics, namely, 1. a theoretical framework for discrete convexity based on row-convexity, 2. row-non-convexity measure: a measure for segmentation complexity, 3. a new Strategy for IMAT plan quality improvement, 4. an I-Map segmentation module and comparative effectiveness of IMRT/IMAT plans, and 5. parallel computation and a high-throughput computing cluster via Condor.
4.1 A Theoretical Framework for Discrete Convexity Based on Row-convexity

In this research, we plan to explore and establish a new theoretical framework based on row-convexity defined in Section that can be useful for discrete optimization. Let $\Gamma \subseteq \mathbb{R}^n$ and a function $f : \Gamma \rightarrow \mathbb{R}$. When $\Gamma$ and $f$ are convex, a convex program may be expressed as follows:

$$\text{Minimize } f(x) \text{ subject to } x \in \Gamma.$$ 

Convex programs constitute a class of optimization problems that are tractable both theoretically and practically, with a firm theoretical base provided by traditional “convex analysis.” The tractability of convex programs is largely based on the following properties of the convex functions.

1. Local optimality or minimality guarantees global optimality. This implies that a global optimum can be found by a descent algorithm.
2. Duality, such as min-max relation theorem or the separation theorem, holds good.

When $\Gamma \subseteq \mathbb{Z}^n$, we have a discrete optimization problem. Except for network flow problems, discrete optimization problems arising from practical applications are generally difficult to solve efficiently. Previous discrete convexity results based on matroid theory Murota (2003) show that the tractability of the network flow problems is due to the optimality property of discrete convex functions. Still our understanding of convexity in discrete optimization seems to be only partial and exploring discrete convexity in different settings may lead to valuable results. We plan to compare our new framework with previous properties and theorems in discrete convexity (e.g., mid-point convexity, $L$ and $M$ convex functions in matroid theory, optimality and duality, conjugacy based on the Legendre-Fenchel transformation, separation between convex and concave functions, etc.). We focus on identifying and classifying sets and functions that are row-convex but do not satisfy previous properties.

4.2 Row-non-convexity Measure: a Measure for Segmentation Complexity

Here we propose to construct measures based on row-convexity (row-non-convexity measures) to predict the complexity and quality of an IMAT plan to be generated under a fixed set of treatment plan parameters. For our framework, our goal is to find segmentations that are solutions to the following error minimization problem.

**Error Minimization Problem (EMP):** For a given intensity matrix $A$ and a positive integer $K$, find a segmentation $S$ that minimizes the relative error over the set of apertures satisfying machine specific constraints:

$$\min_S \text{err}(A, S) \text{ such that } S = \{(\alpha_k, S_k) : k \in [K]\}, \text{ where } \alpha_k \in \mathbb{R}^+ \text{ and } S_k \in FS.$$ 

Here $FS$ represents the set of binary matrices (i.e., apertures) that satisfy the machine-specific constraints. Note that the variables of this problem are $\alpha_k$ and $S_k$. Since this problem is NP-hard Baatar et al. (2005), we have to rely on heuristics to generate approximate solutions. For a given $m \times n$ intensity map $A$ and a non-negative real number $\varepsilon < 1$, the projection map $P_\varepsilon(A)$ is the $m \times n$ binary matrix where the $(i, j)$th entry of $P_\varepsilon(A)$ is one if and only if the $(i, j)$th entry of $A$ is greater than $\varepsilon \max\{a_{ij}\}$. The projection $P_\varepsilon(A)$ is called row-convex if it satisfies the single-interval constraint. It can be shown that an intensity map $A$ is row-concave if $P_\varepsilon(A)$ is row-convex for all $\varepsilon > 0$. Thus a shape matrix $S$ that satisfies the single-interval constraint is an example of a partially row-concave or partially row-convex intensity map. It can be intuitively...
seen that row-non-convexity of the projections of $A$ results in higher segmentation errors in feasible IMAT solutions since the solution components (shapes) themselves must be partially row-convex. For a positive integer $K$, we will define the $K$th row-non-convexity measure ($\rho_K(A)$) ($K = 1$ case can be found in Gunawardena et al. (2010)) which will be used to identify the amount of row-non-convexity of an intensity map in advance. An error minimization problem of the form EMP is called unconstrained if $F S$ is the set of all $m \times n$ binary matrices $B$. Let $u\ell_K$ be $\min \{ |A - \sum_{k \in [K]} \alpha_k S_k|_1 / |A|_1 : S_k \in B, \alpha_k \in R^+ \}$ (the unconstrained solution). Let $F^*$ be all the binary matrices that satisfy the single-interval constraint and $\ell^*_K$ be $\min \{ |A - \sum_{k \in [K]} \alpha_k S_k|_1 / |A|_1 : S_k \in F^*, \alpha_k \in R^+ \}$ (the single-interval constrained solution). We define the $K$th row-non-convexity measure $\rho_K(A)$ of the intensity map $A$ as $\ell^*_K - u\ell_K$, the difference between the single-interval constrained solution and the unconstrained solution. Since $F^* \subseteq B$, it can be shown that $0 \leq \rho_K(A) \leq 1$. The prior knowledge of arc-segments that contain row-non-convexity intensity maps helps us focus effort on difficult to approximate intensity maps. In Gunawardena et al. (2010), using a sequential algorithm, we were able to calculate $\rho_1(A)$ for clinical cases, but such an algorithm is not feasible for $K > 1$. In this research, we plan to develop a parallel algorithm to calculate $\rho_K(A)$ for $K \geq 1$ and use these measures to predict the segmentation complexity of an IMAT plan at a pre-processing stage and to help clinicians to choose $K$ at the beginning without generating time consuming plans for different $K$. Figure 5 shows examples of clinical intensity maps $A$ and their row-non-convexity measures $\rho_1(A)$ for a prostate case for the angles $140^\circ$, and $180^\circ$. The examples given in the figure show that the maps with higher values of $\rho_1(A)$ are more irregular.

\begin{table}[h]
\centering
\begin{tabular}{c|c|c}
\hline
& $140^\circ$ & $180^\circ$
\hline
$\rho(A) = 0.0053$ & $\rho(A) = 0.0912$
\hline
\end{tabular}
\caption{Integer approximations of real-valued intensity maps $A$ with entries scaled between 0 and 100 for the angles $140^\circ$, and $180^\circ$ and their row-non-convexity measures $\rho(A)$ for a prostate case}
\end{table}

4.3 A New Strategy for IMAT Plan Quality Improvement

We plan to investigate net dose distributions of opposing radiation beams of an IMAT plan and use those results to introduce a new strategy for IMAT plan quality improvement. Let $A$ and $B$ be two $m \times n$ intensity maps delivered from opposing angles ($180^\circ$ apart) of an IMAT plan. Therefore the beam corresponding to the $(i, j)$th entry of $A$ is directly opposed to the $(i, n-j)$th entry of $B$. Let the actual doses delivered using segmentations through these beams be $\tilde{a}$ and $\tilde{b}$. If the delivered dose is higher(or lower) than the original dose of a beam, it will produce an overdose(or underdose). It seems that overdose from a beam can be neutralized by an underdose from its opposing beam. Instead of minimizing absolute errors $|a_{i,j} - \tilde{a}|$ and $|b_{i,n-j} - \tilde{b}|$, one may consider...
combining an opposing beam pair’s error as $|a_{i,j} - \bar{a} + \beta(b_{i,n-j} - \bar{b})|$, where $\beta$ is a positive scalar, at various stages of IMAT error optimization. The selection of the parameter $\beta$ may depend on the location and anatomy of a tumor. Although focusing on opposing beam pairs for error minimization is an intuitively natural strategy, it has not been discussed in IMAT literature. We plan to explore error optimization based on opposing beam pairs and produce results for plan quality comparison.

Figure 6: Present IMRT planning system

4.4 An I-Map Segmentation Module and Comparative Effectiveness of IMRT/IMAT plans
We plan to generate multiple IMRT/IMAT plans for each tumor site under a fixed set of treatment plan parameters for given prostate, head and neck and pancreas cases, and check their comparative effectiveness with available plan comparators and study the correlations among clinical results and their predictions based on row-non-convexity measures. We can represent a presently available commercial RTP system for IMRT/IMAT by the block diagram shown in Figure 6. Note that IMAT is generally available in a single sweep case (Varian Rapid Arc, Elekta VMAT) and, due to optimization complexity, multiple arc IMAT as covered in this investigation is still under research and development. Both IMRT and IMAT can be delivered with a conventional LINAC with the capability for dynamic delivery. According to Figure 6, a clinician chooses either IMRT or IMAT subjectively. Then a treatment planner chooses a set of beam angles manually (in the IMRT case) and goes through very time consuming steps in: (1) pencil beam dose calculation/optimization, (2) intensity map segmentation, and (3) aperture weight optimization and final dose calculation. A short term goal of our proposed research is to provide algorithms and decision support systems that will provide clinicians with comparative effectiveness measures that will assist in choosing between IMRT and IMAT for a given tumor case. Our long term goal is to improve (i.e., avoid human errors) the present RTP system by developing a fully automated RTP system as shown in Figure 7. In this direction, the proposed research will develop an I-Map segmentation module and a plan comparator for a future automated RTP system based on HTC. Our collaborators Prof. Warren D’Souza, Prof. Robert Meyer, and Dr. Howard Zhang are actively pursuing research (D’Souza et al., 2004, 2008; Ferris et al., 2006b) in automated beam angle selection and dose calculation/optimization. Their results can be incorporated with the results from this research to produce a fully automated RTP system as shown in Figure 7.

We plan to use Pinnacle software on Varian machines to generate intensity maps for our RTP plans. These maps will be generated at the University of Maryland through collaboration with Professor Warren D’Souza. The project will be conducted as a retrospective analysis of deidentified data. The top-down structure for plan generation through this system is given in Figure 8.

We propose to use 3 tumor sites, prostate, head/neck, and pancreas. At the top level (IMRT/IMAT) in Figure 8, we develop a user friendly GUI to receive input data (i.e., I-Maps, plan type, tumor
Pencil beam dose calculations for possible angles

Beam angle selection and pencil beam dose optimization

Intensity Map Segmentation (multiple IMRT/IMAT plans)

Aperture weight optimization and final OAR/PTV dose calculations

Plan Comparator

Figure 7: Future Fully Automated IMRT/IMAT Planning System

IMRT/IMAT
Input: I-Maps

IMRT

Prostate
Head/Neck
Pancreas

IMAT

Prostate
Head/Neck
Pancreas

Figure 8: The top-down structure of the IMRT/IMAT plan generation

identity) and deliver output. Depending on the plan type, IMRT/IMAT module provides patient data to the IMRT or IMAT module at the second level. At the third level, we develop 6 sub-modules (3 for IMRT, 3 for IMAT) to work with each type of tumor. These 3rd level IMRT sub-modules will be developed by using our DMM algorithm and the IMAT sub-modules will be based on our NIMAT algorithm. Both DMM and NIMAT frameworks have flexibility to allow us to incorporate different strategies and segment counts. Thus each 3rd level sub-module will be refined to the 4th level sub-modules (leaf arrows) which represent plans generated with different segment counts for segmentations (see Table 1 for an example). Those leaf sub-modules can be run simultaneously in a HTC platform and produce a collection of plans which is used as input for a plan evaluator.

For a given disease site, we use the IMRT/IMAT segmentation module to generate the IMRT and IMAT plans exemplified in Table 1. We have selected our parameters so that each row in Table 1 has comparable IMRT and IMAT plans with respect to segment count (the number of apertures). These plans correspond to the 4th level sub-modules in Figure 8 and will be used to decide the best treatment for a given disease site by using a plan evaluator. These results will be compared with the values of row-non-convexity measures ($\rho_K(A)$) and establish correlations between the evaluator results and $\rho$ values. We plan to obtain from our collaborators at the University of Maryland data for 10 prostate cases, 10 head and neck cases and 10 pancreas cases and use our system to make tentative conclusions regarding comparative effectiveness of IMRT/IMAT for the considered disease sites.

We propose to develop a web based decision support system (DSS) that includes a plan evaluator based on final dose distributions on the planning target volumes (PTVs) and OARs and a friendly graphical user interface (GUI) to analyze and select the best plan from the cases corresponding to Table 1. The utility of multi-criteria approaches with respect to radiation treatment planning has been the subject of previous studies (Gopal and Starkschall, 2002; Rosen et al., 2005; Zhang et al.,
Table 1: Sweep/ Segment count $K$ for IMRT plans with the number of beam angles $P = 5, 7, 9$ comparable to an IMAT plan with $P = 36$

<table>
<thead>
<tr>
<th>Sweep/Segment Count ($K$)</th>
<th>IMAT</th>
<th>IMRT</th>
</tr>
</thead>
<tbody>
<tr>
<td>$P = 36$</td>
<td>$P = 5$</td>
<td>$P = 7$</td>
</tr>
<tr>
<td>1</td>
<td>7</td>
<td>5</td>
</tr>
<tr>
<td>2</td>
<td>14</td>
<td>10</td>
</tr>
<tr>
<td>4</td>
<td>29</td>
<td>21</td>
</tr>
<tr>
<td>6</td>
<td>43</td>
<td>31</td>
</tr>
</tbody>
</table>

2006; Yu, 1997; Xing et al., 1999; Romeijn et al., 2004; Meyer et al., 2009). We plan to extend the framework given in Meyer et al. (2009) and implement the GUI using Java applets in HTML so that the DSS can be used through a web browser. The plan evaluator will consider cutoffs in terms of dose-volume levels for the most important OAR(s). Following that we will consider the level of overdose (or underdose) for the next most important OAR. Such importance assignments will be based on disease sites. For example, in the case of a pancreas case where the OARs are liver, kidneys, spinal cord and stomach, the kidneys would be considered the most important OAR due to their proximity to the pancreas followed by the spinal cord due its function. Quadratic penalties could also be used in this context if large deviations from targets are to be highly penalized. The GUI also includes the capability of using equivalent uniform dose (EUD) values (Emami et al., 1991; Wu et al., 2002; Aggarwal, 2001; Thieke et al., 2009) for plan comparison. We plan to make this GUI as general as possible so that it will be compatible with all major treatment planning systems. While the power of this GUI will be manifested in a multi-plan paradigm generated via an HTC environment, it will not preclude operators from using this in conjunction with multiple plans generated serially via a commercial planning system. We plan to develop our GUI so that multiple plans may be evaluated simultaneously using dose volume metrics.

4.5 Parallel Computation and a High-throughput Computing Cluster via Condor

Present RTP systems are implemented with sequential algorithms on a single machine and try to balance the tradeoff between planning time (usually limited to 2-4 hours) and plan quality. Extra complexities in writing parallel algorithms and using cluster computing may have prevented previous RTP systems from exploiting parallel platforms for reducing planning times and improving plan quality. The recent advances in parallel and grid computing made by commercial and academic projects (i.e., Condor pool support in Red Hat Linux Fedora systems, Microsoft Cloud Computing etc.) provide very cost effective platforms for parallel computation with multiple CPUs distributed across a network. We plan to implement and test the compute-intensive tools and algorithms proposed in this research on a Condor platform using parallel computation. This project will demonstrate the potential of parallel platforms in implementing RTP systems to the commercial vendors. The first phase of implementing a Condor cluster at the PI’s campus involves a computer lab with 28 Linux work stations. These Linux machines will be installed with Red Hat Fedora systems which will be able to simply join and create Condor pools. This cluster will contain software optimization tools (e.g., GAMS/CPLEX, MATLAB) needed to implement the proposed algorithms.
5 Time Line and Management Plan
The proposed project is planned to be completed in a three year period as shown in the following table. The PI plans to guide an interdisciplinary (Computer Science, Management Computer Systems, and Mathematics) team of 4 undergraduate students at the University of Wisconsin-Whitewater to complete the proposed tasks.

<table>
<thead>
<tr>
<th>Year 1</th>
<th>Year 2</th>
<th>Year 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Conduct basic research in discrete convexity</td>
<td>Test the segmentation module for efficiency and robustness.</td>
<td>Analyze the row-convexity and RTP results.</td>
</tr>
<tr>
<td>Develop and implement HTC parallel algorithms.</td>
<td></td>
<td>Extend the module to other tumor sites.</td>
</tr>
<tr>
<td>Preliminary results with one specific site.</td>
<td></td>
<td>Develop and implement a plan evaluator.</td>
</tr>
</tbody>
</table>

6 Summary: Significance of proposed work
The proposed work is aimed at contributing important new results in both basic and applied research.

6.1 Intellectual Merit
This proposal fits well with the PI’s strong interdisciplinary background in Mathematics, and Computer Science. It introduces row-convexity as a new framework for discrete convexity and provides row-non-convexity measures that can be used to predict the segmentation complexity in IMRT/IMAT. The PI expects that these new convexity tools will also be useful in other similar applications where convexity properties in a given direction are more of a concern than those on the entire domain. IMRT brings the advantages of radiation therapy to a wider range of cancer patients, including patients with difficult-to-reach tumors or tumors located close to vital organs. The results produced in this research will improve the present IMRT framework in many ways and strengthen its theoretical foundation with new optimization results. The short term contribution of our proposed research is to provide clinicians with tools to help them decide the best choice among IMRT/IMAT possible plans at an early stage of RTP. The present RTP system shown in Figure 6 is vulnerable to errors due to the planner’s intervention at various stages. Although a fully automated system as shown in Figure 7 is the answer, highly time consuming optimization steps prevent its implementation. Our HTC approach is feasible and cost effective for this long term goal.

6.2 Broader Impact
The theoretical results in row-convexity in this research will enhance the role of convexity in discrete optimization. The applied results of this research will help improve the quality and efficiency of radiation treatment planning and hence make a positive impact on the quality of life of cancer patients who receive radiation treatments. The proposed research crosses boundaries of several traditional disciplines at the University of Wisconsin-Whitewater (UWW), and hence a wide audience of science and mathematics students at UWW will be impacted directly or indirectly through the implementation of this project. Some of the benefits to UWW are industry oriented undergraduate research, undergraduate preparation for graduate programs in optimization, curricular impact, contribution of new research tools to solve compute-intensive problems, participation of students from underrepresented groups (ethnic minorities, disabled students, women, first generation college students), and undergraduate employment.
References cited


Fenchel, W., 1953: *Convex Cones, Sets and Functions*. Lecture Notes, Princeton University, Armed Services Technical Information Services, AD Number 22695.


