2009

# A Direct Compensator Profile Optimization Approach For Intensity Modulated Radiation Treatment Planning

Kevin Erhart
*University of Central Florida*

# A DIRECT COMPENSATOR PROFILE OPTIMIZATION APPROACH FOR INTENSITY MODULATED RADIATION TREATMENT PLANNING

by

KEVIN J. ERHART
B.S. University of Central Florida, 2004
M.S. University of Central Florida, 2006

A dissertation submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy
in the Department of Mechanical, Materials, and Aerospace Engineering
in the College of Engineering and Computer Science
at the University of Central Florida
Orlando, Florida

Summer Term
2009

Major Professor: Eduardo A. Divo

# ABSTRACT

Radiation therapy accounts for treatment of over one million cancer patients each year in the United States alone, and its use will continue to grow rapidly in the coming years. Recently, many important advancements have been developed that greatly improve the outcomes and effectiveness of this treatment technique, the most notable being Intensity Modulated Radiation Therapy (IMRT). IMRT is a sophisticated treatment technique where the radiation dose is conformed to the tumor volume, thereby sparing nearby healthy tissue from excessive radiation dose. While IMRT is a valuable tool in the planning of radiation treatments, it is not without its difficulties. This research has created, developed, and tested an innovative approach to IMRT treatment planning, coined Direct Compensator Profile Optimization (DCPO), which is shown to eliminate many of the difficulties typically associated with IMRT planning and delivery using solid compensator based treatment. The major innovation of this technique is that it is a direct delivery parameter optimization approach which has adopted a parameterized surface representation using Non-Uniform Rational B-Splines (NURBs) to replace the conventional beamlet weight optimization approach. This new approach brings with it three key advantages: 1) a reduced number of parameters to optimize, reducing the difficulty of numerical optimization; 2) the ability to ensure complete equivalence of planned and actual manufactured compensators; and 3) direct inclusion of delivery device effects during planning with no performance penalties, eliminating the degrading fluence-to-delivery

parameter conversion process. Detailed research into the affects of the DCPO approach on IMRT planning has been completed and a thorough analysis of the developments is provided herein. This research includes a complete description of the DCPO surface representation scheme, inverse planning process, as well as quantification of the manufacturing constraint control procedure. Results are presented which demonstrate the performance and innovation offered by this new approach and show that the resulting compensator shapes can be manufactured to nearly 100 percent of the designed shape.

*To Marisa, Madison, and the rest of my family and friends whose love and support provide the real significance of this accomplishment*

# ACKNOWLEDGMENTS

Special thanks to:

Eduardo Divo and Alain Kassab for introducing me to numerical modeling and inspiring my love of research.

Salvadore Gerace for providing support in several areas throughout this entire effort.

.decimal, Inc. for providing the inspiration and funding for portions of this research effort.

# TABLE OF CONTENTS

# LIST OF FIGURES

# LIST OF TABLES

# CHAPTER 1
# INTRODUCTION

Radiation therapy is a widely used and highly effective technique for the treatment of cancer, however the commissioning and delivery of a course of external beam radiation is a complex process subject to numerous challenges. It is the intent of this research to develop new innovations that act to improve both the planning and delivery of this cancer treatment technique. Specifically, this research develops a new approach for solid compensator intensity modulated radiation therapy planning coined, *Direct Compensator Profile Optimization* (DCPO). The following research report will detail the development and verification of this unique approach, as well as demonstrate its effectiveness by comparison to current state-of-the-art planning techniques.

In order to understand the benefits and implications of the Direct Compensator Profile Optimization approach, a reasonable understanding of the field of radiation therapy is needed. Therefore, Chapters 2 & 3 provide a thorough discussion of the history and relevant background information in the area of radiation therapy. Chapter 2 is devoted to providing a short history of radiation therapy. This history should bring to light the competing goals of radiation treatment while also highlighting numerous improvements that have led to development of the latest radiation therapy procedures. Appropriately, Chapter 3 will then complete the groundwork with a discussion of Intensity Modulated Radiation Therapy (IMRT), which is an advanced delivery and planning technique used in modern radiation

treatment. It is intended that these two chapters are detailed enough so that the remainder of this research may be followed by those without existing experience in the field of radiation treatment planning. Following this background information, the specific details of this new approach will be presented in Chapter 4, including a concise hypothesis statement. As this work offers improvements in several different areas which combine to provide a unified approach that simplifies and improves both the planning and delivery of radiation therapy, Chapter 4 is divided into several appropriate sections. Each of these sections thoroughly details a particular research development and clearly demonstrates its place within the state-of-the-art of the field. After this description of the research methodology, Chapter 5 continues the research developments by providing details of the initial performance testing process. This initial testing provides the data necessary for a thorough discussion and analysis of the competing schemes outlined in Chapter 4, so that the most appropriate approaches may be identified and adopted. Presentation and discussion of the manufacturing constraint determination process is then undertaken and finally Chapter 5 concludes with a concise description of the final Direct Compensator Profile Optimization process. The benefits of DCPO are then quantified by presentation of several example problems in Chapter 6. These examples demonstrate the application of these techniques within the context of a basic radiation treatment planning system, developed specifically for this work. These examples include comparison to current state-of-the-art planning techniques for realistic patient geometries. This report then concludes with a thorough discussion of the observations and findings of this research effort in Chapter 7, including final remarks regarding avenues of exploration for future research within the field.

# CHAPTER 2
# HISTORY OF RADIATION THERAPY

External beam radiation has been used as a treatment for malignant disease almost since the discovery of the x-ray by Roentgen in 1895, with clinical treatment tests well underway within one year of this discovery. Radiation therapy has progressed a long way from these initial treatments and today the process of treatment involves determination of a well defined treatment plan along with finely tuned delivery procedures. This is not to say that the process is perfect as many patients still experience undesired complications and side effects as a result of radiation treatment. In order to appreciate the difficulties of radiation treatment and to arrive at a point where a description of the latest techniques is appropriate, a short history of external beam radiation therapy is necessary. As the purpose of this history is to provide a practical understanding of radiation therapy, an example patient and treatment will be used throughout the chapter. This example will involve treatment of a spherical abdominal tumor as this is representative of a very simple patient, yet will prove useful to facilitate a practical and application driven understanding of this history.

The earliest forms of radiation treatment typically involved simple exposure to a fixed strength beam, often to a wide area of the patient's body. While these treatments did provide some benefit to patients, the low levels of tumor dose were often insufficient to offer a complete cure to most patients. Additionally, many researchers in the field quickly learned the downside of such treatments, as many patients (and the researchers themselves) often

suffered from severe radiation poisoning or burns. Researchers and oncologists thus realized that they could only provide a useful treatment by shielding normal tissue from exposure as much as possible, while at the same time providing sufficient dose to destroy the tumor cells (this goal forms the foundation of the field of radiation therapy as a curative cancer treatment). Early forms of shielding often utilized lead sheets or plates laid directly on the patient, a practice which still exists for certain types of treatment today.

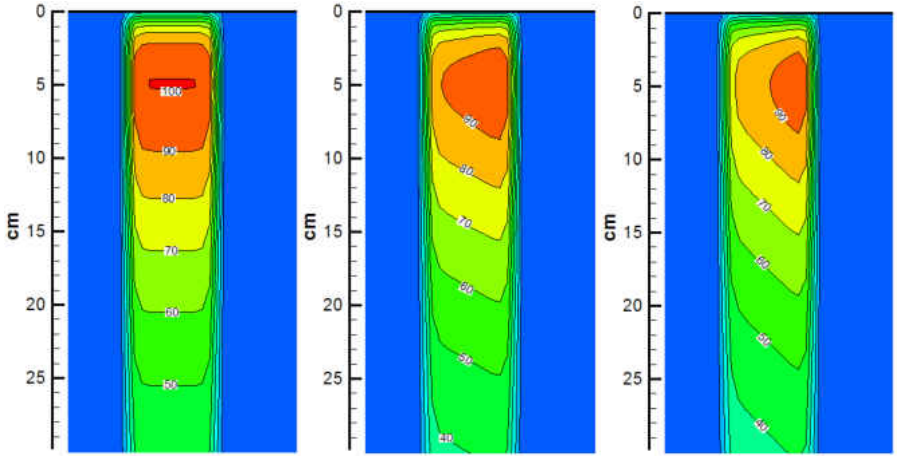
These shielding practices did allow for simple, two-dimensional control of the radiation dose, clearly limiting radiation exposure for much of the adjacent area and a majority of the body. However, this type of "shaping" of the radiation field does not account for the variation of dose with depth inside the patient. Using the abdominal tumor example, suppose that treatment is attempted using a single radiation beam placed directly above the patient, with the patient laying flat on their back. When viewing down the beam's axis (viewing the patient from above) it is clear that a lead shield with a circular hole above the tumor will limit the dose to the desired spherical tumor region as much as possible. However, shifting to a view perpendicular to the beam's axis (viewing the patient from the side) and looking at a two-dimensional slice through the center of the tumor, the dose field shown in Figure 2.1a is observed. Such a figure is referred to as an isodose (constant dose) plot and it illustrates the variation of the delivered dose with depth inside the patient's body. It should be noted that radiation dose fields such as this are generally displayed in terms of relative dose and are normalized so that a value of 100% occurs at the location of maximum dose (or at the isocenter, as will be seen later for multiple beam treatments). Combining these two views shows that the actual three-dimensional dose is being confined to a cylindrical portion of

the patient's body (although the tumor is spherical), the tumor dose is highly non-uniform, and depending upon the location (depth) of the tumor, the maximum dose is not necessarily within the tumor region. This results in a relatively poor radiation treatment as normal tissue receives significant dose and the tumor coverage is difficult to control. Typically, the only means for controlling this depth dose distribution (and therefore the tumor dose in this example) is to vary the energy (quality) of the radiation beam. Unfortunately, this offers little control over the general shape and in practice really only allows adjustment of the location (depth) of maximum dose. Therefore, another means needs to be found if the dose is to be further "shaped" to the tumor volume.

The next historical progression came as an attempt to alter the dose distribution with depth and involved the use of beam modifying devices; often metallic wedges. The dose distributions which can be achieved using two common wedge designs are shown in Figures 2.1b & 2.1c. While these modifying devices do increase dose control to some degree and prove useful for certain tumor shapes and locations, in the hypothetical spherical tumor example they provide little solution improvement because control over the depth of maximum dose is still not possible.

The next advancements came as researchers found that significant dose shaping could be achieved by using multiple treatment beams. The most common initial attempts utilized parallel opposed beams. Simply put, this means using pairs of beams which enter the patient exactly opposite of one another. A sample dose field from four such equally spaced beams (two horizontal and two vertical) is shown in Figure 2.2. Inspection of this figure shows that such a treatment allows the high dose region to be confined to a small box shaped

area around the tumor. An important observation regarding the dose within this box is the relative uniformity of dose that is achieved within the region, meaning that the majority of the tumor will receive an almost equal dose (assuming the entire tumor is within this region). This achievement was a monumental step forward and practitioners soon began to vary the treatment angles and number of beams in order to achieve even more conformal dose distributions.




(a) Open Beam (b) $30^o$ Wedged Beam (c) $45^o$ Wedged Beam

Figure 2.1: Isodose Contours for Single Beam Treatments

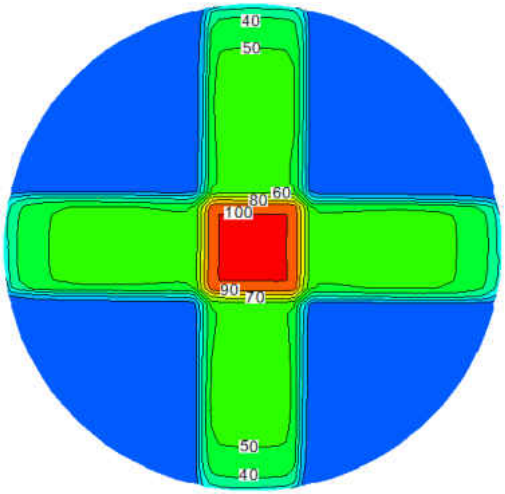



Figure 2.2: Relative Dose Field for Four Parallel Opposed Beams

Despite all of these advancements, the problem of conforming dose to a simple spherical tumor, has still not been fully addressed, as parallel opposed beams are generally limited to providing high dose to geometric regions with sharp, crisp corners such as rectangles, diamonds, and hexagons (10-12 beams would need to be used to begin to achieve a roughly circular shape, however this large number of beams greatly increases the time and expense of the radiation treatment). Two solutions were developed to solve this deficiency: rotational (arc) therapy and three-dimensional conformal radiation therapy (3DCRT). Both techniques remove a major downside of the existing parallel opposed beam treatments: the presence of significant overlapping of the beams outside the tumor (target) volume. Removing these overlapping regions allows the magnitude of the normal tissue dose to be significantly reduced, although this comes at the expense of spreading dose over a wider range of the nearby normal tissue. This is generally a very worthwhile trade-off and reduced clinical complications are usually observed for such treatments. The idea behind the first solution, rotational therapy, is simple: treat from every possible angle. This became possible with the development of electronically controlled delivery machines (a discussion of the delivery technology will be undertaken shortly). Once the delivery mechanism could be precisely controlled, it became possible to rotate a uniform strength beam around a stationary patient, thereby creating a dose field which contains a nearly cylindrical region of uniform high dose at the rotational center (isocenter). A schematic representation of this high dose region is shown in Figure 2.3. Additionally, by controlling the size and shape of the treatment beam this region can be made larger or smaller and can approach a spherical shape (using a round rather than rectangular beam opening).

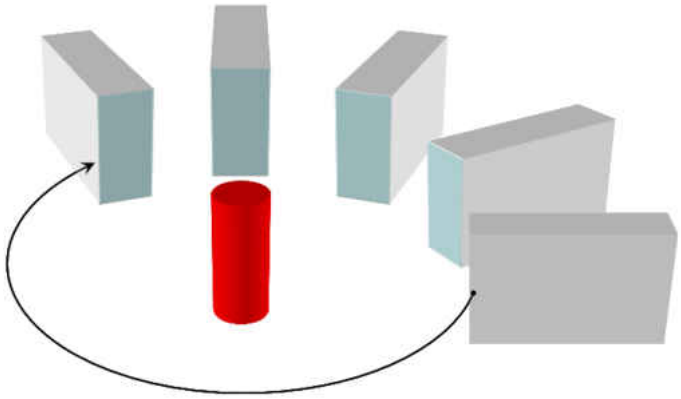


Figure 2.3: Sketch of High Dose Region for Standard Rotational Therapy

At first glance, rotational therapy may seem like the most ideal solution to conforming high dose regions and spreading normal tissue dose, but 3DCRT incorporates an extra mechanism of control that provides the ability to achieve more complicated dose conformation. By adding control over the size and shape of each treatment field, 3DCRT allows for exceptional dose conformity without the need for a more sophisticated delivery mechanism. This process uses a small, odd number of fixed angle treatment beams (often 3, 5, or 7) in order to allow more specific control of the shaping parameters as compared to rotational therapy. The process of 3DCRT is illustrated in Figure 2.4. This figure shows that each beam is shaped specifically according the geometric projection of the tumor in the beam axis direction and results in a beam that is reasonably well shaped to the tumor volume. Once again it should be apparent that such shaping is based only on a two-dimensional view and does not account for the variation of the dose with depth inside the patient. However, shaping each beam to the "correct" geometric configuration confines the overlapping region of the multiple beams much more closely to the actual tumor volume, thereby achieving a nearly uniform high dose in that volume. Application of this process was a major milestone in the history of external beam

radiation therapy, but required an experienced and skilled physicist/dosimetrist (and a lot of time) to develop quality treatment plans as the effects of various angles, shaping parameters, beam weights, and also wedges had to be assessed "by hand". Advancements in computer technology soon allowed this tedious planning process to be aided by specially designed computerized treatment planning systems (TPS). However, adjustment of the parameters and evaluation of the resulting plans was still largely the task of the cancer treatment team.

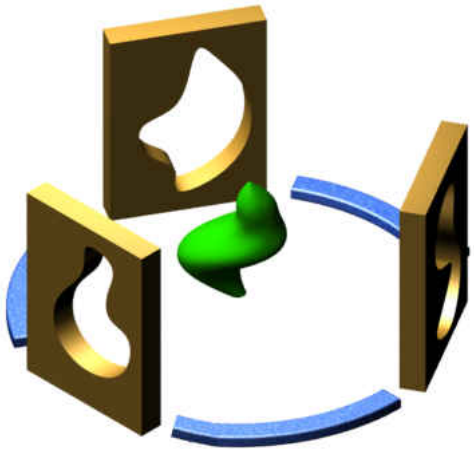


Figure 2.4: Three-Dimensional Conformal Therapy

The benefits of computerized planning were astonishing and numerous researchers immediately attempted to improve their treatment plans by developing automated numerical optimization procedures for determining planning parameters. It was then discovered that even greater plan control could be gained by allowing an extra degree of freedom (variability) within the numerical optimization procedure: the variation of intensity within a single treatment beam. This allowed the elimination of wedges and missing tissue compensators and was the start of the age of Intensity Modulated Radiation Therapy (IMRT). This process is quite complex and the next chapter is devoted solely to the details of this breakthrough

technique. However, before moving on to IMRT a description of the traditional radiation delivery devices and machines will first be provided to conclude this chapter.

The historical developments described above focused largely on the planning of radiation treatment procedures. However, it is important to note that even with such sophisticated planning processes, treatment would not be possible without today's modern equipment. The equipment needed to complete a treatment of radiation therapy is not limited to the radiation delivery devices themselves and also includes patient immobilization devices, medical imaging devices, and dose measurement instrumentation. These additional needs become apparent as one reflects upon the idea of dose conformity, and how the accuracy and effectiveness of the entire process relies heavily upon accurate medical imaging and patient positioning. Therefore, the important inventions that led to improvements in radiation therapy include delivery devices such as the linear accelerator, as well as other technologies such as MRI and CT imaging.

Initially, delivery of radiation therapy involved exposure to radium filled tubes. The strength of these initial radiation treatments was not great enough to provide useful treatment for anything but superficial cancers. However, it was soon discovered that small tubes could be surgically implanted into the cancerous tissue allowing for the treatment of more deep seated tumors, a process called brachytherapy which still exists today. However, since many patients are too weak to undergo surgery or their tumors are too close to vital organs to operate, this is not the ideal solution for every patient. Therefore companies such as General Electric soon developed higher energy tubes which allowed external beam treatment of a wider range of disease. By the mid 1930's research had lead to the development of radiation

tubes reaching an energy of 1MeV [1]. These initial radiation devices all relied upon natural mined radium as the primary radiation source. One of the first major advancements came with the development of $^{60}Co$ (an isotope of Cobalt containing a total of 60 neutrons and protons, pronounced Cobalt 60) treatment units in the early 1950's. These devices utilize synthetic $^{60}Co$ and produce radiation with a maximum energy near 1.2MeV. These $^{60}Co$ machines are essentially the precursor to today's modern delivery units. The success of these machines at treating deep seated tumors led to development of the modern isocentric form for treatment units (isocentric refers to the ability of the treatment head to rotate around a fixed point in space, so that patients can be treated from many different angles as needed). Since the establishment of this isocentric device little variation has occurred in the overall delivery unit configuration and the physical similarities between $^{60}Co$ devices and modern delivery units are quite substantial (see Figures 2.5 & 2.6).

The next major breakthrough came with the invention of the linear accelerator, which is a device that propels charged particles through a vacuum tunnel into a heavy metal target. This target absorbs the particles energy and releases high energy x-rays as a result. This breakthrough was more of a technological advancement and did not initially generate major changes to the treatment and delivery approaches used for radiation therapy. In addition to no longer requiring a fixed radiation source material, another major benefit of the linear accelerator (linac) is its capability to generate very high energy x-rays (18MeV is commonplace), which improves the ability to treat very deep lesions (higher energy increases the penetrating power of the beam, increasing the depth of the location of maximum dose).

Figure 2.5: $^{60}Co$ Therapy Unit (image courtesy of .decimal, Inc. [2])

Figure 2.6: Linear Accelerator (Varian Medical [3])

The major benefits of these improved delivery devices were not fully realized however until medical imaging technology made similar advancements. The proliferation of computer technology and the invention of Computed Tomography (1971) are possibly the greatest innovations to affect the field of radiation therapy. These two tools allowed physicians to precisely locate the cancerous tissue without surgery and contributed immensely to the success and growth of the radiation therapy field. Accurate imaging combined with high energy linacs form the basic package utilized for modern radiation delivery.

# CHAPTER 3
# INTENSITY MODULATED RADIATION THERAPY

The previous chapter provided a brief journey through the advancements made in the field of radiation therapy, specifically for the treatment of cancer. The discussion thus far has emphasized the stages of improvement that treatments have undertaken and has laid the foundation needed for discussion of a highly sophisticated approach to radiation treatment known as Intensity Modulated Radiation Therapy (IMRT). This chapter begins with a detailed overview of the IMRT process, departing from the point where it become a natural extension of the previous treatment approaches. After a description of the general aspects of IMRT, details regarding the specific delivery mechanisms currently employed in treatment clinics is presented. Finally, this chapter concludes with a discussion of the challenges and difficulties associated with IMRT treatment planning and delivery which provides insight into the areas of deficiency that this research addresses.

## 3.1 IMRT Overview

Like many scientific and medical innovations, IMRT is a natural extension of past developments. As with any radiation treatment process, the goal of IMRT is to provide a lethal radiation dose to the cancerous cells with as little normal tissue dose as possible. Numerous references provide complex descriptions of the entire IMRT process, see [4, 5] for example,

however, a much more basic description will suffice herein (a simple, concise description of IMRT is provided by Webb [6] as well). Essentially, IMRT is an extension of the 3DCRT process, whereby further dose conformity is achieved by allowing additional variation of the beam delivery parameters. Specifically, the added variables (degrees of freedom) come from allowing individual beam intensities to vary, shown schematically in Figure 3.1. This simple sketch shows how the individual beam strengths may be modulated in order to conform the region of high dose to the planning target volume (PTV), which includes the tumor and necessary margins to account for uncertainties, while limiting the dose to the adjacent organ-at-risk (OAR). IMRT necessarily has become a complex process, because addition of this extra variability precludes the possibility of simple "forward planning" by hand as used in 3DCRT. IMRT treatment plans must be developed using a process called "inverse planning", which is essentially a numerical (computerized) optimization process where numerous plans are compared (in an intelligent manner) in order to obtain the best possible dose conformity.

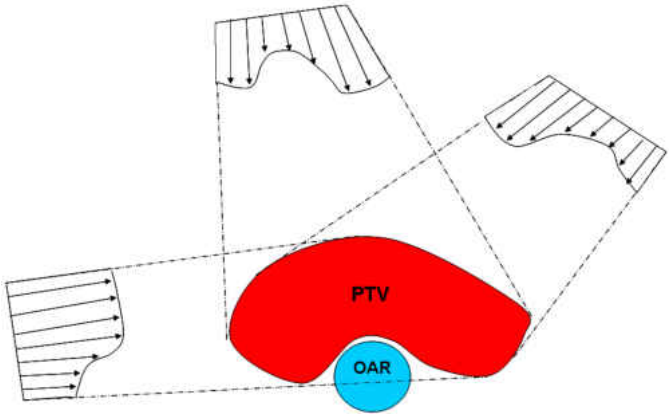



Figure 3.1: Two-Dimensional Intensity Modulated Beam Profiles

Understanding all the details of the IMRT process can be a daunting task, especially when one discovers the large body of research currently being conducted in the field. While

reviews of this literature are readily available from authors such as Webb [7, 8], only a few papers, see [6, 9, 10] for instance, are truly needed to provide the level of understanding necessary for the majority of those in the field. Along with the huge body of literature, the importance of this technique in the fight against cancer is further evidenced by the fact that over one million cancer patients are treated with external radiation in the U.S. each year [11]. This ever growing number stems from reduced treatment complications and increased cure rates that are often associated with IMRT. As this work aims to improve the planning and delivery of IMRT, further details of the IMRT process, including a description of the inverse planning procedure, are given below. Finally, the discussion of IMRT herein, is completed by a thorough comparison of the two most common delivery techniques.

While IMRT is a powerful tool in the treatment of cancer, it is not necessary or ideal for all cancer cases and the decision for use of this technology lies in the hands of trained medical professionals, typically the attending physician or a qualified medical physicist. Once the decision to treat using IMRT has been made, the process will start by setting the patient into a specific treatment position (utilizing various fixation devices as needed, including vacuum formed masks, hold downs, bite plates, and other patient specific devices) so that consistent treatment can be achieved. Once an appropriate position has been found, the patient skin is semi-permanently marked so that re-alignment can occur in a precise manner. Medical images of the patient are then obtained in this treatment position using either computed tomography (CT) or magnetic resonance imaging (MRI) technology. The images are digitally recorded and the patient is no longer needed until treatment planning is completed and delivery is ready to commence. The treatment planning team (physician,

physicist, and dosimetrist) will use the images to determine a patient specific treatment plan. Development of this treatment plan begins with the physician outlining the critical anatomical structures on the digital patient images, a process referred to as contouring. Contouring is used to delineate the planning target volume (PTV), which consists of the visible cancerous tissue as well as an appropriate margin to account for unseen microscopic disease, from the healthy normal tissue. Additionally, depending upon the location and extent of the disease, vital organs such as the heart, bladder, brain stem, or spinal cord, may be contoured so that radiation dose may be further limited in these organs-at-risk (OARs). Figure 3.2 shows contouring of a single CT slice for a typical prostate cancer treatment setup.

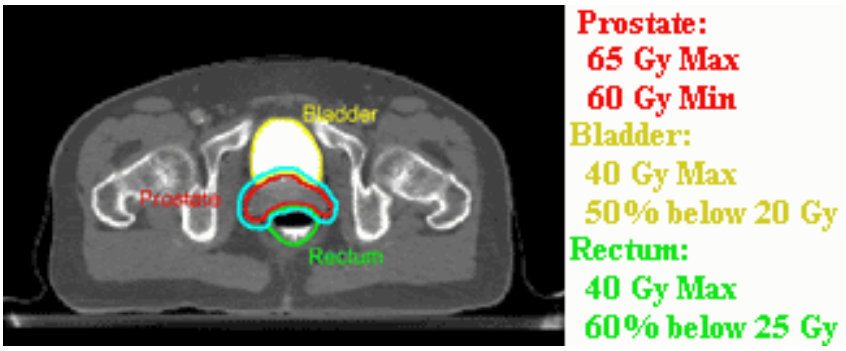



Figure 3.2: Typical IMRT Patient Contours and Prescription (adapted from http://www.radonc.uchicago.edu [12])

The oncologist then completes their part in the contouring phase by providing an appropriate prescription for each of the outlined structures. These prescriptions may be a range of acceptable dose levels in the case of a PTV, or in the case of an OAR a maximum dose or dose-to-volume constraint; in addition to the contouring of critical structures an example prescription can also be seen in Figure 3.2. This prescription serves as the constraint needed to perform the inverse planning optimization process. Next, the patient contours are

converted into a discrete set of points (voxels) to facilitate numerical calculation of the dose field and comparison to prescription values on a point-by-point basis. This is important as the remainder of the planning process is largely automated relying upon numerical dose calculation and optimization routines. The computerized treatment planning system, the foundation of which is simply a combination dose calculation and optimization engine, then adjusts the treatment parameters and compares the resulting dose field to the prescribed values until an optimal (or acceptable) plan is reached. While it is possible to include parameters such as beam number, beam angle, and beam energy into the optimization process [13, 14, 15, 16], such functionality is generally considered an academic exercise as the extra planning time is not justified by the slight increase in plan qualities that are obtained. Thus, the planner will often directly choose the number of beams for treatment as well as the energies and angles, with odd numbers of equally spaced, coplanar beams (5, 7, or 9) being the most common treatment setups. A typical five coplanar beam setup is illustrated in Figure 3.3. Each treatment field is generally a rectangular shape and the size is often determined using a simple projection of the PTV onto a beams-eye-view (BEV) for each treatment angle. This beam setup information is entered into the treatment planning system, leaving only the intensity level variation for each beam as the remaining unknown to be identified by the numerical optimization routines.

Description of the beam intensity profiles is typically accomplished in a discrete fashion whereby each treatment beam is discretized into beamlets; a beamlet being a small rectangular portion of a single beam. Optimization then consists of determining the best weight for each beamlet so that the prescribed dose is adequately matched and conformation to the

PTV obtained. The beamlet weight can be regarded as a blocking factor for the individual beamlet, with a weight of 1 indicating the beamlet is unblocked and 0 indicating that it is fully blocked (although in some cases a relative weighting is used so that values above 1 are permissible). A schematic of a set of optimized beamlet profiles and the resulting relative dose field (for a single two-dimensional slice) is shown in Figure 3.4. Finally, the planning stage is completed by converting the optimized beamlet weights into the necessary parameters for the chosen delivery modality.

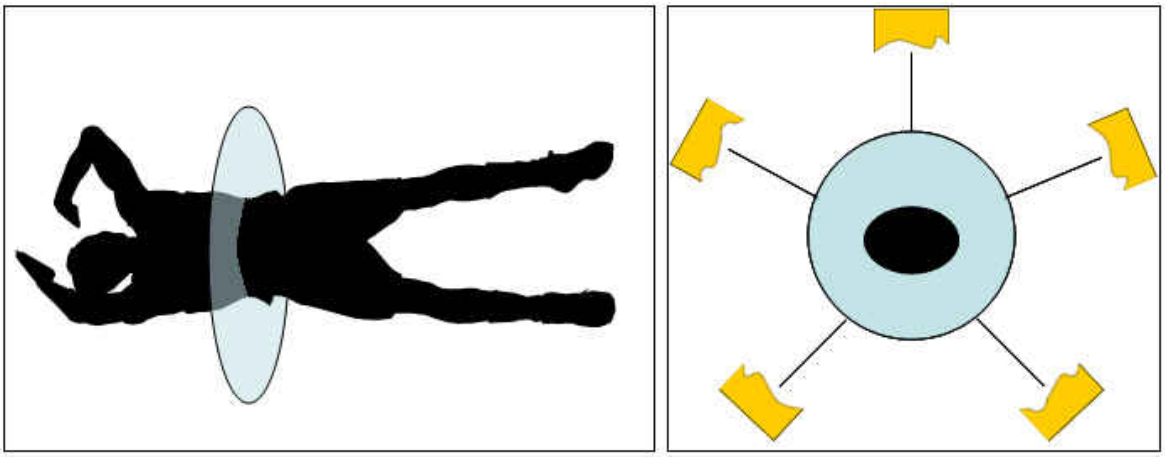


Figure 3.3: Typical Coplanar IMRT Beam Arrangement (Five Beams shown)

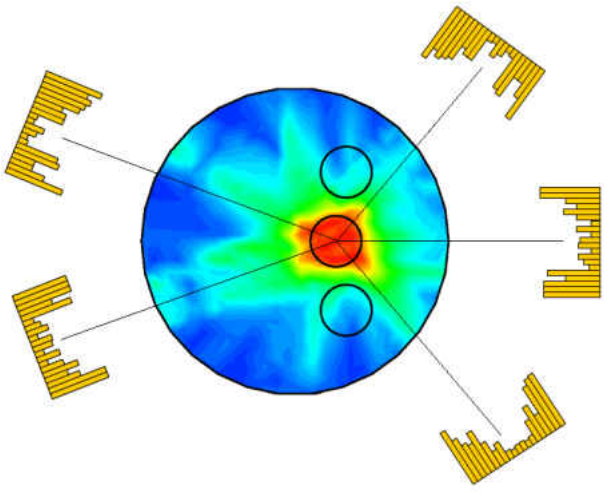
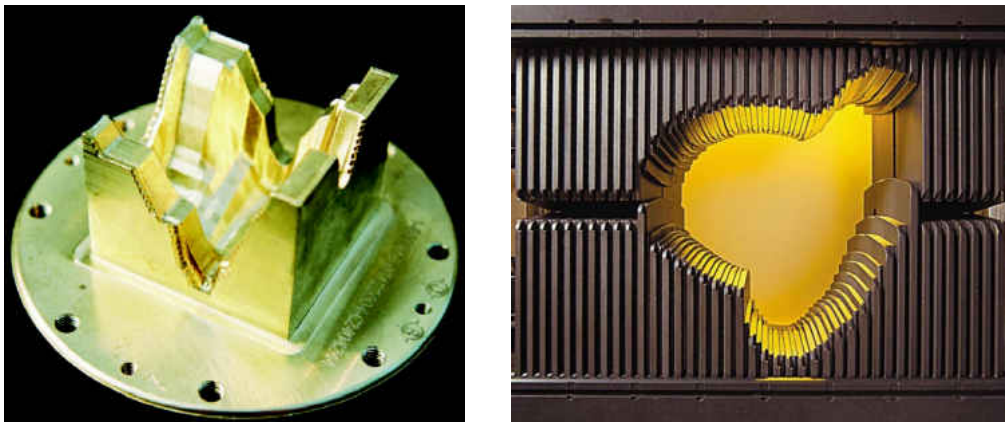


Figure 3.4: Optimized Beamlet Profiles with Corresponding Relative Dose Field

## 3.2 IMRT Delivery

Delivery of IMRT is accomplished by either spatial fluence (dose rate) variation via a solid modulating/compensating filter (commonly referred to as a solid compensator) or by temporal variation of fluence (time-varying) via a Multi-Leaf Collimator (MLC) device. A solid compensator achieves fluence variation by placing a block of certain thickness directly in the path of the beam. In regions where the block is thick, the beam intensity reaching the patient is greatly reduced, while in regions where the block is thin the beam intensity remains almost unchanged (a typical solid metal compensator is shown in Figure 3.5a). In contrast, the MLC accomplishes the fluence variation by opening different areas of the beam for different lengths of time. Put another way, the MLC always turns the beam on to "full strength" and simply varies the exposure time to each area to accomplish the required modulation. If a large dose is required in an certain area, then the area will be exposed for a long time. However, if an area of the field requires a smaller dose, the MLC leaves are adjusted (these leaves are electronically controlled to position themselves into the necessary field shapes during the course of the treatment), so that this area is blocked during a portion of the beam-on time, thereby reducing the dose that this area receives (a standard 80 leaf MLC is shown in Figure 3.5b).



(a) Solid Compensator (b) Multi-Leaf Collimator

Figure 3.5: Comparison of IMRT Delivery Devices (images courtesy .decimal, Inc. [2])

The compensator is actually an older modulation technique, which has largely been replaced by the multi-leaf collimator in the United States, but the solid compensator is beginning to see a resurgence in clinical usage due to new research findings. Before explaining such findings and describing the motivation for focusing this work on compensator based treatment, a thorough comparison of the two techniques is needed. The brief descriptions above should have shown that the MLC is obviously a step forward in technological terms compared to the solid compensator, as the MLC provides a means of treating all patients with a single movable device. However, this technological advancement does not necessarily translate into additional treatment capabilities or improved clinical outcomes. In fact, there are generally only four points that one may raise in favor of the MLC:

1. ease of clinical use
2. immediate availability
3. environmentally more favorable
4. increased clinical safety

The first point is an unarguable truth and one which compensator based treatment could never surpass due to the patient specific nature of radiation treatments. Due to the unique size and shape of each patient and their tumor(s) a specific solid compensator must be manufactured and placed in the treatment delivery unit for every patient (actually several are needed per patient as most are treated from multiple angles). However, this point should really have no grounds for justifying clinical usage of a device, as ease of use has no direct correlation to clinical outcomes or effects. The only possible benefit that this advantage could provide is decreased operational costs or times, thereby allowing treatment of more patients at reduced expense. However, recent retrospective studies [17, 18, 19] have shown that the reverse is actually true, with MLC delivery leading to increased costs due to large up-front expenses, higher maintenance costs, greater down times, and longer beam-on times.

The second point raised in favor of the MLC is that a single MLC is able to treat any patient due to the presence of the movable leafs, and because commercial planning systems can interface directly with the MLC control system, allowing the prescribed treatment plan to be immediately delivered through the MLC for each patient. This advantage has been greatly reduced in recent years, due to rapid growth and automation within the manufacturing industry, and two viable paths exist to allow nearly immediate treatment using solid compensators, thereby matching the capabilities of the MLC. Some clinics employ an in-house manufacturing center to create the custom compensator devices, while others utilize mail-order compensator manufacturing services. Mail-order services eliminate the need for a clinic to staff, maintain, and oversee a manufacturing facility and have recently gained popularity. Companies such as .decimal, Inc. have greatly stream-lined the manufacturing

process and offer 12-24 hour turn-around times on compensator fabrication and delivery, making the difference between compensators and MLCs practically irrelevant in this respect.

The third point seems quite sound at first glance, however a deeper look into the operating, manufacturing, and maintaining of the MLC is needed before a true claim of superiority is possible in this regard. This author is unaware of any study to-date comparing such issues. Also, with recent advancements in recycling practices, compensator materials can be conserved and re-used with a high level of efficiency.

The fourth and final point regarding clinical safety is one that stems from concerns regarding the installation of the compensators in the linear accelerator head by the radiation therapist before each treatment. It is possible that a block may be installed improperly or that the wrong block may be used, however similar concerns regarding the loading of a wrong plan into the MLC control system can also be raised. Each compensator also contains an indexing feature which prohibits installation in the wrong direction. An additional safety concern is that of therapist and patient injury due to dropping a large heavy compensator (some may weigh up to 20 pounds) or due to the exertion of the therapist in loading numerous heavy blocks day-after-day. Once again however, this point offers a rather improper justification for the use of MLCs over compensators as these issues can be mitigated by proper care on behalf of treatment professionals.

The overall theme for the motivation to use MLC based treatment has been shown to be its ease of use. This attitude is clearly evident in the field with well respected texts such as that of Khan [4] advertising the MLC as such, stating, "the impact of this [MLC] technology is in the *automation* of field shaping and modulation ...," and dismissing compensators

because, "carrying so many heavy blocks, patient after patient, creates a nuisance for therapists who have to guard against dropping a block accidentally or using a wrong block." Although this discussion has demonstrated that the MLC has no significant advantage over the solid compensator, it has not established any real motivation for increasing compensator usage by proving that compensators actually provide clinically relevant benefit. There are several key points that can be raised in this regard, namely:

1. compensators can provide increased spatial dose resolution
2. compensators provide more consistent dose
3. compensators impose no limitations on the dose delivery rate
4. compensators reduce skin surface doses

The first point is addressed by noting that the spatial resolution of an MLC is always limited in one direction by the available leaf width size. A two-dimensional, beams-eye-view of the MLC as shown in Figure 3.6 illustrates this resolution concern. It can be argued that this is a device resolution concern, which will not translate directly to delivered dose differences, however, experiments [2, 20, 21] have shown that the dose resolution is indeed degraded for an MLC. The level of spatial degradation is apparent by inspection of Figure 3.7, which shows the fluence distribution achievable for both an MLC and solid compensator when attempting to produce a radially symmetric ring of high fluence (shown in white) while maintaining low fluence in the outer and central regions.

The second point regarding consistency of delivered dose by compensator treatments stems from the fact that most compensators in clinical use are made from solid metallic materials and contain no moving parts. This fully static nature of the compensator eliminates

some of the day-to-day variability of the treatment as the fluence modulation produced by the compensator is completely repeatable and constant for every treatment fraction. This is important to emphasize, due to the fact that MLC delivery does contain some slight day-to-day variability due to statistical variation in the operation of the electronic leaf motion control system. While this may not be a noticeable issue for most treatments, it is a major concern when failure occurs (i.e. leaf sticking or seizing or another operational malfunction) as this causes severe errors in delivered dose. An even more important benefit of compensator dose consistency results from the way in which the modulation is accomplished. Compensators provide a true fluence modulation as they adjust the beam strength in a spatial manner, whereas an MLC provides an indirect fluence modulation by varying the exposure time only, not the beam strength. This has great relevance on received dose for moving organs and tumors [22] even with a properly drawn PTV which includes margins to account for tumor motion. Figure 3.8 shows the high dose region (shaded pink area) at three separate times for both compensator and MLC delivery. The clinical target volume, CTV, (i.e. the actual cancerous tissue volume) is shown shaded green and the extents of the motion are shown by the dashed curves. This figure provides a clear observation that the temporal modulation of the MLC can lead to improper dosing of the tumor as the high dose region changes throughout time. An excellent description and quantification of this issue is also given by Bortfield et al [23]. As also shown in this figure this does not occur with compensator treatments as the high dose region is constant throughout the entire treatment time (for each beam).

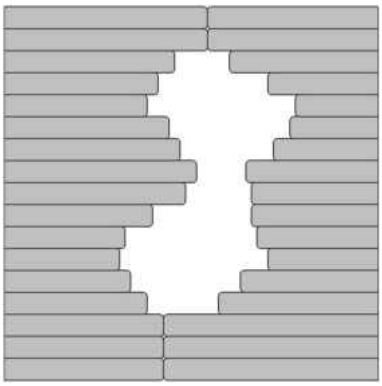
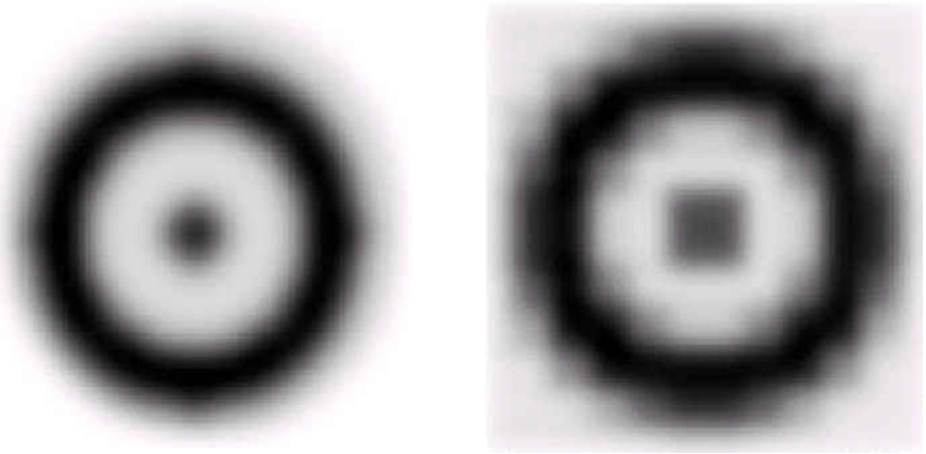


Figure 3.6: Beams-Eye-View of Standard MLC Showing Stair-Case Resolution Issue



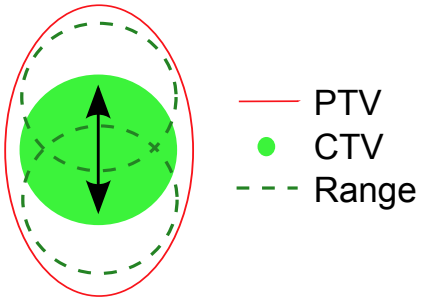
(a) Compensator Delivery (b) MLC Delivery

Figure 3.7: Delivered Dose Resolution Comparison (.decimal, Inc. [2])

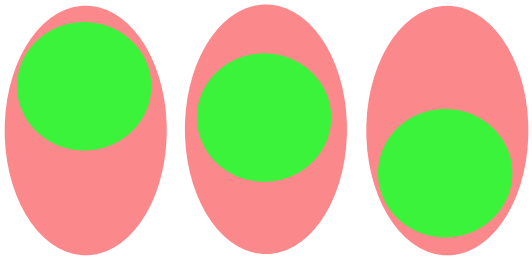
The third claim relates to a recently developing benefit which appears due to the necessity of MLC leaf motion. The maximum leaf speed of a given MLC machine is a finite, fixed value which is actually the current factor limiting the dose delivery rate on modern linear accelerators. Put more simply, the leafs cannot move fast enough to keep up with the delivery rate of modern equipment. Therefore, the output dose rate must be reduced in order to allow MLC treatment to succeed. Compensators have no inherent limitation in this regard and will always allow machine maximum dose rates to be used. This benefit

allows patients to spend less time on the treatment table, thereby experiencing less fatigue and feeling less stressed by the procedure. This increased dose rate also translates directly into reduced beam-on times, which reduces power consumption and lowers the wear-and-tear placed on the linear accelerator devices, especially for quality assurance, QA, testing (proper QA for MLC requires a full, real time simulation in order to ensure proper temporal dose variation, whereas compensators require only enough simulation time to ensure accurate dose measurements are achieved as fluence is modulated spatially, not temporally).

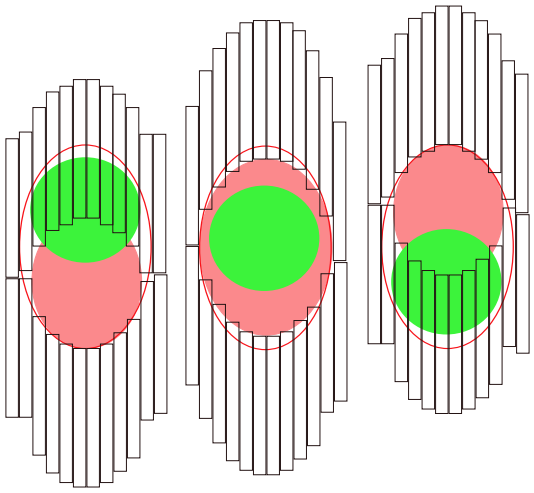
The final point listed above occurs due to the low energy filtration that high density compensator materials provide (this is referred to as beam hardening and will be discussed in detail in the following chapter). This filtration removes a large percentage of the low energy photons, thereby increasing the skin sparing effects of high energy IMRT treatments. Experiments by Jiang [24] and more recently by Stathakis [25] have shown a 5%-10% reduction in skin dose for many patients, with some receiving reductions as high as 40% when compensator treatment is used over MLC. Each of these advantages is significant enough to justify an increase in compensator based treatments, as they all have the potential to provide actual clinical benefit, therefore the clinical need for improved compensator based treatment planning is quite large.




(a) Treatment Contours



(b) Compensator Based Treatment



(c) Multi-Leaf Collimator Based Treatment

Figure 3.8: Impact of Tumor Motion on Delivered Dose

On a final note, a patient survey was recently undertaken by Yaeger and Cost [26], where a sample of clients from their treatment facility were polled, all of which received IMRT treatment at the same facility, under the same treatment staff. The results of this

poll showed that of all patients receiving MLC IMRT less than 50% were satisfied with their treatment, whereas 100% of those receiving compensator based treatment were satisfied with the procedure. It should be understood that this survey takes no account of clinical outcomes or treatment success, and is simply patient opinion. Those authors claim that this 100% satisfaction rate comes from the increased human interaction given to those receiving compensator based treatment. Such interaction was not purposeful to skew the data, it was simply commonplace for the therapist to check on the patient while they entered the room to manually change the compensator after each treatment angle. Since MLC delivery does not require such a manual change, the clinician generally does not enter the treatment room between portals so patients receive less interaction and are generally more anxious during the process. This study and the clinical benefits described above naturally bring about the conclusion that available technology for compensator based treatment forces an increase in *perceived* treatment quality, while simultaneously providing an *actual* increase in treatment quality.

## 3.3 IMRT Challenges

Radiation therapy is a complex field with significant treatment challenges. Much of the difficulty stems from the competing objectives of radiation therapy: 1) deliver a lethal radiation dose to the cancerous tissue and 2) provide non-harmful radiation dose to healthy tissue. Meeting these goals is difficult because radiation therapy is a non-invasive technique, which requires the radiation beam to pass directly through healthy tissue in order to reach

the tumor seated inside the patient's body (this allows it to be used without requiring dangerous or difficult surgery). Additionally, the tumor is typically not large enough or dense enough to absorb all of the incoming radiation, so tissue beyond the tumor will also receive some exposure. Clinicians and researchers developed IMRT as a sophisticated technique for balancing the two competing treatment goals by conforming radiation dose to the tumor as closely as possible, which has been widely shown to maximize treatment efficacy. However, even with IMRT there are difficulties that still exist and areas which could benefit from new innovation.

One particularly difficult aspect of IMRT is the inverse planning process. Recall that inverse planning is the process by which the necessary treatment parameters are determined such that an acceptable dose field is achieved. This process is difficult for many reasons, but one notable area is the representation of the intensity modulation. As described in Section 3.1, the modulation of the fluence is most often accomplished numerically by use of small individual beamlets. Many beamlets will exist for any useful plan (typically several hundred per beam), making the optimization process challenging due to weak parameter correlations and low parameter interaction and sensitivity. Further details of these inverse planning concerns are provided in the following chapter as they pertain to the specific details presented therein.

A more recently emphasized area of concern with IMRT is tumor localization. Numerous studies have shown that tumor motion due to respiration and digestion processes can have significant effects on the location of the tumor at any given time. Digestive system concerns such as stomach, bladder, and bowel fullness can cause day-to-day variability,

while respiratory motion of the diaphragm and lungs can cause continuous variability in tumor location. These concerns are particularly poignant for IMRT since the dose is highly conformal; therefore a small error in tumor location can lead to a miss in the delivered dose [22]. Ehler, Nelms, and Tome [22] have shown that using compensators for treatment of moving tumors can increase dose uniformity and clinical effectiveness. They concluded that the static nature of the solid compensator simplifies the delivery process for such cases and is a key ingredient to its success. Additionally, Markman et al [27] have shown that producing smoother fluence modulation profiles results in decreased sensitivity to patient positioning errors. Since a positioning error is analogous to an internal motion error, their work also pertains to the case of tumor localization concerns. Direct Compensator Profile Optimization is designed to increase the smoothness of the compensators for manufacturing considerations, therefore this will have the beneficial side effect of reducing positioning and tumor motion sensitivity.

A final area of challenge specifically for solid compensator IMRT is that of manufacturing. Manufacturing of solid compensators is generally viewed as the most inconvenient aspect of their use, but it is obviously unavoidable as each beam requires its own distinct modulation profile for every patient. Like most parts of the IMRT process, several choices exist for this step as well. The three most practical approaches include casting, powder/pellet molding, and CNC milling. Casting and molding are typically accomplished in similar fashion and each of these processes begins by creating a foam or wax plug, which is the reverse shape of the desired compensator (the plug may be drilled, milled, or cut to the appropriate shape). Following creation of the plug, the actual compensator is made by either pouring melted

material (typically Cerrobend) for casting or packing powder/pellets (often copper, tin, or stainless steel) for molding into a rectangular box containing the plug [28, 29]. For casting, the material hardens and can be removed from the box for use. For molding, the packed box is needed to contain the material, therefore the box itself, complete with the powder *and plug* will become the compensator [29]. The third option, milling, is generally performed on metallic materials, with brass being the most common choice. Milling is generally the most precise and offers the greatest manufacturing flexibility; it is also the most time efficient as only a single manufacturing step is required. Because of this single step nature, milling also offers the highest possibility for automation of the manufacturing process. For these reasons this work will focus solely on milling for compensator fabrication.

Despite the benefits of the milling procedure as compared to other manufacturing processes, there are still limitations regarding its use. Specifically, the major difficultly comes from the fact that manufacturing is not addressed until *after* inverse planning is complete. The typical process for the design and manufacture of a radiation compensator is shown in Figure 3.9. This chart shows the major deficiency, namely, that milling concerns are not known until after the compensator order has been placed and analyzed. This problem can cause significant delays in the radiation treatment process, as the treatment plan must be modified and inverse planning completed again in the hopes of providing a more manufacturable compensator. This is not an overly alarming concern and informal estimates from .decimal, Inc. state that this issue affects less than five percent of their compensators. What is more common however, is that the matching is not poor enough to necessitate a new plan (due to time and cost considerations), and a mismatch of 3%-5% is

often accepted. It should be noted that the actual milled surface can be ascertained from the milling simulation software, so that the expected patient dose can be re-computed, however these discrepancies generally reduce the modulation range and degrade the plan further from the "optimal" plan provided by the treatment planning system.

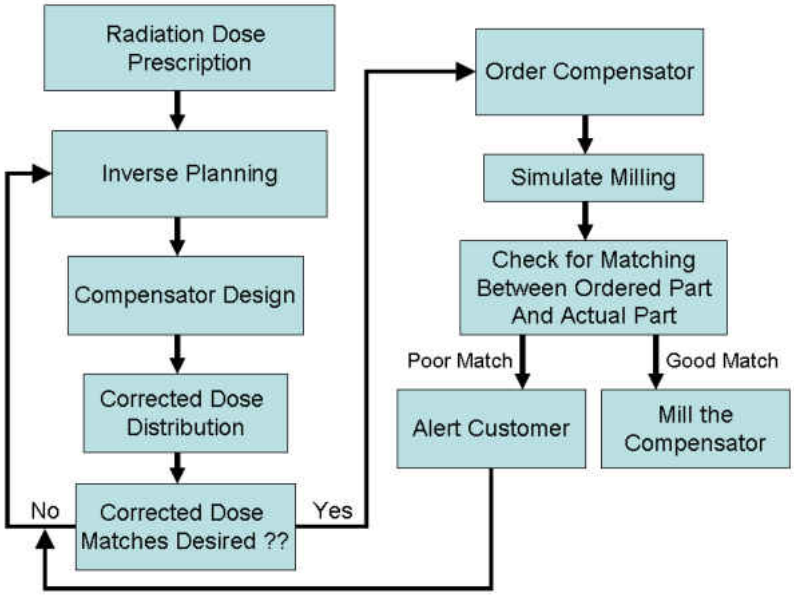



Figure 3.9: Standard Compensator Design and Manufacturing Processes

The remainder of this document describes the Direct Compensator Profile Optimization process and includes detailed descriptions of how it is able to overcome many of the challenges presented in this section. In addition, numerous examples are provided to justify the claims regarding the abilities of the DCPO procedure.

# CHAPTER 4
# METHODOLOGY

By now, the process of Intensity Modulated Radiation Therapy should be sufficiently understood so that the characteristics of an ideal treatment planning approach can be discerned and the major challenges to implementation appreciated. The preceding discussions have mentioned the areas of deficiency in regards to current solid compensator IMRT treatment planning, however the work completed herein to resolve these issues has not yet been thoroughly described. It is the focus of this chapter to provide such a description for the *Direct Compensator Profile Optimization* (DCPO) approach. The premise of DCPO is quite straight forward: generate the treatment plan by directly optimizing the shape of the compensator surface. The following hypothesis statement highlights the major benefit that this approach allows:

> *Direct surface optimization reduces the number of optimization parameters and allows manufacturing and delivery device limitations to be addressed during the inverse planning stage. This parameter reduction and manufacturing/delivery limitation inclusions will increase the computational efficiency of the inverse planning process, reduce the oscillations along the compensator surfaces, and increase the agreement between planned and delivered radiation dose, without affecting the quality of the resulting treatment plans.*

There are several difficulties associated with such a process that must be addressed, the most important of which is how to describe and control the compensator surface shape. DCPO utilizes a three-dimensional analytical surface representation to provide a smooth, continuous representation of the surface profile using a relatively small number of control parameters. This analytical representation also acts as a unique parametrization of the underlying field beamlets and avoids the need for direct optimization of the numerous beamlet strength values (it is this parametrization that provides the many benefits associated with this planning approach). Since DCPO operates using the actual physical shape of the compensator, the planning process is able to incorporate two new features which are quite expensive or impossible to include with other more traditional approaches. First, manufacturability of the compensator surface can be ensured *during planning* by appropriately limiting the surface representation scheme. This guarantees that the planned surface is nearly 100% machinable, thereby ensuring that the delivered dose matches the planned dose (note that such an ability is not truly possible without a method that directly optimizes the physical delivery parameters, as does DCPO). Secondly, dose calculations may now include beam perturbations caused by the presence of the compensator within the optimization process as the actual, local material thickness is known at all stages of planning. The incorporation of these features, the implementation of the surface representation scheme, and important changes to the inverse planning (optimization) process are described in detail in the remainder of this chapter. The description of each approach also helps to further establish the relationship of DCPO to current state-of-the-art approaches used for treatment planning.

### 4.1 Surface Representation and Beamlet Parametrization

As previously described, almost all current approaches to IMRT treatment planning begin by dividing each treatment beam into a set of small rectangular beamlets. While the beamlet size is often adjustable, most practical systems limit resolution between $5 - 10\,mm$. Beamlet resolution (size) is important as it is the determining factor in two competing aspects of treatment planning: dose calculation accuracy and optimization performance. In terms of dose calculation accuracy, small beamlets are more desirable as it improves the numerical approximations made within most dose calculation algorithms. As for optimization performance, since most planning systems use the beamlet strengths directly for inverse planning (recall that it is the strength of each beamlet that provides the fluence modulation and facilitates the development of conformal treatment plans), small beamlets are undesirable as it leads to increased numbers of optimization parameters and reduced sensitivity amongst the beamlet strengths. It is possible that improvements can be made by allowing for non-uniform beamlet discretizations, but implementation of such an approach has not yet been shown in the published literature.

An alternative procedure, shown by Markman et al [27], demonstrated a novel interpolation approach which was used to parametrize the beamlet strengths. Parametrization simply means that rather than varying the strength of each beamlet individually, one instead varies the values of some "control parameters". The control parameters are used in some mathematical relationship to facilitate calculation of the actual strength of each beamlet, then dose calculations can proceed as usual using the individual beamlet strengths. Such

a parametrization has several important implications in the planning process, two of which were well recognized by Markman et al. First, the number of planning variables (degrees of freedom) can be significantly reduced, so that planning times are shortened. Second, the resulting plans will generally have smoother fluence fields which reduces some unnecessarily sharp gradients, thereby decreasing dose errors associated with patient positioning and/or organ motion. What was missed in this work however, was the applicability of this process to solid compensator based treatments, as well as the effects of such a parametrization on the inverse planning algorithms. DCPO utilizes a parametrization approach very similar to that proposed by Markman, however, DCPO takes the procedure several steps beyond this previous work by moving the interpolation from fluence to the physical compensator surface. The process is illustrated for a two-dimensional geometry in Figure 4.1, where the red dots indicate the control parameters (surface heights in this sketch) the solid black line shows the resulting surface representation, and the yellow bars indicate the finely discretized beamlet heights (this sketch provides a view of a single slice of a full compensator, however it should be noted that this work has developed and verified the DCPO process for full three-dimensional patient cases).

This type of direct physical delivery parameter optimization has an intuitive simplicity and the theoretical advantages are quite broad, with the most notable being the elimination of the secondary optimization step (this process is described below), as well as the possibility for more accurate agreement between planned and delivered doses (due to inclusion of delivery device effects *during planning*). While DCPO is the first procedure of its kind for compensator based treatments, it is not the first approach which has attempted to use

physical delivery parameters during the optimization process. Direct Aperture Optimization (DAO) is a recently developed technique for MLC based treatments that also follows such a path. DAO trades the ideal fluence optimization for a direct MLC positioning optimization, thereby removing the time-consuming, plan degrading process of leaf segmentation. DAO also allows for the incorporation of specific MLC delivery concerns into the planning system during the inverse optimization stage. Effects and limitations such as leaf width, leaf edge shape, and inter-leaf leakage can all be included so that dose calculations are improved. Further benefits of DAO and delivery parameter optimization in general are discussed by Yu et al [30]. DAO has gained some clinical success through Philips Pinnacle $P^3$IMRT and their specific version of DAO referred to as Direct Machine Parameter Optimization (DMPO). Clinical studies by van Asselen et al [31] have shown that DMPO is able to provide improved tumor dose homogeneity for breast cancer treatment. Much like DAO and DMPO, this new process of DCPO allows for the key aspect of delivery device effects to be addressed during inverse planning with no major computational penalty (in the DCPO approach efficiency is actually improved). Specifically for compensator based treatments, these delivery device effects include dose calculation perturbations due to beam hardening and delivered dose variability due to manufacturability problems. Incorporation of both of these effects will be addressed in the corresponding sections below, but first a detailed description of the secondary optimization process is needed.

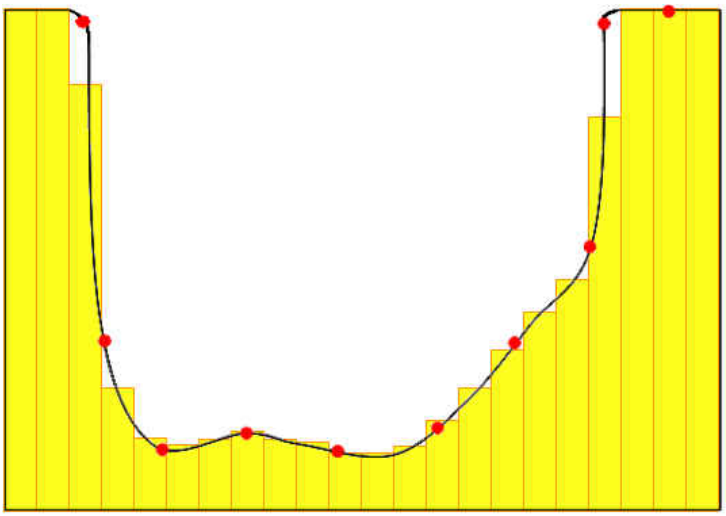


Figure 4.1: Two-Dimensional Illustration of Direct Compensator Profile Optimization

Secondary optimizations are often needed in the IMRT planning process due the substantial changes in modulation that occur when the "ideal" fluence (solution from the optimization routine) is converted to a deliverable fluence. Both MLC and solid compensator IMRT have specific considerations regarding this conversion from ideal to deliverable fluence. For solid compensators there are three concerns that necessitate a secondary optimization process for most planning approaches: beam hardening, non-uniqueness of fluence-to-compensator conversion, and manufacturability. Beam hardening makes the calculation of compensator thickness an iterative process since the attenuation coefficient, $\mu$, is a function of beam energy and/or material thickness. Therefore, the following equation is used in a simple iteration loop to determine the appropriate compensator thickness, $t$, at numerous positions across the beam (usually at the center-line of each beamlet):

$$I = I_o \, e^{-\mu(t)\, t} \tag{4.1}$$

where $I_o$ refers to the unmodulated beam intensity and $I$ is the desired intensity value obtained from the ideal fluence map (this applies to both relative and absolute intensity values). The exponential behavior of attenuation complicates attempts to smooth the ideal fluence profile before converting to compensator thicknesses as this may result in undesirable effects. Thus the most effective way to smooth the compensator is to do so after conversion to thickness values. Existing programs recognize this and perform such smoothing after compensator shape generation, however such programs also often include another overall beam weight optimization after the smoothing in order to help eliminate the plan degradation resulting from these processes [32]. The second issue regarding the non-uniqueness of conversion is often addressed by an assumption attributed to Jiang [24] which gives "optimal" compensator designs as those which use the minimum possible thickness to achieve the calculated ideal fluence modulation. These are optimal is the sense that the required compensator thickness and beam-on times are minimized. The DCPO approach avoids this problem altogether since the compensator surface, not the ideal fluence, is provided by the optimization. This approach also has the benefit that it may allow plans to be improved due to beam hardening effects. Although substantial effects are not possible from such inclusion, doses to the skin and superficial tissues may be reduced by an appreciable extent in some cases. Also, it is understood that such effects come at the expense of increased beam-on time, however this increase is not significant for the typical ranges of modulation that are expected in common planning situations.

In order to accomplish the transition to Direct Compensator Profile Optimization, a means of appropriately describing each compensator surface is required. Numerous schemes

exist which can be used for this purpose and several were tested by Markman et al [27] for the case of fluence representation, including polynomial spline and radial basis function interpolations. However, this research has some additional objectives beyond those of Markman; therefore it is important to take a deeper look at these goals, so that the most fitting surface representation scheme can be chosen. One important goal is that the planning process should restrict its efforts to working only with compensator surfaces that are 100% manufacturable with standard techniques. This restriction quickly provides the additional constraint that smooth, continuous surfaces are required. Another important goal is to remove unnecessary surface oscillations so that excessive surface gradients are reduced and smoother compensators result. Prior experience with parametrization and interpolation by the author [33, 34, 35] has shown that polynomial regression can provide viable solutions for noisy (oscillatory) data. However, low order polynomials would not allow representation of the types of surfaces expected for radiation compensators and high-order global polynomials are too oscillatory for this application. Therefore, a spline approach rather than global regression is a better choice. Two approaches are prominent candidates to satisfy the given objectives: bi-cubic spline and Non-Uniform Rational B-spline (NURB) surfaces.

Bi-cubic spline interpolation is appealing as it satisfies the continuity and smoothness requirements, while also allowing for reasonably steep field gradients to exist where needed. Additionally, Markman [27] found this approach to be quite promising in his work. The process of cubic spline interpolation is well known and provides sufficient smoothness by enforcing continuity of the field as well as the first and second derivatives at each of the spline control points. Variation of the surface is accomplished by simply adjusting the heights and

possibly the horizontal positions of the spline control points. The implementation procedure for two-dimensional bi-cubic splines is thoroughly detailed in texts such as Numerical Recipes [36]. Cubic spline and bi-cubic spline surfaces are shown in Figure 4.2 for an arbitrarily chosen set of control points, so that smoothness and steepness can both be well visualized.

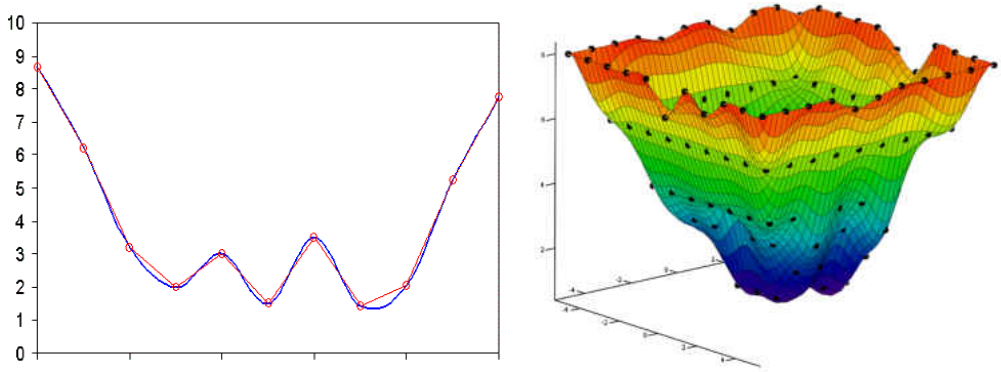



Figure 4.2: Cubic Spline Interpolated Surfaces (1D and 2D)

Cubic spline interpolation does seem to satisfy the major requirements for successful surface representation, however, upon testing it was found that adequate control over the maximum and minimum surface heights was not feasible. This problem occurs due to the fact that the splines are forced to pass exactly through each control point, therefore the surface is forced to overshoot/undershoot in the neighborhood of steep gradients. The resulting surface concerns are highlighted by Figure 4.3.

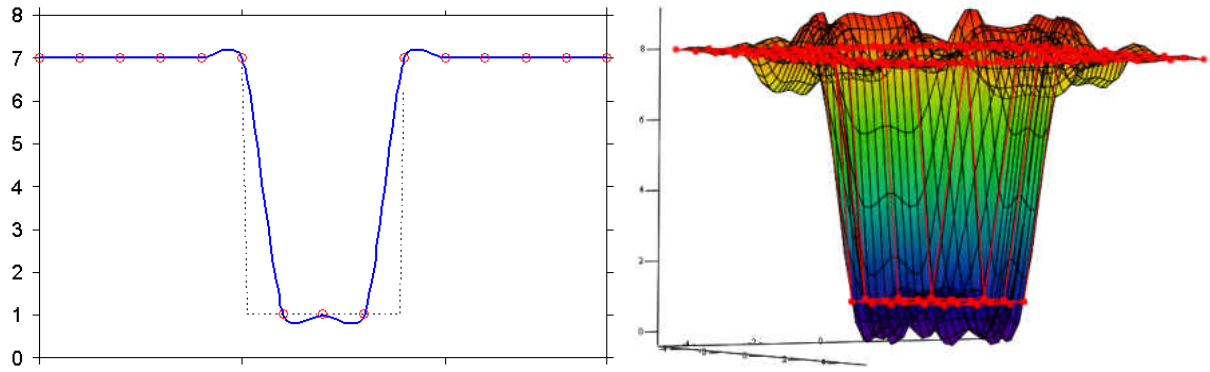



Figure 4.3: Cubic Spline Surface Control Issues

Non-uniform rational B-splines use a robust control technique which allows complex surfaces to be described with fewer parameters than many alternative approaches, making them an attractive choice for this work. Also, as NURBs are not interpolating functions they tend to reduce the oscillations as compared to many other polynomial based methods. Another important property of the NURBs is that they are guaranteed to be bounded by their underlying control point range [37]. This bounding is a primary motivation for using NURBs since a useful compensator design must respect limits on both maximum and minimum surface heights. The surface complexity possible with NURBs is often a desirable property, however since smoothness and steepness are important concerns of this work, the NURBs must be sufficiently limited in order to ensure only appropriate surfaces are able to represented. Therefore, some simplifications have been made to the typical NURB surface representation so that the available surface complexity is reduced. These simplifications also allow for a more straight forward parametrization and a more computationally efficient surface calculation process. The specifics of these simplifications are presented below along with a brief description of NURB implementation process.

Two-dimensional NURB surfaces are direct extensions of their one-dimensional NURB curve counter-parts. Therefore, the description of the reduced NURB representation used herein will be explained for NURB curves and then the necessary extensions for multi-dimensional surfaces will be directly provided. This discussion will not attempt to explain the full implementation procedure for NURBs as this exists in numerous references (see [37] for example). Instead the goal is to discuss the relevant qualities that make NURBs ideal for this work as well as describe the specific restrictions and reductions used on the NURBs

herein. A NURB curve typically consists of four major components, a basis function of specified order, $k$, a knot vector, $X$ , a weight vector, $CW$, and a control point vector, $CP$. The basis functions, $\theta$, are defined in a recursive manner as:

$$\theta\left(u,i,k,X\right)=\frac{\left(u-X_i\right)\cdot\theta\left(u,i,k-1,X\right)}{X_{i+k-1}-X_i}+\frac{\left(X_{i+k}-u\right)\cdot\theta\left(u,i+1,k-1,X\right)}{X_{i+k}-X_{i+1}} \tag{4.2}$$

for $k>1$, while for $k=1$:

$$\theta\left(u,i,1,X\right)=\begin{cases}1 & if \quad X_i\leq u<X_{i+1}\\ 0 & otherwise\end{cases} \tag{4.3}$$

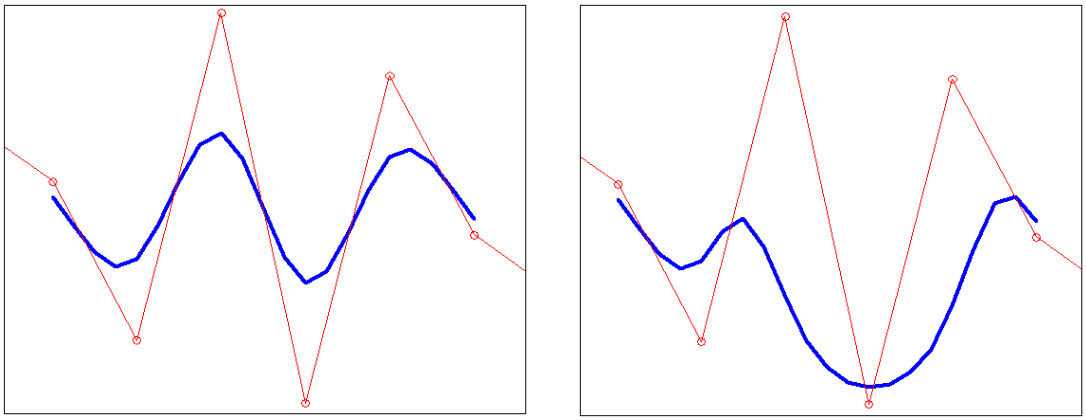
where $k$ is the order of the NURB curve, $X$ is the knot vector, $u$ is the evaluation point along the curve, and $i$ is the index of the control point vector. The expression for $k=1$ shows that each basis function is non-zero only over a brief window of the curve. This means that changes to a single basis function or control point will affect only a small area of the curve, allowing for fine tuning of localized sections of the curve. Also, since each basis function is non-zero only over $k$ knots, the spacing of the knot vector values can be used to control the influence region of each basis function. This is one of the features that makes NURBs so robust. Techniques such as duplicating knot values can be used to create sharp cusps and corners or force interpolation if so desired. However, since this work wishes to avoid such features, a uniformly spaced knot vector will always be used. This proved to be a very useful restriction for this work, not only due to the concerns just described, but also because the knot vector must always be monotonic and the resulting NURB curve is very sensitive to even small changes in the knot vector values. This extreme sensitivity makes surface optimization

difficult, as even slight changes in the knot vector can lead to significant surface changes. Additionally, maintaining the monotonicity of the knot vector throughout the optimization proved difficult to handle well (re-ordering the knot vector to ensure monotonicity can cause drastic surface changes). Fixing the knot vectors leaves full specification of the NURB curves directly in the weight and control point vectors. Understanding the purpose of these vectors is quite straight forward as each has an intuitive physical meaning. Simply put the control points specify locations at which the curve should approach and the weights specify how strongly the control point should "pull" the curve toward itself. The role of these points and weights can also be seen in the expression of the complete NURB curve, which is generally a parametric relation as shown by:

$$q\left(u\right) = \frac{\sum_{i=1}^{Nx}\left(CP_{i_q}\, CW_i\, \theta\left(u, i, k, X\right)\right)}{\sum_{i=1}^{Nx}\left(CW_i\, \theta\left(u, i, k, X\right)\right)} \tag{4.4}$$

where $q$ refers to any particular coordinate direction (x, y, or z), $CP$ and $CW$ are the vectors of control points and weights, respectively, $k$ is the order of the NURB curve, and $Nx$ refers to the number of control points/weights. Figure 4.4 further illustrates the roles of the points and weights for an arbitrary curve in two-dimensions. The red points show the control point positions (connected by linear splines for clarity) and the blue line shows the resulting NURB curve. Note that the control point positions are identical for each curve and dictate the major trends of the resulting curve. In Figure 4.4a equal weights have been used, while in Figure 4.4b the lowest point has a significantly increased weight value, which acts to "pull" the curve toward this point. This figure thus highlights the two different means of

controlling a NURB curve that are utilized herein: adjustment of the control point heights and weights.



(a) Equal Weights (b) Unequal Weights

Figure 4.4: NURBs Curves in Two-Dimensions

There is one final reduction that has been implemented to simplify the use of the NURB surfaces as representations for radiation compensators. The typical parametric NURB representation allows the x, y, and z coordinates to be specified somewhat independently using only a single knot vector (even for a multi-dimensional curve), with each direction having an independent control point value. A difficulty with this representation arises, however, when one wishes to evaluate the resulting NURB curve at intervals which are constant in any particular coordinate direction. This occurs because the NURB curve must be evaluated using the parametric variable, $u$, not an individual x, y, or z position. Therefore, in order to determine the height of the compensator surface at the center of any particular beamlet (a known x, y location) a one-dimensional minimization would need to be performed in each direction to find the correct values of the parametric variables u and v that yield the desired x and y values. While this is not especially difficult, it is time

consuming and therefore very undesirable for this application. To circumvent this issue, the NURB parametrization has been shifted directly to x-y space, meaning the knot vectors are actually x and y coordinate locations. This forces equal spacing of the control points in the x-y plane, however, it is shown below that this does not pose any limitations beyond that of using constant, uniform knot vector spacing and does not hinder representation of the types of surfaces that are desired in this application.

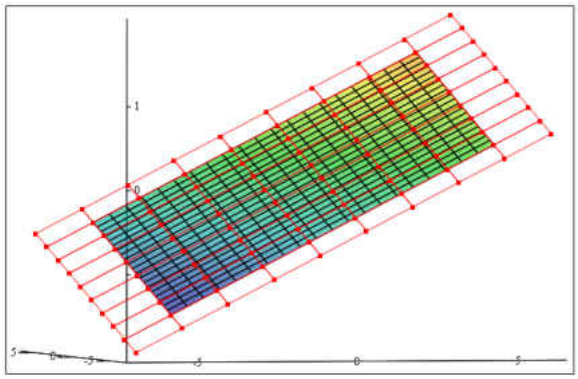
The extension to two-dimensional surfaces (as opposed to curves described above) requires adding another knot vector for the second dimension and generation of a matrix of control points and weights for each valid region of space specified by the knot vectors. Additionally, the extension of Equation 4.4 for a two-dimensional reduced NURB surface in x-y space is given as:

$$z\left(x,y\right)=\frac{\sum_{i=1}^{Nx}\left[\sum_{j=1}^{Ny}\left(CP_{i,j}\,CW_{i,j}\,\theta\left(x,i,K,X\right)\;\theta\left(y,j,K,Y\right)\right)\right]}{\sum_{i=1}^{Nx}\left[\sum_{j=1}^{Ny}\left(CW_{i,j}\,\theta\left(x,i,K,X\right)\;\theta\left(y,j,K,Y\right)\right)\right]} \tag{4.5}$$

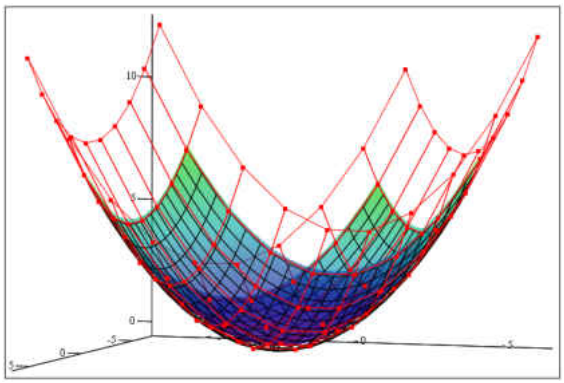
where $CP$ and $CW$ are now matrices, $X$ and $Y$ are the knot vectors for the $x$ and $y$ directions, respectively, and $Nx$ and $Ny$ refer to the number of control points/weights in the $x$ and $y$ directions, respectively.

In order to address the concern of whether this NURB surface scheme provides enough influence and control to specify the types of surfaces expected for a real three-dimensional radiation compensator, several initial surface studies were performed. First, simple smooth surfaces were chosen and the NURB control points were placed directly on the desired surface. Linear, quadratic, and trigonometric surfaces proved to be quite easily captured as evidenced by the results in Figure 4.5. The error present in the sinusoidal surface shown in Figure 4.5c

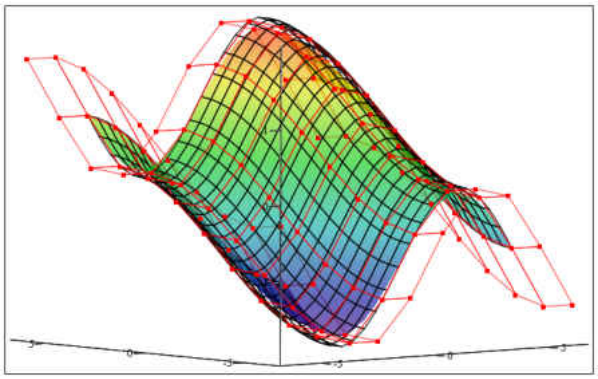
is difficult to visualize directly, therefore a simple average deviation has been computed (2.6% of the maximum surface height) to verify the high level of surface agreement. It should be noted that these surfaces all used equal weights for each control point, therefore further surface agreement could be achieved through adjustment of these values. Modeling of more complex surfaces containing steps and abrupt changes was then attempted using the same direct control parameter prescription (i.e. the control points and weights were adjusted manually in an attempt to match the desired surface shape). Once again such surfaces were able to be represented to an acceptable degree with the reduced NURB surface representation. Results for some sample abrupt surfaces are provided in Figures 4.6 - 4.8. In each of these figures the black mesh represents the prescribed (desired) surface, the red mesh shows the NURB control points, and the colored surface is the actual resulting NURB surface (this surface is translucent in order to increase the visibility of the three-dimensional image). Also, some surfaces are shown as two-dimensional plots, taken at representative slice(s); in such plots the NURB surface is shown as the solid blue line(s) with all other colors remaining consistent. It should also be noted that the reduced form of the NURBs used herein does not provide a direct x-y location for the control points and each is valid over a small range of the surface corresponding to the knot vector values; the control points are therefore shown at the center of their respective ranges.



(a) Bi-Linear

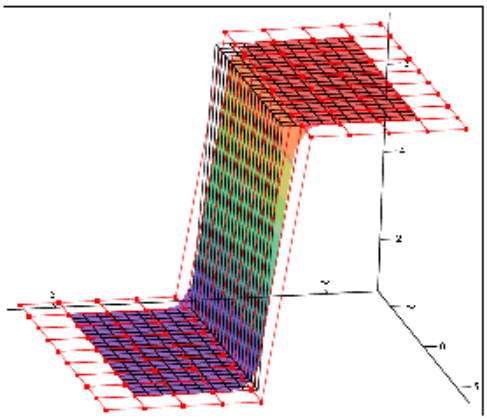


(b) Bi-Quadratic

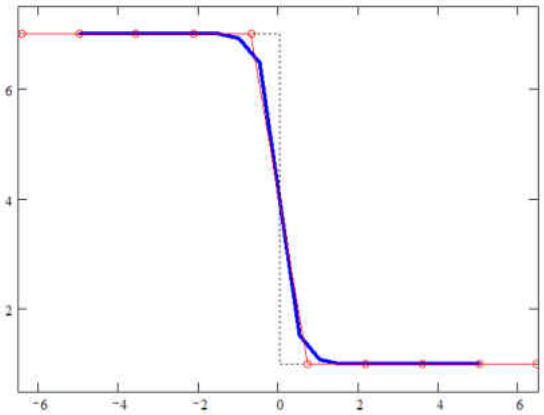


(c) Sinusoidal

Figure 4.5: Three-Dimensional NURB Surface Testing: Smooth Surfaces

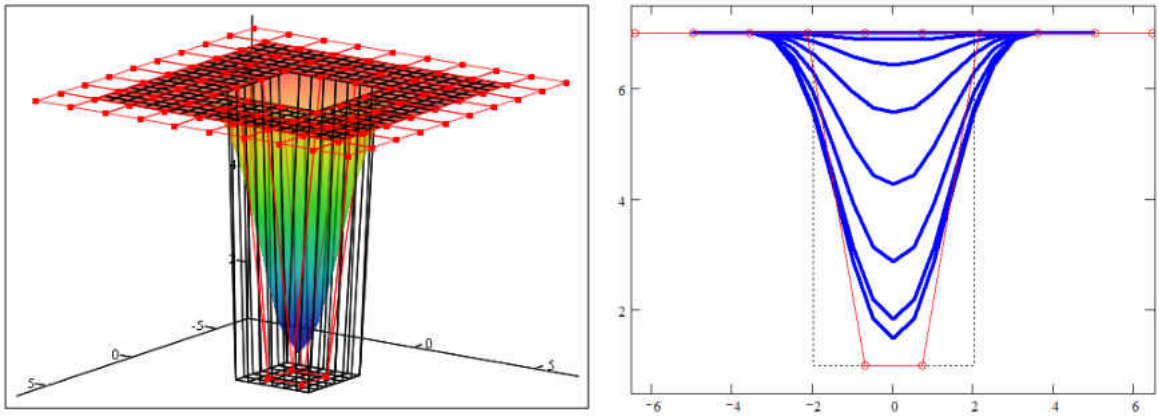


(a) Complete 3D Surface



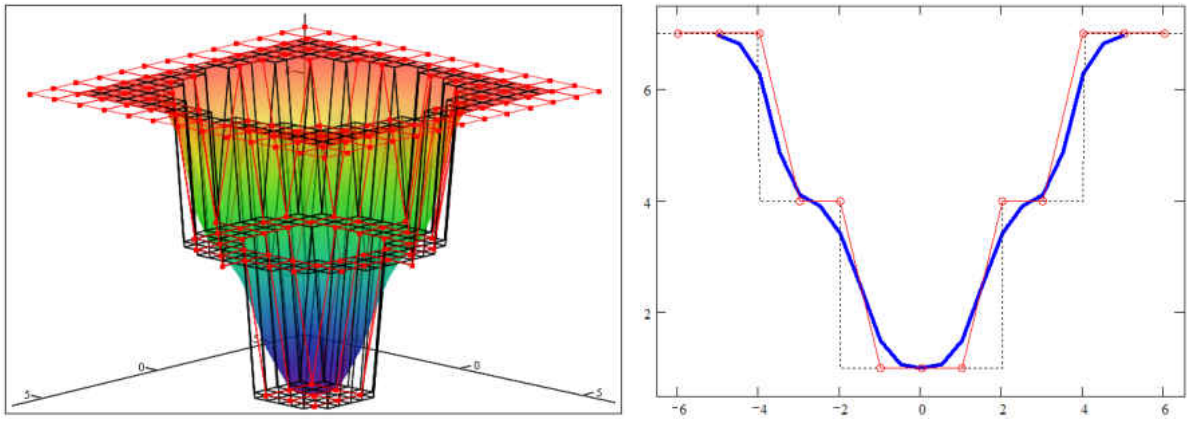

(b) Center-line 2D Slice

Figure 4.6: Three-Dimensional NURB Surface Testing: Single Abrupt Step



(a) Complete 3D Surface (b) Several 2D Slices

Figure 4.7: Three-Dimensional NURB Surface Testing: Abrupt Square Pocket



(a) Complete 3D Surface (b) Center-line 2D Slice

Figure 4.8: Three-Dimensional NURB Surface Testing: Abrupt Tiered Pocket

Overall these results show that this reduced NURB representation does provide adequate resolution and control such that the major features of typical radiation compensators can be well expressed. Additionally, these results demonstrate the important qualities that NURBs bring to the surface representation, including the elimination of overshoot and reduction of surface oscillation, as well as greater surface control. The superiority of the NURBs can be

further visualized by comparing the bi-cubic spline interpolation for an abrupt square pocket surface (shown previously in Figure 4.3) to that of the NURB surface given in Figure 4.7. Clearly, the cubic spline is not producing an acceptable surface representation. It could be argued that adjustment of the control points could produce a more desirable surface match for the cubic splines; one which does not contain overshoot or undershoot, however the cubic splines provide no means of analytically controlling the maximum and minimum values (due to overshoot) so avoidance of this issue cannot be guaranteed in a general sense. Therefore, the exclusive use of the reduced NURB surface representation is quite justified, but for the sake of completeness one full treatment planning case is shown in the next chapter to give a final validation for the exclusion of the cubic spline approach.

## 4.2 Dose Calculation

Development of computerized dose calculation algorithms is a major reason for the progression of radiation treatment planning to where it is today. This section will provide on overview of the physics of radiation dose calculations and describe the particular dose calculation procedure utilized throughout this work.

Radiation dose deposition in tissue is truly a stochastic (random) process. When considering a single photon it is impossible to say with certainty where its energy will be released due to the inherent randomness of atomic arrangement and electron position at any given instant. However, since a typical radiation treatment involves a huge number (easily trillions) of photons being delivered to the patient, statistical probabilities become

nearly exact, making dose calculation a deterministic quantity for any realistic treatment. This difference in view point for radiation dose deposition has led to development of two distinct approaches to dose calculation, Monte Carlo methods and deterministic methods. Monte Carlo methods are a class of techniques where individual particle paths (photons and necessarily electrons) are simulated using a probabilistic approach (random walks). As the particles are tracked, collisions with nuclei and electrons, as well as other interactions occur, and the resulting energy depositions are recorded. Once a statistically significant number of particles have been traced, the relative energy deposition becomes nearly constant and a converged solution is achieved (a more detailed description is given by Dupree and Fraley [38]). Such approaches are regarded as the most accurate as they are modeling the true physics of the problem in quite a realistic manner. Additionally, experimental results [39, 40] have shown the excellent accuracy that such methods can obtain and the sentiment that Monte Carlo calculations are actually more accurate than most experimental measurements is often observed from those in the field (especially for superficial doses less than $1\,cm\ in$ depth). The major downside to Monte Carlo methods is the relatively large computational times that they require, due to the need to simulate a very large number of particles. This fact alone is enough to render Monte Carlo methods inapplicable for routine treatment planning calculations, especially within the context of IMRT inverse planning, since a large number of competing plans must be evaluated quickly. While much research [41, 42, 43] has attempted to improve of efficiency of Monte Carlo techniques there is currently no approach that renders the process efficient enough for routine IMRT inverse planning with common computational hardware.

Deterministic radiation dose calculations encompass a much wider range of techniques including empirical, semi-empirical, and first-principle based approaches. Empirical approaches are the most simplistic and rely heavily upon experimental or Monte Carlo generated data as kernels for the dose calculation. Semi-empirical techniques utilize a combination of empirical dose kernels and scatter corrections in order to obtain more accurate dose for inhomogeneous tissue. Pencil beam calculations, such as those described by Plessis [44], are the most common semi-empirical schemes. The most accurate and robust of the deterministic approaches are those derived from first principles. Three-Dimensional Superposition/Convolution [45] and Collapsed-Cone Convolution [46] calculations are very commonly used in treatment planning calculations. As expected, the increased accuracy of each of these techniques comes at the added expense of computational performance. Many approaches to treatment planning therefore utilize semi-empirical dose calculations during a majority of the inverse planning process, and limit the use of more accurate methods to final dose calculations or only at selected intervals throughout the planning process. Such approaches are commonplace in treatment planning, however some researchers discount this practice and have attempted to demonstrate the affects of accuracy level during planning on final plan quality [47, 48]. Scholz, Nill, and Oelfke [48] showed a slight improvement in plan quality by using superposition compared to pencil beam calculations during planning. Their work used Monte Carlo simulations to show that the pencil beam solutions were in general no more than 10% less accurate for highly inhomogeneous tissues such as lungs. However, the final treatment plans were not 10% lower in quality when planning was completed with pencil beams versus full Monte Carlo dose calculations. This shows that trends are preserved even

with lower accuracy dose calculations so that inverse planning is not significantly affected (except for the case of lung tissue) by the use of lower accuracy models. This observation helps justify the use of pencil beams throughout the majority of the inverse planning process for many treatment planning situations.

The main purpose of this research is to present and demonstrate a new concept for direct delivery parameter optimization with compensators that includes manufacturing considerations during inverse planning, which, as the above discussion has provided, can be accomplished using a basic dose calculation algorithm. Therefore, simple pencil beam calculations are used throughout, with no corrections for heterogeneous tissues. Also, beam spectral changes and hardening have been included in a fully empirical and qualitative sense only. While these two simplifications are a major concern for realistic treatment planning, they do not detract from the overall goal of this research effort. It should also be noted that incorporation of more accurate and robust dose calculation routines is a relatively minor endeavor once the basics of the DCPO process have been developed and established herein. The pencil beam dose calculation process implemented herein is based upon that provided by Computerized Medical Systems (CMS) for their XiO treatment planning system [32]. The dose calculation is accomplished by summing the dose from each beamlet (pencil beam) to a set of discrete calculation points (voxels) within the patient. Computing the dose from a beamlet to a single point requires knowledge of the *t*otal *e*nergy *r*eleased per unit *ma*ss, *terma*, the lateral dose factor, and the relative beamlet strength. The *terma* is a function of depth below the patient surface, $d$, as given by:

$$terma(d) = 4 - \frac{a_1 \exp\left(-a_2\, d^2\right)}{a_3 + d} + a_4 \exp\left(-a_5 d\right) \tag{4.6}$$

where the coefficients $a_1 - a_5$ may be tuned specifically for a given beam energy and linear accelerator. In addition, this work has used these coefficients to incorporate beam hardening, by allowing each to be a function of the beamlet compensator thickness. It was found that a linear variation of each coefficient produced a reasonable estimation of the effects of beam hardening for the range of interest needed herein. The lateral dose factor is computed based on the beamlet size and shape and the off-axis distance to the point of calculation, $r$, (i.e. the perpendicular distance from the beamlet central axis to the current voxel). The beamlet shape is designed such that the beamlet strength fall-off matches that of the overall beam and that the sum of the beamlets at any point within the beam is unity. The beamlet profiles for $5mm$ spaced beamlets is shown in two-dimensions in Figure 4.9, where the dashed curves are individual beamlets and the solid red line represents the overall beam. The lateral dose for such beamlets is given by:

$$LatDose(r) = \begin{cases} \sinh\left(B\,H\right)\, \exp\left(-B\,r\right) & if\ r > H \\ 1 - \cosh\left(B\,r\right)\, \exp\left(-B\,H\right) & otherwise \end{cases} \tag{4.7}$$

where $B$ is a constant parameter used to control beamlet shape and $H$ is the halfwidth of the beamlet (radius at which beamlet strength is 50%). The relative beamlet strength, $S$, is computed from Equation 4.1 as the ratio $I/I_o$ using the thickness at the center of the beamlet, which is found by evaluating the analytical surface at this location. Finally, the dose from a beamlet to a single point, $p$, is computed from the following expression:

$$dose(p) = terma\,(d(p))\ LatDose\,(r(p))\ S \tag{4.8}$$

where $d(p)$ and $r(p)$ are computed geometrically as previously described.

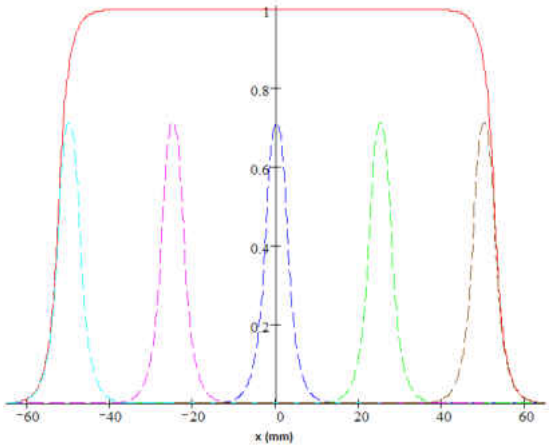
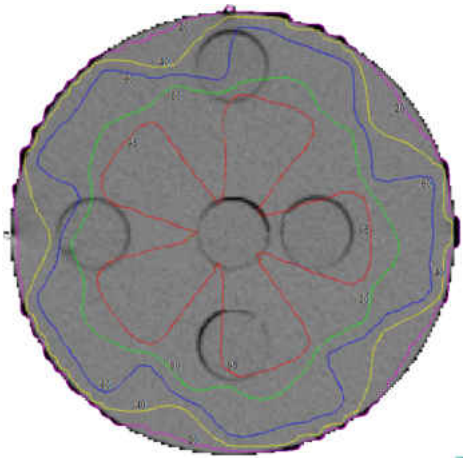


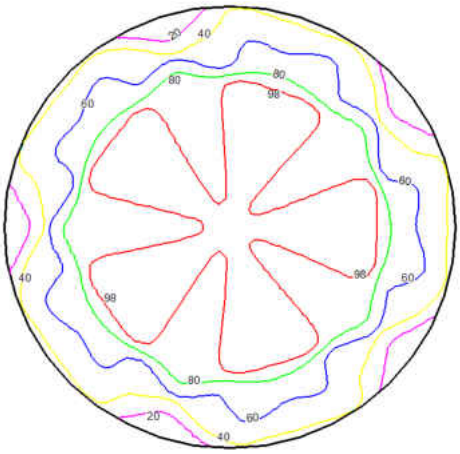

Figure 4.9: Beamlet Profiles in Two-Dimensions

Despite the simplistic nature of the dose calculation algorithm used for this work, the quality of the solutions for homogeneous tissue is very good. Figure 4.10 qualitatively demonstrates the agreement of dose calculation for the pencil beam approach described above with that of the Philips Pinnacle Treatment Planning System version 8.1. This figure provides a dose comparison on a cylindrical water phantom ($8\,cm$ radius) using 5 coplanar equally spaced treatment fields. Each field is $10\,cm \times 10\,cm$ and is blocked on the left half ($x < 0$) by a $3\,cm$ thick brass compensator. Dose has been normalized to 100% at the isocenter and contour lines for 98%, 80%, 60%, 40%, and 20% relative dose are shown for Pinnacle and this work (DCPO) separately in Sub-Figures 4.10a and 4.10b, and overlaid in Sub-Figure 4.10c (DCPO dose contours are shown as broken/dashed lines in Sub-Figure 4.10c). The general trends in dose are quite well preserved, which provides the confidence

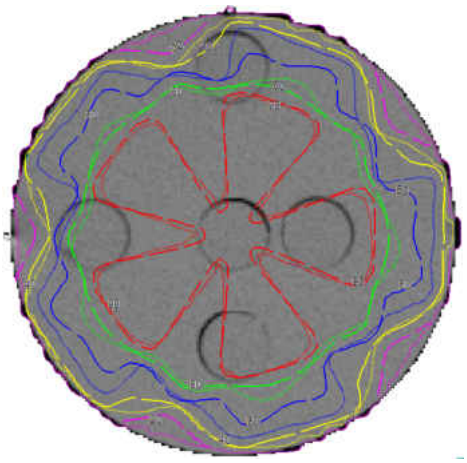
needed to ensure that the conclusions drawn in this research would be equally applicable if a more accurate dose calculation were utilized.



(a) Philips Pinnacle TPS Dose Contours




(b) DCPO Calculated Dose Contours



(c) Isodose Contour Overlay

Figure 4.10: Dose Calculation Comparison: Five Beam Half Blocked *3cm* Compensators

## 4.3 Inverse Planning and Optimization

In basic terms, intensity modulated radiation therapy is simply an improved means of planning and delivering highly conformal radiation treatments. What makes IMRT unique is that the strength of the individual treatment beams is varied in order to achieve this conformity. The addition of this intra-beam intensity variation greatly increases the difficulties of planning, as the number of variables to control tends to become quite large. This large number of variables (control parameters) makes traditional planning (i.e. direct trial-and-error via experience and intuition) almost impossible. Therefore IMRT requires a computerized planning approach where the modulation parameters are determined via an intelligent searching process. This is commonly referred to as *inverse planning* since the process involves determining treatment parameters necessary for achieving a desired solution (dose prescription). Traditional inverse planning processes use numerical optimization algorithms to perform the necessary search over the control parameter space. Many options exist for the numerical optimization, but most choices can be lumped into the two broad categories of gradient and stochastic methods. Gradient based schemes are typically regarded as more efficient, however they lack the ability to find global optimums as they are forced to travel in a purely downhill fashion. Therefore, when the solution space contains multiple minima, as shown in Figure 4.11, gradient schemes tend to get "stuck" at the lowest point of whichever hill their initial starting point landed upon. In contrast, stochastic (non-gradient based) methods are more likely to find the true global optimum, as they are not generally forced to maintain only downhill motions; this comes at the expense of increased

run times however. Regardless of the chosen optimization technique, a major difficulty with IMRT inverse planning comes from the relatively small correlation that exists between the strengths of adjacent beamlets. This problem is best described through illustration, provided by Figure 4.12, which demonstrates the lack of independence of some optimization parameters (due to fact that multiple beamlets provide dose to same region) and well as the lack of interdependence (beamlets have no knowledge of the values of their neighbors). The parametrization of DCPO significantly increases both the interdependence of neighbors which actually increases the independence amongst the multiple beams. This helps to improve the efficiency and performance of most optimization schemes, and provides justification for exploration of schemes which currently perform poorly for beamlet based optimization.

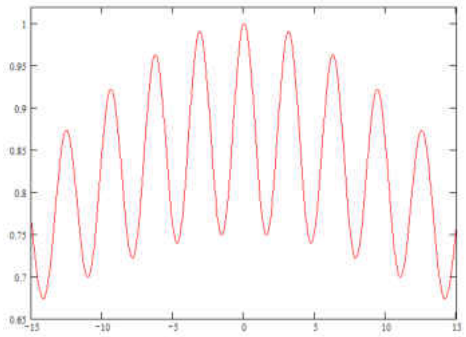


Figure 4.11: Global vs. Local Solution Optimization

Current commercial and academic planning systems use both classes of optimization, however two methods seem to be particularly common, namely conjugate gradient and simulated annealing. Conjugate gradient is a gradient based approach which is an improvement over the steepest descent algorithm. Steepest descent involves determining the local optimal solution by starting from an initial guess and moving in the direction of the highest gradient to find the next (hopefully better) solution. The process is repeated until the local gradient

reaches a very small value, thereby indicating the presence of a local minimum/maximum [36]. Conjugate gradient improves upon this procedure by using information from previous steps along with the local gradient value to determine a more appropriate search direction for each step. The popular commercial system XiO by CMS uses a process quite similar to this conjugate gradient scheme in their implementation [32].

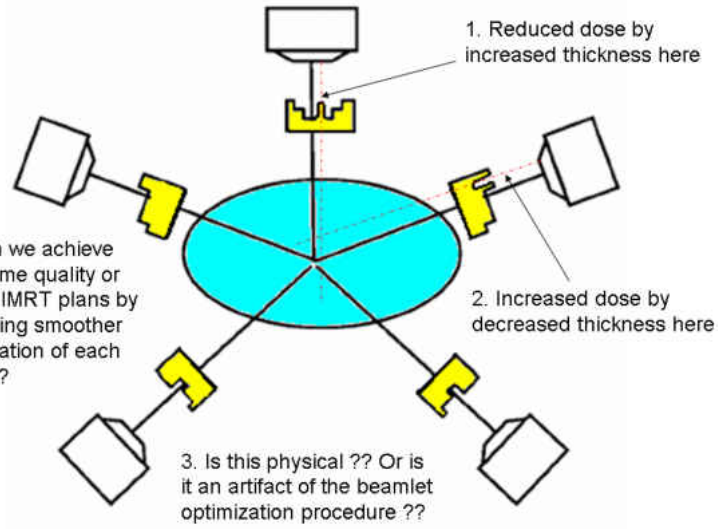



Figure 4.12: Intensity Profile Correlation Sensitivity

Simulated annealing is a non-gradient optimization method which is based upon an analogy to the manufacturing heat treatment process of annealing. The real annealing processes involves controlled cycles or stages of heating and cooling a material, starting at a high temperature. The material is held at various temperatures along a cooling path in order to promote relaxation of internal stresses by allowing controlled recrystallization of the atoms. Crystalline structures appear because they are the lowest energy state for the atoms to arrange themselves in for certain materials, so annealing is a natural energy minimization process and a good basis for an algorithmic optimization scheme. Simulated annealing attempts to follow this idea by starting from some initial guess value and testing

the "energy" (quality of solution) of various small perturbations in the current solution. Based on the current "temperature" (process control parameter) and the change in energy, the solution is either accepted or rejected [49]. This is repeated while slowly lowering the temperature, which adjusts the acceptability criteria for the new trial solutions; as the temperature decreases it becomes less likely that a higher energy trial solution will be accepted. This ensures that the process will proceed in a nearly downhill fashion as the temperature becomes low, thereby facilitating convergence of the algorithm. Generation of the trial solutions, regardless of temperature or energy, is generally undertaken so that only small steps are permissible, a process which favors convergence but limits the ability to "jump" to significantly better solutions. While this technique does not often achieve a true minimization, it is typically able to reach nearly optimal results even for problems where a large number of control parameters are present [50].

Another often utilized IMRT optimization approach (more so in the academic setting) is Genetic Algorithm (GA) optimization. Genetic algorithms were developed in the 1970's by John Holland and have since been highly developed, described, and researched by countless authors. A review of GA details and implementation is provided by Davis [51], but the main premise is to evolve a set (population) of trial solutions (individuals) through operations analogous to human reproduction (mutation and crossover) in a manner such that the most "fit" individuals have the greatest likelihood of reproducing. In the absence of crossover (i.e. pure mutation) the GA can be shown to be very similar to certain forms of simulated annealing [52]. Therefore crossover is the primary difference between the operation of these two techniques and its importance is often noted in the GA literature. Genetic algorithms

are typically considered an effective means of locating a globally optimal solution, although the computational burden to reach such a point is often much larger than other techniques, especially when the number of search parameters is large [53].

As there is no general consensus within the IMRT community as to which optimization scheme is preferable, and also due to the significant parameter correlation sensitivity changes expected from DCPO, several optimization schemes have been explored for this work. Specifically, the performance of genetic algorithms, downhill simplex, and several steepest descent variations have been analyzed. The motivation for choosing each scheme and a brief description of each is presented in the remainder of this section and then results of a detailed performance study comparing each approach are provided in the following chapter.

As described, the surface representation used by DCPO increases the optimization parameter interaction and sensitivity, since unlike a beamlet based approach, each control point affects a region of the surface, as opposed to an isolated location. This increase in parameter inter-dependence (or abstraction) has generally been shown to improve the performance of genetic algorithm optimization [54]. Therefore, GAs should be more likely to perform well for DCPO than for traditional beamlet strength optimization. Additionally, due to the often reported similarity between GA and simulated annealing performance [52] and this author's familiarity with genetic algorithms [33, 34, 35], they have been chosen for exploration herein, exclusive of simulated annealing. Several gradient based optimizers have also been implemented and tested for this work. The motivation for exploring these techniques is the vast body of research and literature that demonstrates the exceptional convergence rates of these approaches (see [36, 55, 56] for example). Due to these fast

convergence rates, acceptable solutions can often be achieved in only a handful of iterations. Therefore, several variations on the classic steepest descent algorithm have been investigated, including conjugate gradient and variable metric optimization. The downside of typical gradient based techniques is unfortunately the computation of the gradient itself. Therefore, gradient approximation techniques, such as Broyden's method, have also been utilized herein. The non-linear downhill simplex method of Nelder and Mead [57] is a gradient-like approach, where direct computation of the gradient is avoided, and instead a simplex ($n+1$ dimensional surface) is allowed to "crawl" along the solution space in a downhill fashion. This technique has been shown to effectively determine local minima with no need to explicitly compute the gradient [36]. With the motivations established, the details of each approach are addressed after a description of the objective function calculation.

Regardless of the type of optimization performed, a means of comparing the quality of each trial solution is necessary. This is accomplished via a fitness function that specifies the relationship between the input parameters and the quality of the solution. For the case of IMRT treatment planning with DCPO the input parameters are the following (for each treatment beam): 1) overall beam weight, 2) matrix of control point heights, and 3) matrix of control point weights. These input parameters are sent to the NURB surface calculation routines described in Section 4.1, which facilitates calculation of the individual beamlet heights. The beamlet heights are then sent to the dose calculation module which computes the resulting dose inside the discretized patient geometry. Finally, the resulting dose at each point is compared with the prescription and the following expression is used to determine a single value for the objective (fitness) function, $\phi$:

$$\phi = \sum_{j=1}^{N} \left\{ \beta_j \sum_{i=1}^{n_j} \left[ (D_i - D_{p_i})^2 \right] \right\} \tag{4.9}$$

where $N$ is the number of important anatomical structures (PTVs and OARs), $n_j$ is the number of discrete points (voxels) within a particular structure, $\beta_j$ is the importance weight for a particular structure, $D_i$ is the computed dose at voxel $i$, and $D_{p_i}$ is the prescribed dose at the same point. It should be noted that this prescribed dose may be a range rather than a single value in which case $D_{p_i}$ takes on the maximum or minimum allowable value, as appropriate. Additionally, in order to make the OAR doses as small as possible and the tumor as homogeneous as possible, a linear contribution is added to the objective function even when the dose, $D_i$, falls within the prescribed range. This new objective function follows that of CMS's XiO planning system [32] and is illustrated for a sample PTV and OAR in Figure 4.13. Additionally, in order to facilitate the use of gradient approximation techniques, such as Broyden's method [36], the form of Equation 4.10 has also been modified, providing a new objective function in the form:

$$\phi = \textstyle\sum_{j=1}^{N} \left\{ \sum_{i=1}^{n_j} \left[ f_j^2(x_i) \right] \right\} \tag{4.10}$$

$$with$$

$$linear = |D_i - D_{p_i}| \qquad over = D_{max_i} - D_i \qquad under = D_i - D_{min_i}$$

$$f_j(x_i) = \begin{cases} \sqrt{\frac{\beta_j}{n_j} (linear + over^2)} & if\ D_i > D_{max_i} \\ \sqrt{\frac{\beta_j}{n_j} (linear + under^2)} & if\ D_i < D_{min_i} \\ \sqrt{\frac{\beta_j}{n_j} (linear)} & otherwise \end{cases}$$

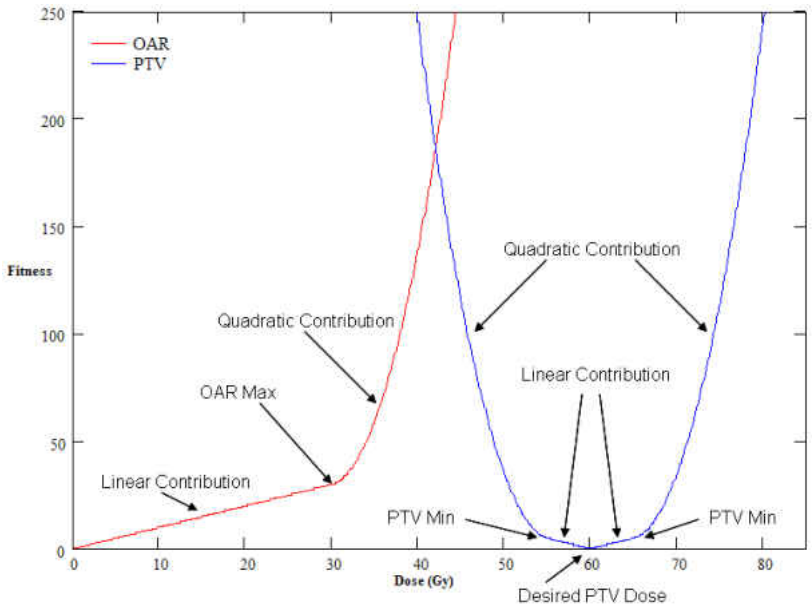



Figure 4.13: Optimization Fitness Functions (PTV and OAR)

The focus of many classic papers on genetic algorithms has been achieving the property of robustness, so that performance is maintained over a wide range of problem types. This is often a worthwhile goal, as the majority of optimization problems generally require only one solution or set of solutions; especially for standard design optimization problems where only one final design is needed. However, it is quite intuitive that specialization can allow performance improvements for a particular class of problems, at the expense of robustness, and many authors (see [51, 58] for example) have corroborated this idea. Such specialization is very important herein as IMRT treatment planning falls into the rare class of problems where numerous optimization runs are needed, since every patient requires an optimization for a highly similar situation; thus the genetic algorithm developed for this work has been tailored to perform specifically for radiation planning problems through a rigorous development and tuning process. This tuning process has led to a final form of

the GA in which real number representation is utilized, along with linear fitness scaling, single-point cross-over, and a diploid style chromosome without dominance (phenotypes are computed as the average of the gene value from each chromosome). A steady-state reproduction style is used in which a small number of children ($\approx 10\%$ of the population size) are generated and replace the worst individuals in the population. This ensures that the best individuals are maintained (elitism) and eases the implementation of disallowing duplicate individuals. Both sexual and asexual forms of reproduction have been utilized. Sexual reproduction involves each parent sending one complete chromosome to the new child. The value of this chromosome is determined in one of three ways: directly send one of the parent chromosomes, crossover the two chromosomes from a parent and send one of the resulting new chromosomes, or directly send the parent phenotype (average) values. Asexual reproduction is accomplished via three mutation operators: traditional random mutation which acts on individual genes, "single creep" mutation which increments a single gene by a small amount, and "multi-creep" mutation which increments the entire chromosome in a problem specific manner. The mutation rates for each operator are adjusted in a linear fashion as the solution progresses in order to achieve a good balance between global and local searching. Random mutations are slowed and creep is increased throughout each run.

Steepest descent is a more straight-forward technique and is the most fundamentally simple gradient based optimization approach, as described above. Each additional gradient technique used in this work is a direct modification to this basic algorithm, therefore, the exact approach to steepest descent taken herein will now be explained in more detail. The process starts with some initial trial solution that can be random, based on prior solution

knowledge, or constructed using geometric rules. The gradient is then calculated numerically using a finite differencing approximation. The gradient of a scalar dependent variable, $\phi$, in *n-dimensional* space is given as:

$$\nabla\phi = \sum_{i=1}^{n} \left[ \frac{\partial \phi}{\partial x_i} \overrightarrow{e_{x_i}} \right] \tag{4.11}$$

where $\overrightarrow{e_{x_i}}$ is a unit vector in the $x_i$ direction. The necessary partial derivatives are approximated using the following one-sided finite difference expression:

$$\frac{\partial \phi}{\partial x_i} \approx \frac{\phi\left(x + \alpha_j\left(r, i\right) \triangle x\right) - \phi\left(x\right)}{\alpha_j\left(r, i\right) \triangle x} \tag{4.12}$$

where $\triangle x$ is some small preset perturbation value $[\Theta\left(10^{-3}\right)]$, $i$ is the current coordinate direction for differentiation, $r$ is a uniformly distributed random number ranging from zero to one, and:

$$\alpha_j\left(r, i\right) = \begin{cases} 1 & if \ \left(\left(r \leq 0.5\right) \& \left(i = j\right)\right) \\ -1 & if \ \left(\left(r > 0.5\right) \& \left(i = j\right)\right) \\ 0 & otherwise \end{cases} \tag{4.13}$$

This form accomplishes a forward difference approximation for $r \leq 0.5$ and a backward difference otherwise, ensuring that the calculation of derivatives does not bias the solution process (central differencing is avoided as this would double the workload of computing the gradient vector). For the case of IMRT treatment planning, $\phi$, is directly the fitness value obtained from the formulation given by Equation 4.10 and $x_i$ corresponds to each of the input parameters. Once the gradient is known, steepest descent utilizes the following equation to

move the trial solution to the local minimum in the direction opposing the gradient:

$$x_i^{(k+1)} = x_i^{(k)} - L^{(k)} \ (\nabla\phi)^{(k)} \tag{4.14}$$

where $L$ is determined via one-dimensional minimization along the gradient direction (an inverse parabolic interpolation approach [36] has been utilized herein). Additionally, a gradient approximation scheme (Broyden's method [59]) has been implemented to eliminate the need to compute the full gradient (via finite differencing) at each iteration (the full gradient is computed via finite differencing only for the first iteration and at some preset iteration counts). For Broyden's method the gradient is approximated using the following two relations:

$$(\nabla\phi)^{(k)} = \left(2\, J^T f\right)^{(k)} \tag{4.15}$$

where

$$J^{(k)} \approx J^{(k-1)} - \frac{\left[J^{(k-1)}\left(x^{(k)} - x^{(k-1)}\right) - \left(f^{(k)} - f^{(k-1)}\right)\right]\left(x^{(k)} - x^{(k-1)}\right)^T}{\left(x^{(k)} - x^{(k-1)}\right)\cdot\left(x^{(k)} - x^{(k-1)}\right)} \tag{4.16}$$

and the initial Jacobian matrix, $J$, is computed using finite differencing (analogous to that shown above) such that:

$$J_{ij} = \frac{\partial f_j(x_i)}{\partial x_i} \approx \frac{f\left(x + \alpha_j\left(r,i\right)\triangle x\right) - f\left(x\right)}{\alpha_j\left(r,i\right)\ \triangle x} \tag{4.17}$$

Despite the intuitive nature of steepest descent it is actually quite inefficient, because the current iteration utilizes only the local gradient information and takes no account of prior search directions. Also, as the previous trial solution was found by minimizing along the

gradient direction, the next step is guaranteed to have the gradient direction perpendicular to the previous. Conjugate gradient was developed to alleviate this concern by ensuring the new search direction is conjugate to the previous directions as much as possible. This generally results in a more efficient and less oscillative iteration process compared to steepest descent [36]. A similar scheme, the variable metric approach, is often shown to provide slightly more stability and efficiency in the selection of search directions [56]. Both these processes have therefore been implemented and tested herein and no consistent benefit was found from either approach, therefore only conjugate gradient has been considered for further investigation.

The non-linear downhill simplex method is based upon the geometric entity known as a simplex, which is an *n-dimensional* polytope, consisting of *n+1* points in $n$-dimensional space. In two-dimensions a simplex is a triangle, and the downhill simplex optimization process is easily visualized as a triangle being flipped, extended (stretched), and contracted (shrunk) in an effort to "crawl" down the surface to reach the minimum point. This crawling is accomplished by attempting to "flip" the simplex with the worst vertex moving along the line connecting it with the centroid. The resulting new vertex is then accepted or rejected based on a set of rules regarding its relationship with the other vertices. If acceptable the search may be extended along this direction in an attempt to move further down the hill. If the process is rejected, the simplex can contract or shrink toward the best vertex in an attempt to squeeze into small local minima. The process generally converges more slowly than gradient based schemes, but the benefit of not requiring calculation of the gradient makes the overall process more computational efficient in some situations [36]. The standard

form of the non-linear downhill simplex method presented by Nelder and Mead [57] and reviewed in [36] has been implemented herein.

This section has provided the details of each numerical optimization approach selected for analyzation herein. The motivation for each selection was also presented and the following chapter will continue with completion of the full performance analysis for each of the three techniques.

## 4.4 Manufacturing

This work proposes to improve the *delivery* of IMRT by incorporating the specific constraints of the milling process into the inverse planning optimization. Incorporating machining considerations into the planning process is unique in that it closes the design loop and allows for greater consistency between actual delivered radiation dose and planned radiation dose compared to current methods. This is made possible because the DCPO approach developed herein is a direct delivery parameter optimization approach, whereby the actual compensator shapes are known during inverse planning. During inverse planning most treatment planning systems, both academic and commercial, consider only ideal fluence modulation (DAO approaches being the exception) and do not consider whether a compensator is capable of being manufactured to achieve the prescribed modulation (or if the plan is deliverable for MLC based techniques as MLCs also have some delivery limitations). Furthermore, most current techniques generally require a smoothing algorithm to ensure that the compensator surface is not jagged or stair cased. This smoothing changes the compensator shape and

with it the delivered dose. Thus smoothing necessitates an additional dose calculation step to ensure that the delivered dose has not been excessively degraded by the smoothing process. It is also not uncommon that even after smoothing, the compensator design produced by a treatment planning software is simply unable to be manufactured to the desired specifications (generally because the smoothing algorithms are not based on a particular manufacturing process). These issues with the current design process are a large contributor to discrepancies between the optimal planned dose, the calculated deliverable dose, and the actual delivered dose. Ensuring that the DCPO based planning system uses only 100% manufacturable compensator designs effectively closes the compensator design loop and helps decrease these differences between planned and delivered radiation doses.

This discussion has shown that two major benefits arise from the incorporation of manufacturing into the planning loop: (1) complete agreement between each trial plan dose and deliverable dose (which also removes the need for secondary deliverable dose calculations) and (2) increased agreement between final (optimal) planned and actually delivered radiation dose. The effects of manufacturing inclusion on the planning process can seen by comparing the flow chart in Figure 4.14 to that of the standard inverse planning process chart given previously in Figure 3.9 (see page 32). This figure shows the simplification offered by this work through elimination of the secondary loop in the current process (Figure 3.9) by moving this step into the main design loop with the DCPO process (Figure 4.14).

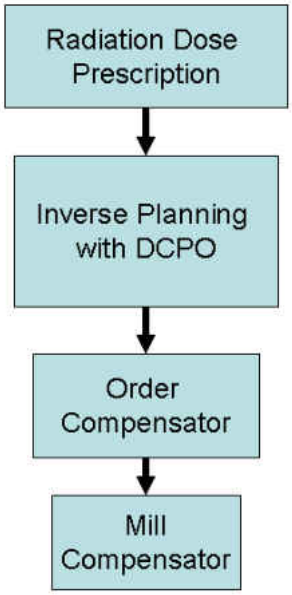



Figure 4.14: New DCPO Compensator Design and Manufacturing Processes

Consideration of compensator machining constraints in this work is undertaken in an indirect fashion. An indirect approach is needed because performing a complete milling simulation on the large number of trial solutions that must be generated and tested during the inverse planning process would be prohibitive. Manufacturability is therefore addressed by imposing limitations on the control parameters used by the compensator surface representation scheme. Appropriate limitations are determined specifically for this work by utilizing milling surface limitations as well as through the use of milling simulation software from compensator manufacturing firm .decimal, Inc. The control parameter limitations are found through an analysis of various compensator designs that contain many features which commonly create difficulties for the milling process, such as steep walls and deep valleys. Deep valleys can lead to under-cutting, see the shaded blue region in Figure 4.15, while steep walls can necessitate the use of five-axis milling (when wall steepness exceeds the slope of the

tapered milling tools). The simulation results and tooling constraints are used to determine bounds for the control parameters such as minimum control point spacing and allowable control point weight ratios. These bounds are used within the optimization portion of the inverse treatment planning process to efficiently enforce manufacturing constraints during treatment planning. The results of this machining study, as well as a thorough discussion of the implications on surface control, are provided in the following chapter.

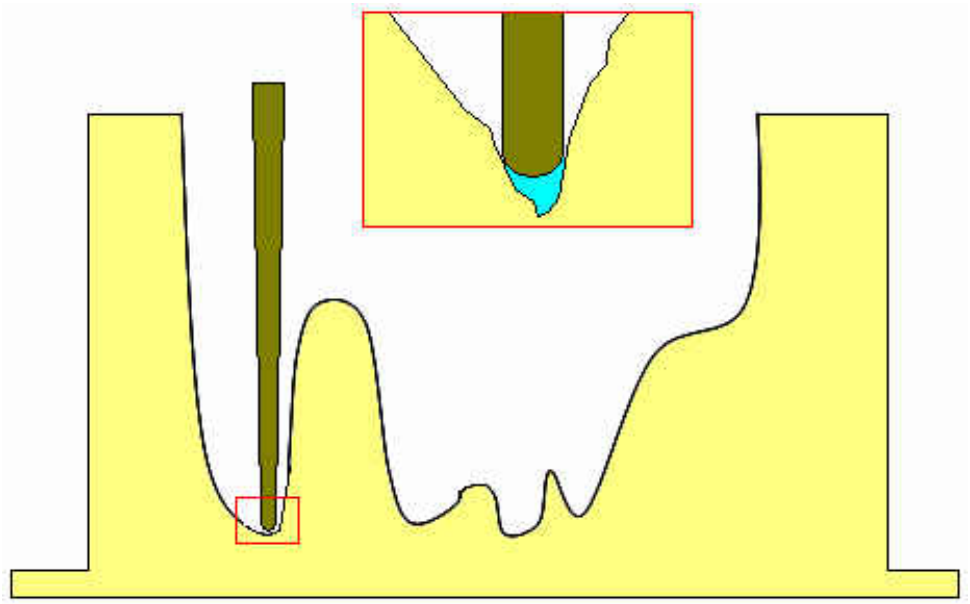


Figure 4.15: Milling Tool Limitation Example

## 4.5 Summary of DCPO

As described above, the process of Direct Compensator Profile Optimization involves several novel ideas and contributions to the field of compensator intensity modulated radiation therapy, including being the first direct delivery parameter optimization scheme for compensator based IMRT treatment planning. The above descriptions have been necessarily broken up, so that the details of each development could be sufficiently explained. This section therefore provides a brief yet complete description of DCPO so that synthesis of the various processes is achieved. A basic flow chart describing the DCPO approach is therefore shown in Figure 4.16.

In summary, DCPO utilizes an analytical surface representation scheme to describe the compensator surfaces (thicknesses) during inverse planning. The analytical surface allows for calculation of individual beamlet strengths. At this point, the process directly follows traditional planning, including the use of standard pencil beam dose calculation algorithms (with appropriate beam hardening corrections) and an iteration loop which can be completed using one of several different optimization schemes. Figure 4.16 highlights the similarity between the DCPO process and standard approach as far as the basic treatment planning algorithm is concerned, with the only high-level difference being the solution representation used in the first two steps. This similarity is quite beneficial as it will allow incorporation of DCPO into existing planning systems with only isolated modifications and additions by later investigators. However, what is difficult in such a high-level description is understanding the fundamental benefits that DCPO offers and provides at each of the stages (boxes)

indicated in Figure 4.16. In the first stage, DCPO provides a reduction in the number of parameters needed to specify the fluence field. The second stage is new to DCPO as traditional approaches optimize the beamlet strengths themselves. Stage three may be performed identically to standard schemes, however, since DCPO provides physical delivery device parameters (i.e. compensator thicknesses) at this time, there is no penalty or added difficulty for including beam spectral changes and compensator-induced scatter at this point in the planning dose calculations. The fourth and fifth stages are no different from existing approaches as the fitness and quality of a treatment plan are judged solely on the patient dose distribution. The final stage again appears to match existing techniques, however with DCPO there is no need for a final dose calculation or quality assurance feedback loop (due to smoothing) as these issues are directly handled in the trial solution representation.

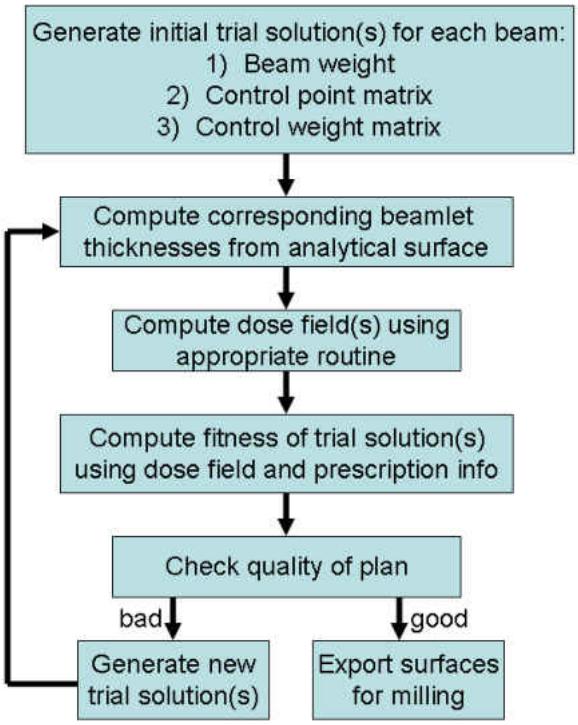



Figure 4.16: Process Flow Chart for DCPO

# CHAPTER 5
# RESEARCH DEVELOPMENTS

The process of Direct Compensator Profile Optimization has been explained and the potential benefits thoroughly described. This chapter therefore focuses on the procedures undertaken to test this new methodology, verify its performance, and determine the most appropriate and effective schemes to utilize for each particular stage of DCPO. This procedure began with a preliminary implementation of the complete DCPO process in a simplified two-dimensional sense such that the fundamentals of the approach could be tested and verified. As a goal of DCPO is to be applicable to realistic patient configurations, complete development and implementation of each of the schemes described above was then performed. Testing and verification was completed in various stages throughout the process in order to ensure each new development was performing as expected. This chapter presents results in a manner consistent with these verification procedures, thereby illustrating the development process and showing the rigor used to ensure proper functioning and performance of each of the DCPO routines. This thorough testing was used to assess the performance, ascertain any areas of deficiency, and aid in the refinement of the solution approaches for each part of the DCPO process. This testing also addressed the issue of manufacturing constraint determination and several example cases are shown which provide a means for quantification of the surface control parameters necessary to ensure complete surface machinability.

## 5.1 Initial Process Verification

The initial process verification began with implementation of cubic spline surfaces for compensator representation, a basic dose calculation module, and genetic algorithm optimization routine. The goal of this initial stage was only to prove the soundness of the DCPO concept (efficiency or accuracy were not addressed at this point), therefore the choice of cubic splines and genetic algorithms was made purely for convenience as the author has considerable experience with these procedures [35, 34, 60, 33]. Two simulated patient test cases were completed so that initial verifications could be made before moving on to more rigorous developments and testing. The details of these two test cases are presented in the remainder of this section.

### 5.1.1 Verification Case #1

The first test case consists of a simple circular patient skin surface with an off-center target volume (PTV) and two symmetrically placed organs-at-risk (OARs). The geometry of this simulated patient is shown in Figure 5.1. The optimization was carried out by setting the cost (objective) function based on maintaining the PTV relative dose range between 100%-94%, while minimizing the dose to the OARs and rest-of-body (note that the dose fields here have been normalized to produce a maximum of 100%, therefore the target PTV dose is 97% with 100% being the maximum allowable value). Results were obtained using 7 and 9 equally spaced treatment beams. The resulting relative dose fields for each case are shown in Figure 5.2, as well as the resulting compensator shapes for the nine beam case in Figure

5.3. A Dose Volume Histogram (DVH) is then provided in Figure 5.4 which allows for a more direct quantitative analysis of the resulting treatment plan.

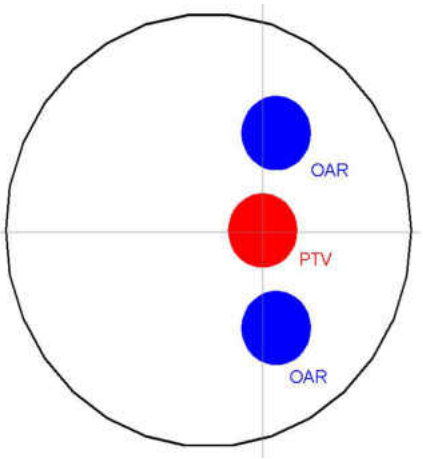
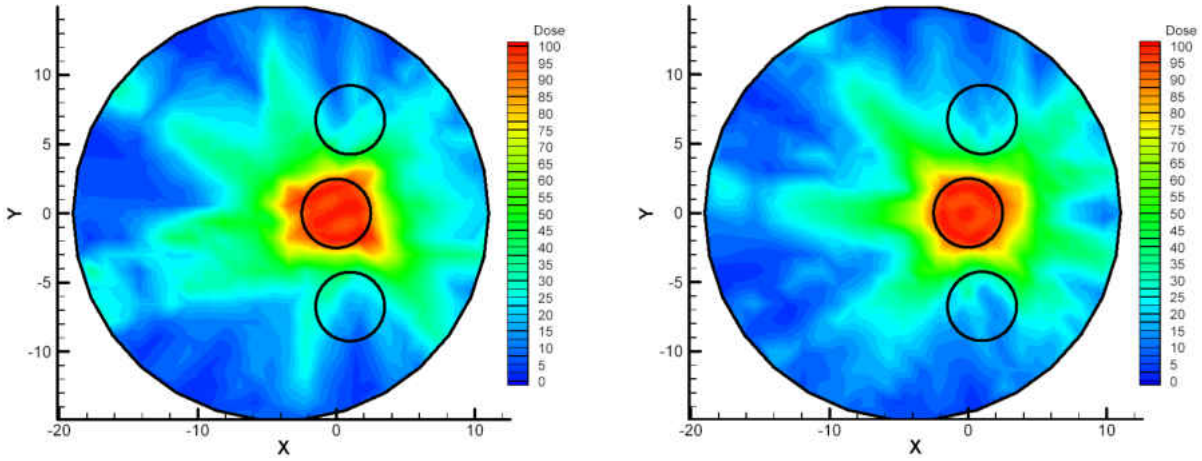



Figure 5.1: Verification Test Case #1: Simulated Patient Geometry




(a) Seven Beams

(b) Nine Beams

Figure 5.2: Verification Test Case #1: Dose Field Results

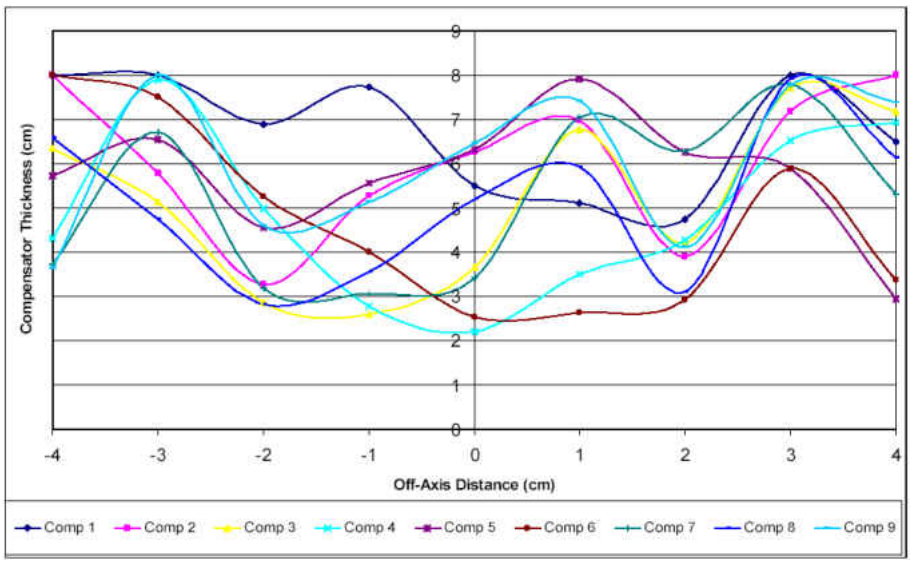



Figure 5.3: Verification Test Case #1: Compensator Results

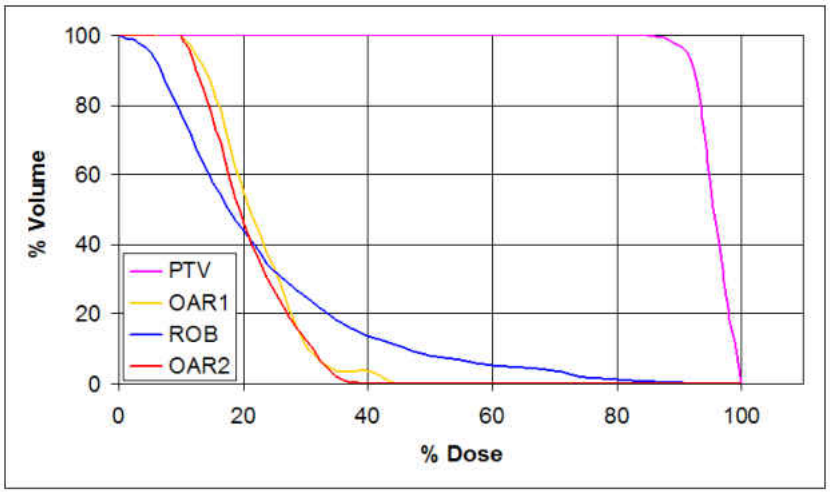



Figure 5.4: Verification Test Case #1: Dose Volume Histogram

### 5.1.2 Verification Case #2

The second example case consists of a "C" shaped target volume adjacent to a small round organ-at-risk, to simulate a more complex spinal region treatment. The simulated patient geometry is shown in Figure 5.5 and the resulting dose field for nine equally spaced treatment

beams can be seen in Figure 5.6. The resulting compensator shapes for each beam are shown in Figure 5.7. Again a relative dose of 100%-94% was used for the PTV and minimization of dose was used for the OAR and rest-of-body. Figure 5.8 shows the DVH plot, showing the reasonable quality of the DCPO planned solution.

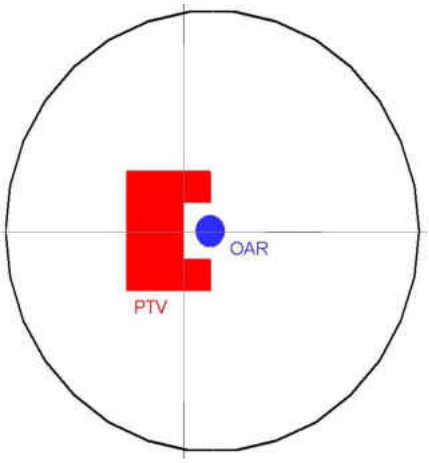



Figure 5.5: Verification Test Case #2: Simulated Patient Geometry

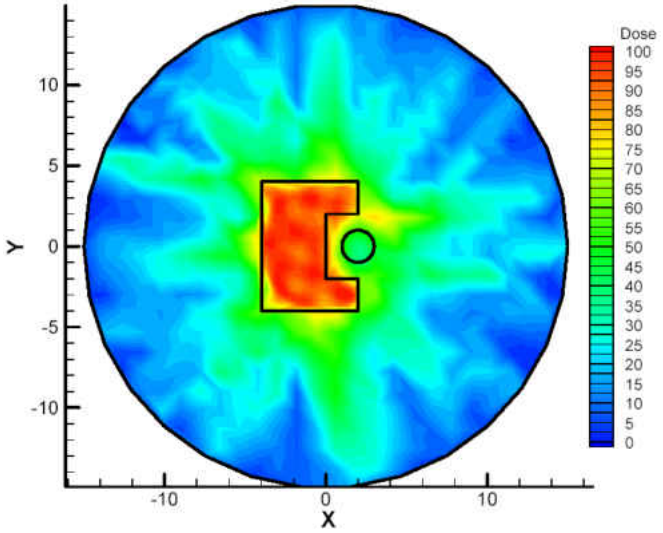



Figure 5.6: Verification Test Case #2: Dose Field Results

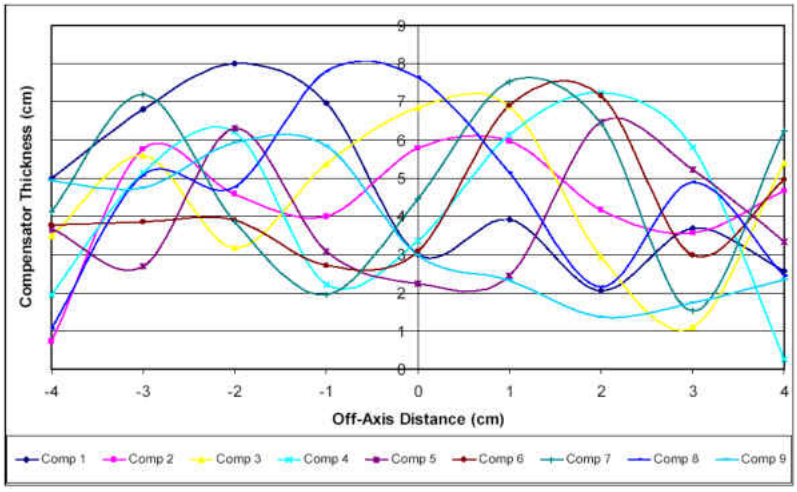



Figure 5.7: Verification Test Case #2: Compensator Results

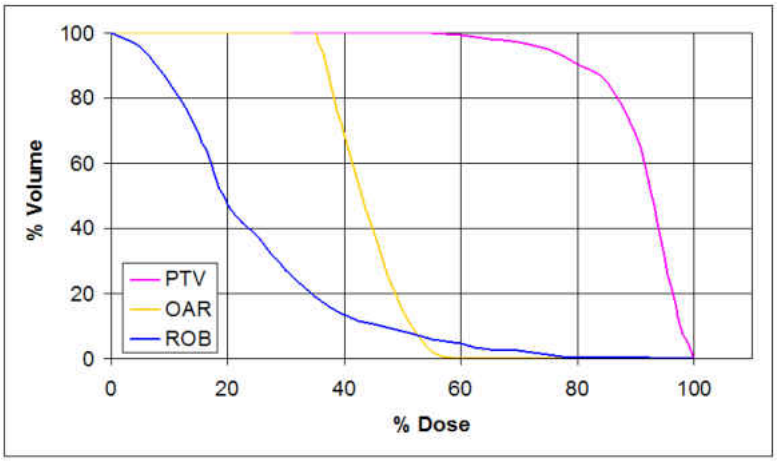



Figure 5.8: Verification Test Case #2: Dose Volume Histogram

After completion of these results it was determined that inclusion of an overall beam weight factor was necessary within the optimization scheme as an added control parameter. An inspection of Figures 5.3 & 5.7 reveals that most of the compensators do not provide modulation over the entire range of interest; in other words single beams do not span from zero thickness to full thickness. This lack of minimum thickness region on most of the beams points to the fact that the compensator is providing the overall beam weighting by limiting the minimum thickness of the surfaces. This is undesirable as it limits the useful modulation

range of the devices. Therefore, the remainder of this work has included an overall beam weight factor for each individual beam (in addition to the standard surface control parameters described previously), which specifies a temporal weighting and was found to eliminate this concern and significantly improve the resulting treatment plans.

Although, these cases are reasonably simplistic in nature they have proven that the DCPO process is sound and that the parametrization scheme does not prevent the ability to produce reasonable treatment plans. It should be understood that these results are provided only to demonstrate the preliminary verification of the DCPO process. Therefore, the quality of these plans should not be overly concerning at this time (as each of the competing processes are further developed below it will be shown that much higher quality results can be achieved with DCPO). Overall, this initial effort provided motivation for continuing development into DCPO and gave insight into the areas which needed extra attention as the research moved forward. The next stage of this research required the complete development and implementation of each of the various schemes described in the previous chapter so that efficiency and accuracy can be addressed within the DCPO process.

## 5.2 Manufacturing Considerations

The determination of manufacturing constraints is a key aspect and major innovation of this work. As such, a thorough performance study has been completed in order to address this area. As described previously, the two major concerns when compensators are manufactured using a milling procedure are under-cutting and surface slope (wall steepness). In order to

prevent problems in these areas from arising in this work, the NURB surface parameters will be controlled through a bounding and limiting scheme. Before this scheme was developed and the bounds/limits quantified it was important to ascertain the typical values which limit the surface machinability. Typical milling limitations were obtained from .decimal, Inc. as they are a leading manufacturer of solid metal compensators. Their standard manufacturing process utilizes three-axis milling with a ball nose (rounded) minimum tool tip diameter of $1/8''$ ( $0.3175\,cm$ ). Additionally, as typical compensators contain quite deep valleys, this tool generally has a tapered shaft to increase its strength and stiffness (such a tool is shown in Figure 5.9). As this is the tool used for finishing operations, it is the smallest tool used and is the main limitation within the milling process. The dimensions of this tool, along with a small safety factor, directly provide the necessary milling constraints. Table 5.1 quantifies the values which are typically employed in manufacturing facilities. In addition to these milling constraints, typical linear accelerators have limitations on the sizes of compensators which can be placed into the treatment head tray. These limitations are also included in Table 5.1.

Figure 5.9: Typical Machine Tool Used for Final Milling

Table 5.1: Milling Constraints and Limitations

| | Minimum Radius | Taper | Maximum Height | Minimum Height |
|---|---|---|---|---|
| Tool | 0.159 cm | $2^o$ | 10 cm | – |
| Surface | 0.159 cm | $3^o$ | 8 cm | 3 mm |

Once the appropriate machining limitations had been found it became necessary to determine the corresponding NURB control parameters that would prevent violation of the limitations (only NURB surfaces were considered during this study as the inapplicability of cubic splines to DCPO has already been described). The most obvious and simple means of satisfying the tray size limitations is to limit the control point values (z-direction heights) between the maximum and minimum allowable thickness values. This is possible due to the bounding property of the NURB surfaces, which means that actual surface thicknesses will *always* remain within the bounds of the control point range.

The remaining two constraints are a bit more challenging to uphold and a two fold approach is used to address both constraints simultaneously, by limiting the x and y spacing of the control points as well as the control weight ratios. An initial estimate on the minimum allowable control point spacing, with regards to surface steepness only, was found by assuming that the surface is locally linear and that it passes directly through each control point. This means for the given maximum and minimum surface heights as well as the surface angle limitation, a simple trigonometric calculation provides a value of $0.419\,cm$ as the absolute lowest allowable horizontal spacing. However, this estimate is quite rudimentary and due to the complex relations between the NURB surface and the control point heights and weights, the surfaces can actually be steeper than this elementary linear estimate provides. A simple half blocked compensator (similar to that shown in Figure 4.6) can be used to demonstrate the effects of control point spacing on the maximum surface slope. Figure 5.10 shows a slice of the NURB surfaces resulting from four different horizontal spacing values.

Analysis of these results shows that the surface slope threshold is broken at a spacing of approximately $0.588\,cm$ not $0.419\,cm$ as predicted by a purely linear assumption. This half blocked field is an important test case as it produces the maximum possible change in height over the minimum possible distance. Therefore, it was decided that a value of $0.600\,cm$ be used as the minimum allowable control point spacing for surface steepness concerns.

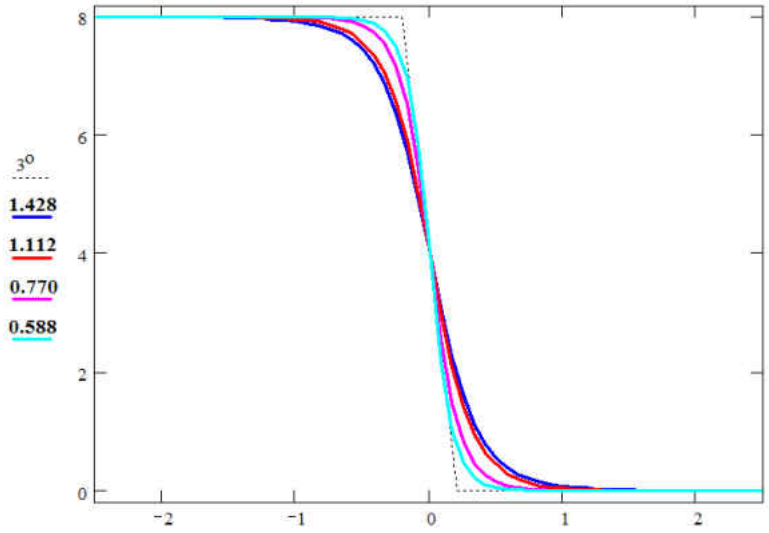



Figure 5.10: Control Point Spacing Effects on Surface Steepness

This process of limiting the x-y control point spacing affects not only the steepness of the surface, but also the size of the internal valleys that can lead to under-cutting of the compensator surface (see Figure 4.15 for a description of under-cutting). It should be recognized that the control point spacing alone could be used to address both steepness and under-cutting concerns, however initial testing proved that such a simplistic approach was far too limiting to the overall resolution of the surface. Therefore, this study also determined guidelines regarding the weights of the NURB control points which allowed for closer spacing while still limiting the slopes and valleys in the desired fashion.

This discussion has shown that the two major goals of this study are to find: (1) the minimum allowable control point spacing to ensure surface slopes are not beyond the threshold and (2) the allowable range of weight ratios and control point heights that ensure valley sizes prevent the possibility of under-cutting. The half blocked case shown above has already provided an appropriate value to satisfy the first of these goals. The second goal is a bit more complex and the first few example problems below have been designed to aid in this determination. The remaining examples demonstrate the performance of the constraining process. This effectively proves that every compensator produced by the complete DCPO process is fully machinable, thereby satisfying one of the major objectives of this research.

The first example considered is the most obvious case which can potentially create an under-cutting situation, a single deep valley (pocket). The simplest version involves having just one control point at the minimum allowable thickness while all others remain at the maximum. The valleys resulting from several different control point spacing and weight combinations are shown in Figures 5.11 - 5.12. The first of these figures shows the effects of control point spacing for uniform weights. These results show that for the range of control point spacing expected in this work the NURBs are unable to violate the machining constraints when equal weighting is utilized. However, Figure 5.12 reveals that non-equal weighting can have a significant impact on the resulting NURB surfaces. This figure displays the results at a fixed control point spacing ($0.625\,cm$) for various weight ratios. An important observation is that the weighting is able to increase the depth of the valley and thus the surface steepness is increased, however, the radius of the valley trough is actually increased which improves the manufacturability. This example proves that for the minimum expected

spacing a weight ratio of up to 20:1 may be used without violating any of the pertinent machining constraints. Also shown in each of these figures (as a black line) is a three degree sloped pocket with a bottom diameter of 0.3175 $cm$ which represents the minimum possible pocket that can be created using the described milling procedure.

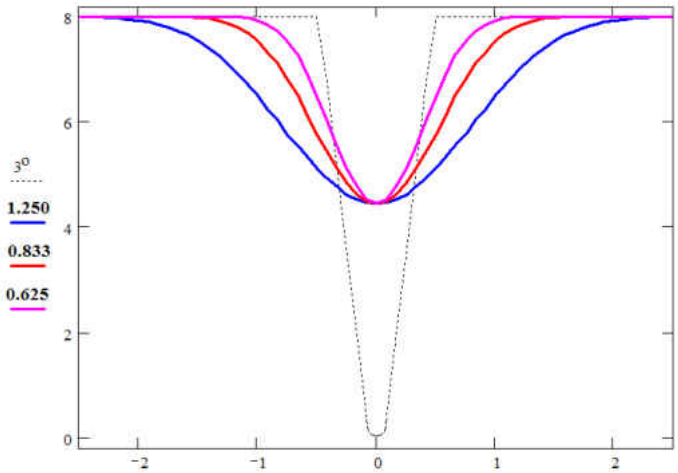



Figure 5.11: Control Point Spacing Effects on Surface Steepness and Valley Radius

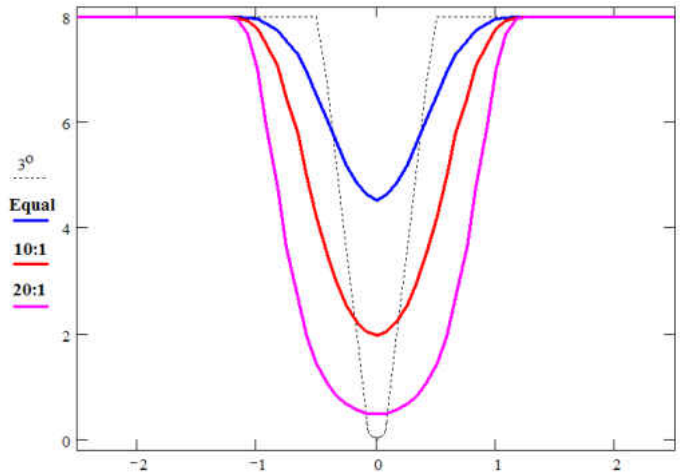
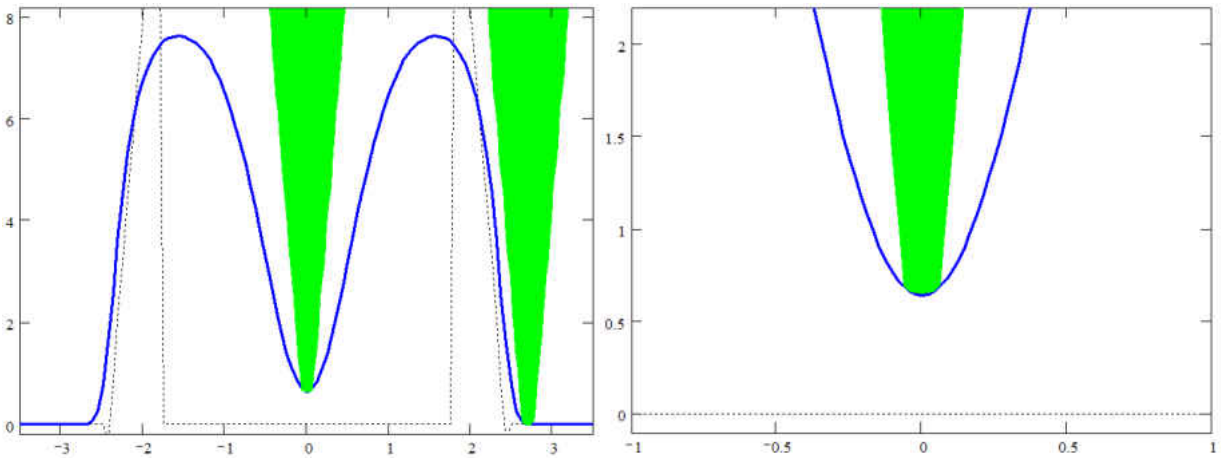



Figure 5.12: Control Point Weight Effects on Surface Steepness and Valley Radius

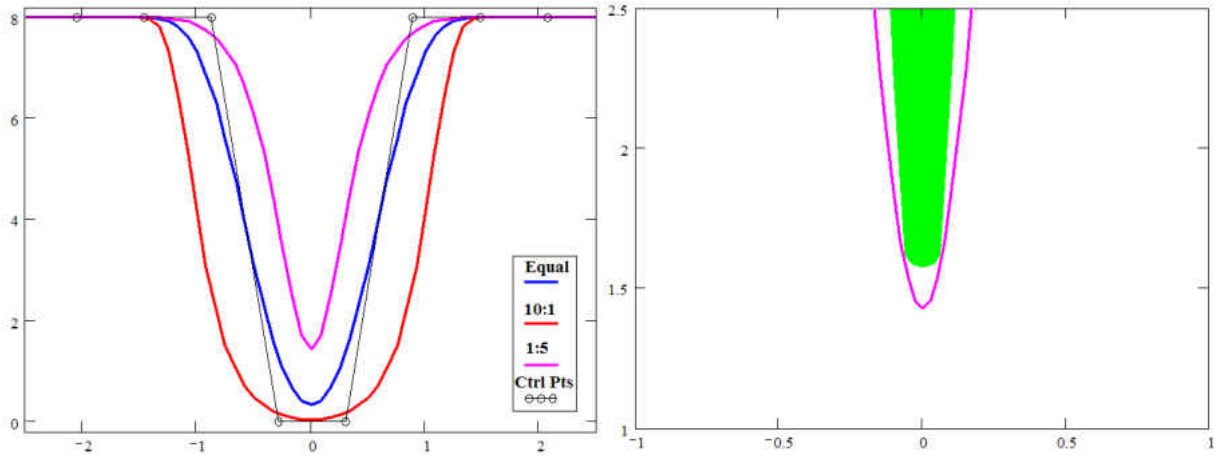
The above examples have provided useful initial values and ranges for the NURBs control parameters, however more rigorous testing is needed before confidence can be established in these results. Therefore, several more complex surfaces have been analyzed to provide

such testing. Figure 5.13 shows the results of a doubly peaked compensator shape. Such a surface has steep grades which nearly approach the maximum allowable limits, as well as deep internal radii which also approach the minimum allowable size. The blue curve shows the surface, while the solid green shapes represent the milling cutting tool for comparison purposes. Figure 5.13b shows that the center section has just reached the tool limitation, which shows that a weight ratio of 20:1 is again satisfactory for the control point heights and spacings utilized above. Figure 5.14 shows a slightly wider channel, which contains two control points located at the minimum allowable height. It was found that this scenario actually is able to produce a more abrupt pocket than is the case of a single centrally located control point. While surface steepness concerns are still unable to be violated for this arrangement, it was found that as the weight ratio between the lower nodes and upper nodes was reduced, the radius of the internal valley decreased significantly. Comparing several weight ratios revealed that a 1:5 ratio (lower node weights smaller than upper) was about the threshold where undercutting become evident. Therefore, it was concluded that the weight ratio limitations will need to check the relative heights between nodes and always compute the adjacent node weight ratio so that the higher node (greater thickness) is in the denominator of the ratio so that consistent bounding is achieved. Several additional configurations were also developed and tested, but none proved as limiting as those presented herein.



(a) Entire Central Axis Slice (b) Expanded Central Valley View

Figure 5.13: Control Weight Effects on Manufacturability (Double Peaked Surface)




(a) Entire Central Axis Slice (b) Expanded Valley View

Figure 5.14: Control Weight Effects on Manufacturability (Two-Point Valley)

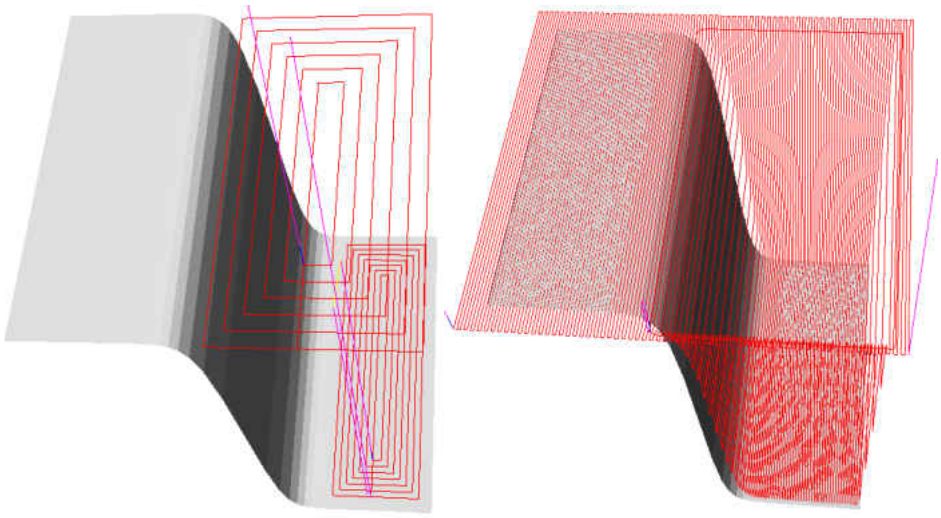
From these results it can be observed that manufacturing limitations can successfully be indirectly included within the treatment planning process by constraining the NURBs control parameters. These results have also shown that this inclusion can be accomplished by simply imposing both upper and lower bounds on the NURB control point heights and weights, as well as limiting the minimum horizontal control point spacing. Additionally,

this study has exposed the fact that these bounds and limitations are not prohibitive to the treatment planning process and in fact these bounds are actually beyond the values expected for routine clinical planning (typical beamlet based planning systems limit the resolution to $0.50\,cm - 1.00\,cm$ so that in order to achieve a reasonable parameter reduction with DCPO, spacing should be limited near $0.80\,cm - 1.25\,cm$). Table 5.2 gives the final values which are used for the remainder of this work regarding limiting the NURBs control parameters.

Table 5.2: Final NURBs Control Values for Milling Constraint

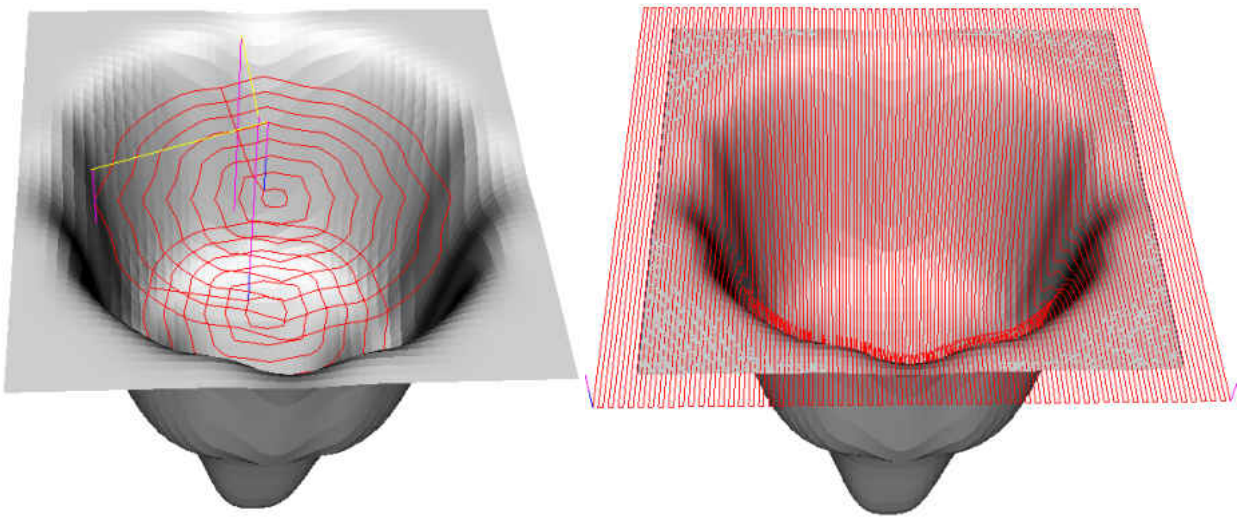
| | Horizontal Spacing | Control Heights | Control Weights | Weight Ratio |
|---|---|---|---|---|
| Minimum | 0.625 cm | 0.3 cm | 0.05 | 1:5 |
| Maximum | 2.0 cm | 8 cm | 1.0 | 20:1 |

As further evidence of the manufacturability of the compensators designed with this approach, complete milling simulation has been completed on several compensator surfaces. This simulation provides complete specification of the milling tool path as well as a quantification of the error between the prescribed and milled surface. Analysis of the results for simulations of the stepped, pocketed, and tiered surfaces showed that the maximum undercutting distance was found to be less than $0.1mm$ (0.0039"), over-cutting was negligible (below $0.01mm$), and the minimum surface angle (steepness) was $4.2^o$ from vertical, within the allowable minimum of $3^o$. These results indicate the success of three-axis milling for these parts and give confidence that compensators developed using the DCPO process are nearly 100% machinable. Figures 5.10 & 5.16 show sample images from these milling simulations complete with tool path lines for small portions of the roughing and finishing cuts.



(a) Roughing Tool Paths (b) Finishing Tool Paths

Figure 5.15: Sample Milling Simulation Results



(a) Roughing Tool Paths (b) Finishing Tool Paths

Figure 5.16: Sample Milling Simulation Results

## 5.3 Optimization (Inverse Planning) Analysis

At this stage of the project initial testing has verified that both dose calculation and surface representation routines were functioning as expected. Additionally, manufacturing limitations had been fully quantified and the DCPO process was theoretically ready for full testing. However, since the optimization process is the main determinant of overall efficiency and plan quality, each competing optimization scheme required intense performance testing before an appropriate decision could be made. Two distinct efforts were carried out in this regard with the aim of isolating two key aspects of the inverse planning processes: large parameter count performance and effects of high parameter interdependence (sensitivity).

The performance of each optimization routine for large parameter counts is important since typical three-dimensional treatment plans will require several hundred independent optimization parameters even with the reduction offered by DCPO (a single $10\,cm \times 10\,cm$ field would require roughly $100-200$ parameters using $1\,cm$ control point spacing for DCPO; in contrast using a standard beamlet size of $0.5\,cm$ would require 400 parameters for a traditional planning approach). Large parameter count optimization was addressed in an efficient manner by directly optimizing to a prescribed compensator shape (momentarily removing the patient and dose calculation concerns). Performance and tuning of each optimization scheme was carried out for several such test cases. Next, the effects of increased parameter interdependence due to the DCPO parametrization were analyzed by examining the performance of each optimization scheme on several two-dimensional test cases. Two-dimensional test cases are used for several reasons: (1) the reduced problem size

allows optimization to be carried out for longer intervals, (2) the comparison of results and plans is more straight forward, and (3) three-dimensional analysis is nearly equivalent to repeated two-dimensional analysis. Also, the optimization routines actually do not require any modification for the transition to three-dimensions, as their use is dependent solely upon an input "string" of values representing the NURB or cubic spline surface control parameters and the resulting objective function. Furthermore, the parameter interaction and sensitivity with respect to one another is expected to be on the same order for both two- and three-dimensional surfaces.

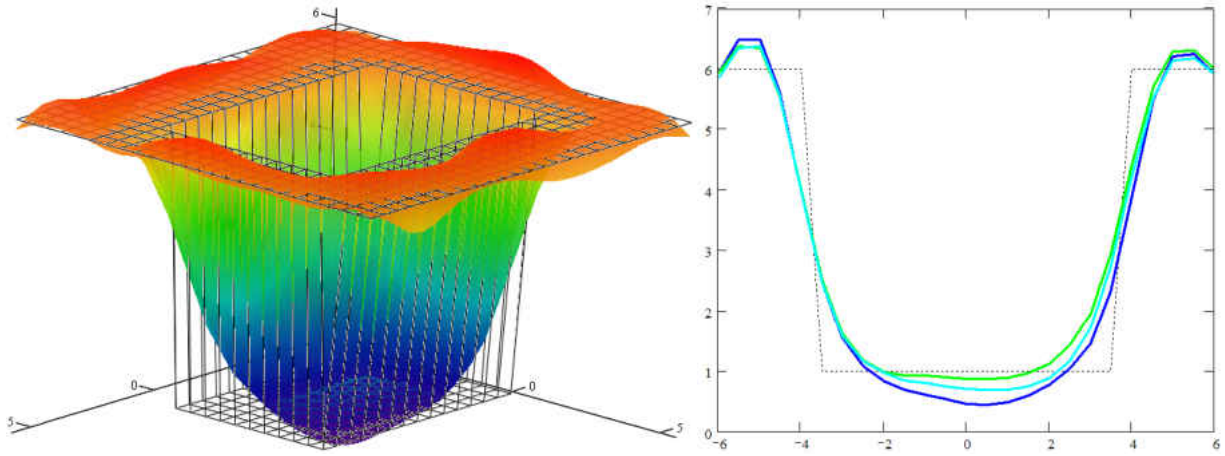
### 5.3.1 Large Parameter Count Optimization

Large parameter counts are often very difficult for certain types of optimization schemes to handle efficiently, therefore this issue received first priority. These tests involved attempting to match a preset surface shape, by optimizing the NURB surface control parameters. A simple quadratic objective function was used for each optimization routine, essentially creating a Least-Squares surface fitting routine. Testing began with simple smooth surfaces such as linear and quadratic functions and then progressed to more complex, realistic compensator shapes. Each of the optimization schemes performed adequately for the simple surface tests and no major difficulties were encountered, although genetic algorithm performance was somewhat less compared to simplex and gradient based approaches. This performance difference became much greater as surface complexity increased. Results for some exemplary cases for each optimization approach are provided below, along with a discussion of the general observations seen during completion of the various test cases.

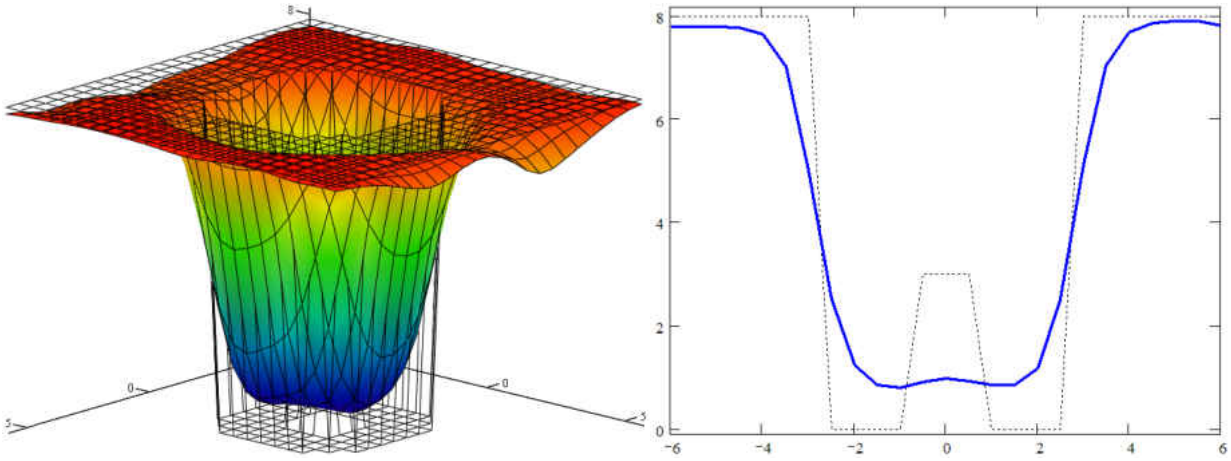
Specifically, results for two test cases are presented. Test case (a) consists of a single $12\,cm \times 12\,cm$ field with a stepped pocket compensator that is $1\,cm$ thick when $|x| < 4 \quad and \quad |y| < 3$, and $6\,cm$ thick *otherwise*, with the origin placed at central axis (this ideal surface is shown in black in the background of each of the results below). The NURB surface is optimized using a $10 \times 10$ field of control points and weights and the objective function (surface agreement) is computed at $0.5\,cm$ spaced locations along the surface. Test case (b) consists of a single $12\,cm \times 12\,cm$ field with a radially symmetric island pocket compensator that has a thickness of $3\,cm$ for $|r| \leq 1$, $0\,cm$ for $1 < |r| \leq 3$, and $8\,cm$ *otherwise*, with the origin placed at the central axis (again this ideal surface is shown in black in the background of each of the results below). The NURB surface is optimized using an $11 \times 11$ field of control points and weights and the objective function (surface agreement) is again computed at $0.5\,cm$ spaced locations along the surface.

Initial three-dimensional, large parameter count testing quickly proved that even with the parameter reductions gained by DCPO, the genetic algorithm optimization schemes did not perform very well. Performance for the GA approach was mediocre, with poor surface agreement and surface oscillations often still persisting even after excessively long optimization runs (millions of trial solutions). It was found that despite the parametrization provided by the DCPO approach, $100 - 200$ parameters are still needed in order to provide the resolution needed to describe a *single* useful compensator. This number of optimization parameters proved too great for direct application of standard genetic algorithms as demonstrated by the extremely slow convergence after the initial rapid improvements. This problem is supported by this author's prior research [33, 35] and is also well known and

supported in the field [61, 62]. Results for the two surface shapes described above are given in Figure 5.17 for genetic algorithm optimization.



(a) Pocket

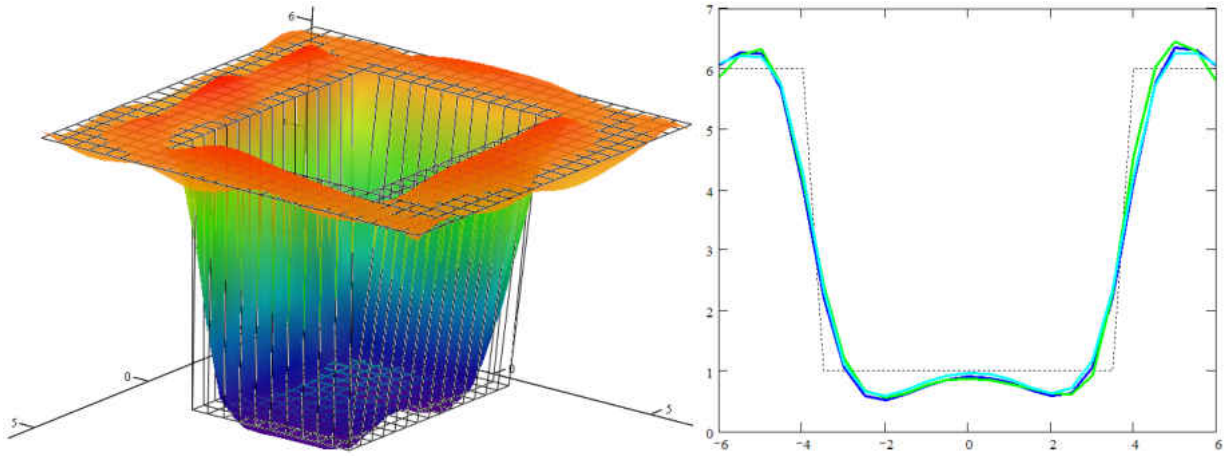


(b) Pocket with Island

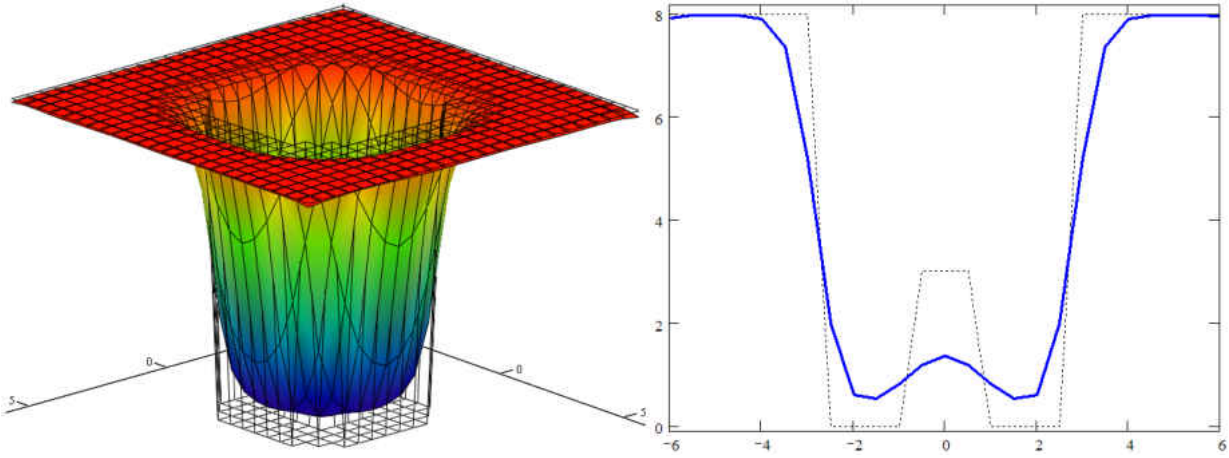
Figure 5.17: Genetic Algorithm Optimization Testing: Direct Surface Optimization

The downhill simplex and gradient-based optimization algorithms were found to be quite similar in behavior and both were effective at locating local solution minima. Initial testing of these approaches found that high quality solutions could be obtained much more quickly than with the genetic algorithms and that each technique seemed to converge to the same (or

nearly the same) solution from several different initial guess values (constant thickness and completely random initial guesses were both tested). Figure 5.18 shows the sample simplex solutions while Figure 5.19 shows the same examples for the conjugate gradient approach (although several gradient-based optimization routines were tested the final results of each were quite similar with efficiency being the only variable; therefore as conjugate gradient was found to be the most efficient only results for this approach are presented).

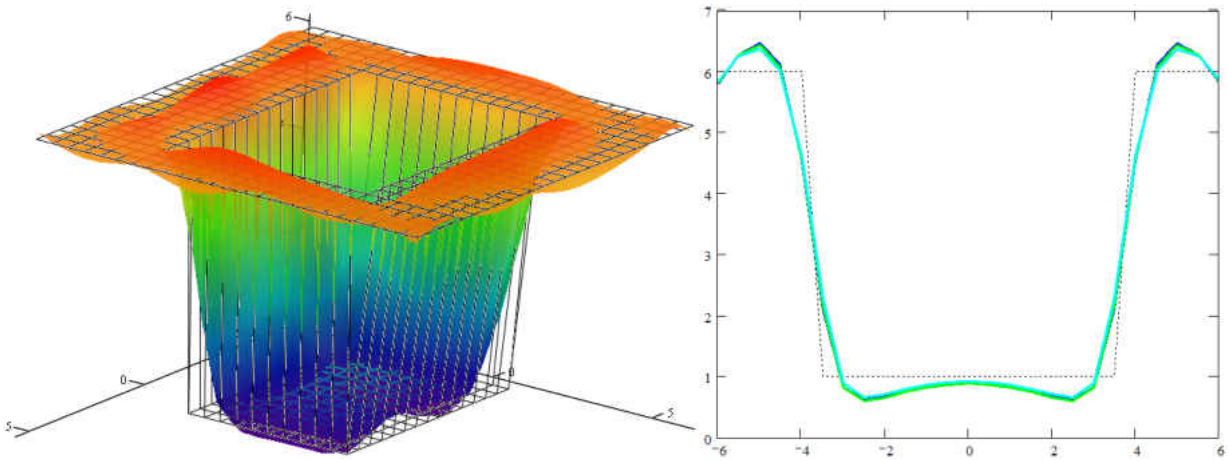


(a) Pocket

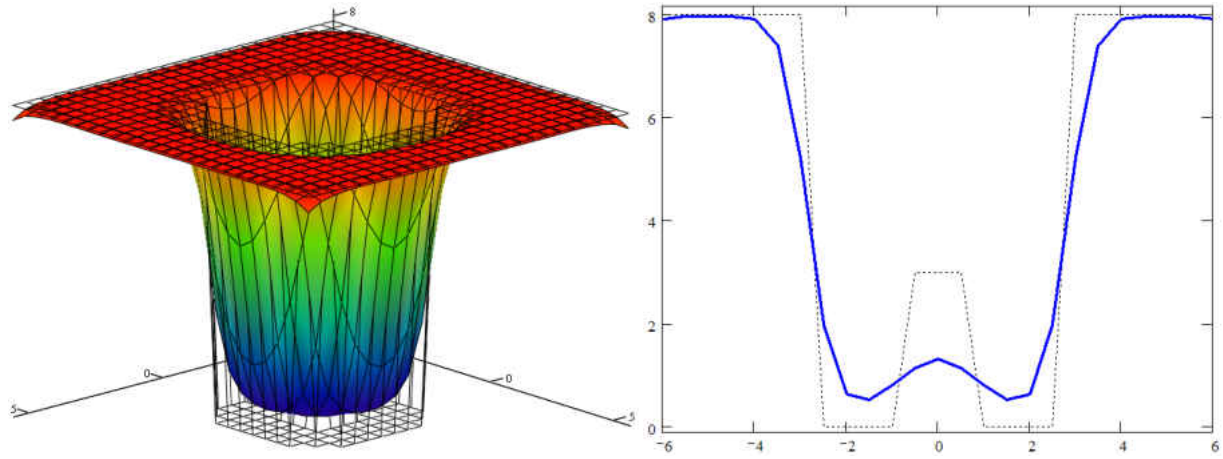


(b) Pocket with Island

Figure 5.18: Simplex Optimization Testing: Direct Surface Optimization



(a) Pocket



(b) Pocket with Island

Figure 5.19: Gradient Optimization Testing: Direct Surface Optimization

The similarities between the downhill simplex and gradient optimization schemes for these cases are quite impressive. As these initial trials involved optimizing directly to a prescribed surface shape, it is likely that only one family of optimal solutions exists (family is used as it is the NURBs weight ratios and not the weights themselves that dictate the surface shape, therefore multiple (identical) solutions can exist as long as the weight ratios are equivalent). The existence of such a small set of optimal solutions leads to the

notion that the search space has only several nearly equivalent local minima (or a single, flat global minima). This is evidenced by the repeatability of the gradient and simplex solutions, since numerous initial guess configurations converged to nearly the same final results. A quantitative comparison of the simplex and conjugate gradient results shows that the maximum difference in surface height at any point is 10%, however this is a vertical difference which occurs only on the steep walls of the pocket; therefore actual surface-to-surface distance (perpendicular distance) is considerably less and the average surface difference is only 1.5%. An effective qualitative comparison of the solution agreement is achieved by overlaying plots of two-dimensional surface slices, which has been done in Figure 5.20. This figure shows that the gradient solution (red lines) is slightly better than the simplex (blue lines) as it is closer to the desired surface (black lines) over a majority of each slice (solid lines depict a slice at the center-line, while the dashed line slices are midway between the center and field edge; the field is not symmetric and the x and y directions are shown on the left and right respectively).

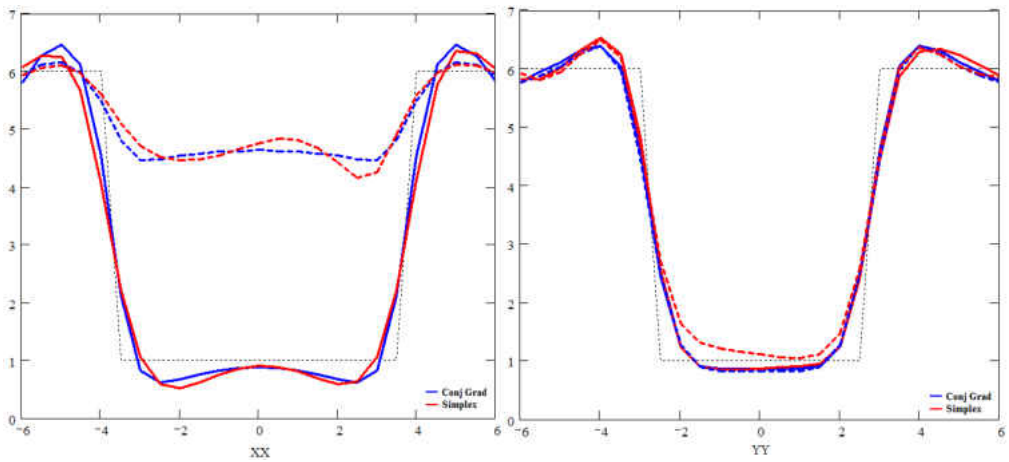



Figure 5.20: Conjugate Gradient vs. Downhill Simplex Direct Surface Optimization

Despite the similarities in solution quality, the efficiency of the techniques was markedly different. The conjugate gradient method consistently converged in fewer iterations and required fewer total function evaluations compared to downhill simplex. This behavior is illustrated by comparison of the curves in Figure 5.21, which shows the objective value as a function of the number of function evaluations for each optimization scheme. The conjugate gradient method converged most rapidly for these, and most all, example problems that were tested, requiring only a small quantity of iterations to converge $(20 - 50)$. Even with this rapid convergence, conjugate gradient still required many trial solutions ($\approx 5,000$) due to the need to calculate the gradient at each iteration (which requires one function evaluation *for each* optimization parameter). Convergence of the genetic algorithm to the same solution as the other two approaches (both gradient and simplex approaches converged to nearly identical solutions for all direct surface optimization examples that were tested) generally required $10 - 100$ times more trial solution evaluations compared to the conjugate gradient scheme. Thus the performance of the genetic algorithm was found to be very poor due to slow convergence rates (the GA solution quality is low due to stoppage of the optimization before complete convergence was reached; all routines were allowed to run for a fixed number of function evaluations to provide a fair comparison). Similar behavior was found for each example tested, providing further proof of the hypothesis that the number of solution parameters is still too great for effective genetic algorithm optimization.

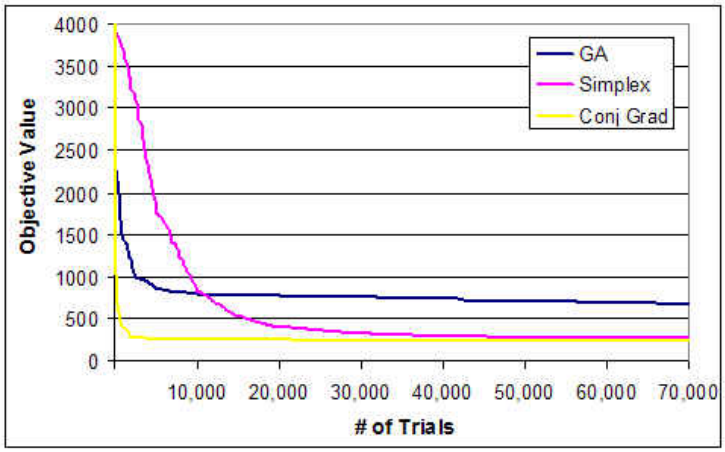



Figure 5.21: Objective Function Convergence

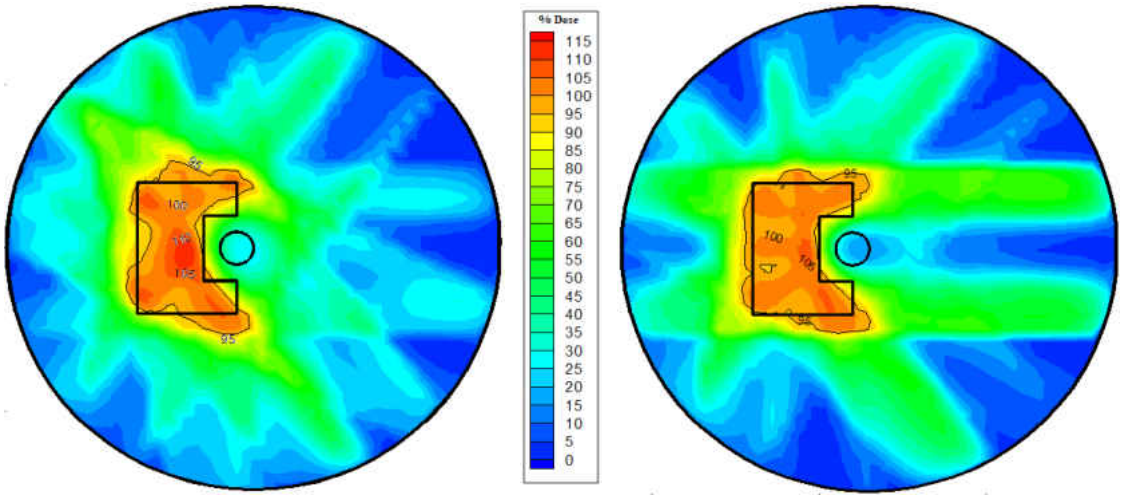
A possible criticism of these tests is the fact that only one (or possibly just a few) minima exist within the given search space. However, prior investigations by [63] have shown that the solution space for complete radiation inverse planning problems is globally convex when only minimum and maximum dose objectives are utilized. Therefore, this investigation is completely applicable to full treatment planning analysis as the gradient based approaches should perform equally well for any convex function (such schemes are also guaranteed to converge to the global minima for fully convex functions [32]). Despite the poor performance of the genetic algorithms in this area, their use is continued in the next high parameter sensitivity study as such sensitivity changes often have a strong influence on the performance of numerical optimization schemes. However, due to the consistent out-performance of conjugate gradient over downhill simplex in these baseline studies, the decision has been made to exclude the downhill simplex approach as a viable technique for this work and the remaining efforts will focus solely on conjugate gradient and genetic algorithm schemes.

### 5.3.2 High Parameter Sensitivity Optimization

Much literature [63, 64] shows that gradient based solution approaches are often preferable for inverse radiation treatment planning due to the convex shape of the majority of the solution space, however, other works [65, 66] show that stochastic methods are preferable. This next stage of research analysis attempts to discern the effects of increased parameter interaction and sensitivity on the performance of the conjugate gradient and genetic algorithm optimization schemes in order to aid in determination of the most appropriate technique for use with DCPO. Test case #2 from section 5.1.2 provides a reasonable challenge to planning due to the closeness of the OAR and PTV. Abutting or nearly abutting situations such as this represent the most difficult planning scenario, because a very steep dose gradient must be achieved in the vicinity of the abutment. Thus, this case has been utilized as a primary means of testing the optimization schemes herein. Also, one final cubic spline surface solution is shown for this case to demonstrate the inapplicability of the scheme for DCPO (as these type of abutting configurations often occur in real IMRT plans, this a somewhat realistic test which clearly highlights the issues of cubic spline control, even for 2D cases). For completeness an additional case is also presented in order to show that these findings can indeed be generalized to other patient configurations.

This final performance analysis begins by computing plans for identical configurations and prescriptions using both the genetic algorithm and conjugate gradient schemes. The geometry for this case has been shown previously in Figure 5.5. For this analysis the number of treatment beams has been reduced to five (equally spaced and coplanar) and

the prescription has been modified slightly (note that the initial results for this problem normalized the dose field to the point of maximum dose, however for the remainder of this work the dose field is always normalized to the isocenter value): PTV target dose is $74\,Gy$ (100%) with maximum dose of $77.7\,Gy$ (105%) and minimum dose of $70\,Gy$ (95%); OAR is minimized with maximum allowable dose of $18\,Gy$ (25%). The resulting dose fields using a $10\,cm$ beam with a NURB surface representation (control point spacing of $1.6\,cm$) for both GA and gradient optimization schemes are given in Figure 5.22. The corresponding dose volume histogram for each scheme is overlaid and shown in Figure 5.23. This DVH comparison clearly shows that the gradient plan is better than the GA plan in terms of PTV dose homogeneity (less cold spots and reduced hot spots) and OAR dose minimization. While the rest-of-body (ROB) dose is slightly greater for a portion of the gradient plan, it is not substantial enough to cause any concern.




(a) Genetic Algorithm (b) Conjugate Gradient

Figure 5.22: Optimization Comparison: Relative Dose Fields

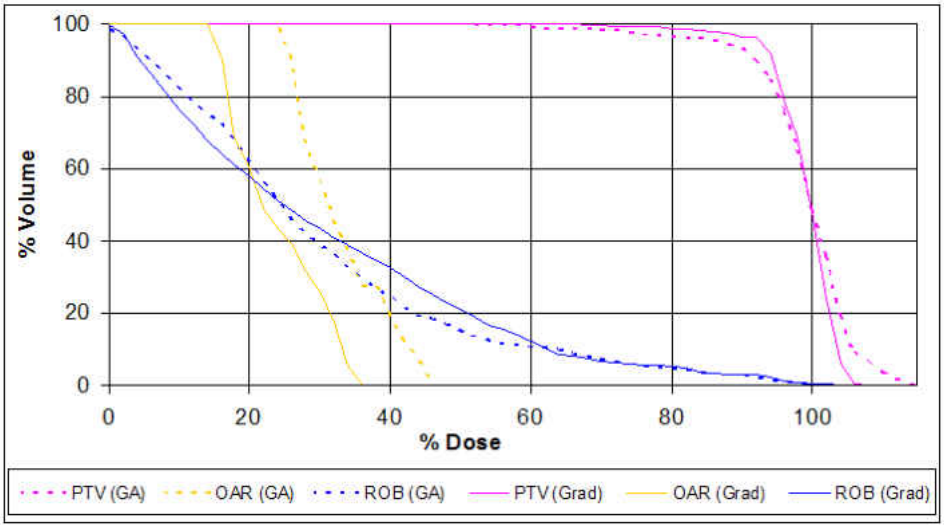



Figure 5.23: Optimization Comparison: Dose Volume Histogram

It can be stated that the findings from the quality comparison of this one plan do not justify the claim that the gradient based approach clearly out performs the GA. However, upon analyzing the efficiency of each approach this claim cannot be disputed. Figure 5.24 shows a comparison of the general trends for the genetic algorithm and conjugate gradient optimization schemes in terms of objective function value versus number of trial solutions tested. While this figure shows only one particular run (roughly an average run of each), it should be noted that the genetic algorithm *never* surpassed the average conjugate gradient performance in any of the runs tested (over 10 runs of each case were tested using random guess initialization for the gradient approach and a combination of random guess and fixed input for the GA initialization).

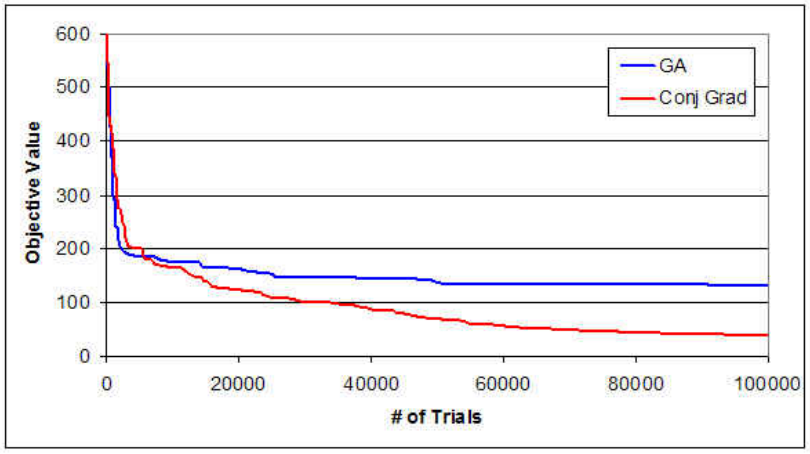



Figure 5.24: Objective Function Convergence for Full Planning Case

As mentioned previously, despite the ease of use and high quality of the cubic spline interpolation approach, it has the major drawback of not providing an analytical means to bound the resulting interpolation. It has been stated that this limitation eliminates the use of this approach for DCPO, which was based largely on hypothetical situations. In order to prove that such simulations were not unfounded and demonstrate the importance of bounding within DCPO this planning case has been repeated again using the latest optimization schemes with the cubic spline surface representation. While these results should be comparable to those provided in Section 5.1.2, much higher quality is obtained here due to the incorporation of overall beam weights and due to the increased efficiency of the numerical optimizations (this allowed the solution to be computed to a higher degree of convergence with less effort). Figure 5.25 shows the dose volume histogram for the cubic spline plan (along with the NURB plan for comparison) which demonstrates the near equivalence of the two procedures in terms of achievable plan quality. However, by analyzing the resulting compensator surfaces in Figure 5.26 (only two of five are shown) one finds that the cubic

spline surface required truncation in order to maintain the surface height above zero thickness and the maximum below the linear accelerator tray limitation. The full cubic spline surface is shown in the figure, with the areas of truncation highlighted in green. It should be noted that the control points never dipped below zero, but the surface does, illustrating the importance of *a priori* surface bounding as the resulting sharp internal corners would not be machinable.

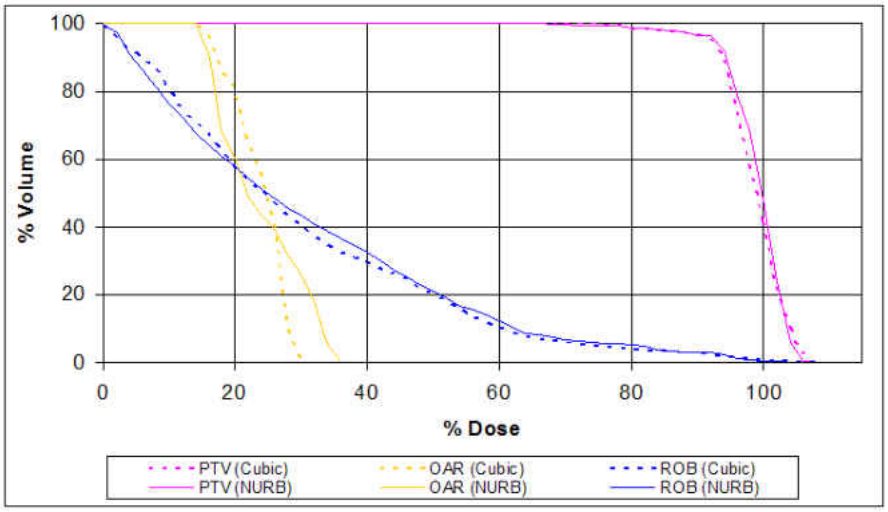



Figure 5.25: Cubic Spline and NURB Surface Dose Volume Histograms

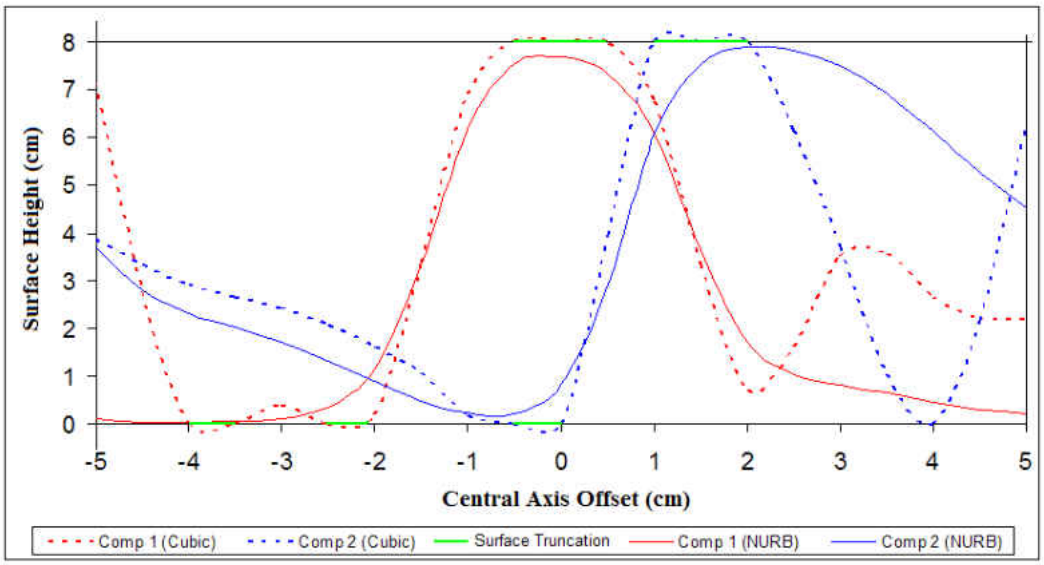



Figure 5.26: Cubic Spline and NURB Compensator Surface Comparison

In order to demonstrate the general applicability and further prove the soundness of the above observations another slightly more complex treatment planning case is presented. This case represents a single slice from a typical measurement phantom (simulated patient apparatus for dose measurements and experimentation) whereby a PTV with both convex and concave boundaries is included, flanked by small circular organs-at-risk on either side. The complete geometry is shown in Figure 5.27. The planning prescription is similar to that used above with the PTV set at 100% dose (minimum 95% and maximum 105%), organ dose minimized (maximum of 50% dose), and the ROB dose is minimized as well. Once again this case was planned utilizing both the conjugate gradient and genetic algorithm approaches. The resulting converged relative dose fields using five equally spaced, coplanar beams are shown in Figure 5.28 and the DVHs are provided in Figure 5.29. The resulting compensator shapes for each of the five fields are shown for both plans in Figure 5.30.

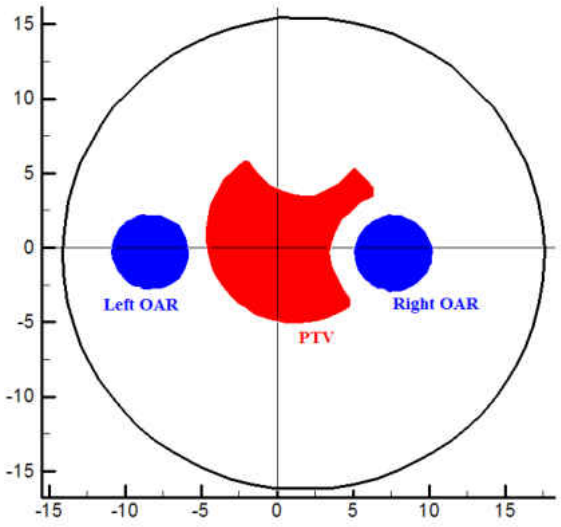
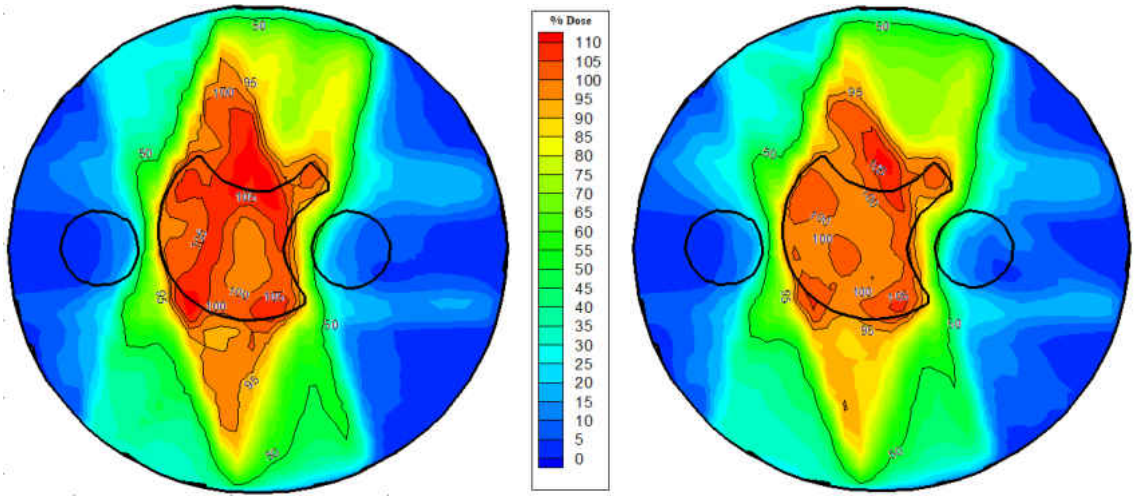



Figure 5.27: Two-Dimensional Phantom: Geometry




(a) Genetic Algorithm (b) Conjugate Gradient

Figure 5.28: Two-Dimensional Phantom: Relative Dose Fields

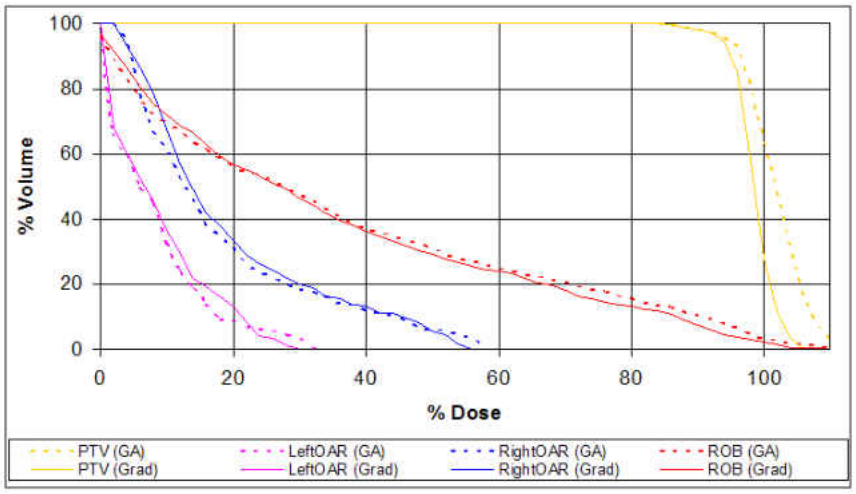



Figure 5.29: Two-Dimensional Phantom: Dose Volume Histogram

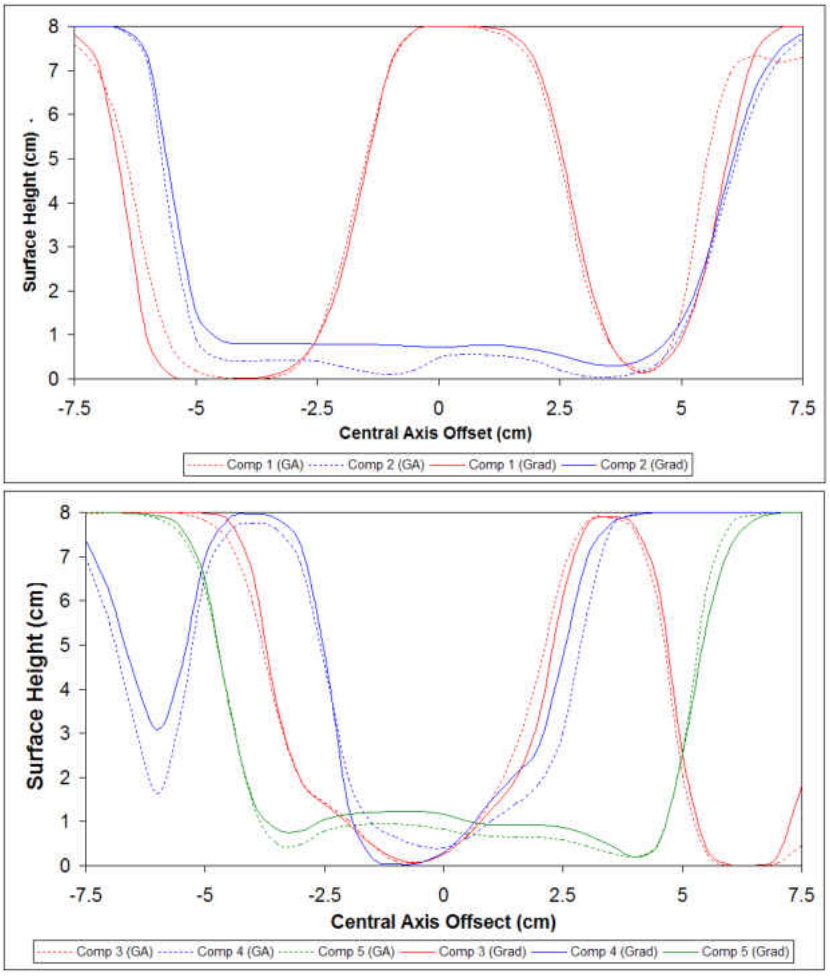
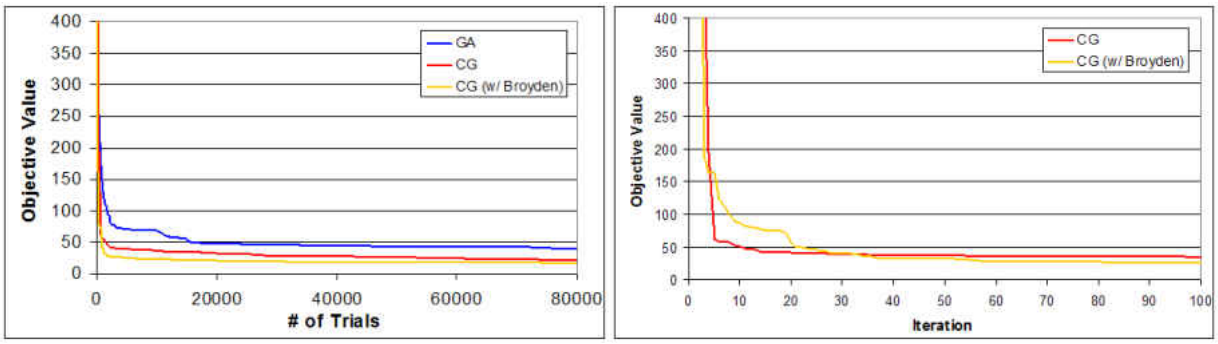



Figure 5.30: Compensators for Gradient and GA Generated Treatment Plans

It can be seen from these results that the genetic algorithm (GA) plan is nearly the same as that found using the conjugate gradient (CG) approach. This is evidenced by the shapes of the resulting compensators for each scheme, DVHs, and relative dose fields. Figure 5.30 shows that for each corresponding beam, the compensator surfaces follow similar general trends, reaching peaks and valleys at roughly the same locations and obtaining close to the same modulation range for each. However, the objective function convergence (Figure 5.31) once again shows that the GA required significantly increased numbers of trial solutions (function evaluations) to reach the provided results. Also interesting to note is that the CG

approach has now been supplemented by use of Broyden's Jacobian estimation technique (see Section 4.3), which significantly reduces the number of trial solutions needed to reach a converged solution. Figure 5.31 highlights this added efficiency by showing typical results for the CG technique, both with and without the use of Broyden's method. In Figure 5.31b it appears as if the standard conjugate gradient approach out performs that of Broyden's method. While this is true on an iteration count basis, it does not hold on a function evaluation count (trial count) basis, as seen in Figure 5.31a. Since the evaluation of an entire dose field is required for each trial solution, it is the trial count and not the iteration count that is the most appropriate determinate of efficiency. Therefore, the trade-off in additional iterations for less solution trials is quite worthwhile herein and Broyden's approach was able to bring about improved solution efficiency for all cases tested.




(a) All

(b) Gradient plus Broyden

Figure 5.31: Two-Dimensional Phantom: Objective Function Convergence

While these results combined with the added efficiency of Broyden's method are likely enough to absolutely recommend the use of gradient based techniques over genetic algorithms for DCPO planning, one final study has been completed to finalize this claim. Since the control parameter search space is very large, even in 2D, it is impossible to visualize the

actual $N$-dimensional surface of the objective function (for this phantom case $N = 155$ control parameters are used, 15 NURBs control points and weights, plus one overall beam weight per treatment beam). Therefore, the only means to reasonably view the solution space is to utilize a series of one-dimensional plots for the control parameters (holding all others constant). Figure 5.32 provides such plots for several arbitrarily chosen control parameter directions at the point of solution convergence (i.e. using the final plan results shown above). These plots all show that the local search space is not oscillatory and only contains a single minima in each direction. While only a few plots are presented herein, a large quantity were generated and tested throughout the course of this research, none of which have demonstrated any different solution behavior. This consistent behavior significantly helps to justify the claim that the overall search space contains only a single overall minima (i.e. the solution surface is always convex). The claim of convexity was also proven mathematically by Deasy [63] for the case of root-mean-square objective functions, which is nearly equivalent to the linear plus quadratic objective used herein. The work of Deasy also concurs with the choice of using gradient based optimization (when DVH constraints are not imposed) claiming that convergence is achieved 1-2 orders of magnitude faster than with stochastic techniques, such as the GA, for convex functions.

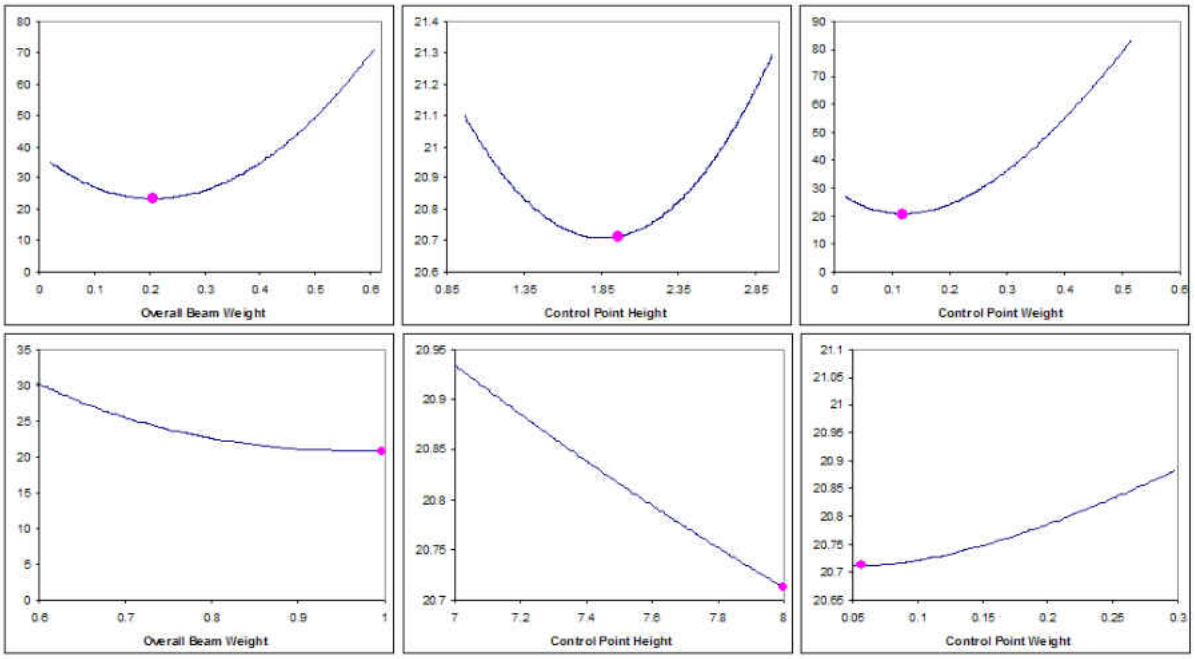



Figure 5.32: One-Dimensional Search Space Plots

## 5.4 Development and Performance Summary

The results gathered above have allowed the determination of the most appropriate optimization scheme and also quantified the process of manufacturing constraining. These results have brought about the observation that the genetic algorithm alone is not the ideal optimization engine for this particular type of problem and have shown that gradient based approaches are most suitable for IMRT inverse planning with DCPO. This observation is in agreement with available literature which has shown that IMRT planning generally results in fully convex objective functions [63]. In addition, the gradient optimization scheme has been accelerated through the use of Broyden's method, which eliminates the need to compute the gradient directly at every iteration. This research has led to a final form of DCPO which therefore

utilizes a conjugate gradient optimization technique, supplemented with Broyden's method. Additionally, the NURB surface control parameter limits have been explicitly quantified through a detailed research study and analysis. This indirect control method has been proven sufficiently adequate for providing nearly complete manufacturability of the IMRT compensator surfaces. The next chapter makes use of this final optimized DCPO process and presents results demonstrating the benefits of DCPO compared to current state-of-the-art compensator treatment planning procedures.

# CHAPTER 6
# RESULTS

The previous chapter has shown development and tuning of the Direct Compensator Profile Optimization process. As the entire process has now been fully described and the most efficient surface representation and optimization schemes have been found, this chapter is now able to serve as a complete demonstration of this research work. This demonstration provides specific examples which quantify the benefits and illustrate the utility and applicability of DCPO for Intensity Modulated Radiation Therapy planning. First, an example from the previous chapter is revisited and the solutions are compared to those obtained using a standard discrete beamlet optimization approach. This example provides data needed for several important comparisons, including plan quality, compensator manufacturability, and inverse planning (optimization) performance. In addition, this example serves to demonstrate the reduced sensitivity to patient positioning errors provided by DCPO based planning. Then a complete treatment plan is developed, using a real fully three-dimensional water phantom configuration. This final case shows the applicability of DCPO for clinical work and should demonstrate the readiness of the procedure for future commercial development.

## 6.1 Discrete Beamlets vs. DCPO

This section continues the previously utilized two-dimensional phantom test case (see Figure 5.27) in order to provide a full comparison of DCPO against existing state-of-the-art planning practices. This comparison is divided into three major areas of focus: plan quality, manufacturing considerations, and patient positioning sensitivity.

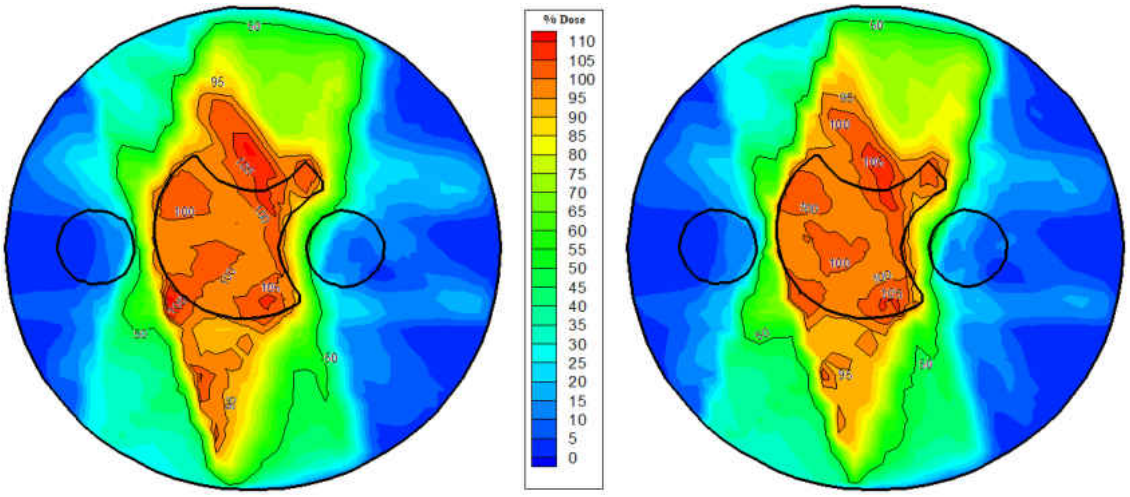
### 6.1.1 Treatment Plan Comparison

In order to provide a realistic comparison between DCPO and standard planning approaches, treatment plans for the phantom test case have been developed using a standard discrete beamlet technique. Plans have been generated using both $10\,mm$ and $5\,mm$ sized beamlets, requiring 85 and 160 optimization parameters, respectively. It should be noted that in two-dimensions the parameter count comparison is very favorable for the standard beamlet approach, but in three-dimensions this is not the case. For example, if this field were square ($15\,cm \times 15\,cm$) a $5\,mm$ beamlet approach would require 1285 optimization parameters, while DCPO would require only 455 ($10\,mm$ spacing has been shown quite sufficient for DCPO). Figure 6.1 shows contours of the resulting relative dose fields for both beamlet spacing values.

The dose field results do not show any significant differences, however, analyzing the DVHs in Figure 6.2 clearly shows the superiority of the $5\,mm$ plan (for example, 35% of the right OAR receives 20% or greater dose for the $10\,mm$ plan, while only 25% receives this level for the $5\,mm$ case, these values are 13% compared to 3% for the left OAR). Therefore, it

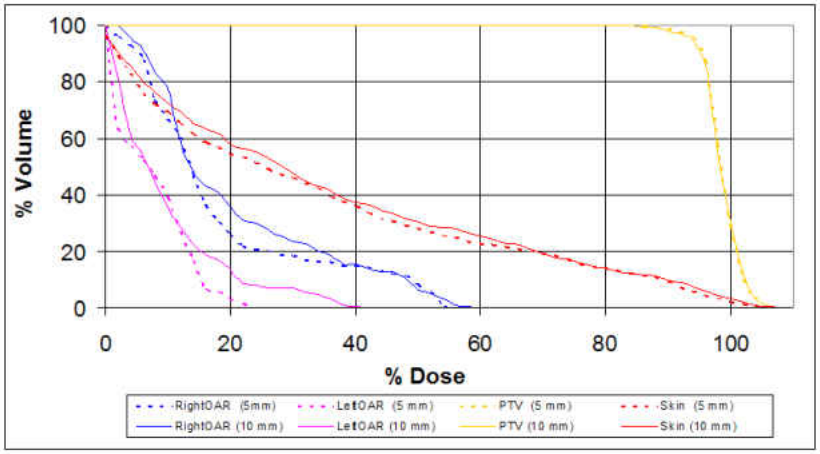
is clear that the 5 $mm$ plan is superior to the larger beamlet case. Therefore, the remainder of this discussion will apply to the 5 $mm$ plan, unless otherwise specified.

Although the plans are reasonably similar, it is observed that the standard beamlet plans provide slightly better organ sparing, while achieving nearly identical tumor dose. Therefore, it appears as if the beamlet plans are superior to those of DCPO, however, it should be noted that the steep modulation gradients present in the beamlet plans are not obtainable in practice due to delivery device limitations. Therefore, the quality of these plans will be diminished once the compensators are smoothed and converted to manufacturable shapes. The plans are therefore not truly comparable, since one (DCPO) is deliverable while the other (beamlet) is not. This issue of deliverability is the main manufacturing concern and is addressed in the following section, where these results are updated and this comparison is repeated after inclusion of the effects of smoothing which is needed to produce machinable compensator surfaces.

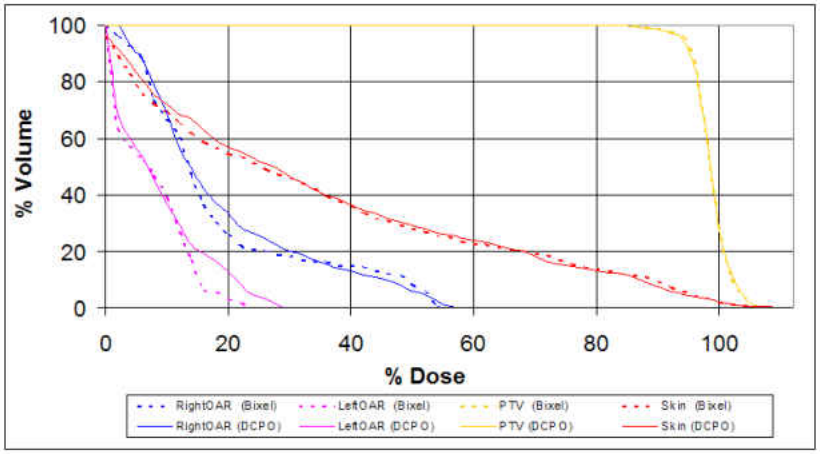



(a) 10 $mm$ Beamlets (b) 5 $mm$ Beamlets

Figure 6.1: Beamlet Plan Dose Fields




(a) Beamlet Plans Only




(b) Beamlet and DCPO

Figure 6.2: Beamlet vs. DCPO Dose Volume Histograms

### 6.1.2 Manufacturing Considerations

The section above has presented results comparing the quality of plans generated using DCPO and standard beamlet approaches. Despite the slightly higher quality of the beamlet based plans, the DCPO approach provides a major benefit: the complete manufacturability of the resulting compensators. Thus the DCPO planned dose is *exactly* equivalent to the

delivered dose (assuming that the dose calculation routines provide this level of accuracy). This section describes the effects of manufacturability on the delivered dose field for the beamlet approach, and highlights the plan degradation that is typical of approaches that account for delivery device considerations *after* optimization.

Figure 6.3 shows the compensator thicknesses for the beamlet approach (note the staircase shape of the results) before smoothing along with those obtained using DCPO. It is evident that the general trends are well preserved by DCPO and that DCPO results have an inherent smoothness compared to those obtained using a direct beamlet approach. Figure 6.4 shows the beamlet based compensators smoothed to meet the previously determined minimum machining requirements, as compared with the "ideal" planned compensators. Additionally, the effects of the smoothing process on the dose field (i.e. the deliverable dose) have been computed and the resulting DVHs are provided in Figure 6.5, along with those before smoothing and with DCPO for comparison purposes. These results highlight the typical plan degradation that results from smoothing the compensators to account for machinability. It is noticed that the "ideal" discrete beamlet plans were slightly superior in organ sparing compared with those generated using DCPO, however, after smoothing the organ sparing became nearly equal but the PTV (tumor) dose has suffered a loss in homogeneity (slight "cold" and "hot" spots now exist within the tumor region). This behavior is common in after-the-fact delivery correction procedures and its avoidance is a major benefit of the DCPO planning process.

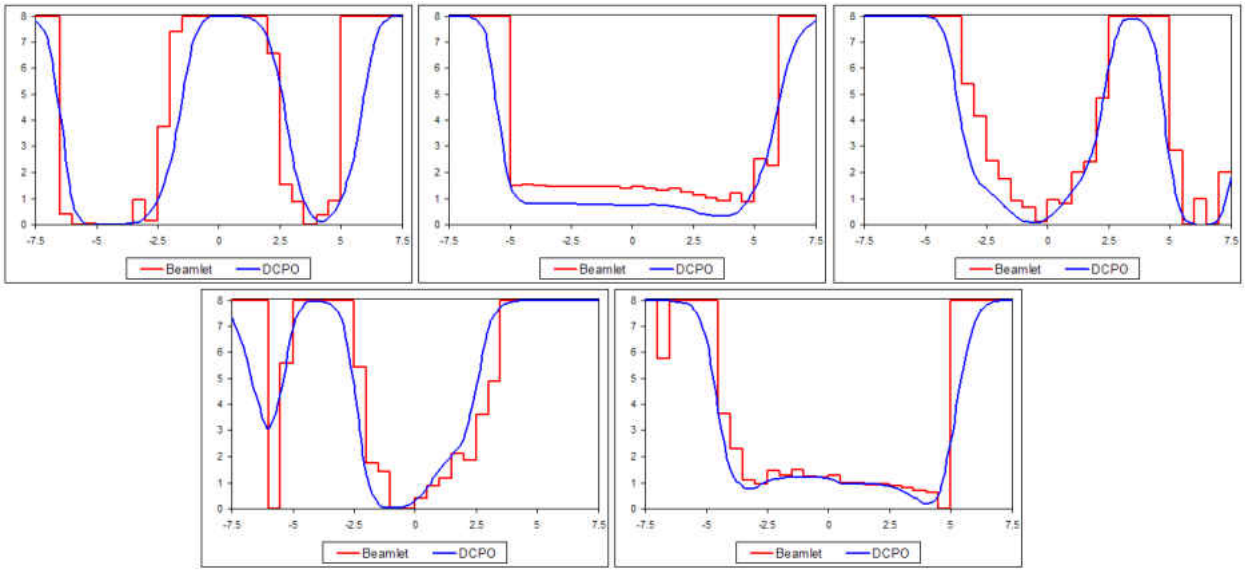



Figure 6.3: Beamlet vs. DCPO Compensator Surfaces

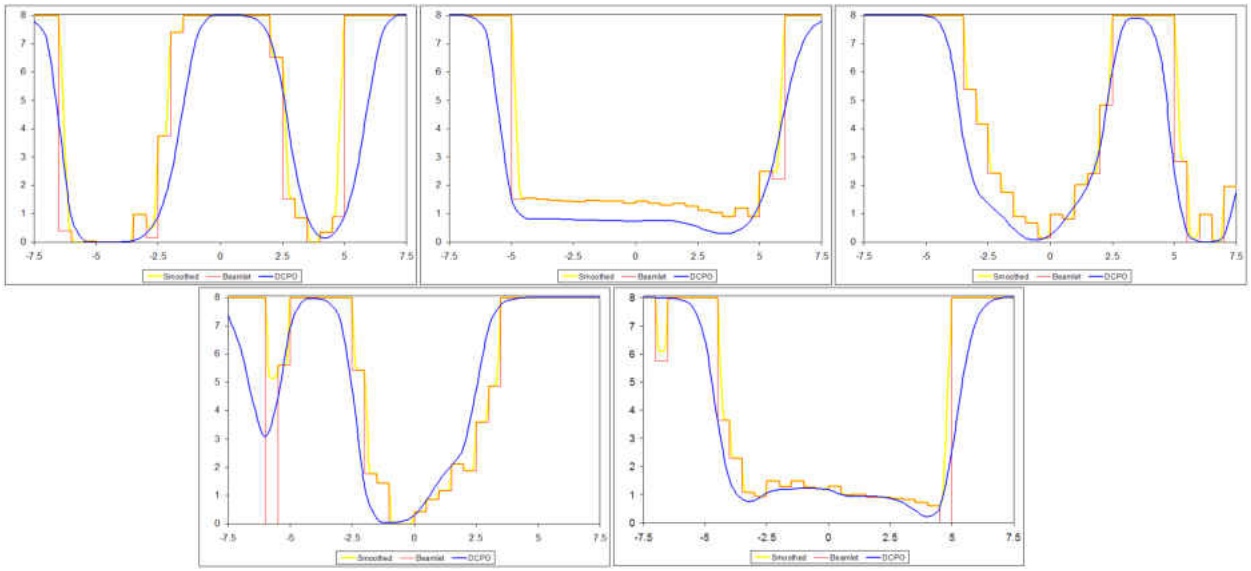
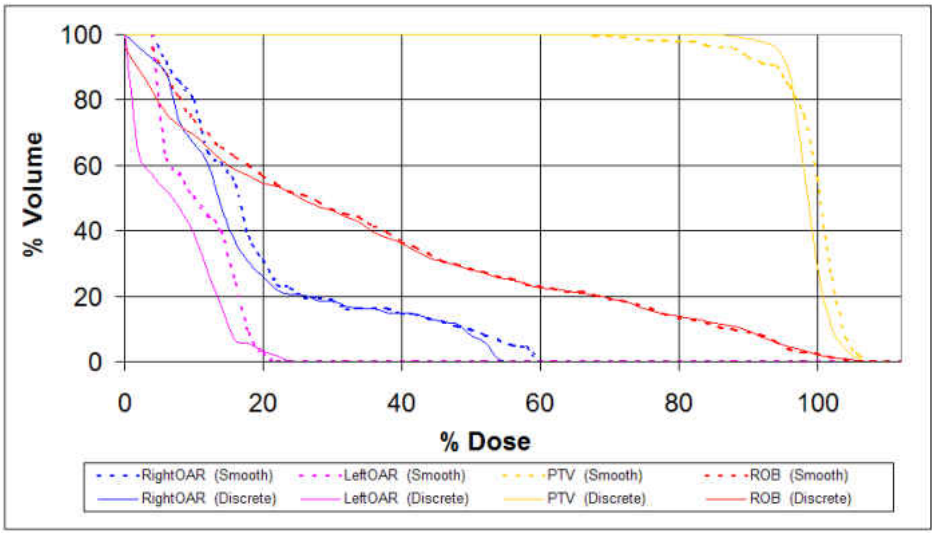
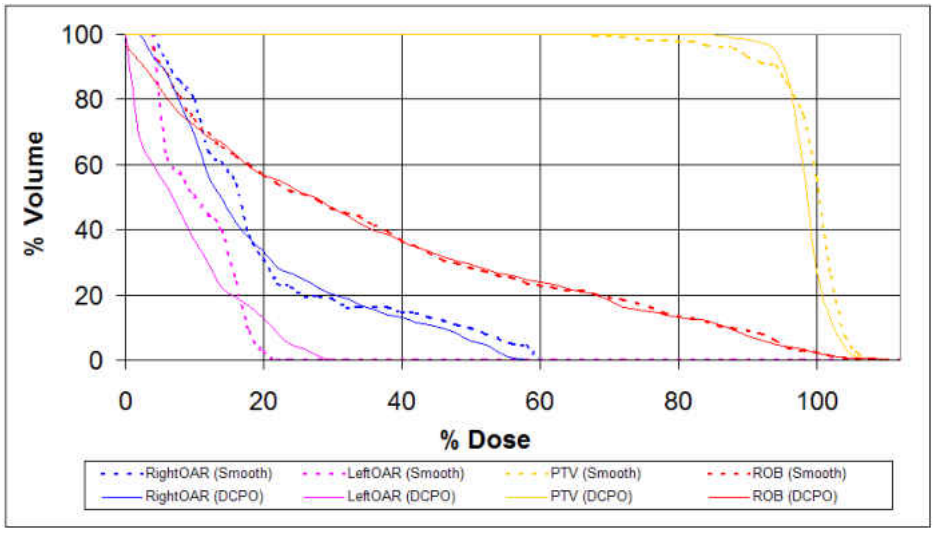



Figure 6.4: Smoothed Beamlet Compensator Surfaces




(a) Discrete vs. Smoothed Beamlets



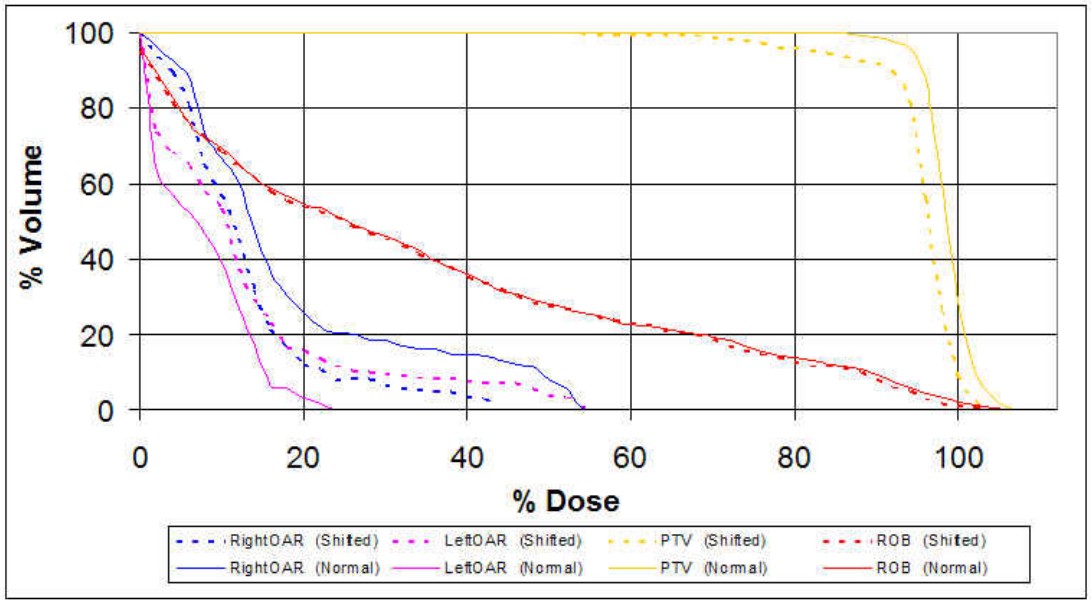

(b) Smoothed Beamlet vs. DCPO

Figure 6.5: Smoothed Beamlet and DCPO Dose Volume Histograms

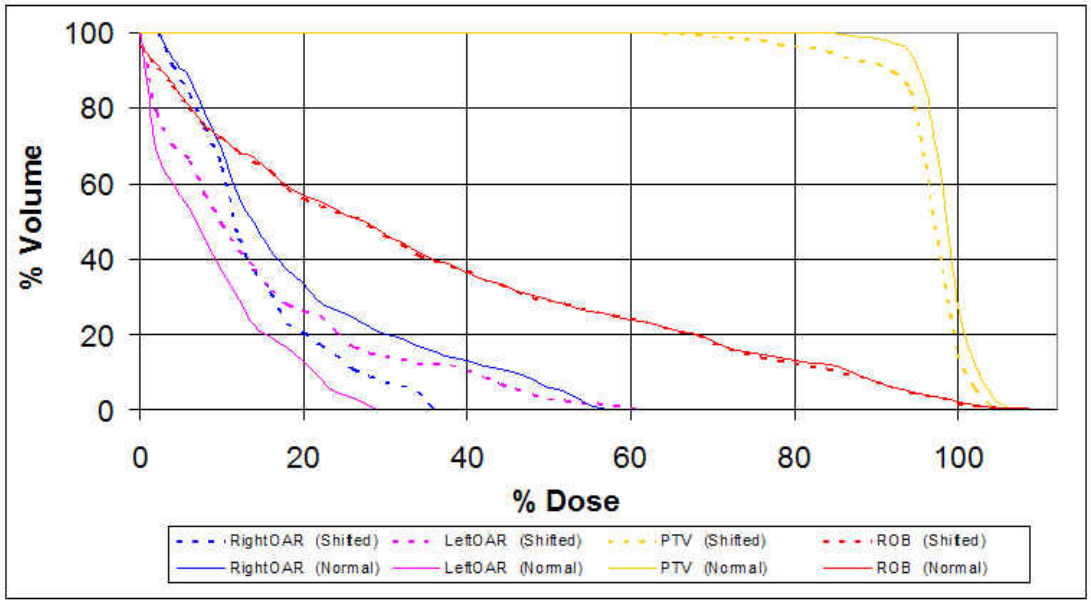
### 6.1.3 Patient Positioning Sensitivity

An important point regarding the fluence modulation (and therefore compensator surface) is that reducing unnecessary oscillations and increasing the surface smoothness substantially

reduces the dosing error associated with patient positioning uncertainties. The effects of smoothing the fluence modulation during planning (through beamlet parametrization) was demonstrated by Markman et al [27] for the case of MLC based treatment planning. A similar study was undertaken herein to prove that this applies equally well to DCPO and compensator based planning with beamlet parametrization. Patient positioning errors are simulated in this study by shifting the patient geometric data by some arbitrary amount and analyzing the effects on the resulting dose distribution through a single forward calculation using the DCPO optimized compensator surfaces. Using both the DCPO and direct beamlet plans computed above, results have been generated and compared for shifts of $3\,mm$ in each of the positive and negative directions of each coordinate axis. The effects on the dose field are completely straight forward since a homogeneous tissue assumption has been used herein. These effects are quite difficult to visualize for the entire dose field, therefore only dose volume histograms are included as they provide a more quantitative comparison of these effects. The DVHs for the positive x-direction are shown in Figure 6.6. Specific data points have been extracted for this positive x-direction as well as each of the remaining three directions and are presented in Table 6.1. The data in this table shows that for three of the four chosen directions the solution sensitivity is nearly equivalent, however, quite striking is the positive x-direction shift, which shows a 12% reduction in PTV volume change for 96% dose coverage. While this single case may not prove complete generality of this effect, when combined with the similar findings of Markman et al [27] one finds sufficient evidence to support the claim.




(a) 5 $mm$ Beamlets




(b) DCPO

Figure 6.6: Patient Positioning Error Analysis

Table 6.1: DVH Comparison for Positioning Error (% Volume at Specified Dose)

| | | PTV @ 96% Dose | | Left OAR @ 20% Dose | | Right OAR @ 20% Dose | |
|---|---|---|---|---|---|---|---|
| | | Bixel | DCPO | Bixel | DCPO | Bixel | DCPO |
| | Normal | 88 | 85 | 3 | 13 | 26 | 33 |
| + | Shifted | 56 | 65 | 16 | 26 | 12 | 21 |
| x | Difference | -32 | -20 | +13 | +13 | -14 | -12 |
| - | Shifted | 85 | 81 | 0 | 0 | 38.5 | 26 |
| x | Difference | -3 | -4 | -3 | -13 | +12.5 | -7 |
| + | Shifted | 88.5 | 85 | 2 | 14 | 32 | 40 |
| y | Difference | +0.5 | 0 | -1 | +1 | +6 | +7 |
| - | Shifted | 85 | 85 | 8 | 9 | 27 | 32 |
| y | Difference | -3 | 0 | +5 | -4 | +1 | -1 |

## 6.2 Full Patient Planning

This final example case is used to demonstrate the complete spectrum of performance of the direct compensator profile optimization technique developed herein. The familiar phantom example above is expanded to its true three-dimensional form. The contours of the critical structures are shown in Figure 6.7, with one image containing labels naming each of the structures. These contours were generated within a basic treatment planning system

developed specifically for this research effort. The dose prescription is as follows: PTV target dose is 100% (minimum 98% and maximum 102%), left and right organ doses are minimized (maximum allowable dose is 45%), and the ROB dose is minimized as well. Five equally space coplanar beams are again used.

The results for this full three-dimensional plan are shown in Figures 6.8 - 6.9, which display the complete dose volume histograms and relative dose fields, respectively. The DVH reveals that the plan is of good quality and shows that DCPO is a viable approach for IMRT treatment planning. Two of the resulting five compensators are shown in Figure 6.11 for illustrative purposes. The smoothness of these compensators reflects that of all five and show that DCPO provides quality plans and machinable compensators. In addition, a complete milling simulation analysis was performed on each of the five resulting compensators (one such simulation is shown in Figure 6.12) yielding a maximum undercut of only $0.3\,cm$, which clearly proves the high level of machinability that DCPO is able to obtain.

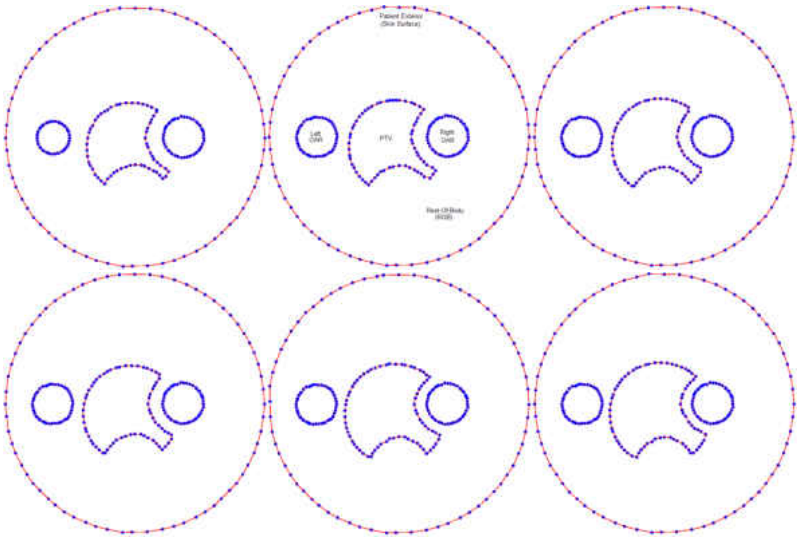


Figure 6.7: Full Patient Contours on Each Slice

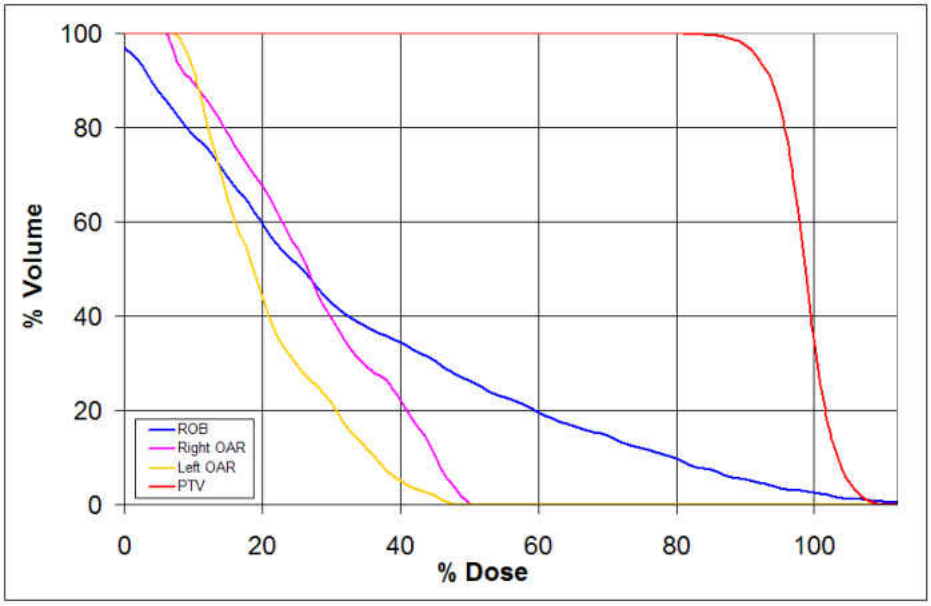



Figure 6.8: Full Patient Dose Volume Histogram

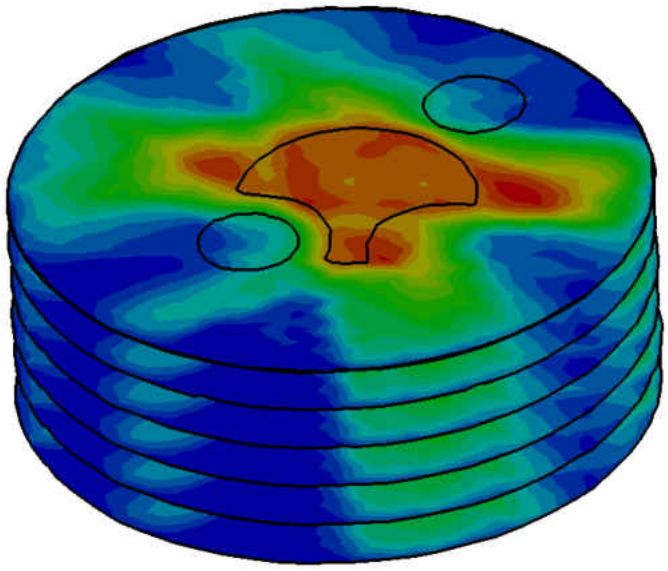


Figure 6.9: Full Patient Relative Dose Field (3D)

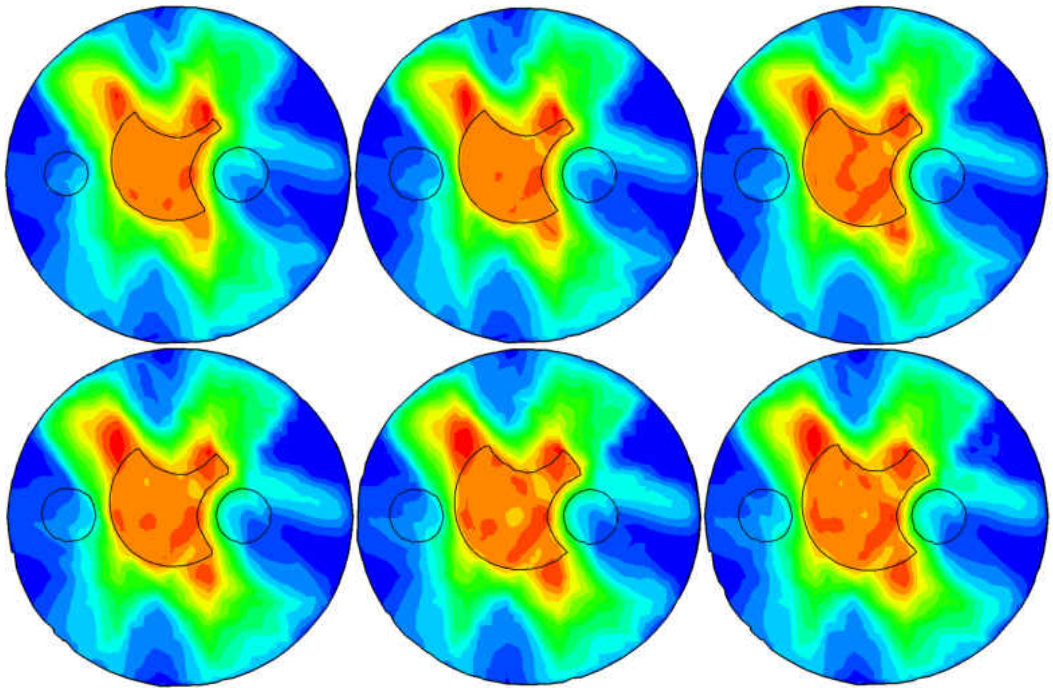


Figure 6.10: Full Patient Relative Dose Field Slices

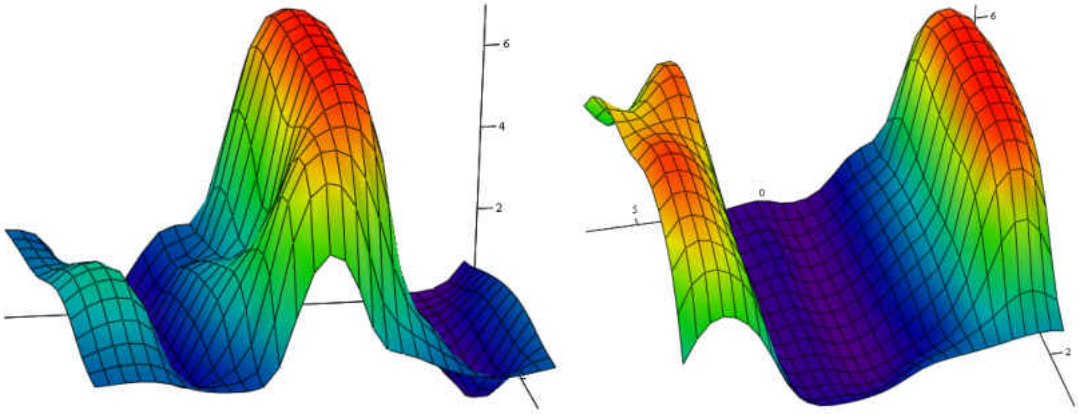



Figure 6.11: Full Patient Compensators (two of five shown)

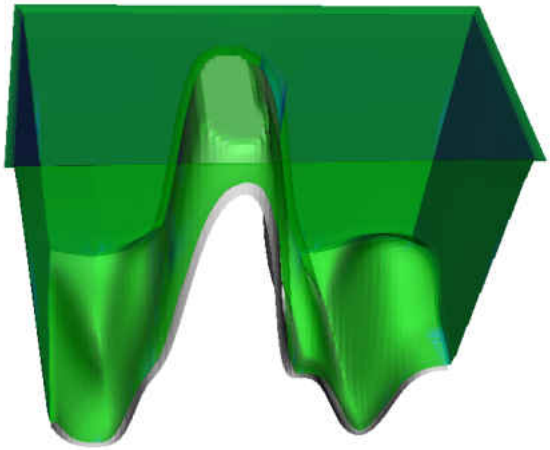


Figure 6.12: Compensator Milling Simulation Analysis

This chapter has provided detailed comparisons with existing treatment planning approaches, provided quantification of DCPO performance, as well as demonstrated the quality achieved by the DCPO process for realistic treatment planning, all of which have been completed with satisfactory results. These examples have provided an in-depth look at DCPO and established the results necessary to prove its utility. Final observations and conclusions regarding this research and directions for future extension of these ideas are provided in the upcoming final chapter.

# CHAPTER 7
# CONCLUSIONS

This research has provided the foundations for the first direct delivery parameter optimization scheme for solid compensator intensity modulated radiation therapy treatment planning. This paper has highlighted the development of this innovative delivery parameter optimization scheme, named Direct Compensator Profile Optimization (DCPO). This approach utilizes analytical Non-Uniform Rational B-spline (NURB) surfaces to represent the compensator during IMRT inverse planning. These NURB surfaces have also been appropriately limited, through a detailed milling constraint analysis, such that the allowable compensator surfaces are completely machinable. The use of standard dose calculation routines has been facilitated by discretization of the analytic surface into individual beamlets whose strengths are provided by the corresponding compensator surface height. In addition, the effects of this new approach on the performance of several numerical optimization schemes has been quantified, and a standard conjugate gradient approach has been adopted as the most suitable technique for the inverse planning process. Overall, the DCPO approach has been shown to provide several advantages compared to conventional beamlet strength optimization approaches while maintaining overall plan quality:

- less parameters are needed to accomplish the fluence modulation
- planned compensator surfaces are nearly 100% machinable using conventional three-axis milling

- effects of the delivery device on dose are included during inverse planning, with no performance penalty
- planned patient dose more accurately reflects the actual deliverable patient dose

This research has demonstrated the performance of DCPO compared to other state-of-the-art treatment planning approaches through direct numerical examples using various simulated patient configurations. A thorough analysis of the benefits and limitations of DCPO has been described showing that DCPO computed plans are comparable in quality to those developed using conventional approaches, while providing the four major benefits listed above. Additionally, it has been shown that the effects of patient positioning errors may be reduced in some cases by use of DCPO.

Although rigorous verification testing was completed and presented herein, further fine tuning of the various process steps may bring about additional performance benefits. Specifically, several areas are recommended for continued development and exploration of this technique. Possible research tasks, as well as a statement of focus for each of these areas is provided below.

- NURB surface representation
  - allow for non-uniform knot vectors
  - allow for variable surface resolution
  - explore the effects of allowing variable x and y coordinates for the control points

  Determine how the overall optimization parameter count can be further reduced without sacrificing plan quality or surface machinability.

- Inverse planning
  - study the effects of prescriptions on plan quality
  - incorporate dose-volume-constraints and analyze their effects on the solution space
  - explore the effects of critical structure weighting on solution quality (possibly Pareto optimal solutions) to assess planning trade-offs between organ sparing and tumor dose homogeneity

  Determine the feasibility of standardizing or automating dose prescriptions in order to increase plan quality and robustness.
- Dose calculations
  - incorporate more accurate dose calculation routines
  - verify and test performance of DCPO for inhomogeneous tissue treatments
  - increase the fidelity of the surface-to-fluence conversion process by allowing non-uniform/adaptive beamlet sizes

  Determine if final plan quality is significantly improved by increasing dose calculation accuracy *during planning*.
- Manufacturing
  - incorporate high-speed machining simulation into the planning process (incorporate machine time into objective function)
  - continue case study analysis to determine mass production appropriateness of NURBs constraint parameters

- expand the constraint analysis for the case of five-axis milling

Determine whether constraint refinement and adjustment will allow for improvements in plan quality.

This research has introduced a new technique for the planning of Intensity Modulated Radiation Therapy using solid compensators. Although several important areas of focus have been listed for future exploration, this research has provided a fundamental start into development of this promising new technique. It is hoped that future investigations can continue this work and bring DCPO into the clinical setting, such that real patients can benefit from the improved treatment plans that this approach is able to provide.

# LIST OF REFERENCES


[1] "RT Answers", American Society for Radiation Oncology (2009), www.rtanswers.com/about/history/history.htm.

[2] "The .decimal Point", .decimal, Inc. (2008), www.dotdecimal.com.

[3] "http://www.varian.com/us/oncology/radiation_oncology/clinac/clinac_cx.html", Varian Medical Systems (2009).

[4] Khan, F.M., *The Physics of Radiation Therapy*, vol. 39, Lippincott Williams & Wilkins (2003).

[5] Washington, C.M. and Leaver, D., *Principles and Practice of Radiation Therapy*, Mosby, 2 edn. (2003).

[6] Webb, S., "Intensity-Modulated Radiation Therapy (IMRT): A Clinical Reality for Cancer Treatment, 'Any Fool Can Understand This"', *The British Journal of Radiology*, **78**, S64–S72 (2005).

[7] Webb, S., *Contemporary IMRT: Developing Physics amd Clinical Implementation*, Institute of Physics Publishing (2004).

[8] Webb, S., *Intensity-Modulated Radiation Therapy*, CRC Press (2001).

[9] Palta, J. and Mackie, T., eds., *IMRT- The State of the Art*, AAPM Medical Physics Monograph Number 29, Medical Physics Publishing (2003).

[10] 2001, I.M.R.T.C.W.G., "Intensity Modulated Radiotherapy: Current Status and Issues of Interest", *International Journal of Radiation Oncology Biology Physics*, **51**, 880–914 (2001).

[11] "www.astro.org", American Society for Therapeutic Radiology and Oncology (2009).

[12] "www.radonc.uchicago.edu", University of Chicago (2009), http://www.radonc.uchicago.edu/typea/pat_new-advancements.html.

[13] Haas, O.C.L., Burnham, K.J. and Mills, J.A., "Optimization of Beam Orientation in Radiotherapy using Planar Geometry", *Physics in Medicine and Biology*, **43**, 2179–2193 (1998).

[14] Hou, Q., Wang, J., Chen, Y. and Galvin, J., "Beam Orientation Optimization for IMRT by a Hybrid Method of the Genetic Algorithm and the Simulated Dynamics", *Medical Physics*, **30**(9), 2360–2367 (2003).

[15] Pugachev, A.B., Boyer, A.L., and Xing, L., "Beam Orientation Optimization in Intensity-Modulated Radiation Treatment Planning", *Medical Physics*, **27**(6), 1238–1245 (2000).

[16] Rowbottom, C.G., Khoo, V.S. and Webb, S., "Simultaneous Optimization of Beam Orientations and Beam Weights in Conformal Radiotherapy", *Medical Physics*, **28**(8), 1696–1702 (2001).

[17] Chang, S., Cullip, T.J., Deschesne, K.M., Miller, E.P. and Rosenman, J.G., "Compensators: An Alternative IMRT Delivery Technique", *Journal of Applied Clinical Medical Physics*, **5**(3), 15–36 (2004).

[18] Chang, S., Deschesne, K., Chen, H., Weeks, K., Sibata, C., Carey, E., Levinson, L. and Potter, L., "A Multi-Institutional Retrospective Study On Clinical IMRT Treatment Delivery Efficiency", *Medical Physics*, **35**(6), 2746 (2008).

[19] Tang, X., Tyagi, N., Wang, J., Rice, R., White, G., Cashon, K., Vanderspek, L., Pawlicki, T., Murphy, K. and Jiang, S., "Dosimetric Comparison of Compensator and Multi-Leaf Collimator Based Intensity Modulated Radiosurgery", *Medical Physics*, **35**(6), 2826 (2008).

[20] Tyagi, A., Nangia, S., Chufal, K., Mishra, M., Ghosh, D. and Singh, M.P., "Quality Assurance and Dosimetric Analysis of Intensity Modulated Radiotherapy Using Compensators for Head and Neck Cancers", *Medical Physics*, **35**(6), 2763 (2008).

[21] Kuros-Zolnierczuk, J., Papiez, E., Papiez, L. and Lu, X., "MLC IMRT Vs. Solid Compensators Based IMRT: Comparison of Various Clinical Cases", *Medical Physics*, **33**(6), 2121 (2006).

[22] Ehler, E.D., Nelms, B.E. and Tome, W.A., "On the Dose to a Moving Target While Employing Different IMRT Delivery Mechanisms", *Radiotherapy and Oncology*, **83**(1), 49–56 (2007).

[23] Bortfeld, T., Jokivarsi, K., Goitein, M., Kung, J. and Jiang, S.B., "Effects of Intra-Fraction Motion on IMRT Dose Delivery: Statistical Analysis and Simulation", *Physics in Medicine and Biology*, **47**, 2203–2220 (2002).

[24] Jiang, S.B. and Ayyangar, K.M., "On Compensator Design for Photon Beam Intensity-Modulated Conformal Therapy", *Medical Physics*, **25**(5), 668–675 (1998).

[25] Stathakis, S., "Skin Dose Measurements for IMRT Delivery", in ".decimal Compensator IMRT Symposium", (2008).

[26] Yaeger, T. and Cost, N., "Compensator Based IMRT for Whole Breast with Simultaneously Integrated Seroma Boost", in ".decimal Compensator IMRT Symposium", (2008).

[27] Markman, J., Low, D., Beavis, A.W. and Deasy, J.O., "Beyond Bixels: Generalizing the Optimization Parameters for Intensity Modulated Radiation Therapy", *Medical Physics*, **29**(10), 2298–2304 (2002).

[28] van Santvoort, J., Binnekamp, D., Heijmen, B. and Levendag, P., "Granulate of Stainless Steel as Compensator Material", *Radiotherapy and Oncology*, **34**, 78–80 (1995).

[29] Chang, S., "Compensator-Intensity-Modulated Radiotherapy- A Traditional Tool for Modern Application", *US Oncological Disease*, 80–84 (2006).

[30] Yu, C., Shepard, D., Earl, M. and Yi, B., "Simplifying IMRT with Direct Aperture Optimization", *Medical Physics*, **32**(6), 2084–2085 (2005).

[31] van Asselen, B., Schwarz, M., van Vliet-Vroegindeweij, C., Lebesque, J., Mijnheer, B. and Damen, E., "Intensity-Modulated Radiotherapy of Breast Cancer Using Direct Aperture Optimization", *Radiotherapy and Oncology*, **79**(2), 162–169 (2006).

[32] CMS Inc., *CMS XiO: Intensity Modulated Radiation Therapy (IMRT) - Technical Reference*, xio release 4.2 edn. (2004).

[33] Erhart, K., Divo, E. and Kassab, A., "An Inverse Meshless Technique For The Determination Of Non-Linear Heat Generation Rates In Living Tissue", in "Proceeding of NHT-05 Eurotherm82", (2005).

[34] Erhart, K., Divo, E. and Kassab, A., "An Inverse Localized Meshless Technique for the Determination of Non-Linear Heat Generation Rates in Living Tissue", in "ASME 2006 Summer Heat Transfer-Fluids Conference", (2006).

[35] Erhart, K., Divo, E. and Kassab, A., "An Evolutionary-Based Inverse Approach for the Identification of Non-Linear Heat Generation Rates in Living Tissues Using a Localized Meshless Method", *International Journal of Numerical Methods in Heat and Fluid Flow*, **18**, 401–414 (2008).

[36] Press, W.H., Teukolsky, S.A., Vetterling, W.T. and Flannery, B.P., *Numerical Recipes in C*, Cambridge University Press, 2 edn. (1992).

[37] Schneider, P.J., "NURB Curves- A Guide for the Uninitiated", *MacTech*, **25**, 1–10 (1996).

[38] Dupree, S.A. and Fraley, S.K., *A Monte Carlo Primer: a Practical Approach to Radiation Transport*, Kluwer Academic-Plenum (2002).

[39] Niven, D. and Lewis, J., "Monte Carlo Investigation of the Effects of an IMRT Compensator on Linear Accelerator Beam Output", *Medical Physics*, **34**, 2559 (2007).

[40] Chibani, O. and Charlie-Ma, C.M., "On the Discrepancies Between Monte Carlo Dose Calculations and Measurements for the 18 MV Varian Photon Beam", *Medical Physics*, **34**, 1206–1216 (2007).

[41] Fippel, M., "Fast Monte Carlo Dose Calculation for Photon Beams Based on the VMC Electron Algorithm", *Medical Physics*, **26**, 1466–1475 (1999).

[42] Deasy, J.O., Wickerhauser, M.V. and Picard, M., "Accelerating Monte Carlo Simulations of Radiation Therapy Dose Distributions Using Wavelet Threshold De-Noising", *Medical Physics*, **29**, 2366–2373 (2002).

[43] Spezi, E. and Lewis, G., "An Overview of Monte Carlo Treatment Planning for Radiotherapy", *Radiation Protection Dosimetry*, **131**, 123–129 (2008).

[44] Plessis, F.C. and Willemse, C.A., "Inclusion of Compensator-Induced Scatter and Beam Filtration in Pencil Beam Dose Calculations", *Medical Physics*, **33**(8), 2896–2904 (2006).

[45] Ulmer, W., Pyyry, J. and Kaissl, W., "A 3D Photon Superposition/Convolution Algorithm and its Foundation on Results of Monte Carlo Calculations", *Physics in Medicine and Biology*, **50**, 1767–1790 (2005).

[46] Ahnesjö, A., "Collapsed Cone Convolution of Radiant Energy for Photon Dose Calculation in Heterogeneous Media", *Medical Physics*, **16**, 577–592 (1989).

[47] Jeraj, R. and Keall, P., "Errors in Inverse Treatment Planning Based on Inaccurate Dose Calculation", in "The Use of Computers in Radiation Therapy", (2000).

[48] Scholz, C., Nill, S. and Oelfke, U., "Comparison of IMRT Optimization Based on a Pencil Beam and a Superposition Algorithm", *Medical Physics*, **30**(7), 1909–1913 (2003).

[49] Kirkpatrick, S., Gelatt, C. and Vecchi, M., "Optimization by Simulated Annealing", *Science*, **220**, 671–680 (1983).

[50] Carr, R., "Simulated Annealing", (2009), from MathWorld–A Wolfram Web Resource, created by Eric W. Weisstein.

[51] Davis, L., *Handbook of Genetic Algorithms*, Van Nostrand Reinhold (1991).

[52] Kohonen, J., "A Brief Comparison of Simulated Annealing and Genetic Algorithm Approaches", www.cs.helsinki.fi (1999).

[53] Ahn, C.W., *Advances in Evolutionary Algorithms: Theory, Design, and Practice*, Studies in Computational Intelligence, Springer (2006).

[54] Kicinger, R., Arciszewski, T. and DeJong, K., "Parameterized versus Generative Representations in Structural Design: An Empirical Comparison", in "Proceedings of the Genetic and Evolutionary Computation Conference (GECCO 2005)", 2007–2015, The Association for Computing Machinery (2005).

[55] Polak, E., *Computational Methods in Optimization*, Academic Press (1971).

[56] Brodlie, K., *The State of the Art in Numerical Analysis*, Academic Press (1977).

[57] Nelder, J. and Mead, R., "A Simplex Method for Function Minimization", *Computer Journal*, **7**, 308–313 (1965).

[58] Bramlette, M.F. and Bouchard, E.E., *Genetic Algorithms in Parametric Design of Aircraft*, chap. 10, 109–123, Van Nostrand Reinhold (1991).

[59] Ralston, A. and Rabinowitz, P., *A First Course in Numerical Analysis*, International Series in Pure and Applied Mathematics, McGraw-Hill (1978).

[60] Erhart, K., Divo, E., Kassab, A. and Bialecki, R., "The Truncated Pod Laplace Inversion Method for Transient Heat Conduction", in "Cimenics VIII - International Congress on Numerical Methods in Engineering and Applied Sciences", (2006).

[61] "An Introduction to Genetic Algorithms", Burns Statistics, www.burns-stat.com (2008).

[62] Chaparro, B., Thuillier, S., Menezes, L., Manach, P. and Fernandes, J., "Material Parameters Identification: Gradient-Based, Genetic and Hybrid Optimization Algorithms", *Computational Materials Science*, **44**(2), 339–346 (2008).

[63] Deasy, J., "Multiple Local Minima in Radiotherapy Optimization Problems with Dose-Volume Constraints", *Medical Physics*, **27**(7), 1157–1161 (1997).

[64] Pflugfelder, D., Wilkens, J.J., Nill, S. and Oelfke, U., "A Comparison of Three Optimization Algorithms for Intensity Modulated Radiation Therapy", *Zeitschrift für Medizinische Physik*, **18**(2), 111–119 (2008).

[65] Zhou, L.H., Tang, M.T., Wang, Z.Y., Chen, C.M., Lü, Q.W. and Jin, H.Y., "Genetic Algorithm-Based Dose Optimization in Intensity Modulated Radiation Therapy", *Journal of Southern Medical University*, **27**, 46–48 (2007).

[66] Fuller, C.D., Choi, M., Forthuber, B., Wang, S.J., Rajagiriyil, N., Salter, B.J. and Fuss, M., "Standard Fractionation Intensity Modulated Radiation Therapy (IMRT) of Primary and Recurrent Glioblastoma Multiforme", *Radiation Oncology*, **2**(26), 1–10 (2007).