Contents

Table of Abbreviations

AD5	abdomen (5mm) problem
AN	acoustic neuroma problem
AN3	acoustic neuroma (3mm) problem
AN5	acoustic neuroma (5mm) problem
AVM	arterial veinous malformation problem
BAO	beam angle optimization
CHIM	convex hull of individual minima
CPU	central processing unit
3DCRT	three dimensional conformal radiation therapy
CT	computed tomography
CUB	critical organ upper bounds
DM	decision maker
DNA	deoxyribonucleic acid
DVH	dose-volume histogram
DVC	dose-volume constraint
EUD	equivalent uniform dose
FMO	fluence map optimization
gEUD	generalized equivalent uniform dose
IMB	intensity modulated beam
IMRT	intensity modulated radiation therapy
ITP	intersection points
LG5	lung (5mm) problem
LP	linear programme

MATLAB	MATrix LABoratory
MIP	mixed-integer programme
MLC	multi-leaf collimator
MO	multiple objective
MOP	multiobjective programme
MOLP	multiobjective linear programme
MONP	multiobjective nonlinear programme
MRI	magnetic resonance imaging
NC	normal constraint
NBI	normal boundary intersection
NDP	nondominated points
NTCP	normal tissue complication probability
NUB	normal tissue upper bounds
OAR	organ at risk
PL	pancreatic lesion problem
PR	prostate problem
PTV	planning target volume
PET	positron emission tomography
RFP	reference points
SA	simulated annealing
TCP	tumor control probability
TLB	tumor lower bounds
TUB	tumor upper bounds
GA	genetic algorithm

Chapter 1

Introduction

Radiation therapy (also called radiotherapy) is one of the major forms of treatment for cancer. About 60 percent of people with cancer are treated with radiation therapy, as the only form of treatment or in combination with surgery or chemotherapy.

Radiation therapy uses high-energy particles, or waves (radiation), such as gamma rays or X-rays to kill cancer cells and stop them from spreading. The radiation injures or destroys cells in the area being treated (the target volume) by damaging their genetic material (DNA), interfering with their ability to grow and divide. Although radiation damages both cancer cells and normal cells, most normal cells that are exposed to moderate amounts of radiation can recover from the effects and function properly. The goal of radiation therapy is to damage as many cancer cells as possible, while limiting harm to nearby healthy tissue.

Radiation therapy can be internal radiation or external beam radiation. Internal radiation therapy is a form of treatment where a source of radiation is put inside the patient's body, it is typically used when a high dose of radiation needs to be delivered to a small area; while in external beam radiation therapy a linear accelerator is used to aim high-energy radiation beams to the patient's tumor site from several directions. External beam radiotherapy is the most frequently used form of radiotherapy.

In external beam radiation therapy, a main research topic is how to realize a high degree of conformality, i.e., restricting the high dose area to the tumor vol-

ume. Over the past twenty five years, major improvements were achieved by the development of three-dimensional conformal radiation therapy (3DCRT), especially intensity modulated radiation therapy (IMRT).

Three-dimensional conformal radiation therapy uses imaging machines (such as computed tomography (CT)) and computer technology to create a three-dimensional picture of the tumor so that each radiation beam can be shaped exactly (conform) to the contour of the treatment area from the beam's eye view (BEV). The radiation beams normally have a uniform fluence across the field, or, where appropriate, have this intensity modified by simple beam fluence (intensity) modifying devices, such as wedges or compensating filters. Because the healthy tissue surrounding the tumor is largely spared by this technique, higher doses of radiation can be used to treat the cancer.

Intensity-modulated radiation therapy (IMRT) is an advanced form of 3DCRT. It is the delivery of radiation to the patient via beams that have non-uniform radiation fluence, also called intensity modulated beams (IMBs). These IMBs allow the beam field to be subdivided into smaller beamlets (bixels), each of which can be assigned deliverable intensity. This technique allows a precise adjustment of radiation doses to the tissue within the target area. As a result, more radiation can be delivered to the tumor while less is delivered to the surrounding normal tissue.

This thesis will concentrate on using optimization approaches to improve the treatment planning process of IMRT. We will introduce the treatment process of IMRT in Section 1.1 and treatment planning methods for external beam radiation therapy in Section 1.2. In Section 1.3, we will look at three main optimization problems of IMRT. Finally, we give the outline of the thesis in Section 1.4.

1.1 The Treatment Process of IMRT

The process of treating a patient with IMRT is an elaborate one. It includes imaging, treatment planning, plan confirmation and treatment. During this process, a treatment team, including a radiation oncologist, radiation physicist and radiation

therapist will be involved.

Imaging IMRT begins with the imaging process. First, the patient is immobilized. Sometimes masks, pads or other devices may be used to help the patient to hold still. The patient is then scanned in the position which is also used for the treatment later on a dedicated high-speed computed tomography (CT) scanner. When necessary, other imaging studies, such as magnetic resonance imaging (MRI) scans or positron emission tomography (PET) scans can be used with the CT images together. Then the radiation oncologist locates the tumor volume, planning target volume (PTV) which is an extension of the tumor volume containing tumorous cells only visible at a microscopic level, the organs at risk (OARs) on these images.

Treatment Planning Once the location and the volume of the tumor are identified, the radiation oncologist and the radiation therapist (the treatment planner) work together to produce a plan to actually treat the patient. The radiation oncologist determines the prescription requirements which includes how much radiation to deliver to the PTV and how much radiation can the normal tissue and OARs tolerate. Then a treatment plan is developed with the aid of a computerized treatment planning system. Two approaches of treatment planning are outlined in Section 1.2. The plan specifies the number, size and shape of the radiation beams, as well as their angle of incidence (direction), and the beam intensity profiles for each beam.

The plan is judged by viewing the level curves of the radiation per slice, called isodose curves, and by the cumulative dose-volume histogram (DVH). A dose-volume histogram is a plot of dose or percent dose (relative to tumor prescription dose) versus the percent volume. An ideal DVH curve for the PTV would have the target at 100% percent dose for the entire volume and then drop immediately to zero, indicating that the target is treated exactly as specified with no under or over dosing. The curves for the critical structures would instead drop immediately to zero, meaning that they receive no radiation.

Plan Confirmation The radiation oncologist and radiation physicist then meet and review the plan together, reviewing carefully the dose to the tumor and each surrounding organ, to ensure that the plan fulfils the prescription requirements. Then the treatment is simulated using a special treatment simulator. This part aims to verify the physical parameters and to acquaint the patient with the procedure.

Treatment Subsequently, the actual treatment is carried out by the radiation therapist. The radiation is delivered by a linear accelerator (see Figure 1.1) with a multileaf collimator (see Figure 1.2) equipped in the head of the treatment unit. The gantry of the linear accelerator can be rotated around the patient so that radiation beams can be sent from different angles. The multileaf collimator, which consists of many metal leaves that can move across channels, is used not only to shape the beams but also to form intensity modulated beams (IMBs).

Figure 1.1: *A linear accelerator (SouthEastMissouriHospital, 2008).*

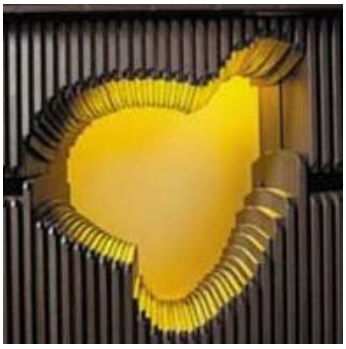

Figure 1.2: *A close-up view of a multileaf collimator (Varian, 2008).*

1.2 Treatment Planning

There are two approaches to treatment planning for external beam radiation therapy, forward planning and inverse planning (see Figure 1.3). In IMRT, inverse planning is used.

In forward planning, a treatment planner enters all necessary treatment param-

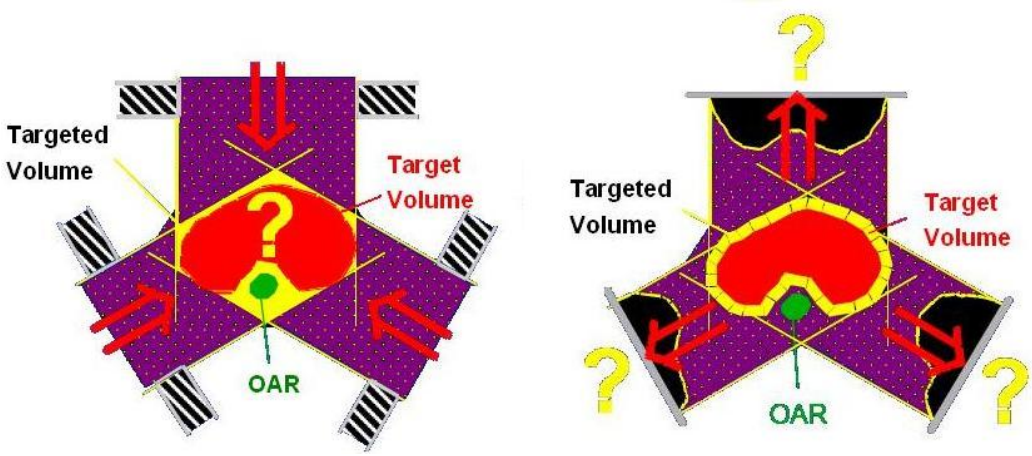

Figure 1.3: *In forward planning (left), treatment parameters are specified before dose calculation. In inverse planning (right), a desired dose distribution is prescribed and appropriate parameters are found for the treatment set up (Schlegel and Mahr, 2001).*

eters such as beam intensity, beam directions and number of beams into computer software and a dose distribution is calculated. If the dose is unacceptable, the task is repeated in a trial-and-error way until a satisfying dose distribution is achieved. Certainly, there is no guarantee that the best possible solution is found. Therefore this approach is time consuming and unsatisfying for complicated anatomical situations.

According to Webb (1997), inverse treatment planning was first discussed by Brahme in the early 1980s. Instead of the planner trying a variety of configurations of beams until a suitable match is found to the dose prescription, the reverse is attempted in inverse planning. The core of an inverse planning system is an optimization model. In fact, the first optimization model for radiotherapy is Bahr *et al.* (1968). Radiotherapy planning is considered as an optimization problem in which treatment goals are formulated either as objective functions or as constraints and the treatment planning parameters are variables. The dose prescription information "drives" the optimization model to find the best solution satisfying the constraints. This approach avoids the time consuming trial-and-error search for appropriate beam parameters. However, usually it is mathematically harder to solve

since it is typically a large dimensional problem that requires complex computational algorithms. In Section 1.3, we will look at three main optimization problems in IMRT.

1.3 Optimization Problems in IMRT

In IMRT planning, a set of parameters, such as number of beams, orientation of beams, intensity of each beam etc. need to be decided for the delivery of a certain radiation dose to the patient. This is done by treatment planner with the aid of an inverse planning system with an optimization engine. Some parameters are chosen by the planner while the others are calculated by the inverse planning system. The computer optimization can potentially simplify the tedious planning procedure and yield the best possible plan.

The decision of the parameters is of crucial importance for the quality of the radiation plan. It can be done by solving three optimization problems sequentially, they are the beam angle optimization (BAO) (Ehrgott *et al.*, 2005; Gong, 2007), the fluence map optimization (FMO) (Lim *et al.*, 2007) and the realization problem (Ahuja and Hamacher, 2004; Ehrgott *et al.*, 2007a), see Figure 1.4.

1. Optimization of beam angles. IMRT involves the delivery of multiple intensity modulated beams at various orientations. The directions selected for delivery of radiation greatly affect radiation dose levels in the tumor and healthy tissue/organs surrounding the tumor. To optimally select the directions, it is necessary to develop functions that measure the quality of a certain combination of directions. Thus optimization can be used to find good combinations of beam angles. This problem is also called BAO problem or geometry problem.

2. Optimization of fluence map (beam intensities). For the treatment planning process a desired dose level can be specified. Given the number of beams and the direction of each beam, the FMO problem (or the beam intensity optimization problem) aims to find the appropriate beam intensities that will deliver the best dose distribution according to objective functions.

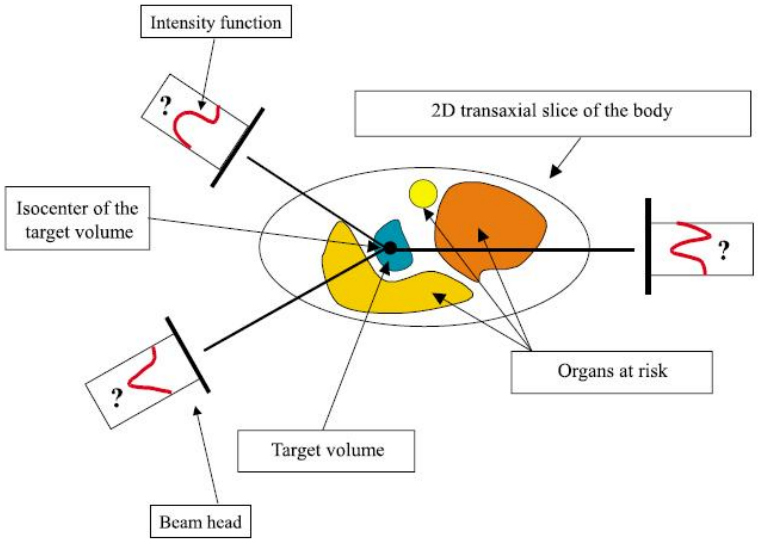

Figure 1.4: *Radiation treatment planning problems: Where to stop the gantry? How much radiation should be delivered? How is the radiation modulated? (from Boland* et al. *(2004)).*

3. Optimization of the realization. Once the fluence maps (IMBs) are determined, one must convert these into MLC leaf sequences that approximately realize them.

   IMBs may be constructed using a sequence of static MLC-shaped fields in which the shape changes between the delivery of quanta of fluence, the so-called multiple-static-field technique. The multiple segment fields are designed for the selected orientations (obtained from BAO) of the gantry under computer control. The radiation is turned on only when the MLC leaves are stopped at each prescribed segment position. This method is also referred to as step-and-shoot or stop-and-shoot. Alternatively, the leaves may define changing shapes with the radiation on, the so-called dynamic multileaf collimator (DMLC) technique. Usually, for a fixed gantry position, the opening formed by the pairs of opposing MLC leaves is swept across the beam volume and also the shape of the opening can be changed under computer control, with the radiation beam on, to produce the desired fluence profiles. This approach is also referred to as the sliding window technique.

Beam-on time is the radiation on time, while treatment time is the duration of the radiation treatment, it includes the beam-on time and the MLC set up time for each beam. The goal of the realization problem is to find a sequence of configurations of a multileaf collimator to deliver the treatment in such a way that the treatment time as well as the beam-on time is as small as possible.

The three problems are related to each other. As the input of one problem is the output of another problem, the solution of one problem affects the solution of the others. Therefore, the ideal is, in fact, to optimize them simultaneously. Unfortunately, the complexity of this integrated approach is too overwhelming, and only some research has been undertaken to combine the first two problems (Schreibmann *et al.*, 2004) or the last two problems together (Cotrutz and Xing, 2003; Li *et al.*, 2003; Preciado-Walters *et al.*, 2004).

1.4 Outline of the Thesis

Optimization in IMRT is too wide to address completely. In this thesis, we focus on the beam intensity optimization problem, i.e., given the number of beams and the direction of each beam, we optimize the intensity of each beam. Since the goals of delivering a sufficiently high uniform dose to the tumor and a small dose to the OARs and normal tissue contradict each other, we cannot achieve them at the same time. This makes the beam intensity optimization problem inherently multicriterial, i.e., there is no single best solution, instead, there are many best compromises. Therefore, we formulate the beam intensity optimization problem as a multiobjective linear programming model and propose solution methods to solve big multiobjective linear programmes (MOLPs). Since there is imprecision in dose caculation and delivery solving the problem approximately with a small fraction of Gy (Gray, the unit of measure for radiation dose) as approximation error is acceptable. We consider exact and approximation methods to solve MOLPs. Most of the work presented in the following chapters is based on working papers by the author and collaborators that have appeared or will appear in the literature (Shao

and Ehrgott, 2006; Ehrgott *et al.*, 2007b; Shao, 2005, 2006; Shao and Ehrgott, 2008, 2007).

This thesis is organized as follows.

In Chapter 2, we will review existing models and solution methods for the beam intensity optimization problem. The advantage and disadvantage of each model will be discussed.

In Chapter 3, we formulate the beam intensity optimization problem as a multiobjective linear programming problem. We introduce Benson's outer approximation method to solve an MOLP in objective space and some improvements are proposed to improve the computational speed.

In Chapter 4, we propose an approximation version of Benson's algorithm. We prove that the approximation version of Benson's algorithm finds a set of ε-nondominated points. Both Benson's algorithm and our approximation version of Benson's algorithm are used to solve two dimensional clinical cases and their results are compared.

In Chapter 5, geometric duality theory for multiple objective linear programmes is used to derive a dual variant of Benson's outer approximation algorithm to solve multiobjective linear programmes in objective space. We prove that solving the dual provides a weight set decomposition. We compare the algorithm with Benson's algorithm on small illustrative and radiotherapy examples.

In Chapter 6, we propose an approximate dual variant of Benson's algorithm to approximately solve an MOLP. It is also proved that ε-nondominated points can be obtained. We apply this algorithm to the same two dimensional clinical cases as in Chapter 4 and we compare the results with those, which have the same approximation quality, obtained by the approximation version of Benson's algorithm.

We address the problem of finding a representative subset of the nondominated set in Chapter 7. We propose a revised normal boundary intersection method which combines the global shooting method and the normal boundary intersection method. This combination overcomes the deficiencies of both the global shooting method and the normal boundary intersection method; the representative subset has the property

of coverage and uniformity. We apply this method to some two dimensional clinical cases.

In Chapter 8, we apply the approaches that we have implemented in the previous chapters on three-dimensional treatment problems. To reduce the computation time, a technique of reducing the resolution in normal tissue is used and the solution quality is measured by comparing the solution with the solution obtained without using the technique. We also show how a representative subset of the nondominated set can help a decision maker to select a treatment plan. Finally, we conclude this thesis with a summary in Chapter 9.

Chapter 2

Optimization Models for the Beam Intensity Optimization Problem

2.1 Introduction

The IMRT beam intensity optimization problem has been extensively studied for a number of years. A survey on this subject from a mathematical point of view has been done by Shepard *et al.* (1999). Two factors affect the optimization performance and outcome, they are the form of the mathematical model and the optimization strategy.

In inverse treatment planning of IMRT, the clinical objectives are specified mathematically in the form of an objective function. The objective function measures the goodness of a treatment plan, so the choice of the objective function is crucial for the optimization of a treatment plan. For beam intensity optimization problem in IMRT, the objective function is a function of the beamlet intensities.

In this chapter, we briefly review the mathematical models and optimization methods for beam intensity optimization with emphasis on more recent publications. We classify the mathematical models into four categories: (1) linear programming models, (2) nonlinear programming models, (3) mixed integer programming models, and (4) multiple objective programming models. The advantage and disadvantage of each model will be discussed.

2.2 Dose Calculation and Prescription

The amount of radiation absorbed by the tissues is called the radiation dose (or dosage). Dose is measured in a unit called a Gray (abbreviated as Gy), absorbed energy per mass, i.e. 1 Gy = 1 J/kg. To evaluate the dose distribution in the patient, we need to discretise the patient body and the beams, see Figure 2.1.

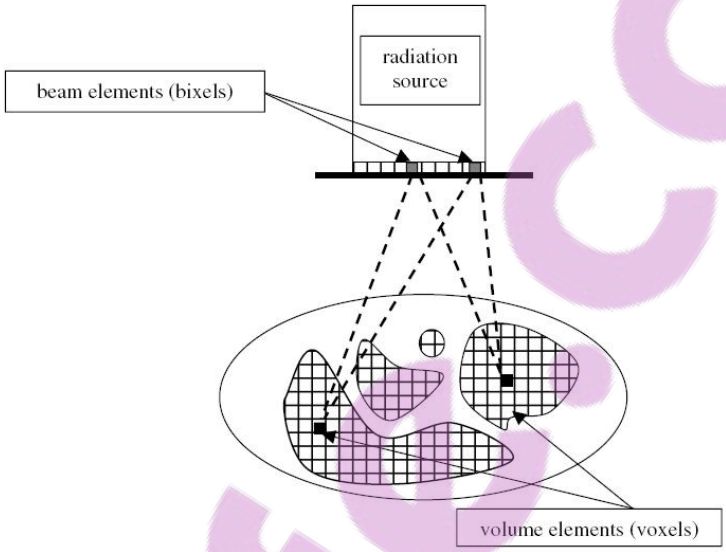

Figure 2.1: *Discretisation of volume elements and beam elements (Hamacher and Küfer, 2002).*

The patient's 3D volume is divided into small cubic voxels (volume elements). Each voxel is represented by a dose-point where deposited dose is calculated and it is assumed the dose value is the same throughout the voxel. Also, each voxel is assigned to a particular structure (tissue type). We use T to represent tumor, C to represent critical organs (K critical organs or OARs are represented by C_1, ..., C_K), and N to represent normal tissue. Then some voxels are marked as part of the tumor, some as part of the critical organ, some as part of the normal tissue, etc. The total number of voxels m is equal to $m_T + m_C + m_N$, where $m_C = m_{C_1} + \ldots + m_{C_K}$. Moreover, each beam at a given gantry angle is discretised into small bixels (beam elements), the edge length of which is determined by the width of an MLC leaf channel and the positions in which the leaf can stop. The number of voxels may

be tens or hundreds of thousands and the number of bixels can be up to 1,000 per beam.

Let n be the number of bixels. We assume the number and direction of beams are given. The dose distribution, described by an m-dimensional dose vector is calculated as

$$d = Ax, \tag{2.1}$$

where $d \in \mathbb{R}^m$ is a dose vector and its elements d_i correspond to the dose deposited in voxel i. Vector $x \in \mathbb{R}^n$ describes the beam intensity, x_j representing the intensity of bixel j. $A \in \mathbb{R}^{m \times n}$ is called dose deposition matrix. The elements a_{ij} of A represent the dose deposited in voxel i due to unit intensity in bixel j. Given the properties of the patient's anatomy and the beamlet source and orientation, a_{ij} is calculated by mathematical models of the physical behavior of radiation as it travels through the body. Sophisticated techniques are available and in clinical use (Nizin *et al.*, 2001) with the gold standard being Monte Carlo simulation (Verhaegen, 2003). We assume that A is given in this thesis. A can be partitioned and reordered into sub-matrices $A_T \in \mathbb{R}^{m_T \times n}, A_C \in \mathbb{R}^{m_C \times n}$ and $A_N \in \mathbb{R}^{m_N \times n}$ according to the rows corresponding to tumor, critical organ and normal tissue voxels, respectively.

Let $a(i)$ be the i-th row vector of dose deposition matrix A, then the dose deposited in voxel i is

$$d_i = a(i)x. \tag{2.2}$$

For treatment planning, the radiation oncologist needs to specify a "prescription dose" for the tumor, each organ at risk and the normal tissue because different tissues can tolerate various amounts of radiation. The "prescription dose" normally consists of a desired dose value for the tumor ($G_T \in \mathbb{R}$) and a specific dose value that we would prefer not to exceed for each organ at risk ($G_{C_k} \in \mathbb{R}$) and the normal tissue ($G_N \in \mathbb{R}$). They are used to construct $TG \in \mathbb{R}^{m_T}$, $TLB \in \mathbb{R}^{m_T}, TUB \in \mathbb{R}^{m_T}, CUB \in \mathbb{R}^{m_C}$ and $NUB \in \mathbb{R}^{m_N}$ representing goal dose to tumor voxels, lower bounds on the dose to tumor voxels, upper bounds on the dose to tumor voxels, upper bounds on the dose to critical organ voxels, and upper bounds on dose to

normal tissue voxels. CUB can be partitioned into $C_1UB, \ldots, C_KUB$ according to the voxels corresponding to different critical organs.

Most modern IMRT inverse planning systems allow the specification of dose-volume limitations. Typically a dose-volume constraint states that only a certain fraction of a structure is allowed to have dose values above some upper threshold or under some lower threshold. For example, in order to improve tumor dose while avoiding serious lung complications, instead of specifying the strict upper dose limit 20 Gy on the lung, the planner could specify that "no more than 40% volume of the lung can exceed a radiation dose of 20 Gy". The advantage of dose-volume constraints stems from the fact that dose-volume histograms (DVHs) are an economical way to represent the entire dose distribution to a structure, although they lose only the spatial information. Ignoring the considerations of location of dose within a structure, a few very carefully chosen points on the DVH curve can give an almost complete measure of the clinical quality of a dose distribution (Halabi *et al.*, 2006).

2.3 Feasibility Problem and Algorithms in IMRT

A special case of the beam intensity optimization problem is that if there is no objective function, then the optimization problem becomes a feasibility problem. It is assumed that every feasible solution is acceptable and all feasible solutions are of more or less equal quality.

To achieve tumor control and avoid normal tissue complication, some linear constraints such as lower and upper bounds to the target volume and upper bounds to the critical organs are imposed. In this thesis we use the notation $y^1 \leq y^2$ to indicate $y^1 \leqq y^2$ but $y^1 \neq y^2$ for $y^1, y^2 \in \mathbb{R}^p$ whereas $y^1 < y^2$ means $y^1_k < y^2_k$ for all $k = 1, \ldots, p$. The mathematical formula of these constraints is as follows:

$$\begin{aligned} TLB &\leqq A_T x \quad \leqq TUB \\ A_C x &\leqq CUB \\ 0 &\leqq x. \end{aligned} \tag{2.3}$$

Feasibility search algorithms seek a set of beam intensities that satisfies all above

dose constraints set by the planner. Projection algorithms are popular approaches for solving convex feasibility problems. A brief overview is found in Xiao *et al.* (2003).

The row action relaxation method of Agmon, Motzkin, and Schoenberg (AMS) (Agmon, 1954; Motzkin and Schoenberg, 1954) is a kind of sequential projection method, which has been proposed by Censor *et al.* (1988a) to solve the beam intensity feasibility problem. Censor *et al.* (1988b) propose to solve the linear feasibility problem by the Cimmino simultaneous projection algorithm. The essence of Cimmino's method is to project the present iterate simultaneously onto all involved sets and then calculate the centroid of the system. Superior to the row action relaxation method, it can give an approximate solution if there is no intersection between the convex sets.

The row action relaxation method and the Cimmino simultaneous projection algorithm are used to solve the linear feasibility problem, which is a special instance of the convex feasibility problem where the convex sets are the half-spaces described by linear inequalities. The convex feasibility problem is to find a point (any point) in the non-empty intersection of the convex sets. A widely used algorithm is projecting onto convex sets. Lee *et al.* (1997) accomplish convergence to a solution by alternately projecting between convex sets, and Cho *et al.* (1998) use alternate projections between convex sets to incorporate dose-volume constraints in which the dose-volume constraint control is achieved by using two convex sets.

A cyclic subgradient algorithm, which is an iterative algorithm, is also used to solve convex feasibility problems. It examines each constraint sequentially, changing the elements of x with each constraint. Starkschall *et al.* (2001) modify the cyclic subgradient projection algorithm to incorporate dose-volume constraints. Michalski *et al.* (2004) solve dose-volume constraint satisfaction problems using the simultaneous subgradient projection algorithm. They state that the algorithm is easy to implement and has minimal memory requirements.

2.4 Linear and Nonlinear Programming Models

The first linear optimization model that was developed to aid radiation therapy design appeared in the literature in 1968 (Bahr *et al.* (1968)). Since then, many researchers have experimented with linear models. See, e.g, Hodes (1974), Langer (1987), Powlis *et al.* (1989), Legras *et al.* (1982), Morrill *et al.* (1990b), Morrill *et al.* (1991a) and Rosen *et al.* (1991). For an overview of these models, the reader is referred to Shepard *et al.* (1999).

Basically, all objectives of linear programming models are variations of: (1) minimize average/maximum dose or deviation from upper bounds on dose to critical organs and normal tissue, (2) maximize average/minimum dose to tumor, and (3) minimize average/maximum deviation from prescribed dose to target.

The constraints might be: (1) nonnegativity constraints for the beam intensity, (2) upper bounds for the critical organs and/or normal tissue, (3) lower and/or upper bounds for the tumor, (4) upper bounds on the ratio between the maximum beamlet intensity and the average beamlet intensity, and (5) upper bounds on the mean dose to the critical organ.

By sensibly combining objectives and constraints, different models can be formulated. Normally, if the objective is on the tumor, then the constraints should be on the critical organs and normal tissue. On the other hand, if the objective is on the critical organs and normal tissue, then the constraints should be on the tumor. However, nonnegativity constraints are always present as a physical constraint.

A simple example for minimizing the weighted sum of maximum deviation on tumor and maximum overdose on critical organs and normal tissue subject to nonnegativity constraints can be described as follows (Lim *et al.* (2002)):

$$\begin{array}{ll} \min & \omega_T \|A_T x - TG\|_\infty + \omega_C \left\|(A_C x - CUB)_+\right\|_\infty + \omega_N \left\|(A_N x - NUB)_+\right\|_\infty \\ \text{s.t.} & x \geqq 0, \end{array} \tag{2.4}$$

where $(\cdot)_+ = \max\{0, \cdot\}$ and ω is a vector of weighting factors also known in the literature as structural importance factors.

The linear programming approach has the advantages of speed and it is guar-

anteed to have positive solutions. However, sometimes linear programming can not find a feasible solution and the source of infeasibility is then unknown. For example, linear programming models with the constraints $A_T x \geqq TLB$ and $A_C x \leqq CUB$ can be infeasible. Furthermore, due to simplex algorithms producing an extreme point solution, linear solvers have the problem that physicians' limits are often attained, that means therapy plans narrowly adhere to the prescription, i.e. either portions of the critical organs are to receive their maximum allowable dose or the tumor is to receive the lowest allowable dose, while none of these two results are desirable. Moreover, it is hard to include dose-volume constraints (which is discussed later).

In order to solve the infeasibility problem of linear programming model, Holder (2003, 2004, 2006) proposes a new linear programming model incorporating elastic constraints. The elastic model is as follows:

$$\begin{array}{lrll} \min & \omega \cdot l^T\alpha + u_C^T\beta + u_N^T\gamma & & \\ \text{s.t.} & TLB - L\alpha \leqq A_T x & \leqq TUB & \\ & A_C x & \leqq CUB + U_C\beta & \\ & A_N x & \leqq NUB + U_N\gamma & \\ & 0 \leqq L\alpha & \leqq TLB & \\ & -CUB \leqq U_C\beta & & \\ & 0 \leqq U_N\gamma & & \\ & 0 \leqq x, & & \end{array} \tag{2.5}$$

where $\alpha \in \mathbb{R}^{qT}$, $\beta \in \mathbb{R}^{qC}$, $\gamma \in \mathbb{R}^{qN}$, $l \in \mathbb{R}^{qT}$, $u_C \in \mathbb{R}^{qC}$, $u_N \in \mathbb{R}^{qN}$, $L \in \mathbb{R}^{m_T \times qT}$, $U_C \in \mathbb{R}^{m_C \times qC}$, $U_N \in \mathbb{R}^{m_N \times qN}$. The constraints $TLB - L\alpha \leqq A_T x \leqq TUB$, $A_C x \leqq CUB + U_C\beta$, and $A_N x \leqq NUB + U_N\gamma$ are called elastic constraints because the bounds are allowed to vary with the vectors α, β, and γ, respectively. The matrices L, U_C, and U_N define how one measures the amount of elasticity, and l, u_C, u_N show how one either penalizes or rewards the amount of elasticity. ω is the weight deciding the importance of the tumor uniformity. As ω increases, we increase the emphasis of finding a plan that achieves a uniform, tumoricidal dose.

Theorem 2.4.1 (Holder (2003))**.** *The interior of the feasible set of (2.5) and the interior of the feasible set of the dual of (2.5) are nonempty.*

Theorem 2.4.2 (Holder (2003))**.** *Let* $(x^*(\omega), \alpha^*(\omega), \beta^*(\omega), \gamma^*(\omega))$ *be an optimal solution to (2.5). For any collection of elastic functions we have that* $l_T\alpha^*(\omega) = O(1/\omega)$*, provided that the prescription allows tumor uniformity.*

According to Theorem 2.4.1, the interior of the feasible set is never empty. When (2.5) is solved with a path following interior point method it terminates with a solution that strictly satisfies as many inequalities as possible.

Different elastic functions lead to different solution analysis, in particular, the author analyzes two collections of elastic functions. One is average analysis, where $l = \frac{1}{m_T}e$, $u_C = \frac{1}{m_C}e$, $u_N = \frac{1}{m_N}e$, $L = I$, $U_C = I$, $U_N = I$ (e is a vector of all ones). The other is absolute analysis, where $l = 1$, $u_C = 1$, $u_N = 1$, $L = e$, $U_C = e$, $U_N = e$.

This model overcomes the disadvantage of infeasibility, also the tumor uniformity can be achieved due to Theorem 2.4.2. However, dose-volume constraints are not included.

Lim *et al.* (2002) develop an optimization framework for conformal radiation treatment planning, in which the authors also summarize two types of linear programming models. The Least-Absolute-Value formulation is very similar to Holder's average analysis. The other one, a min-max formulation, is (2.4).

At present, a basic nonlinear programming model, the weighted least squares model, is one of the most prevalent formulations, see, e.g., Redpath *et al.* (1976); McDonald and Rubin (1977); Starkschall (1984); Webb (1989, 1991, 1994a); Bortfeld *et al.* (1990, 1994); Xing and Chen (1996); Xing *et al.* (1998). Some commercial IMRT systems model their objective as a weighted least squares function, such as Helios and Focus (their models are based on the work of Spirou and Chui (1998) and Xing *et al.* (1998), respectively).

The weighted least squares model is:

$$\begin{array}{ll} \min & \frac{\omega_T}{m_T}\|A_Tx - TG\|_2^2 + \frac{\omega_C}{m_C}\|A_Cx - CUB\|_2^2 + \frac{\omega_N}{m_N}\|A_Nx - NUB\|_2^2 \\ \text{s.t.} & x \geqq 0, \end{array} \tag{2.6}$$

where ω has the same meaning as in (2.4), it is a vector of weighting factors. The function calculates the weighted sum of average squared deviation from the prescribed dose for each organ.

Alternatively, we could only penalize the overdose part for the critical organs and normal tissue, this change would turn the objective into (see Hristov and Fallone (1997, 1998); Spirou and Chui (1998); Wu and Mohan (2000) for reference):

$$\begin{array}{ll} \min & \frac{\omega_T}{m_T}\|A_T x - TG\|_2^2 + \frac{\omega_C}{m_C}\left\|(A_C x - CUB)_+\right\|_2^2 + \frac{\omega_N}{m_N}\left\|(A_N x - NUB)_+\right\|_2^2 \\ \text{s.t.} & x \geqq 0. \end{array} \tag{2.7}$$

Quadratic objective functions have become an accepted standard. In general, they produce satisfactory plans. The convexity properties of this kind of objective have been analyzed by Deasy (1997). Also, Bortfeld (1999) states that it can be shown rigorously that there are no local minima for this simple least squares objective function. But the weakness of quadratic objective functions cannot be neglected. Weighting factors ω have no clinical meaning and the choice is quite arbitrary. Therefore, normally several plans for different choices of weighting factors are considered before a final plan is selected.

All the models discussed so far are "physical" models which are solely based on the dose. Another type of objective function is based on radiobiological models. Radiobiological models argue that optimization should be based on the biological effects produced by the underlying dose distributions.

A common method to express radiobiological objective functions is based on tumor control probabilities (TCP) and normal-tissue complication probabilities (NTCP). The treatment objective is usually stated as the maximization of TCP while maintaining the NTCP within acceptable levels. The key in applying radiobiological models is to determine the appropriate mathematical models to quantify TCP and NTCP.

There are many TCP models in the literature, see, e.g., Brahme and Agren (1987); Källman *et al.* (1992); Webb and Nahum (1993); Webb (1994b); Zaider and Minerbo (2000). Currently, a TCP model that is based on a Poisson dose-response model has received increased interest even though the Poisson model does not consider clonogen proliferation during treatment and the statistics of the number of clonogens is not generally Poisson (Tucker *et al.* (1990)).

A Poisson-based TCP function (Brahme and Agren (1987)) is as follows:

$$TCP = \prod_{i=1}^{M} \exp(-(O/M)\exp(-rd_i)) \tag{2.8}$$

where M is the number of equivolume voxels in the tumor, r is the radiosensitivity of the tumor clonogens, O is the total number of the clonogens in a tumor and d_i is the dose to the i-th voxel. Choi and Deasy (2002) show that this Poisson-based TCP objective function (if one neglects inter-patient heterogeneity) is strictly concave.

Also, NTCP models have been developed by many researchers, see, e.g., Kutcher *et al.* (1991); Niemierko and Goitein (1991); Jackson and Kutcher (1993); Lyman and Wolbrast (1989); Stavrev *et al.* (2003); Alber and Nüsslin (2001). Out of those models, the relative seriality model which is based on the Poisson dose-response model and categorizes volumes into serial, parallel, or both functional subunits attracts more attention. According to this model, an NTCP function with K critical organs can be described as follows (Löf (2000)):

$$\begin{aligned} NTCP &= 1 - \textstyle\prod_{k=1}^{K}(1 - P^k) \\ P^k &= (1 - \textstyle\prod_{i \in C_k}[1 - (P(d_i))^s]^{v_i / \sum_{i \in C_k} v_i})^{1/s} \end{aligned} \tag{2.9}$$

where v_i is the i-th subvolume, d_i is the dose in subvolume v_i and s describes the relative seriality of tissue organization (Källman *et al.* (1992)). If a homogeneous dose is assumed, then $P(d_i)$ is given by $P(d_i) = 2^{-exp(eg(1-d_i/D_{50})}$, where D_{50} is the dose of 50% response, e is the base of the natural logarithm, and g is the normalized dose-response gradient.

Based on TCP and NTCP, radiobiological models can be formulated, see, e.g., Morrill *et al.* (1991b); Källman *et al.* (1992); Webb (1992); Mageras and Mohan (1993); Löf (2000); Wang *et al.* (1995). Currently, one of the most prevalent radiobiological objective functions is the maximization of the uncomplicated tumor control probability P_+, see, for instance, Källman *et al.* (1992) and Löf (2000). P_+ can be described as the probability of tumor control minus the probability of having both tumor control and severe complications. If TCP is independent of NTCP, then

$$P_+ = TCP(1 - NTCP), \tag{2.10}$$

where TCP and $NTCP$ is dependent on the response model, the reader is referred to above TCP and NTCP models.

Löf (2000) mentions that the obvious advantage with this objective function is that the probabilities of curing the tumor and avoiding fatal side effects are combined in one single scalar quantity.

Brahme (2001) describes the benefits of biological optimization and Penagaricano *et al.* (2005) assess biologically-based optimization by comparing the result of biologically-based optimization and physical objective based optimization for three cases. The biological objective of P_+ maximization based on the relative seriality model was used in biologically-based optimization and least squares minimization was used as physical objective in IMRT. By analyzing the result, the author concluded that this form of IMRT appears promising and future research is encouraged.

Based on the TCP model (2.8), Niemierko (1997) develops the original tumor-applicable equivalent uniform dose (EUD), which assumes two dose distributions are equivalent if they cause the same probability for tumor control. The EUD is

$$EUD = -\frac{1}{a}\ln\left(\frac{1}{M}\sum_{j=1}^{M} e^{-ad_j}\right), \tag{2.11}$$

where M is the number of voxels in the anatomic structure of interest, and a is a structure-dependent parameter that depends on the radiation response of the underlying tissue.

Later he developed generalized equivalent uniform dose (gEUD) and extended the application to normal tissue complication (Niemierko (1999)). A gEUD is

$$gEUD = \left(\frac{1}{M}\sum_{i} d_i^a\right)^{\frac{1}{a}}. \tag{2.12}$$

This formulation is also called the power law EUD model. When $0 \leq a \leq \infty$, this definition extends to critical organs. For $a = -\infty, 0, 1$ and ∞, gEUD is equal to the minimal dose, geometric mean and arithmetic mean and maximal dose, respectively. It mimics dose-response reality closely. For example, for the tumor the EUD should be close to the minimal dose and a should be a large negative value; for some parallel

organ such as lungs, the dose response may be more closely related to the mean dose, and a should be near positive 1.

An approach for the determination of the equivalent uniform dose (EUD) for inhomogeneously irradiated normal organs is developed by Thieke *et al.* (2002). The EUD is calculated using convex combinations of mean dose and the maximum dose, it is called max & mean model.

$$EUD = o \cdot d^{max} + (1 - o) \cdot d^{mean} \tag{2.13}$$

where d^{max} is the maximum and d^{mean} is the mean dose of the radiation delivered to the organ, and o is an organ-specific parameter ranging from 0 to 1.

The values of o are determined by a fit to the Emami tables (Emami *et al.* (1991)) for complication levels of 5% and 50%. The authors also compared the quality of the fit with the power law model and found that they have the same error range. The main advantage of this max & mean model is its linearity.

EUD is a hybrid between the physical dose and biological response, it is easy to form an objective function using EUD (Wu *et al.* (2002); Küfer *et al.* (2003)). For example, Wu *et al.* (2002) formulate the objective function as follows:

$$\begin{aligned} F &= F_T \cdot F_C \\ F_T &= \frac{1}{1+\left(\frac{gEUD_0}{gEUD}\right)^{\omega}} \\ F_C &= \frac{1}{1+\left(\frac{gEUD}{gEUD_0}\right)^{\omega}}, \end{aligned} \tag{2.14}$$

where $gEUD_0$ is the the desired dose for target volumes and the maximal tolerable uniform dose for normal structures, and ω is akin to the weight or penalty that indicates the importance of the structure. Alternatively, $\ln(F)$ can be the objective. In addition, the power law EUD model is proved by Choi and Deasy (2002) to be convex or concave depending on the value of a: when $a \geqq 1$, it is convex, otherwise it is concave. Also, EUD can be used as an initial guide for dose-volume optimization (Wu *et al.* (2003b)). However, there is little experience with the use of EUD.

Optimization Algorithms A given problem can be solved using different optimization algorithms. There are gradient-based algorithms, simulated annealing,

genetic algorithms, iterative algorithms, constrained least-squares etc..

In IMRT, the most widely used stochastic algorithms are simulated annealing (SA) and genetic algorithm (GA). They are heuristics for finding a global minimum of a general problem. The underlying principle of SA is to simulate the cooling process of a material in a heat bath, and SA uses this simulation to systematically search for feasible points in a way that makes the generated sequence converge to a global minimum. Webb (1989) first introduced an SA algorithm for the IMRT optimization problem. Since then, SA with different parameters has been tried to optimize beam intensity (Webb (1991, 1992, 1994a); Morrill *et al.* (1990a, 1991b, 1995); Mageras and Mohan (1993)). SA has also been used in optimizing beam orientation and beam intensity simultaneously (Rowbottom *et al.* (2001)). Genetic algorithms (GAs) simulate the natural process of evolution in which the fittest solutions 'survive' and have been proposed. They have been used to optimize external beam treatment plans, see, e.g., Ezzell (1996); Langer *et al.* (1996); Yu (1997); Wu and Zhu (2001); Cotrutz and Xing (2003). However, simulated annealing and genetic algorithm are slow in reaching the optimum, typically needing tens of thousands of iterations.

Compared to stochastic methods such as simulated annealing, gradient-based algorithms (Spirou and Chui (1998)) are fast. They seem to deliver satisfactory treatment plans although not ensuring global optimality. Therefore, a gradient based method seems natural when considering the beam intensity optimization problem. Hristov and Fallone (1997) use an active set method with conjugate-gradient subspace minimization for treatment planning optimization to accommodate the non-negativity constraints. Wu and Mohan (2000) use Newton's method to solve the beam intensity optimization problem with a quadratic objective function. Alber and Reemtsen (2007) use a lagrangian barrier-penalty algorithm to optimize a biological model. Carlsson and Forsgren (2006) use a BFGS quasi-Newton sequential quadratic programming method with diagonal initial Hessian estimate to optimize beam intensities, and they justify that such an approach tends to give smooth beamlet intensities solutions.

Other algorithms such as image reconstruction from projections (Bortfeld *et al.* (1990)), iterative filtered backprojection (Holmes and Mackie (1994)) and fast iterative methods (Xing and Chen (1996); Xing *et al.* (1998)) have been proposed. These algorithms start with an initial approximate solution and generate a sequence of solutions that converge to an optimal solution of the problem. They are simple and straightforward. Constrained least-squares obtains the solution of a linear system by a direct matrix manipulation. It has been applied to conformal radiotherapy by Starkschall (1984) and Crooks and Xing (2002) extended its application to IMRT.

Moreover, a random search algorithm (Niemierko (1992)) and a maximal entropy and maximal likelihood method (Llacer (1997)) have also been proposed.

2.5 Dose Volume Constraints

Many researchers work on the incorporation of dose volume constraints (DVCs) into the mathematical models. Using linear programming to include DVCs was started by Langer (1987), who was the first to incorporate DVCs in the literature. The author produces a sequence of LPs by enumerating all the combinations of constraint points that satisfy dose-volume constraints, solves these LPs and chooses the voxel combination that gives the best objective value. It is difficult to solve the problem if there are more than 20 voxels. For example, consider the dose-volume constraint of "no more than 40% volume of the lung can exceed a radiation dose of 20 Gy". Then, if there are 100 voxels in the lung, in order to satisfy the above dose volume constraint we need to solve $\binom{100}{40}$ LPs. Nevertheless, this work built the foundation for applying mixed integer programming (Langer *et al.*, 1990) to solve the DVC problem.

Morrill *et al.* (1991a) use a collar technique to handle DVCs (the idea of collar was first proposed by Bahr *et al.* (1968)). The authors use the concept that the volumes of normal tissue nearer the target volume will be allowed higher dose constraints than volumes of normal tissue further from the target volume. So they divide normal structures into high-dose and low-dose volumes and set different dose

upper bounds for the high-dose and low-dose volumes. In this way, the DVC is converted into regional dose constraints. However, the best possible result can not be guaranteed due to the partition of high-dose and low-dose volumes being based on the experience.

Merritt and Zhang (2002) use successive linear programming to optimize the beam intensity map. An objective of maximizing minimum tumor dose is used and upper bounds for the tumor and the critical organs are specified. Moreover, dose-volume control is imposed by means of successive relaxation of upper bounds on critical organ dosages until the specified maximum allowable volume has been relaxed. However, the infeasibility problem still exists in this model.

A sequence of LPs is constructed. The k-th LP is:

$$\begin{array}{lrlll} \max & z & & & \\ \text{s.t.} & ez & \leqq A_T x & \leqq TUB & \\ & a\,(i)x & \leqq M & & \text{for all } i \in R_k \\ & a\,(i)x & \leqq CUB_i & & \text{for all } i \in C \backslash R_k \\ & 0 & \leqq x, & & \end{array} \tag{2.15}$$

where $R_k \subseteq C$ is the set of indices corresponding to right-hand sides that have been relaxed after solving the $(k-1)$-th LP and $M > CUB_i$ is some upper bound on those constraints. For $k \geqq 2$, $R_k = \left\{ j \in C : y^*_{j,k-1} > \lambda \right\} \cup R_{k-1}$ and $R_1 = \emptyset$, where $y^*_{j,k-1}$ is a j-th optimal dual variable in the $(k-1)$-th LP and $\lambda > 0$ is some constant parameter.

Basically, at each step, the constraints whose relaxation will produce the largest increase in the objective value will be chosen. Since the dual optimal solution is not unique, it needs to be carefully chosen. By iteratively relaxing the constraints, finally, an LP solution that is close to the mixed integer programming solution will be produced. However, the infeasibility problem still exists in this model.

Romeijn *et al.* (2003, 2006) use a piecewise linear convex function to approximate any convex objective. This model overcomes the apparent limitations of linear programming. In addition, they use the concept of conditional value-at-risk to impose a novel alternative to the traditional dose-volume constraint. The constraint

they formulate bounds the tail averages of the differential dose-volume histograms of structures. It retains linearity to improve dose homogeneity in the target volumes.

There are two ways to deal with dose-volume constraints in nonlinear programming. One way is to impose a penalty term to the original objective function, which is popular (see, Bortfeld *et al.* (1997); Cho *et al.* (1998); Spirou and Chui (1998); Wu and Mohan (2000); Chen *et al.* (2002); Carlsson *et al.* (2006)). All these models adopt a volume sensitive penalty function. For example, Spirou and Chui (1998) add a penalty term $\sum_{i\in C}\zeta_i\omega_i(A_Cx-CUB)_i$ to the objective function, where ω_i is the weight of the constraint, and ζ_i is a flag that can be set to 1 or 0. Only the flags ζ_i corresponding to those voxels, that when sorted in ascending order of dose received, cause dose-volume constraints to be violated are set to 1.

Another way to consider the DVCs in nonlinear programming models is to include them in the optimization algorithm. Cho *et al.* (1998) use a method that is based on the theory of projections onto convex sets in which the dose-volume constraint control is realized by using two separate convex sets. The first set constrains the maximum dose to the whole organ while the second limits the integral dose. Furthermore, Starkschall *et al.* (2001) use a dose-volume feasibility search algorithm in which the cyclic subgradient projection (CSP) algorithm is modified to incorporate dose-volume constraints. Dai and Zhu (2003) discuss two techniques to convert dose-volume constraints to dose limits. One is dose sorting, which is based on the assumption that higher dose limits should be assigned to the constraint points receiving higher dose, and vice versa. The other is a hybrid of the dose-sorting technique and the mixed integer linear programming technique.

The non-convexity of dose volume constraints has been analyzed by Deasy (1997). The author demonstrates that for models with dose-volume constraints there is the possibility of multiple local minima due to the concavity of the feasible set formed by DVCs. Furthermore, Wu and Mohan (2002); Rowbottom and Webb (2002); Llacer *et al.* (2003); Jeraj *et al.* (2003); Wu *et al.* (2003a) verify the existence of local minima by performing case studies. They also show that most of the local minima are very close to the global minimum. However, there still exist some local minima

that have large deviation from the global minimum (Jeraj *et al.* (2003) and Wu *et al.* (2003a)). The severity of the local minima still needs further investigation.

2.6 Mixed Integer Programming

By introducing binary variables into the model it is straight forward to impose dose-volume constraints. By setting these binary variables to 0 or 1, it is possible to enumerate the number of voxels that receive a dose lower than a threshold or higher than a threshold in a specified structure, thus it is possible to model dose-volume constraints. These additional variables turn a linear programme into a mixed integer programme (MIP).

Here, we consider an example to show how to convert a dose constraint into a dose-volume constraint. In linear programming, a single upper bound on the critical organ may be imposed as follows:

$$a\left(i\right)x \leqq CUB_i \quad \forall i \in C. \tag{2.16}$$

Sometimes, such a constraint can impose infeasibility. By specifying an overdose fraction F and a volume fraction P, this constraint can be relaxed to a dose-volume constraint:

$$\begin{aligned} a\left(i\right)x &< (1+y_iF)CUB_i \quad \forall i \in C \\ \textstyle\sum_{i\in C} y_i &< Pm_C, \end{aligned} \tag{2.17}$$

where y_i is a binary variable. If F is big enough, then the problem will be feasible.

The first MIP incorporating dose volume constraints was proposed by Langer *et al.* (1990). In that work, the authors use MIP to find the beam intensities of wedged and open beams in which dose-volume constraints are considered. Subsequently, Langer and Morrill (1996) compare MIP and fast simulated annealing for optimized beam intensities in radiotherapy and find that MIP produced more satisfactory results. Although the application is only for conformal radiotherapy not IMRT, it is straightforward to extend it to IMRT.

Volume-based objective functions can also be used in MIP models. Bednarz *et al.* (2004) propose a volume-based objective function in which the number of under- or

overdosed voxels in selected critical structures and/or targets is to be minimized. By minimizing the objective function, the model obtains a better control of the over- or underdosed volumes in critical structures and targets. It also circumvents the interactive and judicial adjustment of the dose volume constraint values in order to obtain a specific solution.

Moreover, MIP can be used for solving some combined problems. For example, Lee *et al.* (2003) and Wang *et al.* (2003) optimize beam angle and beam intensity at the same time. Bednarz *et al.* (2002) and Preciado-Walters *et al.* (2004) consider beam intensity and beam realization together.

Undoubtedly, MIP frameworks significantly enhance flexibility. This flexibility allows us to build complicated constraints and objectives. However, usually MIP models are solved by branch and bound algorithms with linear programming. For the typical size of clinically relevant MIP problems with many integer variables, finding an optimal solution to such models in an acceptable amount of computational time is not possible with current technology. Therefore, execution time can be a problem for MIP approaches.

2.7 Multiple Objective Programming

In radiotherapy, the desired dose distribution can not always be obtained, due to physical limitations and to the existence of trade-offs between the various conflicting treatment goals. This multiobjective character of inverse planning has been recognized in radiation therapy optimization only in the last 10 years even though radiotherapy has been applied for several decades.

One way to solve a multiple objective (MO) problem is by transforming it into a single objective problem using a specific set of weighting factors for each objective. This is the most prevalent method and is called a priori method for solving MO optimization problems because preference information is incorporated into the model in the form of weights prior to optimization. From this point of view, we can say that the nonlinear models (see (2.6) and (2.7)) and linear programming models (see

(2.4) and (2.5)) are actually a scalarization of multiple objective problems. The problem with this approach is that the weighting factors have no clinical meaning and their relationship to the solution is not known in advance. Therefore, in order to obtain some satisfactory solution, the treatment planner is commonly required to repeat the optimization with varying weighting factors.

Although there are methods (Yu, 1997; Xing *et al.*, 1999; Wu and Zhu, 2001; Cotrutz and Xing, 2002) to choose weighting factors, they each require additional importance factors which are a priori not known. Therefore to date, the inverse planning process is still a trial and error process.

Alternatively, an a posteriori method in which an optimization engine obtains either the nondominated set or a representative subset can be imagined. For a multiple objective problem, a nondominated point in objective space corresponds to an efficient solution in decision space. An efficient solution is defined as a solution for which an improvement in one objective will always lead to a worse result in at least one of the other objectives. The nondominated set is used to analyze the trade-offs between the objectives before selecting a solution that satisfies simultaneously all objectives. Analysis could be done after the optimization and the optimization is not required to be repeated again.

Multiple objective programming (MOP) approaches in IMRT can be divided into two categories, multiple objective nonlinear programming (MONP) and multiple objective linear programming.

MONP is based on nonlinear programming models. The first MOP model for IMRT was proposed by Cotrutz *et al.* (2001). Variance-based objectives were used:

$$\begin{aligned} F &= (f_T, f_{C_1}, \ldots, f_{C_K}) \\ f_T &= \frac{1}{m_T} \frac{\left\| A_T x - \bar{d_T} \right\|_2^2}{\bar{d_T}^2} \\ f_{C_k} &= \frac{1}{m_{C_k}} \frac{\left\| (A_{C_k} x - C_k UB)_+ \right\|_2^2}{C_k UB^2}, k = 1, \ldots, K, \end{aligned}$$

where $\bar{d_T}$ is the current iteration mean dose for the tumor.

The most prevalent solution method for MOP is the weighted sum method.

The relationship between optimal solutions of the weighted sum model and efficient solutions of the MOP is described in Theorems 2.7.1 and 2.7.2.

Theorem 2.7.1. *For a multiobjective problem* $\min \{f_1(x), \ldots, f_p(x) : x \in \mathcal{X}\}$*, let* $\bar{x} \in \mathcal{X}$ *be an optimal solution to the weighted sum problem* $\min \{\bar{\lambda_1} f_1(x) + \ldots + \bar{\lambda_p} f_p(x) : x \in \mathcal{X}\}$ *where* $\bar{\lambda_i} > 0$*, and* $\sum_{i=1}^{p} \bar{\lambda}_i = 1$*. Then,* $\bar{x}$ *is efficient.*

Theorem 2.7.2. *If* $\{f_1, \ldots, f_p\}$ *are convex, let* $\bar{x} \in \mathcal{X}$ *be efficient solution for the multiobjective problem* $\min \{f_1, \ldots, f_p\}$*. Then, there exists a* $\bar{\lambda} = \left(\bar{\lambda_1}, \ldots, \bar{\lambda_p}\right)^T$ *such that* $\bar{x}$ *is a minimal solution of* $\min \left\{\bar{\lambda_1} f_1 + \ldots + \bar{\lambda_p} f_p : x \in \mathcal{X}\right\}$*, where* $\bar{\lambda_i} \geqq 0$ *and* $\sum_{i=1}^{p} \bar{\lambda}_i = 1$*.*

According to Theorem 2.7.1, Cotrutz *et al.* (2001) obtain a set of efficient solutions by using different weighting factors. A conjugate gradient algorithm is used to solve the weighted sum problem.

Based on Cotrutz *et al.* (2001), Lahanas *et al.* (2003b) use slightly different objectives

$$\begin{aligned} F &= (f_T, f_N, f_C), \text{where} \\ f_T &= \frac{1}{m_T} \|A_T x - TG\|_2^2, \\ f_N &= \frac{1}{m_N} \|A_N x\|_2^2, \\ f_C &= \frac{1}{m_C} \left\|(A_C x - CUB)_+\right\|_2^2. \end{aligned} \tag{2.18}$$

Here f_T is the average squared deviation from the prescribed dose to the tumor, f_C is the average squared overdose to the critical organ and f_N is the average squared dose to the normal tissue. To eliminate the negative beam intensity problem, the problem is converted to an unconstrained problem by using the square root of beam intensities as decision variables. Then the authors use the limited memory BFGS algorithm L-BFGS by Liu and Nocedal (1989) to solve the problem and conclude that globally optimal solutions can be found using L-BFGS by comparing the results with the fast simulated annealing algorithm. Moreover, Lahanas *et al.* (2003a) use the evolutionary algorithms NSGA-II and NSGA-IIc to improve the optimization process.

Based on linear programming models, researchers formulate the beam intensity problem using MOLP (see, e.g., Küfer and Hamacher (2000); Ehrgott and Burjony (2001); Hamacher and Küfer (2002); Küfer *et al.* (2003)).

A simple MOLP model can be described as follows (Hamacher and Küfer (2002)):

$$\begin{array}{lrll} \min & F = (f_T, f_C, f_N) & & \\ \text{s.t.} & A_T x & \geqq TLB\,(1 - f_T) & \\ & A_C x & \leqq CUB\,(1 + f_C) & \\ & A_N x & \leqq NUB\,(1 + f_N) & \\ & f_T, f_C, f_N & \geqq 0 & \\ & x & \geqq 0, & \end{array} \tag{2.19}$$

where f_T is the maximal deviation from the prescribed dose for the tumor, f_C is the maximal deviation from the prescribed dose for the critical organ and f_N is the maximal deviation from the prescribed dose for the normal tissue.

Küfer *et al.* (2003), Thieke (2003), and Craft *et al.* (2005, 2006) use the equivalent uniform dose (EUD) (2.13) of all structures as their objectives. The objective functions in Küfer *et al.* (2003) are:

$$\begin{array}{ll} F & = (F_T, F_{C_1}, \ldots, F_{C_K}) \\ F_T & = \frac{G_T - min_T(A_T x)}{G_T} \\ F_{C_k} & = \frac{EUD_{C_k} - G_{C_k}}{G_{C_k}}, k = 1, \ldots, K \end{array} \tag{2.20}$$

where $min_T(A_T x)$ is the minimal dose value to the target and $G_{C_k} \neq 0$.

MOLP is generally easier to solve than MONP due to its linearity. A method to find a balanced solution for MOLP (2.19) is described in Ehrgott and Burjony (2001). A unifying mathematical framework that allows for a comparison of different models via the comparison of the corresponding nondominated set is proposed by Romeijn *et al.* (2004).

Theorem 2.7.3. *The efficient sets of MOP* min $\{f_1(x), \ldots, f_p(x)\}$ *and MOP* min $\{h_1(f_1(x)), \ldots, h_p(f_p(x))\}$*, where* $h_1, \ldots, h_p$ *are strictly increasing functions, are the same.*

Note the nondominated set must change, because $(h(f(x))$ is not the same as $f(x)$. Nonconvex multiple objective functions $g(x) = h(f(x))$ could be replaced by equivalent convex functions if a suitable decomposition into convex functions $f_k(x)$ and increasing functions h_k can be found. The authors apply this method to study the relationship between TCP, EUD, gEUD, the logarithm of the objective function of (2.14) and dose based objective functions (Romeijn *et al.* (2004)). They find that the logarithm of the objective function of (2.14), gEUD and the dose based objective functions that have been studied have the same efficient solution.

Dose-volume objectives have been used in Halabi *et al.* (2006), they are: (1) minimize the number of target voxels receiving less than some prescribed tolerance dose and (2) minimize the number of OAR voxels receiving more than a tolerance dose. They show that the multiobjective formulation with dose-volume objectives is better suited to convex approximation than conventional formulations with dose-volume constraints. They solve the problem by the weighted sum method and the weights are chosen by the PGEN algorithm which is proposed in Craft *et al.* (2006). Also they provide a relaxation of the integer programming formulation which reduces the computation time without significantly compromising the results.

Most of the papers use the weighted sum method such as Cotrutz *et al.* (2001) and Lahanas *et al.* (2003b), which is generally cumbersome. First, it is inefficient because the running time to solve the problem is proportional to the number of solves. For example, if we want 30 efficient plans, it will take around 10 hours if finding one solution requires 20 minutes. Secondly, most models in MONP now use gradient based solution algorithms, which cannot guarantee that the obtained solutions are indeed optimal. Moreover, as the weighting factors have no clinical meanings, it is difficult to choose a set of weights to make the nondominated points evenly distributed. Even if an evenly distributed set of weights is used, it is possible that the points obtained on the nondominated set are not uniformly distributed (Das and Dennis, 1997) because of the unknown relationship between weight and optimal solution. Therefore, these discrete solutions may not fully represent the scope of the nondominated set.

We now summarize other solution methods, which have been used in MOLP formulations of the intensity optimization problem. Hamacher and Küfer (2002) and Küfer *et al.* (2003) describe an idea to generate a subset of the nondominated set based on the concept of neighbor solutions. Craft *et al.* (2005) use the normalized normal constraint method (Messac *et al.*, 2003) to achieve two dimensional trade-offs between tumor dose homogeneity and critical organ sparing and Craft *et al.* (2006) propose a method called PGEN to iteratively choose weights to gradually construct the nondominated set. However, most of these methods cannot give us a comprehensive view of the entire nondominated set. They either find a subset of the nondominated set or try to approximate the whole nondominated set using the nondominated points obtained, usually without guaranteed quality of approximation.

The normalized normal constraint method is based on the normal boundary intersection (NBI) method of Das and Dennis (1998). Both of the methods generate a set of equidistant reference points on the convex hull of the individual minima (CHIM). For each reference point, a corresponding nondominated point is found solving a single objective subproblem. These methods can find evenly distributed nondominated points, but they have the limitation that the solution may overlook a portion of the nondominated set if the normal of the CHIM has both positive and negative components (which may happen for $p > 2$ objectives) (we show this in Chapter 7). The method of Craft *et al.* (2006) uses the idea of sandwiching the nondominated set between a lower and an upper convex approximation. In each iteration it calculates a new weight and updates the lower and upper approximation. However, the CHIM is taken as the upper approximation initially.

Therefore, there is still no effective way for finding a representative subset of the nondominated set.

2.8 Summary

In this chapter, we have reviewed the mathematical models used in the beam intensity optimization problem of radiotherapy. We classified these models into linear programming models, nonlinear programming models, mixed integer programming models and multiple objective models. Recently, there were some improvements in linear programming models addressing the infeasibility problem and introducing new ways to address dose-volume constraints. For nonlinear programming, gradient based solution methods will only achieve local optima, while stochastic algorithms like simulated annealing are slow. MIP frameworks significantly enhance the modeling flexibility, and allow for the incorporation of many different objectives and constraints into the optimization model, but the computation time will be a problem. By analysing these models, we found that the most existing nonlinear, linear programming, and mixed integer programming models are multiobjective in nature due to the weighting factors. In MOP, instead of specifying the weighting factor, a representative set of the nondominated set should be computed for the planner to choose from. Currently, there are no effective solution methods for MOP and new methods need to be developed.

Considering the multiobjective character of radiation therapy we propose to use an MOLP model for the beam intensity optimization problem. In Chapter 3, we present this MOLP model and method to solve MOLPs.

Chapter 3

An MOLP Model for the Beam Intensity Optimization Problem and Solving an MOLP in Objective Space

This chapter is organized as follows. We first introduce some notation in Section 3.1 that is used throughout the rest of the thesis. Then we introduce our MOLP model for beam intensity optimization problem in Section 3.2. In Section 3.3 we provide a brief introduction to multiobjective linear programming. In Section 3.4 we summarize Benson's outer approximation algorithm to solve an MOLP in objective space and illustrate the algorithm with an example. In Section 3.5 we describe some improving modifications to Benson's outer approximation algorithm.

3.1 Notation

In this thesis we use the notation $\mathbb{R}^p_> := \{y \in \mathbb{R}^p : y > 0\}$, $\mathbb{R}^p_\geq := \{y \in \mathbb{R}^p : y \geq 0\}$ and $\mathbb{R}^p_\geqq := \{y \in \mathbb{R}^p : y \geqq 0\}$.

The k-th unit vector in $\mathbb{R}^p$ is denoted e^k and a vector of all ones is denoted by e. Given a mapping $f : \mathbb{R}^n \to \mathbb{R}^p$ and a subset $\mathcal{X} \subseteq \mathbb{R}^n$ we write $f(\mathcal{X}) := \{f(x) :$

$x \in \mathcal{X}\}$.

Let $\mathcal{A} \subseteq \mathbb{R}^p$. We denote the boundary, interior, and relative interior of $\mathcal{A}$ by $\operatorname{bd}\mathcal{A}$, $\operatorname{int}\mathcal{A}$, and $\operatorname{ri}\mathcal{A}$. The convex hull of $\mathcal{A}$ is denoted $\operatorname{conv}\mathcal{A}$.

Let $\mathcal{C} \subseteq \mathbb{R}^p$ be a closed convex cone. An element $y \in \mathcal{A}$ is called $\mathcal{C}$-minimal if $(\{y\} - \mathcal{C} \setminus \{0\}) \cap \mathcal{A} = \emptyset$ and $\mathcal{C}$-maximal if $(\{y\} + \mathcal{C} \setminus \{0\}) \cap \mathcal{A} = \emptyset$. A point $y \in \mathcal{A}$ is called weakly $\mathcal{C}$-minimal (weakly $\mathcal{C}$-maximal) if $(\{y\} - \operatorname{ri}\mathcal{C}) \cap \mathcal{A} = \emptyset$ $((\{y\} + \operatorname{ri}\mathcal{C}) \cap \mathcal{A} = \emptyset)$. We set

$$\begin{aligned} \operatorname{wmin}_{\mathcal{C}}\mathcal{A} &:= \{y \in \mathcal{A} : (\{y\} - \operatorname{ri}\mathcal{C}) \cap \mathcal{A} = \emptyset\} \text{ and} \\ \operatorname{wmax}_{\mathcal{C}}\mathcal{A} &:= \operatorname{wmin}_{(-\mathcal{C})}\mathcal{A}. \end{aligned}$$

In this thesis we consider two special ordering cones, namely $\mathcal{C} = \mathbb{R}^p_{\geqq} = \{x \in \mathbb{R}^p : x_k \geqq 0, k = 1, \ldots, p\}$ and

$$\mathcal{C} = \mathcal{K} := \mathbb{R}_{\geqq}e^p = \{y \in \mathbb{R}^p : y_1 = \cdots = y_{p-1} = 0, y_p \geqq 0\}.$$

For the choice $\mathcal{C} = \mathbb{R}^p_{\geqq}$ the set of weakly $\mathbb{R}^p_{\geqq}$-minimal elements of $\mathcal{A}$ (also called the set of weakly nondominated points of $\mathcal{A}$) is given by

$$\operatorname{wmin}_{\mathbb{R}^p_{\geqq}}\mathcal{A} := \left\{y \in \mathcal{A} : (\{y\} - \operatorname{int}\mathbb{R}^p_{\geqq}) \cap \mathcal{A} = \emptyset\right\}.$$

In case of $\mathcal{C} = \mathcal{K}$ the set of $\mathcal{K}$-maximal elements of $\mathcal{A}$ is given by

$$\max_{\mathcal{K}}\mathcal{A} := \{y \in \mathcal{A} : (\{y\} + \mathcal{K} \setminus \{0\}) \cap \mathcal{A} = \emptyset\}.$$

Note that $\operatorname{ri}\mathcal{K} = \mathcal{K} \setminus \{0\}$ so that weakly $\mathcal{K}$-maximal and $\mathcal{K}$-maximal elements of $\mathcal{A}$ coincide.

Since we will always consider minimization with respect to $\mathbb{R}^p_{\geqq}$ and maximization with respect to $\mathcal{K}$, we sometimes omit the subscripts.

Let us recall some facts concerning the facial structure of polyhedral sets (Webster, 1994). Let $\mathcal{A} \subseteq \mathbb{R}^n$ be a convex set. A convex subset $\mathcal{F} \subseteq \mathcal{A}$ is called a *face* of $\mathcal{A}$ if for all $y^1, y^2 \in \mathcal{A}$ and $\rho \in (0,1)$ such that $\rho y^1 + (1-\rho)y^2 \in \mathcal{F}$ it holds that $y^1, y^2 \in \mathcal{F}$. A face $\mathcal{F}$ of $\mathcal{A}$ is called *proper* if $\emptyset \neq \mathcal{F} \neq \mathcal{A}$. A point $y \in \mathcal{A}$ is called an *extreme point* of $\mathcal{A}$ if $\{y\}$ is a face of $\mathcal{A}$.

A *recession direction* of $\mathcal{A}$ is a vector $d \in \mathbb{R}^p$ such that $y + \rho d \in \mathcal{A}$ for some $y \in \mathcal{A}$ and all $\rho \geqq 0$. The *recession cone* (or asymptotic cone) $\mathcal{A}_\infty$ of $\mathcal{A}$ is the set of all recession directions

$$\mathcal{A}_\infty := \{d \in \mathbb{R}^p : y + \rho d \in \mathcal{A} \text{ for some } y \in \mathcal{A} \text{ and for all } \rho \geqq 0\}.$$

A recession direction $d \neq 0$ is called *extreme* if there are no recession directions $d^1, d^2 \neq 0$ with $d^1 \neq \rho d^2$ for all $\rho > 0$ such that $d = \frac{1}{2}(d^1 + d^2)$.

A polyhedral convex set $\mathcal{A}$ is defined by $\{y \in \mathbb{R}^p : By \geqq \bar{b}\}$, where $B \in \mathbb{R}^{m \times p}$ and $\bar{b} \in \mathbb{R}^m$. A polyhedral set $\mathcal{A}$ has a finite number of faces. A subset $\mathcal{F}$ of $\mathcal{A}$ is a face if and only if there are $\lambda \in \mathbb{R}^p$ and $q \in \mathbb{R}$ such that $\mathcal{A} \subseteq \left\{y \in \mathbb{R}^p : \lambda^T y \geqq q\right\}$ and $\mathcal{F} = \left\{y \in \mathbb{R}^p : \lambda^T y = q\right\} \cap \mathcal{A}$. Moreover, $\mathcal{F}$ is a proper face if and only if $\mathcal{H} := \left\{y \in \mathbb{R}^p : \lambda^T y = q\right\}$ is a supporting hyperplane to $\mathcal{A}$ with $\mathcal{F} = \mathcal{A} \cap \mathcal{H}$ and $\mathcal{F} \neq \mathcal{A}$. We call hyperplane $\mathcal{H} = \left\{y \in \mathbb{R}^p : \lambda^T y = q\right\}$ *supporting* if $\lambda^T y \geqq q$ for all $y \in \mathcal{A}$ and there is some $y^0 \in \mathcal{A}$ such that $\lambda^T y^0 = q$. The proper $(r-1)$-dimensional faces of an r-dimensional polyhedral set $\mathcal{A}$ are called *facets* of $\mathcal{A}$. We say that a supporting hyperplane of $\mathcal{A}$ is degenerate if the supporting hyperplane does not support $\mathcal{A}$ in a facet.

A polyhedral convex set $\mathcal{A}$ can be represented by both a finite set of inequalities and the set of all extreme points and extreme directions of $\mathcal{A}$ (Rockafellar, 1970, Theorem 18.5). Let $\mathcal{E} = \{x^1, \ldots, x^r, d^1, \ldots, d^t\}$ be the set of all extreme points and extreme directions of $\mathcal{A}$ then

$$\mathcal{A} = \left\{ y \in \mathbb{R}^p : y = \sum_{i=1}^{r} \rho_i x^i + \sum_{j=1}^{t} \nu_j d^j \text{ with } \rho_i \geqq 0, \nu_j \geqq 0, \text{ and } \sum_{i=1}^{r} \rho_i = 1 \right\}.$$

For a polyhedral convex set $\mathcal{A}$, the extreme points are called vertices. The set of all vertices of a polyhedron $\mathcal{A}$ is denoted by $\operatorname{vert} \mathcal{A}$. This is often called the "Representation Theorem".

3.2 An MOLP Model for the Beam Intensity Optimization Problem

Based on Holder's linear programming formulation (2.5) in Holder (2003), we formulate the beam intensity optimization problem as a multiple objective linear programme (MOLP). In this model, we minimize the maximum deviation from tumor lower bounds α, critical organ upper bounds β and normal tissue upper bounds γ at the same time. The model is:

$$\begin{array}{lrlll} \min & (\alpha, \beta, \gamma) & & & \\ \text{s.t.} & TLB - \alpha e & \leqq A_T x & \leqq TUB & \\ & & A_C x & \leqq CUB + \beta e & \\ & & A_N x & \leqq NUB + \gamma e & \\ & & 0 & \leqq \alpha & \leqq \alpha_u \\ & -\min CUB & & \leqq \beta & \leqq \beta_u \\ & & 0 & \leqq \gamma & \leqq \gamma_u \\ & & 0 & \leqq x, & \end{array} \tag{3.1}$$

where $\alpha_u \in \mathbb{R}$, $\beta_u \in \mathbb{R}$, and $\gamma_u \in \mathbb{R}$ are upper bounds for α, β, and γ, respectively. They are specified by the radiation oncologist and restrict the search to clinically relevant values.

We can see that the three objectives α, β and γ in (3.1) are limited by upper and lower bounds. The same effect can be achieved by adding upper bounds and lower bounds on the decision variables, i.e., the beam intensity x (Lim *et al.*, 2002). Moreover, we need to point out that this MOLP problem is always feasible as long as appropriate lower bounds and upper bounds for α, β and γ are set, in particular, if these values are set to sufficiently large numbers. (Holder, 2003).

3.3 Introduction to Multiple Objective Linear Programming

In this thesis we consider multiple objective linear programming problems of the form

$$\min\{Cx : x \in \mathcal{X}\}. \tag{3.2}$$

We assume that $\mathcal{X}$ in (3.2) is a nonempty feasible set $\mathcal{X}$ in decision space $\mathbb{R}^n$ defined by $\mathcal{X} = \{x \in \mathbb{R}^n : Ax \geqq b\}$. We have $A \in \mathbb{R}^{m\times n}$ and $b \in \mathbb{R}^m$. The rows of $C \in \mathbb{R}^{p\times n}$ are denoted by c_k and are the coefficients of p linear functions $\langle c_k, \cdot\rangle$, $k = 1, \ldots, p$.

The feasible set $\mathcal{Y}$ in objective space $\mathbb{R}^p$ is defined by

$$\mathcal{Y} = \{Cx : x \in \mathcal{X}\}. \tag{3.3}$$

It is well known that the image $\mathcal{Y}$ of a nonempty, compact polyhedron $\mathcal{X}$ under a linear map C is also a nonempty, compact polyhedron of dimension $\dim\mathcal{Y} \leqq p$ (Rockafellar, 1970).

Definition 3.3.1. *A feasible solution $\hat{x} \in \mathcal{X}$ is an efficient solution of problem (3.2) if there exists no $x \in \mathcal{X}$ such that $Cx \leq C\hat{x}$. The set of all efficient solutions of problem (3.2) will be denoted by $\mathcal{X}_E$ and called the efficient set in decision space. Correspondingly, $\hat{y} = C\hat{x}$ is called a nondominated point and $\mathcal{Y}_N = \{Cx : x \in \mathcal{X}_E\}$ is the nondominated set in objective space of problem (3.2).*

Definition 3.3.2. *A feasible solution $\hat{x} \in \mathcal{X}$ is called weakly efficient if there is no $x \in \mathcal{X}$ such that $Cx < C\hat{x}$. The set of all weakly efficient solutions of problem (3.2) will be denoted by $\mathcal{X}_{WE}$ and called the weakly efficient set in decision space. Correspondingly, the point $\hat{y} = C\hat{x}$ is called a weakly nondominated point and $\mathcal{Y}_{WN} = \{Cx : x \in \mathcal{X}_{WE}\}$ is the weakly nondominated set in objective space of problem (3.2).*

Definition 3.3.3 (Loridan (1984))**.** *Let $\varepsilon \in \mathbb{R}^p_{\geqq}$.*

1. *A feasible solution $\hat{x} \in \mathcal{X}$ is called an ε-efficient solution of problem (3.2) if there does not exist $x \in \mathcal{X}$ such that $Cx \leq C\hat{x} - \varepsilon$. Correspondingly, $\hat{y} = C\hat{x}$ is called an ε-nondominated point in objective space;*

2. *A feasible solution $\hat{x} \in \mathcal{X}$ is called a weakly ε-efficient solution of problem (3.2) if there does not exist $x \in \mathcal{X}$ such that $Cx < C\hat{x} - \varepsilon$. Correspondingly, $\hat{y} = C\hat{x}$ is called a weakly ε-nondominated point in objective space.*

The following theorem and definition are fundamental in multiple objective linear programming. The reader is referred to Ehrgott (2005) for a proof of the theorem.

Theorem 3.3.4. *A feasible solution $x^0 \in \mathcal{X}$ is an efficient solution of the MOLP (3.2) if and only if there exists a $\lambda \in \mathbb{R}^p_>$ such that*

$$\lambda^T C x^0 \leq \lambda^T C x \tag{3.4}$$

for all $x \in \mathcal{X}$.

Definition 3.3.5. *Let $\mathcal{F} \subseteq \mathcal{Y}$ be a face of $\mathcal{Y}$. $\mathcal{F}$ is called nondominated face, if $\mathcal{F} \subseteq \mathcal{Y}_N$. $\mathcal{F}$ is called maximal nondominated face if it is nondominated and there is no other nondominated face that contains $\mathcal{F}$.*

3.4 Benson's Outer Approximation Algorithm

Researchers have developed a variety of methods for generating all or at least part of the efficient set $\mathcal{X}_E$, such as multiobjective simplex methods and interior point methods, see the references in Ehrgott and Wiecek (2005) for more information. Although some of these approaches have had some success in aiding the decision maker (DM) in identifying a preferred solution, this success has been relatively limited due to the heavy computational requirements and the near-impossibility to study the often overwhelming set of efficient solutions $\mathcal{X}_E$.

For an MOLP problem $\mathcal{Y}_N \subseteq \mathbb{R}^p$ and $\mathcal{X}_E \subseteq \mathbb{R}^n$ with p typically much smaller than n and many points in $\mathcal{X}_E$ are mapped to a single point in $\mathcal{Y}_N$. For these reasons Benson (1998b) argues that generating $\mathcal{Y}_N$ should require less computation than generating $\mathcal{X}_E$. Moreover, it is reasonable to assume that a DM will often choose a solution based on the objective values rather than variable values. Therefore, finding $\mathcal{Y}_N$ instead of $\mathcal{X}_E$ is more important for the DM. Benson has proposed an

algorithm to solve an MOLP in objective space in (Benson, 1998b,c). In this section, we summarize his outer approximation algorithm.

For the MOLP problem (3.2), assume that $\mathcal{X}$ is compact and let

$$\mathcal{Y}' = \{y \in \mathbb{R}^p : Cx \leqq y \leqq \hat{y} \text{ for some } x \in \mathcal{X}\}, \tag{3.5}$$

where $\hat{y} \in \mathbb{R}^p$ is chosen to satisfy $\hat{y} > y^{AI}$. The vector $y^{AI} \in \mathbb{R}^p$ is called the anti ideal point for the problem (3.2) and is defined as

$$y_k^{AI} = \max\{y_k : y \in \mathcal{Y}\}. \tag{3.6}$$

Theorem 3.4.1 (Benson (1998b,c))**.** *We have*

1. *The set $\mathcal{Y}' \subseteq \mathbb{R}^p$ is a nonempty, bounded polyhedron of dimension p.*

2. $\mathcal{Y}_N = \mathcal{Y}'_N$.

Theorem 3.4.1 is the basis of the outer approximation algorithm. It works on $\mathcal{Y}'$ to find all nondominated extreme points of $\mathcal{Y}$. In the course of the algorithm, supporting hyperplanes of $\mathcal{Y}'$ are constructed. The following primal dual pair P(y) and D(y) of linear programmes depending on $y \in \mathbb{R}^p$ is needed for that purpose.

$$\begin{aligned} \text{P}(y) \qquad & \min\{z : Ax \geqq b, Cx - ez \leqq y\}, \\ \text{D}(y) \qquad & \max\{b^T u - y^T w : A^T u - C^T w = 0, e^T w = 1, u, w \geqq 0\}. \end{aligned}$$

Theorem 3.4.2 (Benson (1998c))**.** 1. *Let $\hat{p} \in \text{int}\mathcal{Y}'$ and suppose that $s^k \leqq \hat{y}$ and $s^k \notin \mathcal{Y}'$. Let y^k denote the unique point on the boundary of $\mathcal{Y}'$ that belongs to the line segment connecting s^k and $\hat{p}$. Then $y^k \in \mathcal{Y}'_{WN}$.*

2. *Assume that $y^k \in \mathcal{Y}'_{WN}$, and let (u^T, w^T) denote any optimal solution to the dual linear programme D(y^k). Then y^k belongs to the weakly nondominated face $\mathcal{F}(u, w)$ of $\mathcal{Y}'$ given by $\mathcal{F}(u, w) = \{y \in \mathcal{Y}' : \langle w, y\rangle = \langle b, u\rangle\}$.*

If $y^k \in \mathcal{Y}'_{WN}$, then P($y^k$) has the optimal value $z = 0$, and D(y^k) also has the optimal value $b^T u - {y^k}^T w = 0$. The dual optimal solution (u^T, w^T) is used to construct the supporting hyperplane of $\mathcal{Y}'$, $\mathcal{H}(u, w) = \{y \in \mathbb{R}^p : \langle w, y\rangle = \langle b, u\rangle\}$.

Benson's outer approximation algorithm is shown in Algorithm 3.4.3. For details, the reader is referred to (Benson, 1998b,c).

Algorithm 3.4.3 (Benson's outer approximation algorithm)**.**

Initialization*:*	*Compute a point* $\hat{p} \in \text{int}\mathcal{Y}'$ *and construct a p-dimensional simplex* $\mathcal{S}^0 \subseteq \hat{y} - \mathbb{R}^p_{\geqq}$ *containing* $\mathcal{Y}'$*. Store both the vertex set* $\text{vert}\,\mathcal{S}^0$ *of* $\mathcal{S}^0$ *and the inequality representation of* $\mathcal{S}^0$*. Set* $k = 0$ *and go to iteration* k*.*
Iteration k*.*	
Step $k1$	*If, for each* $s \in \text{vert}\,\mathcal{S}^k$, $s \in \mathcal{Y}'$ *is satisfied, then go to Step* $k5$*:* $\mathcal{Y}' = \mathcal{S}^k$*. Otherwise, choose any* $s^k \in \text{vert}\,\mathcal{S}^k$ *such that* $s^k \notin \mathcal{Y}'$ *and continue.*
Step $k2$	*Find the unique value* ρ_k *of* ρ*,* $0 < \rho < 1$*, such that* $\rho s^k + (1-\rho)\hat{p}$ *belongs to the boundary of* $\mathcal{Y}'$*, and set* $y^k = \rho_k s^k + (1-\rho_k)\hat{p}$*.*
Step $k3$	*Set* $\mathcal{S}^{k+1} = \mathcal{S}^k \cap \{y \in \mathbb{R}^p : \langle w^k, y\rangle \geqq \langle b, u^k\rangle\}$*, where* (u^{k^T}, w^{k^T}) *can be found by solving LP* $D(y^k)$*.*
Step $k4$	*Using* $\text{vert}\,\mathcal{S}^k$ *and the definition of* $\mathcal{S}^{k+1}$ *given in Step* $k3$*, determine* $\text{vert}\,\mathcal{S}^{k+1}$*. Set* $k = k+1$ *and go to iteration* k*.*
Step $k5$	*Let the total number of iterations be* $K = k$*. The nondominated extreme points of* $\mathcal{Y}'$ *are* $\mathcal{Y}'_{NE} = \{y \in \text{vert}\,\mathcal{S}^K : y < \hat{y}\}$*.* $\mathcal{Y}_{NE} = \mathcal{Y}'_{NE}$ *is the set of all nondominated extreme points of* $\mathcal{Y}$*. Stop.*

For each $k \geqq 0$, the hyperplane given by $\langle w^k, y\rangle = \langle b, u^k\rangle$ is constructed and $\mathcal{S}^k$ intersects with the halfspace not containing s^k to cut off a portion of $\mathcal{S}^k$, thus $\mathcal{S}^k \supseteq \mathcal{S}^{k+1} \supseteq \mathcal{Y}'$. This is the reason for the name "outer approximation" algorithm, although at termination, the MOLP is solved exactly in objective space. Theorem 3.4.4 proves that Benson's algorithm is finite and it terminates with finding all the nondominated extreme points of $\mathcal{Y}$ in Step $k5$.

Theorem 3.4.4 (Benson (1998b,c))**.** *Algorithm 3.4.3 is finite and at termination* $\mathcal{S}^K = \mathcal{Y}'$. *Let* $\mathcal{Y}'_{NE} = \{y \in \operatorname{vert} \mathcal{S}^K : y < \hat{y}\}$. *Then* $\mathcal{Y}'_{NE}$ *is identical to the set of all nondominated extreme points of* $\mathcal{Y}$, *i.e.,* $\mathcal{Y}'_{NE} = \mathcal{Y}_{NE}$.

The general idea of Benson's algorithm can be explained as follows. First, a simplex cover $\mathcal{S}^0$ that contains $\mathcal{Y}'$ is constructed. $\mathcal{S}^0$ is given by axes parallel hyperplanes defined by the entries of $\hat{y}$ and a supporting hyperplane of $\mathcal{Y}'$ with normal $e = (1, \ldots, 1) \in \mathbb{R}^p$. An interior point $\hat{p}$ of $\mathcal{Y}'$ is found. Then, for each vertex s^k of the cover, it is checked whether s^k is in $\mathcal{Y}'$ or not. If not, $\hat{p}$ and s^k are connected by a line segment that contains a unique boundary point y^k of $\mathcal{Y}'$. A cut (new supporting hyperplane) containing y^k is constructed and the cover $\mathcal{S}^k$ is updated. The procedure repeats until all the vertices of the cover are in $\mathcal{Y}'$. Then the vertices of the cover are the extreme points of $\mathcal{Y}'$ and the nondominated extreme points of $\mathcal{Y}'$ are $\mathcal{Y}'_{NE} = \{y \in \operatorname{vert} \mathcal{S}^k : y < \hat{y}\}$.

We give an example to illustrate Benson's algorithm.

Example 3.4.5. *Consider the MOLP* $\min\{Cx : Ax \geqq b\}$, *where*

$$C = \begin{pmatrix} 3 & 1 \\ -1 & -2 \end{pmatrix}, A = \begin{pmatrix} 0 & -1 \\ -3 & 1 \\ 1 & 0 \\ 0 & 1 \end{pmatrix}, b = \begin{pmatrix} -3 \\ -6 \\ 0 \\ 0 \end{pmatrix}.$$

The feasible set $\mathcal{Y}$ *in objective space is shown in Figure 3.1. Choosing* $\hat{y} = (13, 1)$ *we define* $\mathcal{Y}'$ *as* $\mathcal{Y}' = \{y \in \mathbb{R}^2 : Cx \leqq y \leqq \hat{y},\ Ax \geqq b\}$. *Figure 3.2 shows* $\mathcal{Y}$, $\mathcal{Y}'$, $\mathcal{S}^0$ *and the interior point* $\hat{p} = (12.5, 0.5)$ *of* $\mathcal{Y}'$.

Figures 3.3, 3.4, 3.5, and 3.6 show the first, second, third and fourth hyperplane, respectively. The vertices of $\mathcal{S}^k$ *are shown as filled circles, the vertex being cut off is shown as an empty circle and the boundary point* y^k *is shown as a cross. The first hyperplane cuts off vertex* $(-4, 1)$, *the second cuts off vertex* $(13, -16)$, *the third cuts off vertex* $(0, -3)$, *the fourth cuts off vertex* $(6, -9)$. *We can see the change of* $\mathcal{S}^k$ *after each cut. After the fourth cut, we have* $\mathcal{S}^4 = \mathcal{Y}'$. *Therefore, the vertices of* $\mathcal{S}^4$ *are the extreme points of* $\mathcal{Y}'$. *We obtain all nondominated extreme points by*

$\mathcal{Y}'_{NE} = \{y \in \operatorname{vert} \mathcal{S}^4 : y < \hat{y}\}$. *In this example we obtain the three nondominated extreme points* $(12, -9), (3, -6)$ *and* $(0, 0)$.

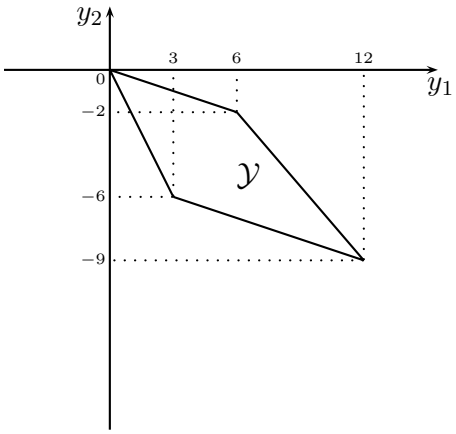

Figure 3.1: *Objective space* $\mathcal{Y}$.

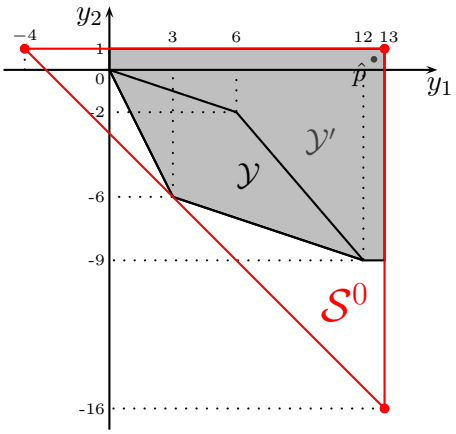

Figure 3.2: $\mathcal{Y}'$, $\mathcal{S}^0$ *and* $\hat{p} \in \operatorname{int}\mathcal{Y}'$.

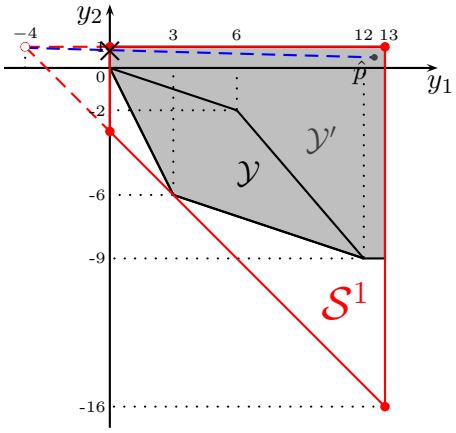

Figure 3.3: *After the first cut.*

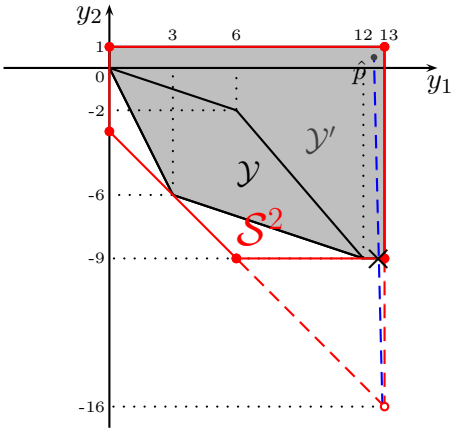

Figure 3.4: *After the second cut.*

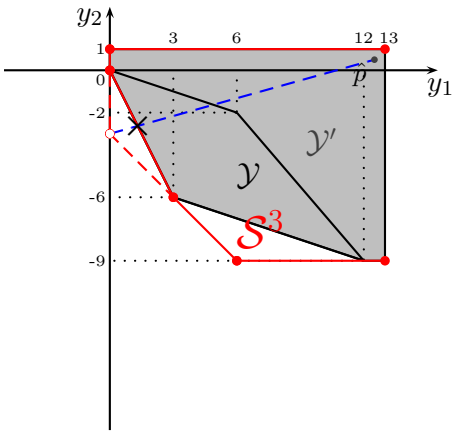

Figure 3.5: *After the third cut.*

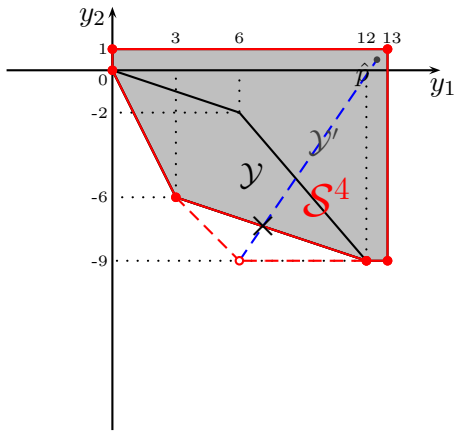

Figure 3.6: *After the fourth cut.*

3.5 Improvements to Benson's Algorithm

In **Step** $k2$ of Algorithm 3.4.3 it is necessary to find the unique ρ $(0 < \rho < 1)$ which determines the boundary point $y^k = \rho s^k + (1-\rho)\hat{p}$ of $\mathcal{Y}'$. Benson (1998c) suggests using a bisection method. However, this requires the solution of many LPs. We show that it is possible to find the appropriate ρ solving just a single LP.

Proposition 3.5.1. *Let s^k be a vertex of $\mathcal{S}^k$ and $s^k \notin \mathcal{Y}'$. Let $\hat{p} \in \text{int}\mathcal{Y}'$ and let $\mu = s^k - \hat{p}$. Then there must exist some $i \in \{1, \ldots, p\}$ such that $\mu_i < 0$.*

Proof. We have $s^k \leqq \hat{y}$. Since $s^k \notin \mathcal{Y}'$, we get $Cx \not\leqq s^k$ for all feasible x. There exists some feasible $\bar{x}$ such that $C\bar{x} \leqq \hat{p}$. Hence $\hat{p} \not\leqq s^k$, which is the desired statement. □

By Proposition 3.5.1 it is possible to choose $l \in \{1, \ldots, p\}$ with $\mu_l = s^k_l - \hat{p}_l < 0$. We choose that $\mu_l < 0$ such that $\mu_l = \min_i \{\mu_i : \mu_i < 0\}$ for $i \in \{1, \ldots, p\}$. Due to the convexity of $\mathcal{Y}'$, among all points of $\mathcal{Y}'$ on the line segment connecting points s^k and $\hat{p}$, the boundary point y of $\mathcal{Y}'$ attains the smallest value of y_l.

Therefore the unique ρ for determining y^k can be found by solving the LP

$$\begin{array}{lrcl} \max & \rho & & \\ \text{s.t.} & \rho s^k + (1-\rho)\hat{p} & \geqq & Cx \\ & Ax & \geqq & b \\ & \rho & \leqq & 1 \\ & \rho & \geqq & 0, \end{array} \tag{3.7}$$

where ρ and x are the variables. Note that, since $s^k \notin \mathcal{Y}'$ and $\hat{p} \in \text{int}\mathcal{Y}$, ρ cannot be 0 or 1 in an optimal solution of (3.7).

This modification dramatically improves the computation time.

Moreover, to calculate the vertices of $\mathcal{S}^{k+1} = \mathcal{S}^k \cap \{y \in \mathbb{R}^p : \langle w^k, y\rangle \geqq \langle b, u^k\rangle\}$ in **Step** $k4$, Benson proposes the method of Horst *et al.* (1988) in (Benson, 1998c) and the simplicial partitioning technique in (Benson, 1998b). We use the on-line vertex enumeration algorithm of Chen and Hansen (1991). This algorithm is based on the algorithm of Horst *et al.* (1988) but its complexity is smaller, as shown in Chen and Hansen (1991). The principle of the on-line vertex enumeration algorithm is to find the vertex sets of $\mathcal{S}^k$ on both sides of the cutting plane $\mathcal{H}(u,w) = \{y \in \mathbb{R}^p : \langle w, y\rangle = \langle b, u\rangle\}$ and then to use adjacency lists of extreme points to identify all edges of $\mathcal{S}^k$ intersecting $\mathcal{H}(u,w)$. The corresponding intersection points are computed and the adjacency lists updated. We found that the on-line vertex enumeration method leads to an improvement in computation speed compared to the simplicial partitioning technique.

3.6 Summary

In this chapter, we have introduced our MOLP model for beam intensity optimization problem. Moreover, we summarized Benson's outer approximation algorithm to solve an MOLP in objective space and we described some improving modifications to Benson's outer approximation algorithm. The modification dramatically improves the computation time, thus it was possible for us to solve some of the radiotherapy problems within an acceptable computation time.

We use Benson's algorithm to determine the entire nondominated set of the MOLP model for the beam intensity optimization problem. Some results are given in Chapter 4. Moreover, in Chapter 4 we propose an approximation version of Benson's algorithm and we also apply it to the beam intensity optimization problem. The results obtained with the approximation method are compared with those obtained with Benson's algorithm.

Chapter 4

Approximately Solving Multiobjective Linear Programmes in Objective Space

This chapter is organized as follows. In Section 4.1 we describe the approximation version of Benson's algorithm and prove that it finds a set of weakly ε-nondominated points in the feasible set in objective space. The rest of the chapter is dedicated to the application in radiotherapy treatment planning. In Section 4.2 we review the MOLP formulation of Chapter 3 for the beam intensity optimization problem and motivate the use of an approximation algorithm by clinical considerations, we provide results on four clinical cases. Finally we draw some conclusions in Section 4.3.

4.1 Approximation Version of Benson's Algorithm

We modify Algorithm 3.4.3 in order to find weakly ε-nondominated points of $\mathcal{Y}'$. In addition to the vertex set vert $\mathcal{S}^k$ we introduce sets $\mathcal{O}$ and $\mathcal{I}$ (initially empty) of points used for the construction of an inner and an outer approximation of $\mathcal{Y}'$. In the algorithm, if an extreme point s^k of vert $\mathcal{S}^k$ is close to $\mathcal{Y}'$, i.e., has a distance less than $\epsilon > 0$ from the boundary point y^k we omit construction of the hyperplane in **Step**

$k3$ but remember both s^k and y^k to construct the inner and outer approximation of $\mathcal{Y}'$.

Our approximation version of Benson's algorithm is identical to Algorithm 3.4.3 except for **Step** $k1$, **Step** $k3$, and **Step** $k5$. Let $\epsilon \in \mathbb{R}, \epsilon \geqq 0$ be a tolerance and let d denote the Euclidean distance. In this thesis we always use $d(x, y) = \|x - y\|_2$, the Euclidean norm. Then the changes are as follows.

Step $k1$ If, for each $s \in \operatorname{vert} \mathcal{S}^k, y \in \mathcal{Y}'$ or $y \in \mathcal{O}$ is satisfied, then go to Step $k5$. Otherwise, choose any $s^k \in \operatorname{vert} \mathcal{S}^k \setminus \mathcal{O}$ such that $s^k \notin \mathcal{Y}'$ and continue.

Step $k3$ If the distance $d(s^k, y^k)$ from s^k to the boundary point y^k of $\mathcal{Y}'$ is at most ϵ, then add s^k to $\mathcal{O}$ and add y^k to $\mathcal{I}$. Go to Step $k1$. Otherwise set $\mathcal{S}^{k+1} = \mathcal{S}^k \cap \{y \in \mathbb{R}^p : \langle w^k, y\rangle \geqq \langle b, u^k\rangle\}$, where $({u^k}^T, {w^k}^T)$ can be found by solving LP D(y^k).

Step $k5$ Let the total number of iterations be $K = k$. Define the set of points of the outer approximation $V_o(\mathcal{S}^K) = \operatorname{vert} \mathcal{S}^K$ and define the set of points of the inner approximation $V_i(\mathcal{S}^K) = (\operatorname{vert} \mathcal{S}^K \setminus \mathcal{O}) \cup \mathcal{I}$. The convex hull $\mathcal{Y}'^i$ of $V_i(\mathcal{S}^K)$ represents the inner approximation of $\mathcal{Y}'$. The convex hull $\mathcal{Y}'^o$ of $V_o(\mathcal{S}^K)$ represents the outer approximation of $\mathcal{Y}'$. Stop.

Since $\mathcal{Y}'$ is convex and $\mathcal{S}^k \supseteq \mathcal{Y}'$ it is clear that $\operatorname{conv} V_o(\mathcal{S}^k) \supseteq \mathcal{Y}' \supseteq \operatorname{conv} V_i(\mathcal{S}^k)$. We apply the modified algorithm to Example 3.4.5.

Example 4.1.1. *In Example 3.4.5, set* $\hat{p} = (12.5, 0.5)$ *and* $\epsilon = 2.0$. *After two cuts there are two points* $s^1 = (0, -3)$ *and* $s^2 = (6, -9)$ *outside* $\mathcal{Y}'$, *see Figure 4.1. The boundary points corresponding to* s^1 *and* s^2 *are* $y^1 = (1\frac{6}{19}, -2\frac{12}{19}) \approx (1.316, -2.632)$ *and* $y^2 = (7\frac{4}{35}, -7\frac{13}{35}) \approx (7.114, -7.371)$, *respectively. The distances between the infeasible points and the boundary points are* $d(s^1, y^1) \approx 1.366$ *and* $d(s^2, y^2) \approx 1.973$. *We accept these two infeasible points for the outer approximation due to the distances to their corresponding boundary points being less than* ϵ. *When the algo-*

rithm terminates, the total number of iterations K is equal to 2, $V_o(\mathcal{S}^2) = \{(13,1), (0,1), (0,-3), (6,-9), (13,-9)\}$ and $V_i(\mathcal{S}^2) = \{(13,1), (0,1), (1.316,-2.632), (7.114,-7.371), (13,-9)\}$. In Figure 4.2, we show the outer approximation $\mathcal{Y}'^o$ and the inner approximation $\mathcal{Y}'^i$ of $\mathcal{Y}'$ and their corresponding sets of points.

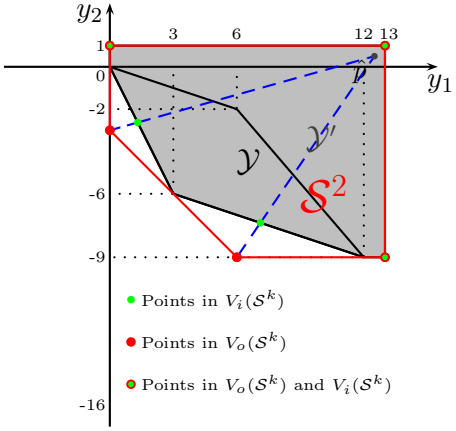

Figure 4.1: *Accepted infeasible points.*

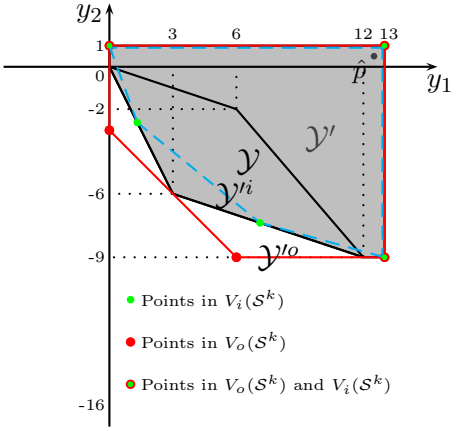

Figure 4.2: *Inner and outer approximation.*

By definition of the approximation version of Benson's algorithm, we have the following observations.

Theorem 4.1.2.

1. *The number of points in $V_o(\mathcal{S}^K)$ is equal to the number of points in $V_i(\mathcal{S}^K)$.*
2. *All points in $V_i(\mathcal{S}^K)$ are on the boundary of $\mathcal{Y}'$. Some points in $V_o(\mathcal{S}^K)$ are outside $\mathcal{Y}'$, while the others are on the boundary of $\mathcal{Y}'$. Moreover, $y \in V_o(\mathcal{S}^K)$ is not on the boundary of $\mathcal{Y}'$ if and only if $y \notin V_i(\mathcal{S}^K)$.*
3. *If y_{ov} is a point in $V_o(\mathcal{S}^K)$, there exists a point y_{iv} in $V_i(\mathcal{S}^K)$ with $d(y_{ov}, y_{iv}) \leqq \epsilon$ and vice versa.*
4. *If $\mathcal{Y}'^i_N$ is the nondominated set of the inner approximation $\mathcal{Y}'^i$ and $\mathcal{Y}'^o_N$ is the nondominated set of the outer approximation $\mathcal{Y}'^o$, then we have $\mathcal{Y}'^i_N + \mathbb{R}^p_{\geqq} \subseteq \mathcal{Y}'_N + \mathbb{R}^p_{\geqq} \subseteq \mathcal{Y}'^o_N + \mathbb{R}^p_{\geqq}$.*

Point 4 in Theorem 4.1.2 means that $\mathcal{Y}_N'^i$ and $\mathcal{Y}_N'^o$ are upper and lower bound sets for $\mathcal{Y}_N$ as defined by Ehrgott and Gandibleux (2007). We would also like to note that the approximation depends on the choice of the interior point $\hat{p}$. Of course, if $\epsilon = 0$ the algorithm is Benson's original algorithm. We proceed to show that $\mathcal{Y}_N'^i$ is a set of weakly ε-nondominated points for $\mathcal{Y}'$.

Proposition 4.1.3. *If y_o is a weakly nondominated point of the outer approximation set $\mathcal{Y}'^o$, then there exists a weakly nondominated point y_i of the inner approximation set $\mathcal{Y}'^i$ such that $d(y_o, y_i) \leqq \epsilon$.*

Proof. Let y_o be a point on $\mathcal{F}_o$, a weakly nondominated face of the outer approximation set with vertices $y_{ov}^1, y_{ov}^2, \ldots, y_{ov}^l \in V_o(\mathcal{S}^K)$. Then y_o can be expressed as a convex combination of the vertices, i.e., $y_o = \sum_{j=1}^l \rho_j y_{ov}^j$ with $\sum_{j=1}^l \rho_j = 1$ and $\rho_j \geqq 0$, $j = 1, \ldots l$.

Let $y_{iv}^1, y_{iv}^2, \ldots, y_{iv}^l \in V_i(\mathcal{S}^K)$ be the corresponding points to $y_{ov}^1, y_{ov}^2, \ldots, y_{ov}^l$ on the inner approximation. Then $d(y_{ov}^j, y_{iv}^j) \leqq \epsilon$, for $j = 1, \ldots, l$. Let $y_i = \sum_{j=1}^l \rho_j y_{iv}^j$, then $d(y_o, y_i) = ||\sum_{j=1}^l \rho_j y_{ov}^j - \sum_{j=1}^l \rho_j y_{iv}^j|| \leqq \sum_{j=1}^l \rho_j ||y_{ov}^j - y_{iv}^j|| \leqq \sum_{j=1}^l \rho_j \epsilon = \epsilon$. If y_i is a weakly nondominated point of the inner approximation we are done. Otherwise, choose the intersection point $\tilde{y}$ of the line connecting y_i and y_o with the boundary of $\mathcal{Y}'^i$. Clearly $d(\tilde{y}, y_o) \leq d(y_i, y_o) \leqq \epsilon$. □

Combining Proposition 4.1.3 with Definition 3.3.3 we obtain our main result.

Theorem 4.1.4. *Let $\varepsilon = \epsilon e$. We have that $\mathcal{Y}_N'^i$ is a set of weakly ε-nondominated points for $\mathcal{Y}'$.*

Proof. Let $y_i \in \mathcal{Y}_N'^i$ and suppose there is $y \in \mathcal{Y}'$ such that $y < y_i - \varepsilon$. Thus $y_i - y > \varepsilon$ and $d(y, y_i) > ||\varepsilon|| = \epsilon||e|| \geqq \epsilon$. By Theorem 4.1.2 we have that $\mathcal{Y}' \subseteq \mathcal{Y}_N'^o + \mathbb{R}^p_{\geqq}$, i.e., there is $y_o \in \mathcal{Y}_N'^o$ such that $y_o \leqq y$. Now observe that the intersection of the hypercube defined by y and y_i with $\mathcal{Y}'^i$ contains the single point y_i because $\mathcal{Y}'^i + \mathbb{R}^p_{\geqq}$ is convex. The hypercube has edge length at least ϵ. Thus we have that $d(y_o, \bar{y}_i) \geqq d(y, \bar{y}_i) \geqq d(y, y_i) > \epsilon$ for any $\bar{y}_i \in \mathcal{Y}_N'^i$, contradicting Proposition 4.1.3. □

Theorem 4.1.4 shows that the approximation version of Benson's algorithm allows a guaranteed approximation quality for the weakly nondominated set of $\mathcal{Y}'$. Because $\mathcal{Y}_N \subseteq \mathcal{Y}'_{WN}$ it is valid for $\mathcal{Y}_N$ as well. But $\mathcal{Y}'^{i}_{N}$ may contain weakly nondominated points of $\mathcal{Y}$ and even points of $\mathcal{Y}' \setminus \mathcal{Y}$, see Figure 4.2.

To approximate the nondominated set of $\mathcal{Y}$ and avoid weakly nondominated points and points of $\mathcal{Y}' \setminus \mathcal{Y}$, we define $\mathcal{Y}^{i}_{NE} = \{y \in V_i(\mathcal{S}^K) : y < \hat{y}\}$ and $\mathcal{Y}^{o}_{NE} = \{y \in V_o(\mathcal{S}^K) : y < \hat{y}\}$. We construct faces using the points in $\mathcal{Y}^{o}_{NE}$ on the same cutting plane (found during the algorithm) and let $\mathcal{Y}^{o}_{N}$ be the union of the faces. Similarly, we can construct $\mathcal{Y}^{i}_{N}$. Then the true nondominated set $\mathcal{Y}_N$ can be approximated from outside by $\mathcal{Y}^{o}_{N}$ and from inside by $\mathcal{Y}^{i}_{N}$.

Example 4.1.5. *For Example 3.4.5, the points in* $\mathcal{Y}^{o}_{NE}$ *are* $(0, -3)$ *and* $(6, -9)$*, while the points in* $\mathcal{Y}^{i}_{NE}$ *are* $(1.316, -2.632)$ *and* $(7.114, -7.371)$*. The set* $\mathcal{Y}^{o}_{N}$ *is the line segment from point* $(0, -3)$ *to point* $(6, -9)$ *and the set* $\mathcal{Y}^{i}_{N}$ *is the line segment from point* $(1.316, -2.632)$ *to point* $(7.114, -7.371)$*, see Figure 4.2. Note that there might exist* $y \in \mathcal{Y}_N$ *which are farther than* ϵ *from any point in* $\mathcal{Y}^{o}_{N}$ *and* $\mathcal{Y}^{i}_{N}$*.*

If $p = 2$, $\mathcal{Y}_N$ can be interpreted as the graph of a convex function of one variable. In that case our algorithm is the same as the sandwich algorithm of Burkard *et al.* (1991) for the approximation of a convex function. However, Burkard *et al.* (1991) do not generalize the method to higher dimension.

4.2 Numerical Results

Recall the MOLP formulation (3.1) for the beam intensity optimization problem we formulated in Chapter 3. We minimize α, β and γ at the same time. They represent the maximum deviation from tumor lower bounds, critical organ upper bounds and normal tissue upper bounds, respectively.

The constraints of (3.1) involve the dose deposition matrix A. As mentioned before, a_{ij} describes the dose deposited in voxel i if unit intensity is applied in bixel j. The coefficients a_{ij} are calculated by mathematical models of the physical behavior of radiation as it travels through the body. While sophisticated techniques

are available and in clinical use (Nizin *et al.*, 2001), with the gold standard being a Monte Carlo simulation (Verhaegen, 2003), the results are always imprecise due to the nonuniform composition of the patient body. Thus, solving (3.1) exactly may give an unwarranted impression of precision, but the result of the optimization can of course not be more precise than the input data. Moreover, considering the imprecision due to discretization of intensity profiles before segmentation as well as the imprecision in delivery due to machine effects and patient movement, we have that for clinical purposes it is perfectly acceptable to solve (3.1) approximately to within a small fraction of a Gy (Gray, the unit of measure for radiation dose). Note that the objectives α, β, γ are commensurate and have the unit Gy and that the tolerance ϵ in the approximation algorithm is absolute, not relative i.e. the error measure is additive, not multiplicative. This is one motivation for solving the problem by an approximation version of Benson's algorithm.

Moreover, we use Benson's algorithm to determine the entire nondominated set of the MOLP model for the beam intensity optimization problem. Due to the size of the MOLP model for clinical cases and despite the improvements described in Section 3.5 it turns out that computation times are excessive. We also observed that for the clinical examples the tradeoffs between the objectives vary widely. Thus the nondominated sets in objective space appear to be "curved" (see the figures later in this section). This means that very many cutting planes are needed to describe $\mathcal{Y}'$. This explains why Benson's algorithm has computational problems and takes very long to terminate. This is the second motivation for using an approximation version of Algorithm 3.4.3.

We solve (3.1) both by Algorithm 3.4.3 and our approximation algorithm described in Section 4.1. Four clinical cases are used, namely an arterial veinous malformation (AVM), an acoustic neuroma (AN), a prostate (PR), and a pancreatic lesion (PL). Simplified CT images that show the outline of the tumor and critical organs at risk are shown in Figure 4.3. These cases have a voxel size of 5mm on a single CT slice. For all examples, a total of 72 evenly spaced beams were used at angles $5^o n$, where $n = 0, \ldots, 71$. The number of voxels and bixels used for optimiza-

tion of each case and the prescription information that defines parameters in (3.1) are shown in Table 4.1. For all the cases, α_u is set to be 20 percent of TLB, β_u is set to be 20 percent of the maximal CUB value. γ_u is chosen to be greater than or equal to TUB. The algorithm was implemented in Matlab 7.1 (R14) using CPLEX 10.0 as LP solver and the tests were run on a dual processor CPU with 1.8 GHz and 1 GB RAM.

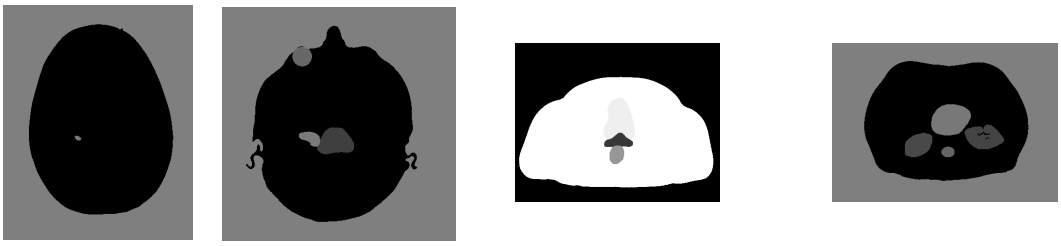

Figure 4.3: *Pictures from left to right are AVM, AN, PR, and PL.*

Table 4.1: *Number of voxels (total = m) and bixels (n). Lower and upper bounds for tumor, critical organs, and normal tissue (in Gy).*

Case	AVM	AN	PR	PL
Tumor voxels	1	9	22	67
Critical organ voxels	0	47	89	91
Normal tissue voxels	1206	999	1182	986
Bixels	319	594	821	1140
TUB	90.64	87.55	90.64	90.64
TLB	85.36	82.45	85.36	85.36
CUB	—	60/45	60/45	60/45
NUB	0.00	0.00	0.00	0.00
α_u	17.07	16.49	42.68	17.07
β_u	—	12.00	30.00	12.00
γ_u	90.64	87.55	100.64	90.64

Note: 60 and 45 are the upper bounds for different critical organs.

For the AVM case, both algorithms find two nondominated extreme points. They are $y^1 = (0, 0, 79.31)$ and $y^2 = (17.07, 0, 63.45)$. Here, β is equal to zero because there is no critical organ in this case. Ignoring the β dimension, we have that the nondominated set is the line segment from point $(0, 79.31)$ to point $(17.07, 63.45)$, see Figure 4.4. The clinical meaning of point $(0, 79.31)$ is that there is a solution for which the (single) voxel in the tumor will receive a dose greater than or equal to the tumor lower bound and smaller than or equal to the upper bound, while some voxel in the normal tissue will receive a dose as high as 79.31 Gy (this is a voxel in immediate proximity of the tumor). The clinical meaning of point $(17.07, 63.45)$ is that there is a solution for which the voxel in the tumor will receive a dose as low as $TLB - 17.07 = 85.36 - 17.07 = 68.29$ Gy, while some voxel in the normal tissue will receive a dose as high as 63.45 Gy. We can explain all the other nondominated points in between those two similarly.

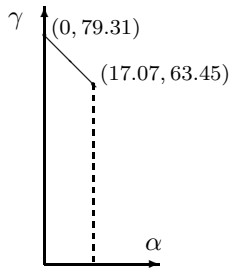

Figure 4.4: *Nondominated set of the AVM case.*

For the other cases, it is not possible to list all nondominated extreme points. We show the set $\mathcal{Y}'$ obtained by Benson's algorithm for the acoustic neuroma in Figure 4.5 and for the prostate in Figure 4.7 side by side with the set $\mathcal{Y}'^o$ of the outer approximation obtained by the approximation version of the algorithm with $\epsilon = 0.1$. The acoustic neuroma is shown in Figure 4.6 and the prostate in Figure 4.8.

The pancreatic lesion case could not be solved exactly within 10 hours of computation. Therefore, we show the sets $\mathcal{Y}'^o$ obtained by the approximation algorithm for various values of ϵ. Figure 4.9 shows the result for $\epsilon = 0.3$, Figure 4.10 is for $\epsilon = 0.1$, Figure 4.11 is for $\epsilon = 0.05$, and Figure 4.12 is for $\epsilon = 0.005$.

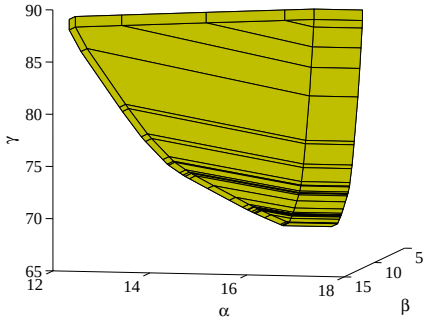

Figure 4.5: *AN:* $\mathcal{Y}'$.

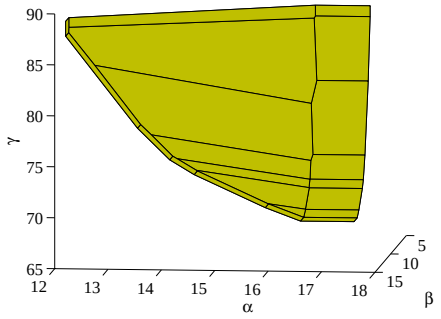

Figure 4.6: *AN:* $\mathcal{Y}'^{o}$ *with* $\epsilon = 0.1$.

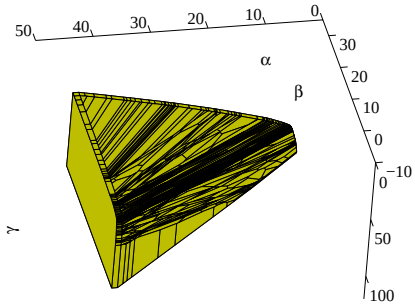

Figure 4.7: *PR:* $\mathcal{Y}'$.

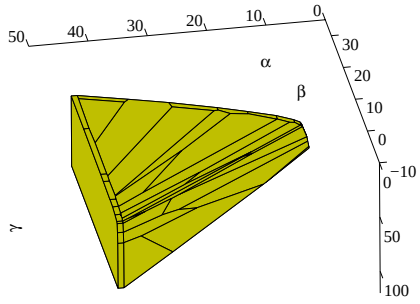

Figure 4.8: *PR:* $\mathcal{Y}'^{o}$ *with* $\epsilon = 0.1$.

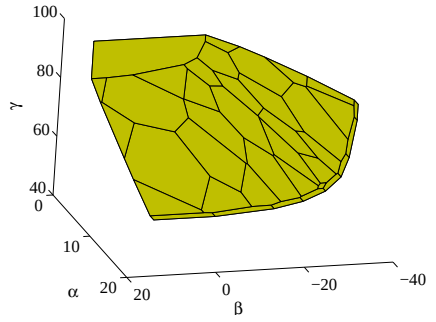

Figure 4.9: *PL:* $\mathcal{Y}'^{o}$ *with* $\epsilon = 0.3$.

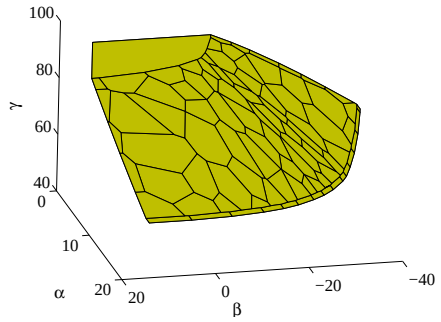

Figure 4.10: *PL:* $\mathcal{Y}'^{o}$ *with* $\varepsilon = 0.1$.

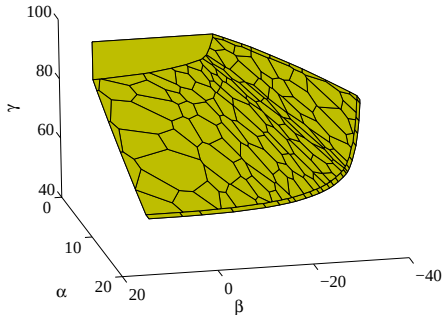

Figure 4.11: *PL: $\mathcal{Y}'^{o}$ with $\epsilon =$ 0.05.*

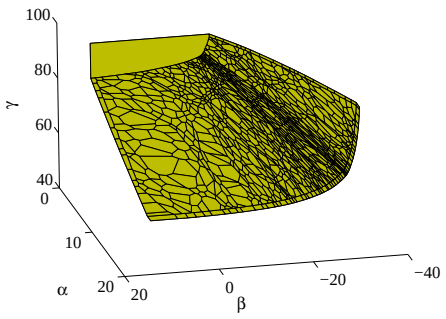

Figure 4.12: *PL: $\mathcal{Y}'^{o}$ with $\varepsilon =$ 0.005.*

Summarizing information comparing the number of nondominated extreme points, the number of cutting planes and the computation time of Benson's algorithm and our approximation version of Benson's algorithm with various values of ϵ is given in Table 4.2.

Table 4.2: *Running time and number of nondominated extreme points and cutting planes for the four cases with different values of ϵ.*

Case	ϵ	Time (seconds)	Nondominated extreme points	Cutting planes
AVM	0.1	0.514	2	3
	0	0.515	2	3
AN	0.1	5.938	27	21
	0	13.984	55	85
PR	0.1	14.781	56	42
	0	995.050	3165	3280
PL	0.3	70.796	57	37
	0.1	164.360	152	90
	0.05	303.630	278	159
	0.005	2147.530	1989	1041

Benson's algorithm can solve the first three clinical cases exactly in less than 1.5

hours. For the pancreatic lesion case, Benson's algorithm did not terminate after 10 hours of computation. On the other hand, the approximation version of Benson's algorithm can solve all four problems within 30 minutes with an error no greater than 0.1 Gy.

For a problem with many nondominated extreme points and a "curved" nondominated surface, such as Figure 4.12 suggests, the approximation version of Benson's algorithm generates fewer extreme points and fewer cutting planes compared with Benson's algorithm. For the prostate example, 3165 nondominated extreme points were found with Benson's algorithm, while the approximation version of Benson's algorithm generates only 56 nondominated extreme points when $\epsilon = 0.1$.

Table 4.2 and the figures clearly show the effect of the choice of ϵ. The smaller the error parameter, the more cutting planes and the more nondominated extreme points are generated and the longer the computation time.

4.3 Summary

In this chapter, we have developed an approximation version of Benson's algorithm to solve MOLPs in objective space. We have shown that the algorithm guarantees to find weakly ε-nondominated points with a specified accuracy ϵ ($\varepsilon = \epsilon e$).

The development of the algorithm was motivated by the beam intensity optimization problem of radiotherapy treatment planning, which can be formulated as an MOLP. The constraint matrix of this model depends on the model of the physical behavior of radiation. Since this calculation is inaccurate the application of an approximation algorithm is justified in the practical application. In this context the parameter ϵ can be chosen by the radiation oncologist, based on his knowledge on how accurately the beam model used by the specific treatment planning system calculates dose deposited in the body.

We have used four different clinical cancer cases to test the algorithm, using only a single CT slice and a voxel size of 5mm. In Chapter 8 we will present results for 3D problems with 3mm voxels.

Our method provides an approximation of the whole nondominated set. Further work is necessary to combine this approach with decision support tools to assist the treatment planner in selecting a treatment plan from this set that is best suited for the individual patient under consideration.

Chapter 5

A Dual Variant of Benson's Outer Approximation Algorithm

5.1 Introduction

We use some new results on duality for multiple objective linear programmes in order to develop a dual variant of Benson's algorithm. Geometric duality (Heyde and Löhne, 2006) defines a dual vector optimization problem which has a completely different outcome set than the primal problem, but it provides a well-defined relationship between the primal and dual outcome set which is easy to handle. The idea of Benson's algorithm can be applied (with some slight modifications) to the dual outcome set. Duality results yield information about the primal outcome set.

The chapter is organized as follows. In Section 5.2 we give an introduction to geometric duality. Section 5.3 is devoted to the original outer approximation algorithm. However, we remove the assumption that $\mathcal{X}$ is compact mentioned in Chapter 4. Section 5.4 deals with our dual variant of the algorithm. In Section 5.5 we prove that the solution of the dual problem provides a weight set decomposition with respect to nondominated extreme points. In Section 5.6 we present numerical results, comparing the primal and the dual algorithm for several examples.

5.2 Geometric Duality

Let $A \in \mathbb{R}^{m\times n}$, $b \in \mathbb{R}^m$, $C \in \mathbb{R}^{p\times n}$, $e \in \mathbb{R}^p$. Consider the vector optimization problem

$$\text{(P)} \qquad \operatorname{wmin}_{\mathbb{R}^p_{\leqq}} C(\mathcal{X}), \quad \mathcal{X} := \{x \in \mathbb{R}^n : Ax \geqq b\},$$

where $\operatorname{wmin}_{\mathbb{R}^p_{\leqq}} \mathcal{A} := \left\{y \in \mathcal{A} : (\{y\} - \operatorname{int} \mathbb{R}^p_{\leqq}) \cap \mathcal{A} = \emptyset\right\}$ as defined in Section 3.1.

Then the dual problem according to the *geometric duality theory* developed in Heyde and Löhne (2006) is

$$\text{(D)} \qquad \max_{\mathcal{K}} D(\mathcal{U}), \quad \mathcal{U} := \left\{(u,\lambda) \in \mathbb{R}^m \times \mathbb{R}^p : (u,\lambda) \geqq 0, A^T u = C^T\lambda, e^T\lambda = 1\right\},$$

where $\max_{\mathcal{K}} \mathcal{A} := \{y \in \mathcal{A} : (\{y\} + \mathcal{K} \setminus \{0\}) \cap \mathcal{A} = \emptyset\}$, $\mathcal{K} := \{y \in \mathbb{R}^p : y_1 = y_2 = ... = y_{p-1} = 0, y_p \geqq 0\}$ as defined before and $D : \mathbb{R}^{m+p} \to \mathbb{R}^p$ is given by

$$D(u,\lambda) := \left(\lambda_1, ..., \lambda_{p-1}, b^T u\right)^T = \begin{pmatrix} 0 & I_{p-1} & 0 \\ b^T & 0 & 0 \end{pmatrix} \begin{pmatrix} u \\ \lambda \end{pmatrix}.$$

Note that for given λ, $\max_{\mathcal{K}} D(\mathcal{U})$ is the dual LP $\max\{b^T u : u \geqq 0, A^T u = C^T\lambda\}$ of the weighted sum LP $\min\{\lambda^T Cx : Ax \geqq b\}$.

The primal problem (P) consists in finding the weakly nondominated points of $C(\mathcal{X})$, the dual problem consists in finding the $\mathcal{K}$-maximal elements of $D(\mathcal{U})$. We introduce the extended polyhedral image sets $\mathcal{P} := C(\mathcal{X}) + \mathbb{R}^p_{\leqq}$ of problem (P) and $\mathcal{D} := D(\mathcal{U}) - \mathcal{K}$ of problem (D). It is known that the $\mathbb{R}^p_{\leqq}$-minimal (nondominated) points of $\mathcal{P}$ and $C(\mathcal{X})$ as well as the $\mathcal{K}$-maximal elements of $\mathcal{D}$ and $D(\mathcal{U})$ coincide, see Heyde and Löhne (2006). An illustration is given in Example 5.2.1.

Example 5.2.1. *Consider problem (P) with the data*

$$C = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \; A = \begin{pmatrix} 2 & 1 \\ 1 & 1 \\ 1 & 2 \\ 1 & 0 \\ 0 & 1 \end{pmatrix}, \; b = \begin{pmatrix} 4 \\ 3 \\ 4 \\ 0 \\ 0 \end{pmatrix}.$$

The extended outcome sets $\mathcal{P}$ and $\mathcal{D}$ of (P) and (D) are shown in Figures 5.1 and 5.2.

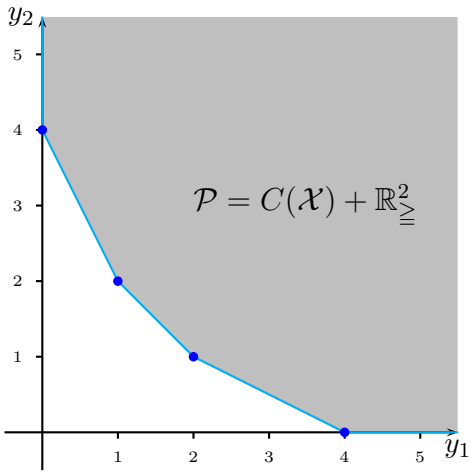

Figure 5.1: *$\mathcal{P}$ in Example 5.2.1.*

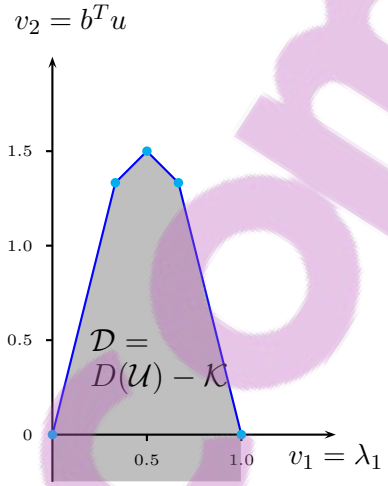

Figure 5.2: *$\mathcal{D}$ in Example 5.2.1.*

The geometric duality theory of Heyde and Löhne (2006) establishes a relationship between the (weakly nondominated) vertices of $\mathcal{P}$ and the $\mathcal{K}$-maximal facets of $\mathcal{D}$ and between the weakly nondominated facets of $\mathcal{P}$ and the $\mathcal{K}$-maximal vertices of $\mathcal{D}$. In this example, the five vertices of $\mathcal{D}$, namely $(0,0)$, $(\frac{1}{3},\frac{4}{3})$, $(\frac{1}{2},\frac{3}{2})$, $(\frac{2}{3},\frac{4}{3})$, and $(1,0)$, correspond to the facets of $\mathcal{P}$ given by $y_2 = 0$, $y_1 + 2y_2 = 4$, $y_1 + y_2 = 3$, $2y_1 + y_2 = 4$, $y_1 = 0$. The four vertices of $\mathcal{P}$, namely $(0,4)$, $(1,2)$, $(2,1)$, and $(4,0)$ correspond to the $\mathcal{K}$-maximal facets of $\mathcal{D}$ given by $4v_1 + v_2 = 4$, $v_1 + v_2 = 2$, $-v_1 + v_2 = 1$, and $-4v_1 + v_2 = 0$, respectively.

Geometric duality is an extension of the well-known duality of polytopes to $\mathcal{P}$ and $\mathcal{D}$. Recall that two polytopes $\mathcal{G}$ and $\mathcal{G}^*$ in $\mathbb{R}^p$ are said to be dual to each other provided there exists a one-to-one mapping Ψ between the set of all faces of $\mathcal{G}$ and the set of all faces of $\mathcal{G}^*$ such that Ψ is inclusion-reversing, i.e. faces $\mathcal{F}_1$ and $\mathcal{F}_2$ of $\mathcal{G}$ satisfy $\mathcal{F}_1 \subseteq \mathcal{F}_2$ if and only if the faces $\Psi(\mathcal{F}_1)$ and $\Psi(\mathcal{F}_2)$ satisfy $\Psi(\mathcal{F}_1) \supseteq \Psi(\mathcal{F}_2)$ (Grünbaum, 2003). The geometric duality theorem (Heyde and Löhne, 2006, Theorem 1) states that there is a similar duality relationship between $\mathcal{P}$ and $\mathcal{D}$.

To be more precise, we introduce the following notation. We consider the coupling function $\varphi : \mathbb{R}^p \times \mathbb{R}^p \to \mathbb{R}$, defined by

$$\varphi(y, v) := \sum_{i=1}^{p-1} y_i v_i + y_p \left(1 - \sum_{i=1}^{p-1} v_i\right) - v_p.$$

Note that $\varphi(\cdot, v)$ and $\varphi(y, \cdot)$ are affine. Choosing the values of the primal and dual objective functions for $x \in \mathcal{X}$ and $(u, \lambda) \in \mathcal{U}$ as arguments, we get

$$\varphi(Cx, D(u, \lambda)) = \lambda^T Cx - b^T u. \tag{5.1}$$

For given λ, $\varphi(Cx, D(u, \lambda))$ caculates the difference between the objective values of the weighted sum LP $\min\{\lambda^T Cx : Ax \geqq b\}$ and its corresponding dual LP. Thus x and u are optimal if and only if $\varphi(Cx, D(u, \lambda)) = 0$ by the complementary slackness condition of linear programming.

Using the coupling function φ, we define the following set-valued maps

$$\begin{aligned} \mathcal{H} : \mathbb{R}^p \rightrightarrows \mathbb{R}^p, \qquad & \mathcal{H}(v) := \{y \in \mathbb{R}^p : \varphi(y, v) = 0\}, \\ \mathcal{H}^* : \mathbb{R}^p \rightrightarrows \mathbb{R}^p, \qquad & \mathcal{H}^*(y) := \{v \in \mathbb{R}^p : \varphi(y, v) = 0\}. \end{aligned}$$

Of course, $\mathcal{H}(v)$ and $\mathcal{H}^*(y)$ are hyperplanes in $\mathbb{R}^p$ for all $v, y \in \mathbb{R}^p$. Using the notation

$$\begin{aligned} \lambda(v) &:= \left(v_1, \ldots, v_{p-1}, 1 - \sum_{i=1}^{p-1} v_i\right)^T \text{ and} \\ \lambda^*(y) &:= \big(y_1 - y_p, \ldots, y_{p-1} - y_p, -1\big)^T \end{aligned}$$

it is easy to see that

$$\begin{aligned} \mathcal{H}(v) &= \big\{y \in \mathbb{R}^p : \lambda(v)^T y = v_p\big\} \text{ and} \\ \mathcal{H}^*(y) &= \big\{v \in \mathbb{R}^p : \lambda^*(y)^T v = -y_p\big\}. \end{aligned}$$

We observe that $\lambda(v) \geqq 0$ if and only if $v_1, \ldots, v_{p-1} \geqq 0$ and $\sum_{i=1}^{p-1} v_i \leqq 1$, a fact we will often use.

The map $\mathcal{H}$ is now used to define our duality map $\Psi : 2^{\mathbb{R}^p} \to 2^{\mathbb{R}^p}$. Let $\mathcal{F}^* \subseteq \mathbb{R}^p$, then

$$\Psi(\mathcal{F}^*) := \bigcap_{v \in \mathcal{F}^*} \mathcal{H}(v) \cap \mathcal{P}.$$

The considerations in the following sections are based on the following geometric duality theorem.

Theorem 5.2.2 (Heyde and Löhne (2006))**.** *Ψ is an inclusion reversing one-to-one map between the set of all proper $\mathcal{K}$-maximal faces of $\mathcal{D}$ and the set of all proper weakly nondominated faces of $\mathcal{P}$ and the inverse map is given by*

$$\Psi^{-1}(\mathcal{F}) = \bigcap_{y \in \mathcal{F}} \mathcal{H}^*(y) \cap \mathcal{D}. \tag{5.2}$$

Moreover, for every proper $\mathcal{K}$-maximal face $\mathcal{F}^*$ of $\mathcal{D}$ it holds $\dim \mathcal{F}^* + \dim \Psi(\mathcal{F}^*) = p - 1$.

We next consider two important consequences.

Corollary 5.2.3 (Heyde and Löhne (2006))**.** *The following statements are equivalent*

(i) *v is a $\mathcal{K}$-maximal vertex of $\mathcal{D}$,*

(ii) *$\mathcal{H}(v) \cap \mathcal{P}$ is a weakly nondominated $(p-1)$-dimensional facet of $\mathcal{P}$.*

Moreover, if $\mathcal{F}$ is a weakly nondominated $(p-1)$-dimensional facet of $\mathcal{P}$, there is some uniquely defined point $v \in \mathbb{R}^p$ such that $\mathcal{F} = \mathcal{H}(v) \cap \mathcal{P}$.

Corollary 5.2.4 (Heyde and Löhne (2006))**.** *The following statements are equivalent*

(i) *y is a weakly nondominated vertex of $\mathcal{P}$,*

(ii) *$\mathcal{H}^*(y) \cap \mathcal{D}$ is a $\mathcal{K}$-maximal $(p-1)$-dimensional facet of $\mathcal{D}$.*

Moreover, if $\mathcal{F}^*$ is a $\mathcal{K}$-maximal $(p-1)$-dimensional facet of $\mathcal{D}$, there is some uniquely defined point $y \in \mathbb{R}^p$ such that $\mathcal{F}^* = \mathcal{H}^*(y) \cap \mathcal{D}$.

The proof of Theorem 5.2.2 in Heyde and Löhne (2006) is based on the consideration of the following two pairs of dual linear programming problems.

$(\mathrm{P}_1(v))$ $$\min_{x \in \mathcal{X}} \lambda(v)^T C x, \quad \mathcal{X} := \{x \in \mathbb{R}^n : Ax \geqq b\}$$

$$(\mathrm{D}_1(v)) \qquad \max_{u\in\mathcal{T}(v)} b^T u, \quad \mathcal{T}(v) := \left\{u \in \mathbb{R}^m : u \geqq 0, A^T u = C^T \lambda(v)\right\}$$

and

$$(\mathrm{P}_2(y)) \qquad \min_{x\in\mathcal{S}(y)} z, \quad \mathcal{S}(y) := \{(x,z) \in \mathbb{R}^n \times \mathbb{R} : Ax \geqq b, Cx - ez \leqq y\},$$

$$(\mathrm{D}_2(y)) \qquad \max_{(u,\lambda)\in\mathcal{U}} (b^T u - y^T \lambda), \quad \mathcal{U} := \left\{(u,\lambda) \in \mathbb{R}^m \times \mathbb{R}^p : (u,\lambda) \geqq 0, A^T u = C^T \lambda, e^T \lambda = 1\right\}.$$

Note that $(\mathrm{P}_2(y))$ and $(\mathrm{D}_2(y))$ have been used before in Chapter 4 and that the other two $(\mathrm{P}_1(v))$ and $(\mathrm{D}_1(v))$ will be essential for the dual variant of Benson's algorithm to be introduced in Section 5.4. Figure 5.2 illustrates the linear programming duality between $(\mathrm{P}_1(v))$ and $(\mathrm{D}_1(v))$. For any $v \in \mathbb{R}^p$ such that $\lambda(v) \geqq 0$ we have that $x \in \mathcal{X}$ and $u \in \mathcal{T}(v)$ are optimal solutions of $(\mathrm{P}_1(v))$ and $(\mathrm{D}_1(v))$, respectively, if and only if $\varphi(Cx, D(u, \lambda(v)) = 0$ in (5.1). Thus, $(v_1, \ldots, v_{p-1}, \lambda(v)^T Cx) = (v_1, \ldots, v_{p-1}, b^T u)$ is a boundary point of $D(\mathcal{U}) - \mathcal{K}$. Feasible values of $(\mathrm{P}_1(v))$ are "above" that point, feasible values of $(\mathrm{D}_1(v))$ are "below": For feasible $x \in \mathcal{X}$ and $u \in \mathcal{T}(v)$ the value of $\varphi(Cx, D(u, \lambda(v))$ measures the duality gap between the two feasible solutions.

The four problems above play a key role in the following algorithms.

5.3 Extension of Benson's Outer Approximation Algorithm

We propose an algorithm, which is essentially Benson's algorithm (see Chapter 3), but involves some slight improvements. We see that it is not necessary to work with bounded simplices as Benson did in the original version. Thus, we compute the nondominated vertices directly and the final step (Benson, 1998c, Theorem 3.2) to check whether a vertex is nondominated or not is superfluous.

In our primal vector optimization problem (P) we assume that the set $\mathcal{P} = C(\mathcal{X}) + \mathbb{R}^p_{\geqq}$ is $\mathbb{R}^p_{\geqq}$-bounded from below, i.e., there exists some $\hat{y} \in \mathbb{R}^p$ such that $\hat{y} \leqq y$ for all $y \in \mathcal{P}$. As a consequence, the ideal point y^I of $\mathcal{P}$ defined by $y^I_k :=$

$\min\{y_k : y \in \mathcal{P}\}$ for $i = 1, \ldots, p$ exists. Of course, this assumption is weaker than the assumption that $\mathcal{X}$ is a bounded set, which is supposed in Benson (1998c). We also assume that $\mathcal{X}$ is nonempty.

The algorithm first constructs a p-dimensional polyhedral set $\mathcal{S}^0 = y^I + \mathbb{R}^p_{\geqq}$ such that $\mathcal{P} \subseteq \mathcal{S}^0$. In every iteration it chooses an extreme point s^k of $\mathcal{S}^{k-1}$ not contained in $\mathcal{P}$ and constructs a supporting hyperplane to $\mathcal{P}$ by solving a linear programme $(\mathrm{D}_2(y^k))$, where y^k is a boundary point of $\mathcal{P}$ on the line segment connecting s^k with an interior point $\hat{p}$ of $\mathcal{P}$. $\mathcal{S}^k$ is defined by intersecting $\mathcal{S}^{k-1}$ with the halfspace of the hyperplane containing $\mathcal{P}$. The algorithm terminates as soon as no such s^k can be found and $\mathcal{S}^{k-1} = \mathcal{P}$.

For the next result we need Lemma 4.1 from Heyde and Löhne (2006).

Lemma 5.3.1. *The following three statements are equivalent.*

(i) $y^0 \in \mathrm{wmin}_{\mathbb{R}^p_{\geqq}} \mathcal{P}$.

(ii) *There is some* $x^0 \in \mathbb{R}^n$ *such that* $(x^0, 0)$ *is an optimal solution to* $(\mathrm{P}_2(y^0))$.

(iii) *There is some* $(u^0, \lambda^0) \in \mathcal{U}$ *with* $b^T u^0 = {y^0}^T \lambda^0$ *that is an optimal solution to* $(\mathrm{D}_2(y^0))$.

Proposition 5.3.2. *Let* $y \in \mathrm{wmin}\mathcal{P}$. *Then there exists an optimal solution of* $(\mathrm{D}_2(y))$ *and for each such solution* $(\bar{u}, \bar{\lambda}) \in \mathcal{U}$, $\mathcal{H}(D(\bar{u}, \bar{\lambda}))$ *is a supporting hyperplane of* $\mathcal{P}$ *with* $y \in \mathcal{H}(D(\bar{u}, \bar{\lambda}))$.

Proof. By Lemma 5.3.1 there exists an optimal solution $(\bar{u}, \bar{\lambda})$ of $(\mathrm{D}_2(y))$ such that $b^T \bar{u} = y^T \bar{\lambda}$. Of course, the latter equality is also valid for any other optimal solution of $(\mathrm{D}_2(y))$. For arbitrary $y \in \mathcal{P}$, there exists some $x \in \mathcal{X}$ such that $y \geqq Cx$. Hence $(x, 0)$ is feasible for $(\mathrm{P}_2(y))$ and duality between $(\mathrm{P}_2(y))$ and $(\mathrm{D}_2(y))$ implies that $\bar{\lambda}^T y \geqq b^T \bar{u}$. Hence $\mathcal{H}(D(\bar{u}, \bar{\lambda})) = \left\{y \in \mathbb{R}^p : \bar{\lambda}^T y = b^T \bar{u}\right\}$ is a supporting hyperplane to $\mathcal{P}$. □

We note that Benson (1998c,a) proves similar results to Lemma 5.3.1 and Proposition 5.3.2 for his original algorithm.

Proposition 5.3.3. *Every vertex of $\mathcal{P}$ is nondominated ($\mathbb{R}^p_{\leqq}$-minimal).*

Proof. Let y be a vertex of $\mathcal{P} = C(\mathcal{X}) + \mathbb{R}^p_{\leqq}$ and assume that y is not $\mathbb{R}^p_{\leqq}$-minimal. Hence, there exists some $z \in \big(\{y\} - \mathbb{R}^p_{\leqq} \setminus \{0\}\big) \cap \mathcal{P}$, i.e., $y \in \{z\} + \mathbb{R}^p_{\leqq} \setminus \{0\} \subseteq C(\mathcal{X}) + \mathbb{R}^p_{\leqq} + \big(\mathbb{R}^p_{\leqq} \setminus \{0\}\big) = C(\mathcal{X}) + \mathbb{R}^p_{\leqq} \setminus \{0\}$. Therefore, there is some $\bar{x} \in \mathcal{X}$ and some $\bar{d} \in \mathbb{R}^p_{\leqq} \setminus \{0\}$ such that $y = C\bar{x} + \bar{d} \in \mathcal{P}$. Hence the points $y - \bar{d}$ and $y + \bar{d}$ belong to $\mathcal{P}$ and $y = \frac{1}{2}(y - \bar{d}) + \frac{1}{2}(y + \bar{d})$. This contradicts y being a vertex of $\mathcal{P}$. □

The following Proposition 5.3.4 shows that we do not need to consider extreme directions but only the vertices (extreme points) in the following algorithm, because the extreme directions are always the same, namely the unit vectors $e^k \in \mathbb{R}^p$.

Proposition 5.3.4. *Let $y \in \mathbb{R}^p$ and let $\mathcal{S} \subseteq \mathbb{R}^p$ be a polyhedral convex set such that $\mathcal{P} \subseteq \mathcal{S} \subseteq \{y\} + \mathbb{R}^p_{\leqq}$. Letting $\mathcal{E}$ be the set of extreme points of $\mathcal{S}$, we have $\mathcal{S} = \operatorname{conv}(\mathcal{E} + \mathbb{R}^p_{\leqq})$.*

Proof. Since $\mathcal{S}$ and $\mathcal{P}$ are closed and convex, we get $\mathbb{R}^p_{\leqq} \subseteq \mathcal{P}_\infty \subseteq \mathcal{S}_\infty \subseteq \mathbb{R}^p_{\leqq}$, hence $\mathcal{S}_\infty = \mathbb{R}^p_{\leqq}$. Now the conclusion follows from (Rockafellar, 1970, Theorem 18.5 and Theorem 19.5). □

In the following algorithm we construct in iteration k a polyhedron $\mathcal{S}^k$, for which we store both a representation by a finite number of points and a representation by a finite number of inequalities. We cannot always ensure that all the points representing the set $\mathcal{S}^k$ are extreme points, i.e., our set may contain some redundant points. Similarly, it may happen that we have redundant inequalities in the inequality representation of $\mathcal{S}^k$. Therefore, we say that a set $\mathcal{E}$ of finitely many points is a *point representation* of $\mathcal{S}$ if $\mathcal{S} = \operatorname{conv}(\mathcal{E} + \mathbb{R}^p_{\leqq})$. If $\mathcal{E}$ only consists of extreme points of $\mathcal{S}$, we say that $\mathcal{E}$ is *nondegenerate*. Otherwise, $\mathcal{E}$ is called *degenerate*.

Analogously, a system of inequalities is called a *nondegenerate inequality representation* of $\mathcal{S}$ if $\mathcal{S}$ is the solution set of the system and if there are no redundant inequalities. An inequality representation of $\mathcal{S}$ is called *degenerate* if there exist redundant inequalities, i.e., there exists a proper subsystem of inequalities having $\mathcal{S}$ as the solution set.

Algorithm 5.3.5.

Initialization ($k = 0$)**.**

(i1) *Choose some* $\hat{p} \in \operatorname{int} \mathcal{P}$.

(i2) *Compute an optimal solution* $\bar{u}_i$ *and the optimal value* y_i^I *of* $(D_1(e^i))$*, for* $i = 1, \ldots, p$.

(i3) *Set* $\mathcal{S}^0 := \{y^I\} + \mathbb{R}^p_{\leqq}$ *and* $k = 1$.

Iteration steps ($k \geqq 1$)**.**

(k1) *If* $\operatorname{vert} \mathcal{S}^{k-1} \subseteq \mathcal{P}$ *stop, otherwise choose a vertex* s^k *of* $\mathcal{S}^{k-1}$ *such that* $s^k \notin \mathcal{P}$.

(k2) *Compute* $\rho^k \in (0,1)$ *such that* $y^k := \rho^k s^k + (1-\rho^k)\hat{p} \in wmin_{\mathbb{R}^p_{\leqq}} \mathcal{P}$.

(k3) *Compute an optimal solution* (u^k, λ^k) *of* $(D_2(y^k))$.

(k4) *Set* $\mathcal{S}^k := \mathcal{S}^{k-1} \cap \{y \in \mathbb{R}^p : \varphi\big(y, D(u^k, \lambda^k)\big) \geqq 0\}$.

(k5) *Set* $k := k+1$ *and go to (k1).*

Results.

(r1) *The set of* $\mathbb{R}^p_{\leqq}$*-minimal vertices of* $\mathcal{P}$ *is* $\operatorname{vert} \mathcal{S}^{k-1}$. *Moreover* $\mathcal{S}^{k-1} = \mathcal{P}$.

(r2) *The set* $\{v \in \mathbb{R}^p : \lambda(v) \geqq 0, \varphi(y,v) \geqq 0$ *for all* $y \in \operatorname{vert} \mathcal{S}^{k-1}\}$ *is defined by a nondegenerate inequality representation of* $\mathcal{D}$.

(r3) *All* $\mathcal{K}$*-maximal vertices of* $\mathcal{D}$ *are contained in the set* $\mathcal{V} := \{D(\bar{u}^1, e^1), D(\bar{u}^2, e^2), \ldots, D(\bar{u}^p, e^p), D(u^1, \lambda^1), \ldots, D(u^{k-1}, \lambda^{k-1})\}$.

(r4) *The set* $\{y \in \mathbb{R}^p : \varphi(y,v) \geqq 0$ *for all* $v \in \mathcal{V}\}$ *is given by a (possibly degenerate) inequality representation of* $\mathcal{P}$.

Details of Algorithm 5.3.5.

(i1) It is obvious that $\operatorname{int} \mathcal{P} \neq \emptyset$. For instance, $Cx + \rho e \in \operatorname{int} \mathcal{P}$ for arbitrary $x \in \mathcal{X}$ and $\rho > 0$.

(i2) Of course, $(\mathrm{D}_1(e^i))$ has an optimal solution because $(\mathrm{P}_1(e^i))$ is bounded.

(i3) From the definition of the ideal point we directly obtain that $\mathcal{S}^0 \supseteq \mathcal{P}$.

(k1) Let $y \in \operatorname{vert} \mathcal{S}^{k-1}$. By computing the optimal value μ of $(\mathrm{P}_2(y))$ or $(\mathrm{D}_2(y))$ it is possible to decide whether $y \in \mathcal{P}$ or not. We have $y \in \mathcal{P}$ if and only if $\mu = 0$.

(k2) Solve the linear programme

$$\rho^k := \max\{\rho : x \in \mathcal{X}, \rho s^k + (1-\rho)\hat{p} \geqq Cx\}. \tag{5.3}$$

Of course, for every $\bar{x} \in \mathcal{X}$, $(\bar{x}, 0)$ is feasible for (5.3). Since $\mathcal{X}$ is nonempty, there exists an optimal solution of (5.3). From $s^k \notin \mathcal{P}$ and $\hat{p} \in \operatorname{int} \mathcal{P}$ we conclude that $y^k \in \operatorname{bd} \mathcal{P}$ and $\rho^k \in (0,1)$. Moreover, we have $\operatorname{bd} \mathcal{P} = \operatorname{wmin}_{\mathbb{R}^p_{\leqq}} \mathcal{P}$ (see e.g. Heyde *et al.* (2007)).

(k3) By Proposition 5.3.2, there exists an optimal solution.

(k4) By Proposition 5.3.2, $\mathcal{H}(D(\bar{u}, \bar{\lambda}))$ is a supporting hyperplane to $\mathcal{P}$ containing y^k. This means, $\varphi\big(y, D(u^k, \lambda^k)\big) \geqq 0$ for all $y \in \mathcal{P}$ and $\varphi\big(y^k, D(u^k, \lambda^k)\big) = 0$. Hence we get $\mathcal{P} \subseteq \mathcal{S}^k \subseteq \mathcal{S}^{k-1}$.

(r1) From (k1) we get $\operatorname{vert} \mathcal{S}^{k-1} \subseteq \mathcal{P}$. By Proposition 5.3.4 we obtain $\mathcal{S}^{k-1} = \operatorname{conv}(\operatorname{vert} \mathcal{S}^{k-1} + \mathbb{R}^p_{\leqq}) \subseteq \mathcal{P}$. As shown in (k4) we have $\mathcal{P} \subseteq \mathcal{S}^{k-1}$. Together we have $\mathcal{P} = \mathcal{S}^{k-1}$. By Proposition 5.3.3 the statement follows.

(r2) By Corollary 5.2.4, $\mathcal{F}^*$ is a $\mathcal{K}$-maximal $(p-1)$-dimensional facet of $\mathcal{D}$ if and only if there exists some $\mathbb{R}^p_{\leqq}$-minimal vertex y of $\mathcal{P}$ such that $\mathcal{F}^* = \mathcal{H}^*(y) \cap \mathcal{D}$. Hence a hyperplane $\mathcal{H}^*(y)$ supports $\mathcal{D}$ in a facet if and only if y is a $\mathbb{R}^p_{\leqq}$-minimal vertex of $\mathcal{P}$. Of course, the corresponding inequalities are not redundant.

(r3) Let v be a $\mathcal{K}$-maximal vertex of $\mathcal{D}$. By Corollary 5.2.3, $\mathcal{F} := \mathcal{H}(v) \cap \mathcal{P}$ is a $\mathbb{R}^p_{\leqq}$-minimal $(p-1)$-dimensional facet of $\mathcal{P}$. Since $\mathcal{S}^{k-1} = \mathcal{P}$ and by the construction of $\mathcal{S}^{k-1}$, for every weakly $\mathbb{R}^p_{\leqq}$-minimal facet $\mathcal{F}$ of $\mathcal{P}$ there exists some $i \in \{0, \ldots, k-1\}$ such that $\mathcal{F} = \mathcal{H}(D(u^i, \lambda^i)) \cap \mathcal{P}$. By Corollary 5.2.3, we get $D(u^i, \lambda^i) = v$.

(r4) This follows from (r3) by the geometric duality theorem.

In Example 5.3.6 we demonstrate the occurrence of degeneracy.

Example 5.3.6. *Consider problem (P) with the data*

$$C = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad A = \begin{pmatrix} 2 & 1 \\ 1 & 2 \\ 3 & 3 \\ 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad b = \begin{pmatrix} 2 \\ 2 \\ 4 \\ 0 \\ 0 \end{pmatrix}.$$

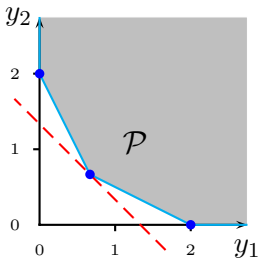

Figure 5.3: *$\mathcal{P}$ and the first supporting hyperplane.*

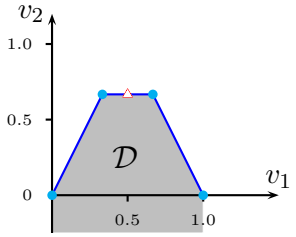

Figure 5.4: *$\mathcal{D}$ and a point in the relative interior of a facet.*

We apply Algorithm 5.3.5 for the choice $\hat{p} = (1,1)^T$. In the initialization ($k = 0$) we obtain $y^I = (0,0)^T$ and $\mathcal{S}^0 = \mathbb{R}^p_{\geqq}$. In the first iteration ($k = 1$) we get $y^1 = (\frac{2}{3}, \frac{2}{3})^T$. We have to solve ($D_2(y^1)$). This problem has three optimal extreme point solutions, namely $(u^1, \lambda^1)^T = (0, \frac{1}{3}, 0, 0, 0, \frac{1}{3}, \frac{2}{3})^T$, $(u^1, \lambda^1)^T = (0, 0, \frac{1}{6}, 0, 0, \frac{1}{2}, \frac{1}{2})^T$ and $(u^1, \lambda^1)^T = (\frac{1}{3}, 0, 0, 0, 0, \frac{2}{3}, \frac{1}{3})^T$. If we choose the second one, we get the redundant inequality $3y_1 + 3y_2 \geqq 4$. The corresponding hyperplane supports $\mathcal{P}$ not in a facet, but just in the vertex y^1, see Figure 5.3. Also, for the choice $(u^1, \lambda^1)^T = (0, 0, \frac{1}{6}, 0, 0, \frac{1}{2}, \frac{1}{2})^T$, the point $D(u^1, \lambda^1)$ is not a vertex of $\mathcal{D}$, see Figure 5.4. This means, Algorithm 5.3.5 yields a degenerate inequality representation of $\mathcal{P}$ and a degenerate point representation of $\mathcal{D}$.

Finally we show the finiteness of the modified algorithm in the same way as in the original variant of Benson (1998c).

Theorem 5.3.7. *The modified outer approximation algorithm is finite.*

Proof. Since $\hat{p} \in \operatorname{int} \mathcal{P}$, the point $y^k \in \mathcal{P}$ computed in iteration k belongs to $\operatorname{int} \mathcal{S}^{k-1}$. We have $\mathcal{S}^k := \mathcal{S}^{k-1} \cap \{y \in \mathbb{R}^p : \varphi\big(y, D(u^k, \lambda^k)\big) \geqq 0\}$ and by Proposition 5.3.2 we know that $\mathcal{F} := \{y \in \mathcal{P} : \varphi\big(y, D(u^k, \lambda^k)\big) = 0\}$ is a face of $\mathcal{P}$ with $y^k \in \mathcal{F}$, where $\mathcal{F} \subseteq \operatorname{bd} \mathcal{S}^k$. This means for the next iteration that $y^{k+1} \notin \mathcal{F}$ (because $y^{k+1} \in \operatorname{int} \mathcal{S}^k$), and therefore y^{k+1} belongs to another face of $\mathcal{P}$. Since $\mathcal{P}$ is polyhedral, it has a finite number of faces, hence the algorithm is finite. □

5.4 The Dual Variant of Benson's Algorithm

As in the previous section we assume that the primal feasible set $\mathcal{X}$ of problem (P) is nonempty and $\mathcal{P}$ is $\mathbb{R}^p_{\geqq}$-bounded from below.

The dual variant of Benson's algorithm first constructs a p-dimensional polyhedral set $\mathcal{S}^0 = \{v \in \mathbb{R}^p : \lambda(v) \geqq 0, \varphi(Cx^0, v) \geqq 0\}$ such that $\mathcal{D} \subseteq \mathcal{S}^0$. Here x^0 is an optimal solution of $(\mathrm{P}_1(\hat{d}))$ for an interior point $\hat{d}$ of $\mathcal{D}$. In every iteration it chooses a vertex s^k of $\mathcal{S}^{k-1}$ not contained in $\mathcal{D}$ and constructs a supporting hyperplane to $\mathcal{D}$ by solving the linear programme $(\mathrm{P}_1(v^k))$, where v^k is a boundary point of $\mathcal{D}$ on the line segment connecting s^k with the interior point $\hat{d}$ of $\mathcal{D}$. $\mathcal{S}^k$ is defined by intersecting $\mathcal{S}^{k-1}$ with the halfspace of the hyperplane containing $\mathcal{D}$ until at termination $\mathcal{S}^{k-1} = \mathcal{D}$.

Proposition 5.4.1. *Let $\bar{v} \in \max_{\mathcal{K}} \mathcal{D}$, then for every solution $\bar{x}$ of $(\mathrm{P}_1(\bar{v}))$, $\mathcal{H}^*(P\bar{x})$ is a supporting hyperplane of $\mathcal{D}$ with $\bar{v} \in \mathcal{H}^*(C\bar{x})$.*

Proof. Let $v \in \mathcal{D}$, i.e., there is some u such that $(u, \lambda(v)) \in \mathcal{U}$ and $v_p \leqq b^T u$. From the weak duality between $(\mathrm{P}_1(v))$ and $(\mathrm{D}_1(v))$ we get that $\lambda(v)^T C\bar{x} \geqq b^T u \geqq v_p$, or equivalently, $\varphi(C\bar{x}, v) \geqq 0$. For $\bar{v} \in \max_{\mathcal{K}} \mathcal{D}$ we get an optimal solution $\bar{u}$ of $(\mathrm{D}_1(\bar{v}))$, and strong duality between $(\mathrm{P}_1(\bar{v}))$ and $(\mathrm{D}_1(\bar{v}))$ implies that $\varphi(C\bar{x}, \bar{v}) = 0$. The result follows from the definition of $\mathcal{H}^*(C\bar{x}) = \{v \in \mathbb{R}^p : \varphi(C\bar{x}, v) = 0\}$. □

Proposition 5.4.2. *Every vertex of $\mathcal{D}$ is $\mathcal{K}$-maximal.*

Proof. Assume there is some vertex $\bar{v} \in \mathcal{D}$ which is not $\mathcal{K}$-maximal. Then there exists some $v \in \bar{v} + \mathcal{K} \cap \mathcal{D}$ with $v \neq \bar{v}$. We get $\bar{v} = \frac{1}{2}v + \frac{1}{2}(\bar{v} - (v - \bar{v}))$, where $v \in \mathcal{D}$

and $(\bar{v}-(v-\bar{v})) \in \mathcal{D}$ are not equal to $\bar{v}$. This contradicts the fact that $\bar{v}$ is a vertex. □

Note that Propositions 5.4.1 and 5.4.2 are the dual variants of Propositions 5.3.2 and 5.3.3.

Similar to Proposition 5.3.4, we can represent the polyhedra approximating the set $\mathcal{D}$ from outside in the following algorithm by a finite number of (extreme) points because the (extreme) directions are always the same.

Proposition 5.4.3. *Let $y \in \mathbb{R}^p$ and let $\mathcal{S} \subseteq \mathbb{R}^p$ be a polyhedral convex set such that $\mathcal{D} \subseteq \mathcal{S} \subseteq \{v \in \mathbb{R}^p : \lambda(v) \geqq 0, \varphi(y,v) \geqq 0\}$. Letting $\mathcal{E}$ be the set of extreme points of $\mathcal{S}$, we have $\mathcal{S} = \operatorname{conv}(\mathcal{E} - \mathcal{K})$.*

Proof. Setting $\mathcal{W} := \{v \in \mathbb{R}^p : \lambda(v) \geqq 0,\ \varphi(y,v) \geqq 0\}$, we have $v \in \mathcal{W}$ if and only if

$$v_1 \geqq 0, \ldots, v_{p-1} \geqq 0, \sum_{i=1}^{p-1} v_i \leqq 1, \lambda^*(y)^T v \geqq -y_p,$$

where the last component of $\lambda^*(y)$ is -1. It follows that $\mathcal{W}_\infty = -\mathcal{K}$. Since $\mathcal{S}$ is closed and convex, we get $-\mathcal{K} \subseteq \mathcal{D}_\infty \subseteq \mathcal{S}_\infty \subseteq -\mathcal{K}$, hence $\mathcal{S}_\infty = -\mathcal{K}$. Now the conclusion follows from (Rockafellar, 1970, Theorem 18.5 and Theorem 19.5.) □

A set $\mathcal{E}$ of finitely many points in $\mathbb{R}^p$ is called a *point representation* of $\mathcal{D}$ if $\mathcal{D} = \operatorname{conv}(\mathcal{E} - \mathcal{K})$. The same notation is used for sets $\mathcal{S}^k$ constructed during the algorithm. Again, we speak about nondegenerate and degenerate point representations depending on whether $\mathcal{E}$ only consists of extreme points of $\mathcal{D}$ or not. With this notation we can say that in the result (r3) of Algorithm 5.3.5 we get a (possibly degenerate) point representation of $\mathcal{D}$.

We propose the following algorithm, subsequently called the *dual variant of Benson's algorithm.*

Algorithm 5.4.4 (Dual variant of Benson's algorithm)**.**

Initialization ($k = 0$)**.**

(i1) *Choose some* $\hat{d} \in \operatorname{int} \mathcal{D}$.

(i2) *Compute an optimal solution* x^0 *of* $(P_1(\hat{d}))$.

(i3) *Set* $S^0 = \{v \in \mathbb{R}^p : \lambda(v) \geqq 0, \varphi(Cx^0, v) \geqq 0\}$ *and* $k = 1$.

Iteration steps ($k \geqq 1$)**.**

(k1) *If* $\operatorname{vert} S^{k-1} \subseteq \mathcal{D}$ *stop, otherwise choose a vertex* s^k *of* $\mathcal{S}^{k-1}$ *such that* $s^k \notin \mathcal{D}$.

(k2) *Compute* $\rho^k \in (0,1)$ *such that* $v^k := \rho^k s^k + (1-\rho^k)\hat{d} \in \max_{\mathcal{K}} \mathcal{D}$.

(k3) *Compute an optimal solution* x^k *of* $(P_1(v^k))$.

(k4) *Set* $S^k := S^{k-1} \cap \{v \in \mathbb{R}^p : \varphi(Cx^k, v) \geqq 0\}$.

(k5) *Set* $k := k + 1$ *and go to (k1).*

Results.

(r1) *The set of* $\mathcal{K}$*-maximal vertices of* $\mathcal{D}$ *is* $\operatorname{vert} \mathcal{S}^{k-1}$.

(r2) *The set* $\left\{y \in \mathbb{R}^p : \varphi(y, v) \geqq 0 \text{ for all } v \in \operatorname{vert} \mathcal{S}^{k-1}\right\}$ *is given by a nondegenerate inequality representation of* $\mathcal{P}$.

(r3) *All* $\mathbb{R}^p_{\leqq}$*-minimal (nondominated) vertices of* $\mathcal{P}$ *are contained in the set* $\mathcal{Y} := \{Cx^0, Cx^1, \ldots, Cx^{k-1}\}$.

(r4) *The set* $\{v \in \mathbb{R}^p : \lambda(v) \geqq 0, \varphi(y, v) \geqq 0 \text{ for all } y \in \mathcal{Y}\}$ *is given by a (possibly degenerate) inequality representation of* $\mathcal{D}$.

Details of Algorithm 5.4.4.

(i1) We show that $\operatorname{int} \mathcal{D} \neq \emptyset$. Since $\mathcal{X}$ is assumed to be nonempty and $\mathcal{P}$ is $\mathbb{R}^p_{\leqq}$-bounded from below, $(\mathrm{P}_1(v))$ has an optimal solution for every $v \in \mathbb{R}^p$ with $\lambda(v) \geqq 0$. By duality, the same is true for $(\mathrm{D}_1(v))$. Denote by $\hat{\gamma}^i$ the optimal value of $(\mathrm{D}_1(e^i))$. Furthermore, set $\hat{\gamma} = \min\{\hat{\gamma}^i : i \in \{1, \ldots, p\}\}$. Then $\hat{\gamma}$ is a lower bound for the optimal values of the problems $(\mathrm{D}_1(v))$ whenever $\lambda(v) \geqq 0$. From the definition of $\mathcal{D}$ we easily obtain

$$\mathcal{D} = \left\{v \in \mathbb{R}^p : \lambda(v) \geqq 0, A^T u = C^T \lambda(v) \text{ and } v_p \leqq b^T u \text{ for some } u \geqq 0\right\}.$$

Hence

$$\mathcal{D} \supseteq \{v \in \mathbb{R}^p : \lambda(v) \geqq 0, v_p \leqq \hat{\gamma}\},$$

which shows that $\operatorname{int} \mathcal{D}$ is nonempty. One possible choice for the point $\hat{d} \in \operatorname{int} \mathcal{D}$ is $\hat{d} = \left(\frac{1}{p}, \ldots, \frac{1}{p}, \hat{\gamma} - 1\right)^T$.

(i2) Since $\mathcal{X}$ is assumed to be nonempty, $\mathcal{P}$ is $\mathbb{R}^p_{\leqq}$-bounded from below and $\lambda(\hat{d}) \geqq 0$, $(\mathrm{P}_1(\hat{d}))$ has an optimal solution.

(i3) It holds that $\mathcal{S}^0 \supseteq \mathcal{D}$. It remains to show that

$$\left\{v \in \mathbb{R}^p : A^T u = C^T \lambda(v), v_p \leqq b^T u \text{ for some } u \geqq 0\right\} \subseteq \left\{v \in \mathbb{R}^p : \varphi(Cx^0, v) \geqq 0\right\}.$$

This follows from weak duality between $(\mathrm{P}_1(v))$ and $(\mathrm{D}_1(v))$ as in the proof of Proposition 5.4.1.

(k1) Compute the optimal value μ of $(\mathrm{P}_1(s^k))$ in order to decide whether s^k belongs to $\mathcal{D}$. We have $s^k \in \mathcal{D}$ if and only if $s^k_p \leqq \mu$.

(k2) Solve the linear programme

$$\rho^k := \max\{\rho : (u, \lambda) \in \mathcal{U}, \rho s^k + (1 - \rho)\hat{d} = D(u, \lambda)\}. \tag{5.4}$$

The existence of an optimal solution of the LP (5.4) can be shown as follows. If there is some $(\bar{u}, \bar{\lambda}) \in \mathcal{U}$ such that $\hat{d} = D(\bar{u}, \bar{\lambda})$, then $(\bar{u}, \bar{\lambda}, 0)$ is feasible for problem (5.4). Otherwise, we have $\hat{d} \notin D(\mathcal{U})$. Let w^i be the i-th unit vector, but let the last component be replaced by $\hat{\gamma}^i$. It is easy to verify that there is some $\bar{\rho} \in (0, 1)$ such that $\bar{\rho} s^k + (1 - \bar{\rho})\hat{d} \in \operatorname{conv}\{w^1, \ldots, w^p\} \subseteq D(\mathcal{U})$, i.e., there is some $(\bar{u}, \bar{\lambda}) \in \mathcal{U}$ such that $(\bar{u}, \bar{\lambda}, \bar{\rho})$ is feasible for problem (5.4). Furthermore, we have $\bar{\rho} < 1$ because otherwise we obtain $s^k \in D(\mathcal{U}) \subseteq \mathcal{D}$, a contradiction.

(k3) Since $\mathcal{X}$ is nonempty and $C(\mathcal{X})$ is $\mathbb{R}^p_{\leqq}$-bounded from below, there exists an optimal solution.

(k4) Analogously to (i3) above we get $\mathcal{S}^k \supseteq \mathcal{D}$.

(r1) From (k1) we get $\operatorname{vert}\mathcal{S}^{k-1} \subseteq \mathcal{D}$. By Proposition 5.4.3 we obtain $\mathcal{S}^{k-1} = \operatorname{conv}(\operatorname{vert}\mathcal{S}^{k-1} - \mathcal{K}) \subseteq \mathcal{D}$. As shown in (k4) we have $\mathcal{D} \subseteq \mathcal{S}^{k-1}$. Together we have $\mathcal{D} = \mathcal{S}^{k-1}$. By Proposition 5.4.2 the statement follows.

(r2) By Corollary 5.2.3, $\mathcal{F}$ is a weakly nondominated $(p-1)$-dimensional facet of $\mathcal{P}$ if and only if there exists some $\mathcal{K}$-maximal vertex v of $\mathcal{D}$ such that $\mathcal{F} = \mathcal{H}(v) \cap \mathcal{P}$. Hence a hyperplane $\mathcal{H}(v)$ supports $\mathcal{P}$ in a facet if and only if v is a $\mathcal{K}$-maximal vertex of $\mathcal{D}$. Thus we have a nondegenerate inequality representation of $\mathcal{P}$.

(r3) Let y be a $\mathbb{R}^p_{\leqq}$-minimal vertex of $\mathcal{P}$. By Corollary 5.2.4, $\mathcal{F}^* := \mathcal{H}^*(y) \cap \mathcal{D}$ is a $\mathcal{K}$-maximal $(p-1)$-dimensional facet of $\mathcal{D}$. Since $\mathcal{S}^{k-1} = \mathcal{D}$ and by the construction of $\mathcal{S}^{k-1}$, for every $\mathcal{K}$-maximal facet $\mathcal{F}^*$ of $\mathcal{D}$ there exists some $i \in \{0, \ldots, k-1\}$ such that $\mathcal{F}^* = \mathcal{H}^*(Cx^i) \cap \mathcal{D}$. By Corollary 5.2.4, we get $Cx^i = y$.

(r4) This follows from (r3) by the geometric duality theorem.

Remark 5.4.5. If $\hat{d} \in \operatorname{int}\mathcal{D}$ such that $\hat{d}_p \leqq \hat{\gamma}$, problem (5.4) is equivalent to the following one having $p+1$ fewer variables and $2p+1$ fewer constraints,

$$\rho^k := \max\left\{ \frac{b^T u - \hat{d}_p}{s^k_p - \hat{d}_p} : u \geqq 0, \tilde{A}u = \tilde{b} \right\}. \tag{5.5}$$

where

$$\tilde{A} := (s^k_p - \hat{d}_p)A^T + C^T\big(\lambda(\hat{d}) - \lambda(s^k)\big)b^T \text{ and } \tilde{b} := C^T\big(s^k_p\lambda(\hat{d}) - \hat{d}_p\lambda(s^k)\big).$$

This equivalence of problem (5.4) and problem (5.5) can be shown in a straightforward way taking into account that $\hat{d}_p \leqq \hat{\gamma}$ implies $\hat{d}_p < s^k_p$.

Example 5.4.6 illustrates the occurrence of degenerate representations of $\mathcal{P}$ and $\mathcal{D}$.

Example 5.4.6. *Consider problem (P) with the data*

$$C = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}, \quad A = \begin{pmatrix} 7 & 21 & 9 \\ 0 & 0 & -1 \\ -7 & -42 & 3 \\ 1 & 7 & 0 \end{pmatrix}, \quad b = \begin{pmatrix} 30 \\ -1 \\ -39 \\ 6 \end{pmatrix}.$$

We apply Algorithm 5.4.4 for the choice $\hat{d} = (\frac{1}{2}, 0)^T$. *In the initialization step* $(k = 0)$ *we solve* $P_1(\hat{d})$ *and obtain the unique optimal solution* $x^0 = (0, 1, 1)^T$. *Hence*

$$\mathcal{S}^0 = \{v \in \mathbb{R}^p : 0 \leqq v_1 \leqq 1,\ v_2 \leqq 1 - v_1\}.$$

There is exactly one vertex of $\mathcal{S}^0$, *namely* $s^1 = (0, 1)^T$, *that does not belong to* $\mathcal{D}$. *Step (k2) yields* $v^1 = (\frac{1}{8}, \frac{3}{4})^T$. *Note that we have the situation that* v^1 *is not in the relative interior of a facet of* $\mathcal{D}$, *because it is a vertex of* $\mathcal{D}$. *In step (k3) we solve* $P_1(v^1)$. *We have exactly three extreme point optimal solutions of* $P_1(v^1)$, *namely* $x^1 = (\frac{3}{4}, \frac{3}{4}, 1)^T$, $x^1 = (3, \frac{3}{7}, 0)^T$ *and* $x^1 = (6, 0, 1)^T$. *In case we choose the second one, we get the redundant inequality* $-\frac{18}{7}v_1 + v_2 \leqq \frac{3}{7}$. *The corresponding hyperplane supports* $\mathcal{D}$ *not in a facet, but just in the vertex* v_1 *(see Figure 5.6). For the choice* $x^1 = (3, \frac{3}{7}, 0)^T$, *the point* Cx^1 *is not a vertex of* $\mathcal{P}$ *(see Figure 5.5). This means, Algorithm 5.4.4 yields a degenerate inequality representation of* $\mathcal{D}$ *and a degenerate point representation of* $\mathcal{P}$.

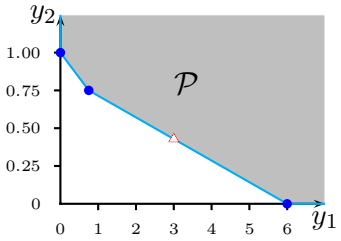

Figure 5.5: *$\mathcal{P}$ and a point in the relative interior of a facet.*

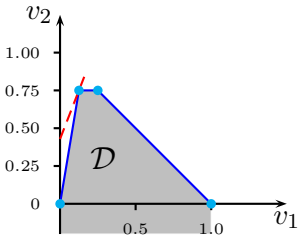

Figure 5.6: *$\mathcal{D}$ and the supporting hyperplane.*

Finally we show that the algorithm terminates after a finite number of steps.

Theorem 5.4.7. *The dual variant of Benson's algorithm is finite.*

Proof. Since $\hat{d} \in \operatorname{int} \mathcal{D}$, the point $v^k \in \mathcal{D}$ computed in iteration k belongs to $\operatorname{int} \mathcal{S}^{k-1}$. We have $\mathcal{S}^k := \mathcal{S}^{k-1} \cap \{v \in \mathbb{R}^p : \varphi(Cx^k, v) \geqq 0\}$ and, by Proposition 5.4.1, we know that $\mathcal{F} := \{v \in \mathcal{D} : \varphi(Cx^k, v) = 0\}$ is a face of $\mathcal{D}$ with $v^k \in \mathcal{F}$, where $\mathcal{F} \subseteq \operatorname{bd} \mathcal{S}^k$. This means for the next iteration that $v^{k+1} \notin \mathcal{F}$ (because $v^{k+1} \in \operatorname{int} \mathcal{S}^k$), and therefore v^{k+1} belongs to another face of $\mathcal{D}$. Since $\mathcal{D}$ is polyhedral, it has a finite number of faces, hence the algorithm is finite. □

5.5 Weight Set Decomposition

It is well known that $\mathbb{R}^p_{\geqq}$-minimal points of $\mathcal{P}$ can be characterized by weighted sum scalarization (Isermann, 1974). A point $y \in \mathcal{P}$ is $\mathbb{R}^p_{\geqq}$-minimal if and only if there exists $w \in \mathbb{R}^p_> = \{w \in \mathbb{R}^p : w > 0, k = 1, \ldots, p\}$ such that $w^T y \leqq w^T y'$ for all $y' \in \mathcal{P}$.

Considering, for fixed $y \in \mathcal{P}$, all w with this property leads to the idea of weight set decomposition, e.g. Benson and Sun (2000). Let $y \in \mathcal{P}$ and define

$$\mathcal{W}(y) := \left\{ w \in \mathbb{R}^p_{\geqq} : w^T y \leqq w^T y' \text{ for all } y' \in \mathcal{P} \right\}.$$

Using the equivalence relation $w^1 \sim w^2$ if and only if $w^1 = \rho w^2$ for some $\rho > 0$ it is clear that we can identify $\mathbb{R}^p_{\geqq} \setminus \{0\}$ with $\Lambda = \{\lambda \in \mathbb{R}^p : \lambda \geqq 0, \sum_{k=1}^p \lambda_k = 1\}$ and $\mathcal{W}(y)$ with $\Lambda(y) = \left\{\lambda \in \Lambda : \lambda^T y \leqq \lambda^T y' \text{ for all } y' \in \mathcal{P}\right\}$.

The following function was already considered in Section 5.2

$$\lambda : \mathbb{R}^p \to \mathbb{R}^p, \quad \lambda(v) := \left(v_1, \ldots, v_{p-1}, 1 - \sum_{i=1}^{p-1} v_i \right)^T.$$

Proposition 5.5.1. *Let $\mathcal{P}$ be nonempty and $\mathbb{R}^p_{\geqq}$-bounded below. Let $\{y^1, \ldots, y^q\}$ be the nondominated extreme points of $\mathcal{P}$ and let $\{\mathcal{F}_1^*, \ldots, \mathcal{F}_q^*\}$ be the corresponding $\mathcal{K}$-maximal facets of $\mathcal{D}$ according to the geometric duality theorem. Then for all $i = 1, \ldots, q$ it holds*

$$\Lambda(y^i) = \lambda(\mathcal{F}_i^*) := \{\lambda(v) : v \in \mathcal{F}_i^*\}$$

and $\{\lambda(\mathcal{F}_i^*) : i = 1, \ldots, q\}$ is a weight set decomposition, that is,

$$\Lambda = \bigcup_{i=1}^{q} \lambda(\mathcal{F}_i^*) \text{ and } \operatorname{ri} \lambda(\mathcal{F}_i^*) \cap \operatorname{ri} \lambda(\mathcal{F}_j^*) = \emptyset \text{ whenever } i \neq j.$$

Proof. Of course, $\lambda(\cdot)$ is a one-to-one map from $\max_{\mathcal{K}} \mathcal{D}$ onto Λ. The inverse map is $v(\lambda') := \lambda^{-1}(\lambda') = (\lambda'_1, \ldots, \lambda'_{p-1}, v_p)^T$ where v_p is the optimal value of the linear programme $(\mathrm{D}_1(\lambda'))$. Moreover, $\lambda(\cdot)$ is affine on convex subsets of $\max_{\mathcal{K}} \mathcal{D}$, in particular on each $\mathcal{K}$-maximal facet of $\mathcal{D}$.

Let $\lambda' \in \Lambda(y^i)$. Determine $v(\lambda')$. By duality between $(\mathrm{P}_1(\lambda'))$ and $(\mathrm{D}_1(\lambda'))$ we get $\varphi(y^i, v(\lambda')) = 0$. Moreover, we have $v(\lambda') \in \mathcal{D}$. Hence $v(\lambda') \in \mathcal{H}^*(y^i) \cap \mathcal{D} = \mathcal{F}_i^*$ and so $\lambda' \in \lambda(\mathcal{F}_i^*)$.

Let $\lambda' \in \lambda(\mathcal{F}_i^*)$, i.e., $v(\lambda') \in \mathcal{F}_i^*$. Then $\mathcal{H}(v(\lambda'))$ supports $\mathcal{P}$ in y^i. This implies that $\lambda' \in \Lambda(y^i)$.

The second statement follows from the properties of $\lambda(\cdot)$ and the fact that

$$\max_{\mathcal{K}} \mathcal{D} = \bigcup_{i=1}^{q} \mathcal{F}_i^* \text{ and } \operatorname{ri} \mathcal{F}_i^* \cap \operatorname{ri} \mathcal{F}_j^* = \emptyset \text{ whenever } i \neq j.$$

This completes the proof. □

Proposition 5.5.1 shows that both Algorithms 5.3.5 and 5.4.4 can be used to compute a weight set decomposition with respect to the nondominated extreme points of $\mathcal{P}$. This result is relevant in the context of multiobjective *integer* linear programmes. These are often solved using a two phase algorithm (Ulungu and Teghem, 1995), where the first phase consists in identifying the nondominated extreme points and the second phase finds all other nondominated points. The major problem in Phase 1 is the determination of a weight set decomposition (Przybylski *et al.*, 2007). It can be expected that the algorithms to solve (D) proposed in this chapter lead to progress in multiple objective integer linear programming algorithms for problems such as network flow problems, where the single objective counterparts can be solved by linear programming.

5.6 Numerical Results

In this section, we solve several multiple objective linear programmes by both the primal and the dual algorithm. We start with some small examples in order to illustrate the relationship between the primal outcome set $\mathcal{P}$ and the dual outcome

set $\mathcal{D}$. Then we address some radiotherapy problems. In each example, the primal problem is solved by (our slightly modified) primal outer approximation algorithm and the dual problem is solved by the dual variant of the algorithm. We show the primal and dual sets $\mathcal{P}$ and $\mathcal{D}$ and list the vertices and the facets for some of the smaller examples. We also compare the computation time of solving the primal and the dual problem. As seen in the considerations above it is sufficient to solve one problem to obtain the outcome set of both the primal and dual problems.

Both algorithms were implemented in Matlab 7.1(R14) using CPLEX 10.0 as LP solver and the tests were run on a dual processor CPU with 1.8 GHz and 1 GB RAM. We used the dual simplex method to solve the LPs. At step (k4), the method of Chen and Hansen (1991) for on-line vertex enumeration by adjacency lists was used to calculate a vertex representation from the inequality representation of $\mathcal{S}^k$.

As Benson's algorithm and its dual variant have steps of the same type, the number of facets of the primal, respectively dual, outcome set seems to correlate with the computation time. We observe in each of our examples that the dual variant is faster if the dual outcome set has fewer facets than the primal. Otherwise the primal method is faster. This means that it depends on the structure of the problem whether the primal or the dual algorithm is the better choice.

Example 5.6.1. *In this example we consider the LP relaxation of an assignment problem with three objectives. The cost matrices of the three objectives are*

$$\begin{pmatrix} 3 & 6 & 4 & 5 \\ 2 & 3 & 5 & 4 \\ 3 & 5 & 4 & 2 \\ 4 & 5 & 3 & 6 \end{pmatrix}, \begin{pmatrix} 2 & 3 & 5 & 4 \\ 5 & 3 & 4 & 3 \\ 5 & 2 & 6 & 4 \\ 4 & 5 & 2 & 5 \end{pmatrix}, \begin{pmatrix} 4 & 2 & 4 & 2 \\ 4 & 2 & 4 & 6 \\ 4 & 2 & 6 & 3 \\ 2 & 4 & 5 & 3 \end{pmatrix}.$$

Figures 5.7 and 5.8 show the weakly nondominated set of $\mathcal{P}$ and the $\mathcal{K}$-maximal subset of $\mathcal{D}$. The four nondominated vertices of $\mathcal{P}$ are $(11, 11, 14)$, $(19, 14, 10)$, $(15, 9, 17)$, and $(13, 16, 11)$. They correspond to the $\mathcal{K}$-maximal facets of $\mathcal{D}$ given by $3v_1+3v_2+v_3 = 14$, $-9v_1-4v_2+v_3 = 10$, $2v_1+8v_2+v_3 = 17$, and $-2v_1-5v_2+v_3 = 11$, respectively. $\mathcal{D}$ has nine vertices which we list along with the corresponding facets of $\mathcal{P}$ in Table 5.1.

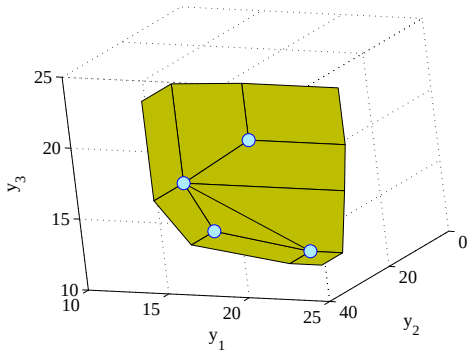

Figure 5.7: *$wmin_{\mathbb{R}^3_{\leqq}} \mathcal{P}$ in Example 5.6.1.*

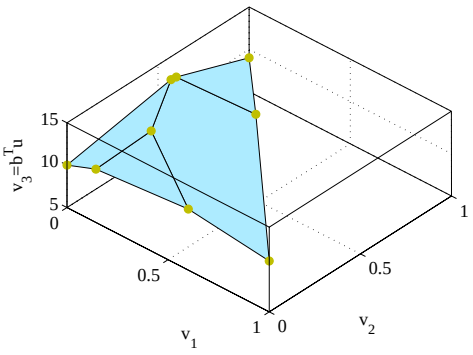

Figure 5.8: *$\max_{\mathcal{K}} \mathcal{D}$ in Example 5.6.1.*

Table 5.1: *Example 5.6.1: Vertices of $\mathcal{D}$ and corresponding facets of $\mathcal{P}$.*

Vertices of $\mathcal{D}$			Facets of $\mathcal{P}$								
v_1	v_2	v_3									
1	0	11	y_1						$=$	11
0	1	9			y_2				$=$	9
0	0	10					y_3		$=$	10
$\frac{1}{3}$	$\frac{2}{3}$	11	$\frac{1}{3}y_1$	$+$	$\frac{2}{3}y_2$				$=$	11
$\frac{3}{5}$	0	$12\frac{1}{5}$	$\frac{3}{5}y_1$	$+$			$\frac{2}{5}y_3$		$=$	$12\frac{1}{5}$
0	$\frac{4}{7}$	$12\frac{2}{7}$			$\frac{4}{7}y_2$	$+$	$\frac{3}{7}y_3$		$=$	$12\frac{2}{7}$
$\frac{1}{7}$	0	$11\frac{2}{7}$	$\frac{1}{7}y_1$	$+$			$\frac{6}{7}y_3$		$=$	$11\frac{2}{7}$
0	$\frac{3}{5}$	$12\frac{1}{5}$			$\frac{3}{5}y_2$	$+$	$\frac{2}{5}y_3$		$=$	$12\frac{1}{5}$
$\frac{11}{61}$	$\frac{16}{61}$	$12\frac{41}{61}$	$\frac{11}{61}y_1$	$+$	$\frac{16}{61}y_2$	$+$	$\frac{34}{61}y_3$		$=$	$12\frac{41}{61}$

Example 5.6.2. *The next small example has again three objectives. The data are*

$$C = \begin{pmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{pmatrix}, \; A = \begin{pmatrix} -1 & -1 & -1 \\ -1 & -3 & -1 \\ -3 & -4 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \; b = \begin{pmatrix} -5 \\ -9 \\ -16 \\ 0 \\ 0 \\ 0 \end{pmatrix}.$$

Figures 5.9 and 5.10 show the weakly nondominated subset of $\mathcal{P}$ and the $\mathcal{K}$-maximal subset of $\mathcal{D}$. Seven vertices of $\mathcal{P}$ and their corresponding $\mathcal{K}$-maximal facets of $\mathcal{D}$ are shown in Table 5.2. $\mathcal{D}$ has nine vertices, which we list with their corresponding facets of $\mathcal{P}$ in Table 5.3.

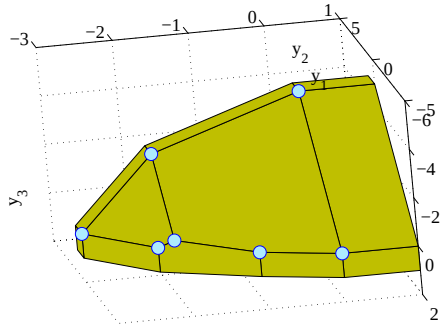

Figure 5.9: *$wmin_{\mathbb{R}^3_{\geqq}} \mathcal{P}$ in Example 5.6.2.*

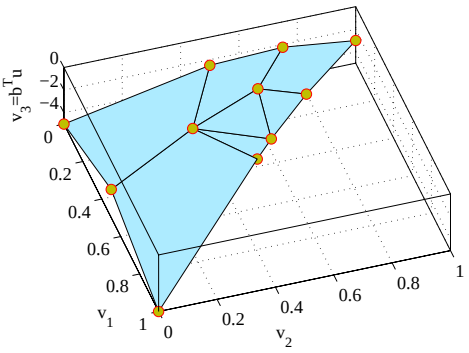

Figure 5.10: *$\max_{\mathcal{K}} \mathcal{D}$ in Example 5.6.2.*

Table 5.2: *Example 5.6.2: Vertices of $\mathcal{P}$ and corresponding $\mathcal{K}$-maximal facets of $\mathcal{D}$.*

Vertices of $\mathcal{P}$			$\mathcal{K}$-maximal Facets of $\mathcal{D}$								
v_1	v_2	v_3									
-5	0	0	$5v_1$	$+$			v_3	$=$	0
0	-3	0			$3v_2$	$+$	v_3	$=$	0
0	0	-5	$-5v_1$	$-$	$5v_2$	$+$	v_3	$=$	-5
$-2\frac{2}{5}$	$-2\frac{1}{5}$	0	$2\frac{2}{5}v_1$	$+$	$2\frac{1}{5}v_2$	$+$	v_3	$=$	0
0	-2	-3	$-3v_1$	$-$	v_2	$+$	v_3	$=$	-3
-4	-1	0	$4v_1$	$+$	v_2	$+$	v_3	$=$	0
$-2\frac{2}{3}$	-2	$-\frac{1}{3}$	$2\frac{1}{3}v_1$	$+$	$1\frac{2}{3}v_2$	$+$	v_3	$=$	$-\frac{1}{3}$

Table 5.3: *Example 5.6.2: Vertices of $\mathcal{D}$ and corresponding facets of $\mathcal{P}$.*

Vertices of $\mathcal{D}$			Facets of $\mathcal{P}$							
v_1	v_2	v_3								
$\frac{1}{3}$	$\frac{1}{3}$	$-1\frac{2}{3}$	$\frac{1}{3}y_1$	+	$\frac{1}{3}y_2$	+	$\frac{1}{3}y_3$	=	$-1\frac{2}{3}$
1	0	-5	y_1					=	-5
0	1	-3			y_2			=	-3
0	0	-5					y_3	=	-5
$\frac{1}{2}$	$\frac{1}{2}$	$-2\frac{1}{2}$	$\frac{1}{2}y_1$	+	$\frac{1}{2}y_2$			=	$-2\frac{1}{2}$
$\frac{1}{2}$	0	$-2\frac{1}{2}$	$\frac{1}{2}y_1$	+			$\frac{1}{2}y_3$	=	$-2\frac{1}{2}$
0	$\frac{3}{4}$	$-2\frac{1}{4}$			$\frac{3}{4}y_2$	+	$\frac{1}{4}y_3$	=	$-2\frac{1}{4}$
0	$\frac{1}{2}$	$-2\frac{1}{2}$			$\frac{1}{2}y_2$	+	$\frac{1}{2}y_3$	=	$-2\frac{1}{2}$
$\frac{1}{4}$	$\frac{3}{4}$	$-2\frac{1}{4}$	$\frac{1}{4}y_1$	+	$\frac{3}{4}y_2$			=	$-2\frac{1}{4}$
$\frac{3}{7}$	$\frac{4}{7}$	$-2\frac{2}{7}$	$\frac{3}{7}y_1$	+	$\frac{4}{7}y_2$			=	$-2\frac{2}{7}$
$\frac{1}{5}$	$\frac{3}{5}$	$-1\frac{4}{5}$	$\frac{1}{5}y_1$	+	$\frac{3}{5}y_2$	+	$\frac{1}{5}y_3$	=	$-1\frac{4}{5}$

Example 5.6.3. *In this example the primal solves faster than the dual. The data are*

$$C = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \; A = \begin{pmatrix} 1 & 1 & 1 \\ 2 & 1 & 2 \\ 1 & 2 & 2 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \; b = \begin{pmatrix} 3 \\ 4 \\ 4 \\ 0 \\ 0 \\ 0 \end{pmatrix}.$$

The weakly $\mathbb{R}^3_{\leqq}$-minimal set of $\mathcal{P}$ and the $\mathcal{K}$-maximal set of $\mathcal{D}$ are shown in Figures 5.11 and 5.12, respectively. $\mathcal{P}$ has seven vertices, they are $(0,0,3)$, $(2,0,1)$, $(0,2,1)$, $(0,4,0)$, $(4,0,0)$, $(1,2,0)$ and $(2,1,0)$. The corresponding $\mathcal{K}$-maximal facets of $\mathcal{D}$ are $3v_1 + 3v_2 + v_3 = 3$, $v_1 - v_2 + v_3 = 1$, $-v_1 + v_2 + v_3 = 1$, $-4v_2 + v_3 = 0$, $-4v_1 + v_3 = 0$, $-v_1 - 2v_2 + v_3 = 0$, and $-2v_1 - v_2 + v_3 = 0$, respectively. The six vertices $(1,0,0)$, $(0,1,0)$, $(0,0,0)$, $(\frac{1}{3},\frac{1}{3},1)$, $(\frac{2}{5},\frac{1}{5},\frac{4}{5})$ and $(\frac{1}{5},\frac{2}{5},\frac{4}{5})$ of $\mathcal{D}$ correspond to

the facets $y_1 = 0$, $y_2 = 0$, $y_3 = 0$, $\frac{1}{3}y_1 + \frac{1}{3}y_2 + \frac{1}{3}y_3 = 1$, $\frac{2}{5}y_1 + \frac{1}{5}y_2 + \frac{2}{5}y_3 = \frac{4}{5}$, *and* $\frac{1}{5}y_1 + \frac{2}{5}y_2 + \frac{2}{5}y_3 = \frac{4}{5}$ *of* $\mathcal{P}$.

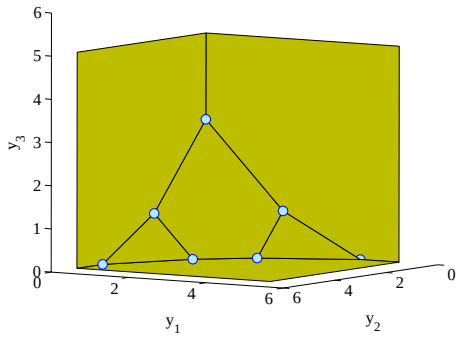

Figure 5.11: *wmin*$_{\mathbb{R}^3_{\geqq}}\mathcal{P}$ *in Example 5.6.3.*

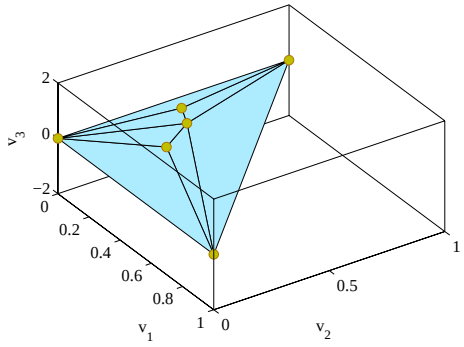

Figure 5.12: $\max_{\mathcal{K}}\mathcal{D}$ *in Example 5.6.3.*

Example 5.6.4. *This problem is a portfolio selection problem (example 2050 in Steuer (1989)) with three objectives, 21 variables and 45 constraints.* $\mathcal{P}$ *has 52 nondominated extreme points.* $\mathcal{D}$ *has 99 extreme points and 52 facets, see Figures 5.13 and 5.14.*

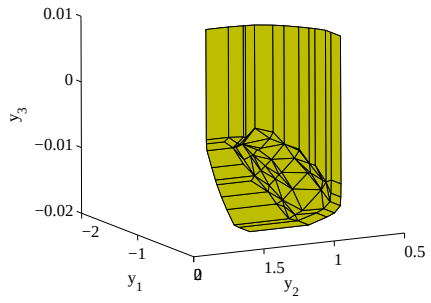

Figure 5.13: *wmin*$_{\mathbb{R}^3_{\geqq}}\mathcal{P}$ *in Example 5.6.4.*

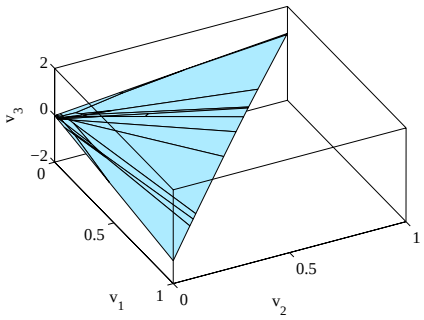

Figure 5.14: $\max_{\mathcal{K}}\mathcal{D}$ *in Example 5.6.4.*

Example 5.6.5. *The problem of intensity optimization in radiotherapy treatment planning is formulated as a multiobjective linear programme in Chapter 3. We use one of the examples from Chapter 4, an acoustic neuroma. The problem has three objectives,* 597 *variables and* 1664 *constraints.* $\mathcal{P}$ *(see Figure 5.15) has 55 vertices and 85 facets,* $\mathcal{D}$ *(see Figure 5.16) has 85 vertices and 55 facets.*

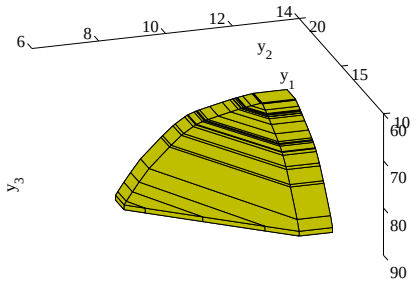

Figure 5.15: *$wmin_{\mathbb{R}^3_{\geqq}} \mathcal{P}$ in Example 5.6.5.*

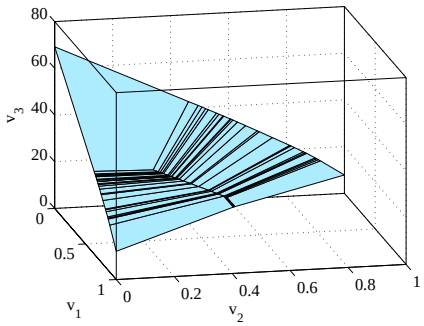

Figure 5.16: *$\max_{\mathcal{K}} \mathcal{D}$ in Example 5.6.5.*

Example 5.6.6. *Our second radiotherapy treatment planning example concerns a prostate case (see Chapter 4). In this three-objective problem $\mathcal{P}$ has 3165 nondominated extreme points and 3280 facets. $\mathcal{P}$ and $\mathcal{D}$ are shown in Figures 5.17 and 5.18, respectively.*

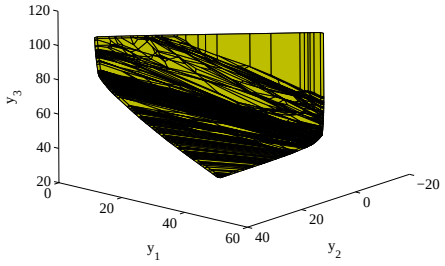

Figure 5.17: *$wmin_{\mathbb{R}^3_{\geqq}} \mathcal{P}$ in Example 5.6.6.*

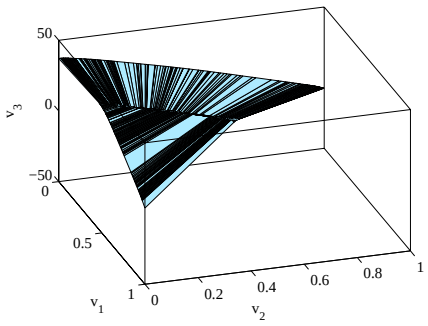

Figure 5.18: *$\max_{\mathcal{K}} \mathcal{D}$ in Example 5.6.6.*

Finally, we compare the computation times of our examples (see Table 5.4). We only give the number of vertices and the number of facets of $\mathcal{P}$, because the number of facets and the number of vertices of $\mathcal{D}$ correspond to them by geometric duality theory.

We see that the dual variant of the algorithm may have a computational speed advantage. It can be regarded as an alternative method which is preferable depending on the structure of the problem.

Table 5.4: *Computation times for the examples.*

		Size of A		$\mathcal{P}$		CPU time (seconds)	
Example	p	m	n	Vertices	Facets	primal	dual
5.2.1	2	5	2	4	5	0.0310	0.0310
5.6.1	3	16	16	4	9	0.0620	0.0470
5.6.2	3	6	3	7	11	0.0940	0.0470
5.6.3	3	6	3	7	6	0.0460	0.0630
5.6.4	3	45	21	52	99	0.7660	0.5310
5.6.5	3	1664	597	55	85	13.9840	8.8640
5.6.6	3	2142	824	3165	3280	995.0500	792.3900

5.7 Summary

In this chapter, we extended Benson's original outer approximation algorithm to solve unbounded problem. Then we developed a dual variant of Benson's algorithm. We proved that the solution of the dual problem provides a weight set decomposition with respect to nondominated extreme points. We applied both the primal algorithm and the dual algorithm to several examples. The numerical results show that the dual variant of Benson's algorithm can be regarded as an alternative method to solve MOLPs.

Chapter 6

Approximating the Nondominated Set of an MOLP by Approximately Solving its Dual Problem

6.1 Introduction

For an MOLP, although it is theoretically possible to identify the complete nondominated set using the primal or dual methods described in Chapters 3 and 5, finding an exact description of this set often turns out to be practically impossible or at least computationally too expensive (see the examples in Chapter 4). Therefore, many researchers focus on approximating the nondominated set, see Ruzika and Wiecek (2005) for a survey. In the literature, the concept of ε-nondominated points has been suggested as a means to account for modeling limitations or computational inaccuracies.

In Chapter 4, we have proposed an approximation version of Benson's algorithm to sandwich the extended feasible set of an MOLP with an outer approximation and an inner approximation. The nondominated set of the inner approximation is proved to be a set of ε-nondominated points. In this chapter, we propose to

solve the dual MOLP approximately, then to calculate a corresponding polyhedral set in objective space of the primal MOLP using the coupling function $\varphi(y,v) = \sum_{i=1}^{p-1} y_i v_i + y_p \left(1 - \sum_{i=1}^{p-1} v_i\right) - v_p$ defined in Section 5.2. The weakly nondominated subset of this polyhedral set can be proved to be a set of weakly ε-nondominated points of the original primal MOLP.

Finally, we apply this approximate method to solve the beam intensity optimization problem of radiotherapy treatment planning. Three clinical cases were used and the results are compared with those obtained by the approximation version of Benson's algorithm which directly approximates the truncated extended feasible set of the primal MOLP.

6.2 Further Analysis of the Dual Variant of Benson's Algorithm

We review the dual variant of Benson's algorithm.

Dual variant of Benson's algorithm.

Initialization ($k = 0$)**.**	
(i1)	*Choose some* $\hat{d} \in \operatorname{int} \mathcal{D}$.
(i2)	*Compute an optimal solution* x^0 *of* $(P_1(\hat{d}))$.
(i3)	*Set* $S^0 = \{v \in \mathbb{R}^p : \lambda(v) \geqq 0, \varphi(Cx^0, v) \geqq 0\}$ *and* $k = 1$.
Iteration steps ($k \geqq 1$)**.**	
(k1)	*If* $\operatorname{vert} S^{k-1} \subseteq \mathcal{D}$ *stop, otherwise choose a vertex* s^k *of* $\mathcal{S}^{k-1}$ *such that* $s^k \notin \mathcal{D}$.
(k2)	*Compute* $\rho^k \in (0,1)$ *such that* $v^k := \rho^k s^k + (1-\rho^k)\hat{d} \in \max_{\mathcal{K}} \mathcal{D}$.
(k3)	*Compute an optimal solution* x^k *of* $(P_1(v^k))$.
(k4)	*Set* $S^k := S^{k-1} \cap \{v \in \mathbb{R}^p : \varphi(Cx^k, v) \geqq 0\}$.
(k5)	*Set* $k := k+1$ *and go to (k1).*

When the algorithm terminates, we have the following results (see Chapter 5 for

a proof).

Proposition 6.2.1.

(1) The set of $\mathcal{K}$-maximal vertices of $\mathcal{D}$ is $\operatorname{vert}\mathcal{S}^{k-1}$.

(2) The set $\left\{y \in \mathbb{R}^p : \varphi(y,v) \geqq 0 \text{ for all } v \in \operatorname{vert}\mathcal{S}^{k-1}\right\}$ *is a nondegenerate inequality representation of $\mathcal{P}$.*

(3) All $\mathbb{R}^p_{\leqq}$-minimal (nondominated) vertices of $\mathcal{P}$ are contained in the set $\mathcal{W} := \{Cx^0, Cx^1, \ldots, Cx^{k-1}\}$.

(4) The set $\{v \in \mathbb{R}^p : \lambda(v) \geqq 0, \varphi(y,v) \geqq 0 \text{ for all } y \in \mathcal{W}\}$ *is a (possibly degenerate) inequality representation of $\mathcal{D}$. "Possibly degenerate" means that this inequality representation may include redundant inequalities (a redundant inequality is produced during the iteration if a supporting hyperplane supports a face but not a facet of $\mathcal{D}$).*

(1) and (4) in Proposition 6.2.1 give the vertex set and the inequality representation of $\mathcal{D}$ while (2) gives the inequality representation of $\mathcal{P}$ and (3) gives a set of nondominated points of $\mathcal{P}$ which includes all the vertices of $\mathcal{P}$. Proposition 6.2.1 suggests that we can get both $\mathcal{P}$ and $\mathcal{D}$ when the algorithm terminates.

At each iteration k, the hyperplane given by $\mathcal{H}^*(Cx^k) = \{v \in \mathbb{R}^p : \varphi(Cx^k, v) = 0\}$ is constructed so that it cuts off a portion of $\mathcal{S}^k$ containing s^k, thus $\mathcal{S}^0 \supseteq \mathcal{S}^1 \supseteq \mathcal{S}^2 \supseteq \ldots \supseteq \mathcal{S}^{k-2} \supseteq \mathcal{S}^{k-1} = \mathcal{D}$.

Geometric duality theory establishes a relationship between $\mathcal{P}$ and $\mathcal{D}$. $\mathcal{P}$ has the property that $\mathcal{P} = \mathcal{P} + \mathbb{R}^p_{\leqq}$ while $\mathcal{D}$ has the property that $\mathcal{D} = \mathcal{D} - \mathcal{K}$ and the projection to its first $p-1$ components is the polytope $\{t \in \mathbb{R}^{p-1}, t \geqq 0, \sum_{i=1}^{p-1} t_i \leqq 1\}$. Now we apply the geometric duality theorem to polyhedral convex sets that have some of the properties of $\mathcal{P}$ and $\mathcal{D}$, respectively.

Property 6.2.2. *In this thesis, we consider special convex polyhedral sets $\mathcal{S} \subseteq \mathbb{R}^p$ with the property that $\mathcal{S} = \mathcal{S} - \mathcal{K}$ and the projection to its first $p-1$ components is the polytope* $\{t \in \mathbb{R}^{p-1}, t \geqq 0, \sum_{i=1}^{p-1} t_i \leqq 1\}$.

Lemma 6.2.3. *For $\mathcal{S}$ with Property 6.2.2, $\mathcal{S}_\infty = -\mathcal{K}$.*

Proof. The proof is part of the proof in Proposition 5.4.3. □

Definition 6.2.4. *For a polyhedral convex set $\mathcal{S} \subseteq \mathbb{R}^p$ with Property 6.2.2, we define $\mathcal{D}(\mathcal{S}) = \{y \in \mathbb{R}^p : \varphi(y, v) \geqq 0,$ for all $v \in \operatorname{vert} \mathcal{S}\}$, where $\varphi(y, v) = \sum_{i=1}^{p-1} y_i v_i + y_p \left(1 - \sum_{i=1}^{p-1} v_i\right) - v_p$.*

Proposition 6.2.5. *Let $\mathcal{S} \subseteq \mathbb{R}^p$ with Property 6.2.2. Then $\mathcal{D}(\mathcal{S}) = \mathcal{D}(\mathcal{S}) + \mathbb{R}^p_{\geqq}$.*

Proof. It is obvious that $\mathcal{D}(\mathcal{S}) \subseteq \mathcal{D}(\mathcal{S}) + \mathbb{R}^p_{\geqq}$, therefore we only need to show $\mathcal{D}(\mathcal{S}) \supseteq \mathcal{D}(\mathcal{S}) + \mathbb{R}^p_{\geqq}$.

Let $w + d \in \mathcal{D}(\mathcal{S}) + \mathbb{R}^p_{\geqq}$ with $w \in \mathcal{D}(\mathcal{S})$ and $d \in \mathbb{R}^p_{\geqq}$. Since $w \in \mathcal{D}(\mathcal{S})$, we have $\varphi(w, v) \geqq 0$ for all $v \in \operatorname{vert} \mathcal{S}$. $\varphi(w + d, v) = \varphi(w, v) + \langle d, \lambda(v)\rangle$ where $\lambda(v) = (v_1, \ldots, v_{p-1}, 1 - \sum_{i=1}^{p-1} v_i)^T$. Since $\varphi(w, v) \geqq 0$ and it is obvious that $\langle d, \lambda(v)\rangle \geqq 0$ because both d and $\lambda(v) \in R^p_{\geqq}$, we have $\varphi(w + d, v) \geqq 0$ and $w + d \in \mathcal{D}(\mathcal{S})$. □

Corollary 6.2.6. *For $\mathcal{S} \subseteq \mathbb{R}^p$ with Property 6.2.2, Theorem 5.2.2 holds for $\mathcal{D} = \mathcal{S}$ and $\mathcal{P} = \mathcal{D}(\mathcal{S})$.*

Proposition 6.2.7. *Let $\mathcal{S}^1$ and $\mathcal{S}^0$ be polyhedral convex sets with property 6.2.2 and $\mathcal{S}^1 \subseteq \mathcal{S}^0$, then $\mathcal{D}(\mathcal{S}^1) \supseteq \mathcal{D}(\mathcal{S}^0)$.*

Proof. By Lemma 6.2.3 we have $\mathcal{S}^1_\infty = \mathcal{S}^0_\infty = -\mathcal{K}$. This means that $\mathcal{S}^0$ and $\mathcal{S}^1$ have only one extreme direction $d = -e^p = (0, \ldots, 0, -1)$.

Suppose $\mathcal{S}^0$ has r vertices, $v^1, \ldots, v^r$. We need to show that for $y \in \mathcal{D}(\mathcal{S}^0)$, i.e., $\varphi(y, v) \geqq 0, v = v^1, \ldots, v^r$, it holds that $y \in \mathcal{D}(\mathcal{S}^1)$.

Let v^* be a vertex of $\mathcal{S}^1$, then $v^* \in \mathcal{S}^0$ as $\mathcal{S}^1 \subseteq \mathcal{S}^0$. Therefore, v^* can be expressed as $v^* = \sum_{i=1}^r \rho_i v^i + \nu d$ with $\rho_i \geqq 0, \nu \geqq 0, \sum_{i=1}^r \rho_i = 1,$ and $d = -e^p$. We calculate

$\varphi(y, v^*)$ as follows.

$$\begin{aligned}
\varphi(y,v^*) &= \sum_{k=1}^{p-1} y_k v_k^* + y_p(1-\sum_{k=1}^{p-1} v_k^*) - v_p^* \\
&= \sum_{k=1}^{p-1} y_k(\sum_{i=1}^{r}\rho_i v_k^i + \nu d_k) + y_p(1-\sum_{k=1}^{p-1}(\sum_{i=1}^{r}\rho_i v_k^i + \nu d_k)) - \sum_{i=1}^{r}\rho_i v_p^i - \nu d_p \\
&= \sum_{k=1}^{p-1} y_k \sum_{i=1}^{r}\rho_i v_k^i + y_p(1-\sum_{k=1}^{p-1}\sum_{i=1}^{r}\rho_i v_k^i) - \sum_{i=1}^{r}\rho_i v_p^i + \sum_{k=1}^{p-1} y_k \nu d_k \\
&\quad -y_p\sum_{k=1}^{p-1}\nu d_k - \nu d_p \\
&= \sum_{i=1}^{r}\rho_i\varphi(y,v^i) + 0 - 0 + \nu \\
&= \sum_{i=1}^{r}\rho_i\varphi(y,v^i) + \nu
\end{aligned}$$

Since $\varphi(y, v^i) \geqq 0$, $\rho \geqq 0$ and $\nu \geqq 0$, we have $\varphi(y, v^*) \geqq 0$. This means that any $y \in \mathcal{D}(\mathcal{S}^0)$ is also contained in $\mathcal{D}(\mathcal{S}^1)$. This proves $\mathcal{D}(\mathcal{S}^0) \subseteq \mathcal{D}(\mathcal{S}^1)$. □

For the dual variant of Benson's algorithm, Proposition 6.2.7 indicates that with the iteration $\mathcal{D}(\mathcal{S}^k)$ enlarges and when the algorithm terminates at iteration k, $\mathcal{D}(\mathcal{S}^{k-1}) = \mathcal{D}(\mathcal{D}) = \mathcal{P}$. We give an example to illustrate the dual variant of Benson's algorithm and we show the changing of $\mathcal{S}^k$ and $\mathcal{D}(\mathcal{S}^k)$ with iteration k.

Example 6.2.8. *Consider the MOLP* $\min\{Cx : Ax \geqq b\}$*, where*

$$C = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, A = \begin{pmatrix} 2 & 1 \\ 1 & 1 \\ 1 & 2 \\ 1 & 0 \\ 0 & 1 \end{pmatrix}, b = \begin{pmatrix} 4 \\ 3 \\ 4 \\ 0 \\ 0 \end{pmatrix}.$$

$\mathcal{P}$ and $\mathcal{D}$ are shown in Figure 6.1. The 5 vertices of $\mathcal{D}$ are $(0,0)$, $(\frac{1}{3},\frac{4}{3})$, $(\frac{1}{2},\frac{3}{2})$, $(\frac{2}{3},\frac{4}{3})$, $(1,0)$*. Their corresponding facets (supporting hyperplanes) of $\mathcal{D}$ are* $y_2 = 0$, $y_1 + 2y_2 = 4$, $y_1 + y_2 = 3$, $2y_1 + y_2 = 4$, $y_1 = 0$*. The four vertices of $\mathcal{P}$,* $(0,4)$, $(1,2)$,

$(2,1)$ *and* $(4,0)$ *correspond to the facets (supporting hyperplanes) of* $\mathcal{D}$ $4v_1 + v_2 = 4$,$v_1 + v_2 = 2$, $-v_1 + v_2 = 1$ *and* $-4v_1 + v_2 = 0$, *respectively.*

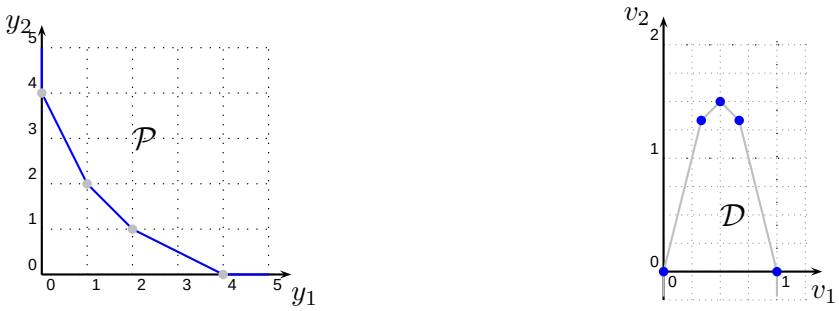

Figure 6.1: *$\mathcal{P}$ and $\mathcal{D}$ for Example 6.2.8.*

Figure 6.2 shows the change of $\mathcal{S}^k$ *with each iteration* k. *As can be seen, with iteration* k, $\mathcal{S}^k$ *becomes smaller and smaller until at termination it is the same as* $\mathcal{D}$. *The vertices of the initial cover* $\mathcal{S}^0$ *are* $(0, \frac{3}{2})$ *and* $(1, \frac{3}{2})$. *The first hyperplane cuts off vertex* $(0, \frac{3}{2})$, *the vertices of* $\mathcal{S}^1$ *are* $(1, \frac{3}{2})$, $(0,0)$, *and* $(\frac{3}{8}, \frac{3}{2})$. *The second hyperplane cuts off vertex* $(1, \frac{3}{2})$, *thus the vertices of* $\mathcal{S}^2$ *are* $(0,0)$, $(\frac{3}{8}, \frac{3}{2})$, $(\frac{5}{8}, \frac{3}{2})$ *and* $(1,0)$. *The third hyperplane cuts off vertex* $(\frac{3}{8}, \frac{3}{2})$, *thus the vertices of* $\mathcal{S}^3$ *are* $(0,0)$, $(\frac{1}{3}, \frac{4}{3})$, $(\frac{1}{2}, \frac{3}{2})$, $(\frac{5}{8}, \frac{3}{2})$ *and* $(1,0)$. *The fourth hyperplane cuts off vertex* $(\frac{5}{8}, \frac{3}{2})$ *and the vertices of* $\mathcal{S}^4$ *are* $(0,0)$, $(\frac{1}{3}, \frac{4}{3})$, $(\frac{1}{2}, \frac{3}{2})$, $(\frac{2}{3}, \frac{4}{3})$ *and* $(1,0)$. *After the fourth cut, we have* $\mathcal{S}^4 = \mathcal{D}$. *Therefore, the vertices of* $\mathcal{S}^4$ *are also the vertices of* $\mathcal{D}$.

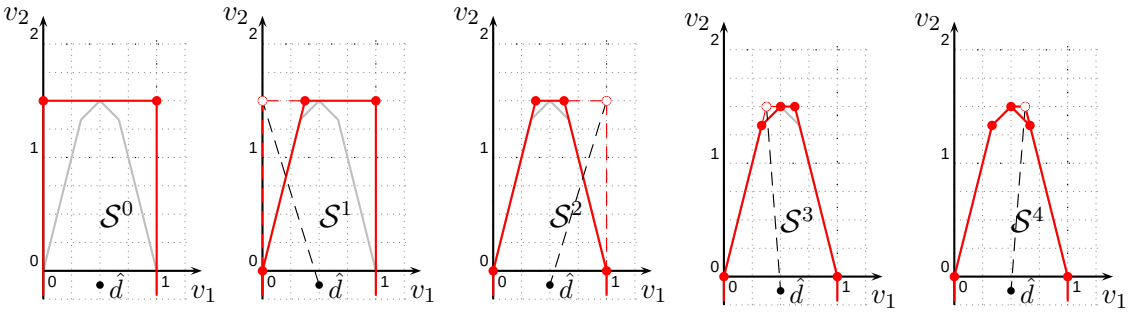

Figure 6.2: *The reduction of* $\mathcal{S}^k$ *with iteration* k.

The change of $\mathcal{D}(\mathcal{S}^k)$ *after each iteration* k *can be seen in Figure 6.3. The calculation of* $\mathcal{D}(\mathcal{S}^k)$ *is according to the definition* $\mathcal{D}(\mathcal{S}^k) = \{y \in \mathbb{R}^p : \varphi(y,v) \geqq 0 \text{ for all } v \in \operatorname{vert} \mathcal{S}^k\}$. *For example,* $\mathcal{D}(\mathcal{S}^0) = \{y \in \mathbb{R}^p : \varphi(y,v) \geqq 0 \text{ for } v =$

$(0, \frac{3}{2}), (1, \frac{3}{2})\}$, *i.e.* $\mathcal{D}(\mathcal{S}^0) = \{y_1 \geqq \frac{3}{2}\} \bigcap \{y_2 \geqq \frac{3}{2}\}$. *In contrast to the reduction of* $\mathcal{S}^k$, $\mathcal{D}(\mathcal{S}^k)$ *enlarges with iteration* k. *When the dual variant of Benson's algorithm terminates,* $\mathcal{S}^4 = \mathcal{D}$ *and* $\mathcal{D}(\mathcal{S}^4) = \mathcal{D}(\mathcal{D}) = \mathcal{P}$.

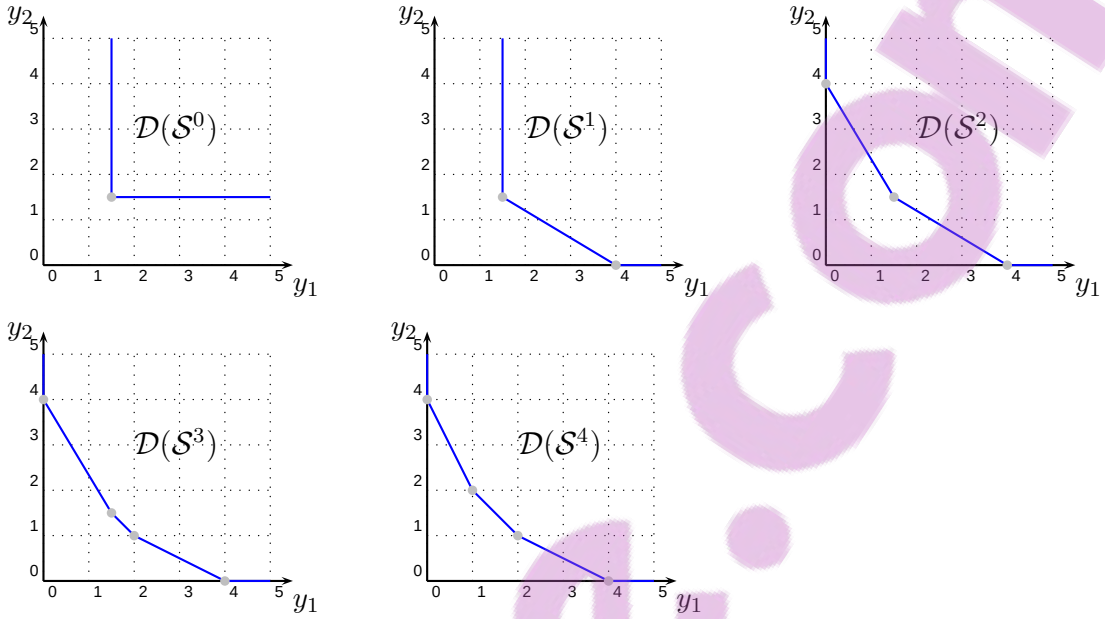

Figure 6.3: *The enlarging of* $\mathcal{D}(\mathcal{S}^k)$ *with iteration* k.

With Proposition 6.2.7 the process of "outer approximation" of $\mathcal{D}$ can be interpreted as a process of "inner approximation" of $\mathcal{P}$.

6.3 Obtaining the Nondominated Facets of $\mathcal{P}$ from $\mathcal{D}$

As seen in Proposition 6.2.1 (2), vertex $v = (v_1, v_2, \ldots, v_p)$ of $\mathcal{D}$ corresponds to a supporting hyperplane of $\mathcal{P}$ that supports $\mathcal{P}$ in a weakly nondominated facet. The hyperplane is $\lambda(v)^T y = v_p$, where $\lambda(v) = (v_1, \ldots, v_{p-1}, 1 - \sum_{i=1}^{p-1} v_i)^T$ and $\lambda(v) \geqq 0$. If $\lambda(v) > 0$, then the supporting hyperplane supports $\mathcal{P}$ in a nondominated facet instead of a weakly nondominated facet. We call v an *inner vertex* of $\mathcal{D}$ if $\lambda(v) > 0$, otherwise, we call it a *boundary vertex* of $\mathcal{D}$. To calculate the nondominated facets of $\mathcal{P}$, we only need to consider inner vertices of $\mathcal{D}$.

Now we are going to show how to calculate the nondominated facet $\mathcal{F}$ that corresponds to an inner vertex v of $\mathcal{D}$.

When the algorithm terminates, Proposition 6.2.1 (1) gives us the vertex set of $\mathcal{D}$ and (4) gives us the inequality representation of $\mathcal{D}$. The inequality representation of $\mathcal{D}$ is $\mathcal{D} = \{v \in \mathbb{R}^p : \lambda(v) \geqq 0, \varphi(y, v) \geqq 0 \text{ for all } y \in \mathcal{W}\}$ where $\mathcal{W} = \{Cx^0, Cx^1, \ldots, Cx^{k-1}\}$.

As we mentioned in Chapter 5, at step (k4), we use the on-line vertex enumeration algorithm of Chen and Hansen (1991). Therefore, we have the adjacency list for each vertex of $\mathcal{D}$. The adjacency list of a vertex includes the adjacent vertices and the adjacent supporting hyperplanes of $\mathcal{D}$. We say that a supporting hyperplane is adjacent to a vertex if the vertex is on the hyperplane. The set of adjacent supporting hyperplanes of an inner vertex is a subset of $\{v \in \mathbb{R}^p : \varphi(y, v) = 0 \text{ for all } y \in \mathcal{W}\}$. For an inner vertex of $\mathcal{D}$, if all its adjacent supporting hyperplanes (cuts) are non degenerate, i.e., each cut supports $\mathcal{D}$ in a facet, then we can find all the points in $\mathcal{P}$ that correspond to the cuts of $\mathcal{D}$ by geometric duality theory. These points are the vertices of $\mathcal{F}$. If not all of its adjacent cuts support $\mathcal{D}$ in facets, i.e., some cuts are degenerate, then we need to use the following result.

Proposition 6.3.1. *Let inner vertex v of $\mathcal{D}$ correspond to nondominated facet $\mathcal{F}$ of $\mathcal{P}$. Suppose v has a degenerate adjacent cut (supporting hyperplane), then the cut corresponds to a nondominated point $p \in \mathcal{F}$, but p is not a vertex of $\mathcal{F}$.*

Proof. Suppose vertex v has k non degenerate cuts (supporting hyperplanes). The k non degenerate cuts correspond to the k vertices of $\mathcal{F}$. The degenerate cut can be expressed by a linear combination of the k non degenerate cuts. Correspondingly, the point that the degenerate cut corresponds to can also be expressed by the same linear combination of the k vertices. Therefore, we have $p \in \mathcal{F}$. □

Proposition 6.3.1 shows that the points of $\mathcal{P}$ which correspond to the adjacent cuts of a vertex of $\mathcal{D}$ are on the same facet of $\mathcal{P}$. Moreover, all non degenerate cuts of $\mathcal{D}$ correspond to the vertices of the facet of $\mathcal{P}$ and all degenerate cuts correspond to points that are not vertices of the facet. Therefore, we can use the convex hull of

the points to find the facet.

We show an example with a degenerate cut.

Example 6.3.2. *For Example 6.2.8, vertex* $(\frac{1}{2}, \frac{3}{2})$ *of* $\mathcal{D}$ *is adjacent to three cuts, they are* $v_1 + v_2 = 2$, $-v_1 + v_2 = 1$ *and* $v_2 = \frac{3}{2}$. *Among them,* $v_2 = \frac{3}{2}$ *is a degenerate cut because it only supports* $\mathcal{D}$ *in vertex* $(\frac{1}{2}, \frac{3}{2})$ *instead of a facet. This can be seen in Figure 6.4.*

Vertex $(\frac{1}{2}, \frac{3}{2})$ *of* $\mathcal{D}$ *corresponds to the line segment between* $(1, 2)$ *and* $(2, 1)$, *a facet* $\mathcal{F}$ *of* $\mathcal{P}$, *as shown in Figure 6.5. Cut* $v_1 + v_2 = 2$ *corresponds to* $(1, 2)$ *and cut* $-v_1 + v_2 = 1$ *corresponds to* $(2, 1)$; *while* $v_2 = \frac{3}{2}$ *corresponds to* $(\frac{3}{2}, \frac{3}{2})$, *a nondominated point of* $\mathcal{P}$. *The two points* $(1, 2)$ *and* $(2, 1)$ *are the vertices of* $\mathcal{F}$, *while point* $(\frac{3}{2}, \frac{3}{2})$ *is in* $\mathcal{F}$, *but it is not a vertex of* $\mathcal{F}$.

The degenerate cut $v_2 = \frac{3}{2}$ *can be obtained by simply adding the two adjacent non degenerate cuts together with equal weights, i.e.,* $\frac{1}{2}(v_1 + v_2) + \frac{1}{2}(-v_1 + v_2) = \frac{1}{2} \times 2 + \frac{1}{2} \times 1$ *or* $v_2 = \frac{3}{2}$. *On the other hand,* $(\frac{3}{2}, \frac{3}{2})$ *can also be obtained by simply adding the two vertices* $(1, 2)$ *and* $(2, 1)$ *together with the same equal weights, i.e.,* $\frac{1}{2} \times (1, 2) + \frac{1}{2} \times (2, 1) = (\frac{3}{2}, \frac{3}{2})$.

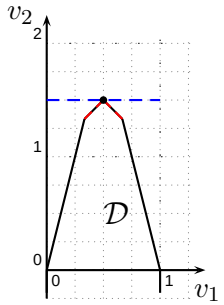

Figure 6.4: *Vertex* $(\frac{1}{2}, \frac{3}{2})$ *and its degenerate cut.*

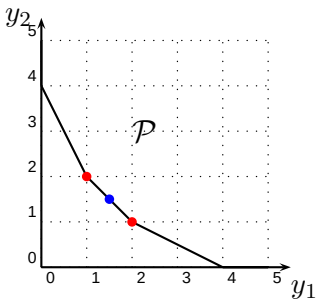

Figure 6.5: *$\mathcal{P}$ and the point corresponding to a degenerate cut.*

6.4 Solving the Dual MOLP Approximately

If the nondominated set of an MOLP is "curved", meaning that there are many facets and many vertices of $\mathcal{P}$, then there will be many vertices and facets of $\mathcal{D}$ according

to Corollary 5.2.3. Therefore, whether we are solving the primal problem with Benson's algorithm or the dual problem with dual variant of Benson's algorithm, computation time may be a problem.

In this section, we propose approximately solving the dual MOLP but controlling the approximation error to get an approximate extended feasible set in objective space $\mathcal{D}^o$, and then finding the polyhedral set $\mathcal{D}(\mathcal{D}^o)$. $\mathcal{D}(\mathcal{D}^o)$ is an inner approximation of $\mathcal{P}$, the original extended primal feasible set. Finally we show that the weakly nondominated set of $\mathcal{D}(\mathcal{D}^o)$ is actually an weakly ε-nondominated set of the original MOLP.

Our approximate dual variant of Benson's algorithm is identical to Algorithm 5.4.4 except for step **(k1)**. Let $\epsilon \in \mathbb{R}, \epsilon \geqq 0$ be a tolerance, then the changes are as follows.

(k1) If, for each $s \in \operatorname{vert} \mathcal{S}^{k-1}$, $s \in \mathcal{D} + \epsilon e^p$ is satisfied, then the outer approximation of $\mathcal{D}$ denoted by $\mathcal{D}^o$ is equal to $\mathcal{S}^{k-1}$. Stop. Otherwise, choose any $s^k \in \operatorname{vert} \mathcal{S}^k$ such that $s^k \notin \mathcal{D} + \epsilon e^p$ and continue.

To check $s \in \mathcal{D} + \epsilon e^p$, we need solve $D_1(s)$ and get its objective value μ. If $s_p - \mu \leqq \epsilon$, then $s \in \mathcal{D} + \epsilon e^p$.

Since $\mathcal{D}^o \supseteq \mathcal{D}$, Proposition 6.2.7 implies $\mathcal{D}(\mathcal{D}^o) \subseteq \mathcal{D}(\mathcal{D}) = \mathcal{P}$. This means that $\mathcal{D}(\mathcal{D}^o)$ is an inner approximation of $\mathcal{P}$. We write $\mathcal{P}^i = \mathcal{D}(\mathcal{D}^o)$.

We illustrate the algorithm continuing Example 6.2.8.

Example 6.4.1. *In Example 6.2.8, let us set* $\epsilon = \frac{3}{20}$. *After two cuts there are two vertices of* $\mathcal{S}^2$ *(*$s^1 = (\frac{3}{8}, \frac{3}{2})$ *and* $s^2 = (\frac{5}{8}, \frac{3}{2})$*) outside* $\mathcal{D}$. *The boundary points of* $\mathcal{D}$ *which have the same first element value as* s^1 *and* s^2 *are* $v^1 = (\frac{3}{8}, \frac{11}{8})$ *and* $v^2 = (\frac{5}{8}, \frac{11}{8})$, *respectively (see Figure 6.6). Both the Euclidean distance between* s^1 *and* v^1 *and the Euclidean distance between* s^2 *and* v^2 *are equal to* $\frac{1}{8}$. *We accept these two infeasible points* s^1, s^2 *for the outer approximation of* $\mathcal{D}$ *since the distances to the corresponding boundary points are less than* ϵ, *i.e.,* s^1, $s^2 \in \mathcal{D} + \epsilon e^q$. *When the algorithm terminates, the total number of iterations* k *is equal to* 3 *and* $\mathcal{S}^2 = \mathcal{D}^o$.

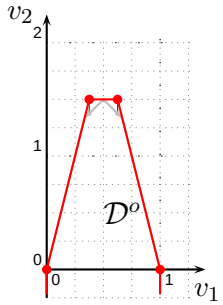

Figure 6.6: $\mathcal{D}^o$.

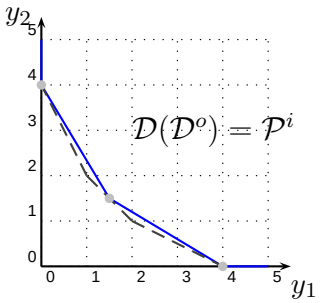

Figure 6.7: $\mathcal{P}^i$.

Figures 6.6 and 6.7 show $\mathcal{D}^o$ and its corresponding polyhedral set $\mathcal{P}^i = \mathcal{D}(\mathcal{D}^o)$.

We evaluate the approximation quality of $\mathcal{P}^i$ as an inner approximation of $\mathcal{P}$ and show that the weakly nondominated set of $\mathcal{P}^i$ is actually a set of weakly ε-nondominated points of $\mathcal{P}$.

First let us move $\mathcal{D}$ along e^p by ϵ, then we get $\mathcal{D}^u = \mathcal{D} + \epsilon e^p$. When the approximate algorithm terminates, the set of $\mathcal{K}$-maximal points of $\mathcal{D}^o$ lies "in between" the set of $\mathcal{K}$-maximal points of $\mathcal{D}$ and the set of $\mathcal{K}$-maximal points of $\mathcal{D}^u$, i.e., $\mathcal{D} \subseteq \mathcal{D}^o \subseteq \mathcal{D}^u$.

For $\mathcal{P} = \mathcal{D}(\mathcal{D})$ and $\mathcal{P}^u = \mathcal{D}(\mathcal{D}^u)$, according to Proposition 6.2.7, $\mathcal{P} \supseteq \mathcal{P}^i \supseteq \mathcal{P}^u$. Therefore, the set of weakly nondominated points ($\mathbb{R}^p_{\leqq}$-minimal points) of $\mathcal{P}^i = \mathcal{D}(\mathcal{D}^o)$ is "in between" the set of weakly nondominated points ($\mathbb{R}^p_{\leqq}$-minimal points) of $\mathcal{P}$ and the set of weakly nondominated points of $\mathcal{P}^u$.

Theorem 6.4.2. *Suppose the dual approximation error is ϵ. Let $\varepsilon = \epsilon e$, then the weakly nondominated set of $\mathcal{P}^i$ is a set of weakly ε-nondominated points of $\mathcal{P}$.*

Proof. First we show that the weakly nondominated set of $\mathcal{P}^u$ is actually a set of weakly ε-nondominated points of $\mathcal{P}$.

According to Corollary 6.2.6, Theorem 5.2.2 applies to $\mathcal{D} = \mathcal{D}^u$ and $\mathcal{P} = \mathcal{P}^u$, thus we can use duality theory to find $\mathcal{P}^u$.

Let $v = (v_1, v_2, \ldots, v_p)$ be a vertex of $\mathcal{D}$. Then there is a corresponding vertex v^u of $\mathcal{D}^u$ and $v^u = v + \epsilon e^p$. Suppose vertex v of $\mathcal{D}$ corresponds to facet $\lambda(v)^T y = v_p$ of $\mathcal{P}$, where $\lambda(v) = (v_1, \ldots, v_{p-1}, 1 - \sum_{i=1}^{p-1} v_i)^T$ and suppose vertex $v^u = (v_1^u, v_2^u, \ldots, v_p^u)$ of $\mathcal{D}^u$ corresponds to facet $\lambda(v^u)^T y = v_p^u$ of $\mathcal{P}^u$, where $\lambda(v^u) = (v_1^u, \ldots, v_{p-1}^u, 1 -$

$\sum_{i=1}^{p-1} v_i^u)^T$. As $v + \epsilon e^p = v^u$, we have $\lambda(v^u) = \lambda(v)$. Therefore, the facet of $\mathcal{P}$ defined by $\lambda(v)^T y = v_p$ is parallel to the facet of $\mathcal{P}^u$ given by $\lambda(v^u)^T y = v_p^u$.

Let $\mathcal{F}$ be a facet of $\mathcal{D}$, then there is a corresponding facet $\mathcal{F}^u$ of $\mathcal{D}^u$ and $\mathcal{F}^u$ is parallel to $\mathcal{F}$. Suppose $\mathcal{F}$ corresponds to vertex $y = (y_1, y_2, \ldots, y_p)$ of $\mathcal{P}$, then the equation of the hyperplane spanned by $\mathcal{F}$ is $(y_p - y_1)v_1 + (y_p - y_2)v_2 + \ldots + (y_p - y_{p-1})v_{p-1} + v_p = y_p$. Suppose $\mathcal{F}^u$ corresponds to vertex $y^u = (y_1^u, y_2^u, \ldots, y_p^u)$ of $\mathcal{P}^u$, then the equation of the hyperplane spanned by $\mathcal{F}^u$ is $(y_p^u - y_1^u)v_1 + (y_p^u - y_2^u)v_2 + \ldots + (y_p^u - y_{p-1}^u)v_{p-1} + v_p = y_p^u$. $\mathcal{F}^u$ and $\mathcal{F}$ are parallel and $\mathcal{F}^u$ is obtained by moving $\mathcal{F}$ along e^p by ϵ, therefore, we have $y_p^u = y_p + \epsilon$, and $(y_p^u - y_1^u) = (y_p - y_1)$, $\ldots$, $(y_p^u - y_{p-1}^u) = (y_p - y_{p-1})$, i.e., $y^u = y + \epsilon e$.

Hence $\mathcal{P}^u$ can be obtained by moving every point y of $\mathcal{P}$ to $y + \epsilon e$. Therefore, the weakly nondominated set of $\mathcal{P}^u$ is a set of weakly ε-nondominated points of $\mathcal{P}$.

Moreover, as $\mathcal{P}^u \subseteq \mathcal{P}^i \subseteq \mathcal{P}$, so the weakly nondominated set of $\mathcal{P}^i$ is also a set of weakly ε-nondominated points of $\mathcal{P}$. □

Example 6.4.3. *For Example 6.2.8,* $\epsilon = \frac{3}{20}$. *Figure 6.8 and Figure 6.9 show* $\mathcal{D}$, $\mathcal{D}^u$, $\mathcal{P}$ *and* $\mathcal{P}^u$. *The vertices of* $\mathcal{P}$ *are* $(0, 4)$, $(1, 2)$, $(2, 1)$ *and* $(4, 0)$. *The corresponding vertices of* $\mathcal{P}^u$ *are* $(\frac{3}{20}, 4\frac{3}{20})$, $(1\frac{3}{20}, 2\frac{3}{20})$, $(2\frac{3}{20}, 1\frac{3}{20})$ *and* $(4\frac{3}{20}, \frac{3}{20})$. *The weakly nondominated set of* $\mathcal{P}^u$ *is a set of weakly* ε*-nondominated points of* $\mathcal{P}$ *where* $\varepsilon = \frac{3}{20}e$.

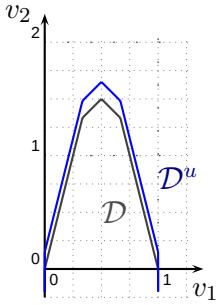

Figure 6.8: *$\mathcal{D}$ and $\mathcal{D}^u$.*

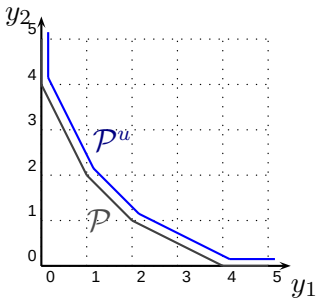

Figure 6.9: *$\mathcal{P}$ and $\mathcal{P}^u$.*

6.5 Numerical Results

We have formulated the beam intensity optimization problem as an MOLP in Chapter 3. The objectives of the MOLP are to minimize the maximum deviation α, β, γ of delivered dose from tumor lower bounds, from critical organ upper bounds and from the normal tissue upper bounds, respectively. We used both Benson's algorithm and an approximation version of Benson's algorithm to solve the MOLP of an acoustic neuroma (AN), a prostate (PR) and a pancreatic lesion (PL) case, respectively in Chapter 4.

In this chapter, we solve the dual problems of the same three clinical cases as above, i.e., the acoustic neuroma (AN), the prostate (PR) and the pancreatic lesion (PL) both by the dual variant of Benson's algorithm and the approximate dual variant of Benson's algorithm. Both algorithms were implemented in Matlab 7.1(R14) using CPLEX 10.0 as LP solver and the tests were run on a dual processor CPU with 1.8 GHz and 1 GB RAM.

Both the acoustic neuroma case and the prostate case can be solved exactly with the dual variant of Benson's algorithm. We show the results of solving the dual problem exactly with the dual variant of Benson's algorithm and approximately with the approximate dual variant of Benson's algorithm. The results for the acoustic neuroma case are shown in Figure 6.10. The set $\max_{\mathcal{K}} \mathcal{D}$ and the union of the nondominated facets of $\mathcal{P}$ obtained by the dual variant of Benson's algorithm are on the top left and right, respectively. The other pictures show the results of solving the dual problem approximately with the approximate dual variant of Benson's algorithm. Pictures on the left are $\max_{\mathcal{K}} \mathcal{D}^o$ while pictures on the right are the union of the nondominated facets of $\mathcal{P}^i$. From top to bottom, they are for approximation error $\epsilon = 0$, $\epsilon = 0.01$, and $\epsilon = 0.1$, respectively.

The results for the prostate case are in Figure 6.11. The pictures on the top left and right are $\max_{\mathcal{K}} \mathcal{D}$ and the union of the nondominated facets of $\mathcal{P}$, they were obtained by the dual variant of Benson's algorithm. The rest of the pictures show the results of solving the dual problem approximately with the approximate dual

variant of Benson's algorithm. Pictures on the left show $\max_{\mathcal{K}} \mathcal{D}^o$ while pictures on the right show the union of the nondominated facets of $\mathcal{P}^i$. From top to bottom, they are for approximation error $\epsilon = 0$, $\epsilon = 0.001$, $\epsilon = 0.01$, and $\epsilon = 0.1$, respectively.

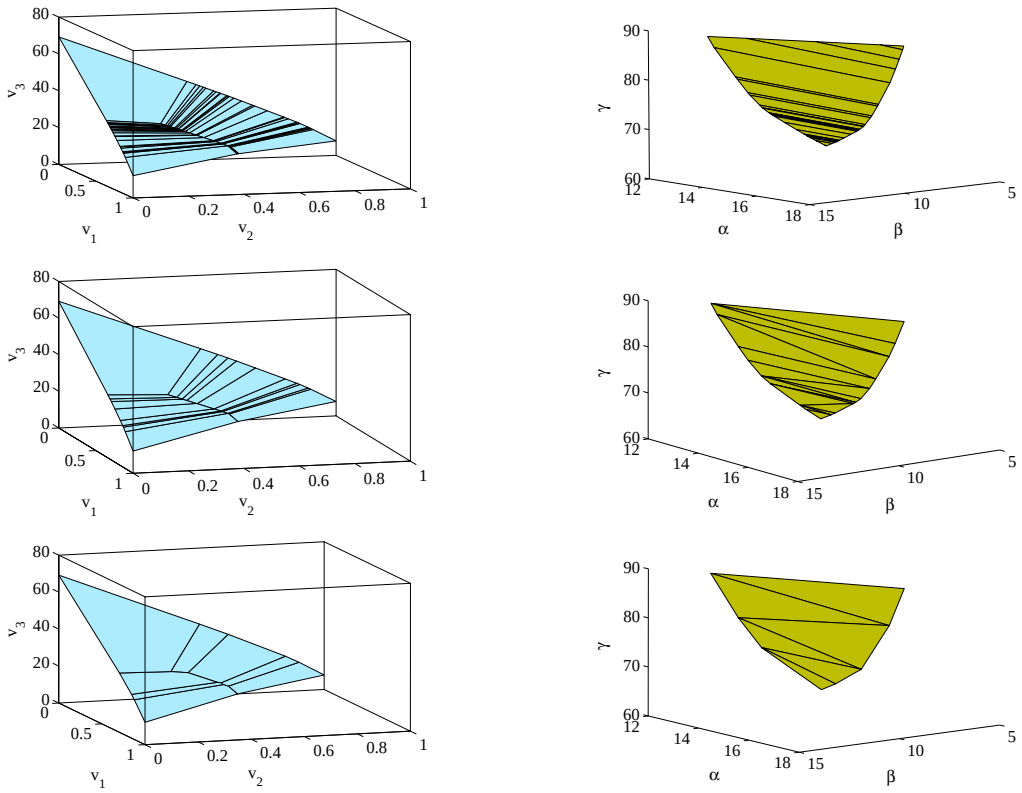

Figure 6.10: *The results for the AN case.*

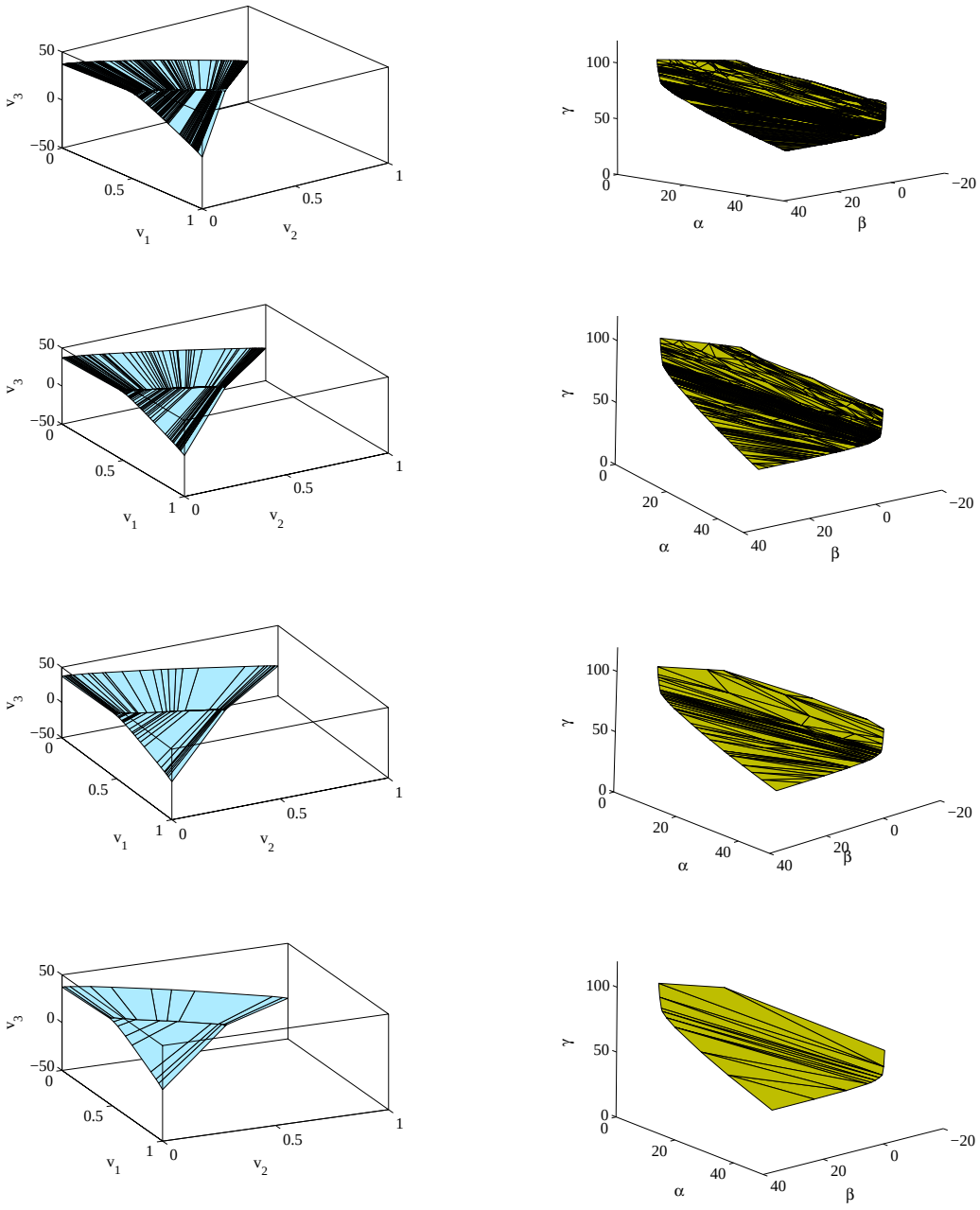

Figure 6.11: *The results for the PR case.*

The dual variant of Benson's algorithm cannot solve the problem of pancreatic lesion case within 10 hours of computation. Therefore, we show the results solved by the approximate dual variant of Benson's algorithm in Figure 6.12. Pictures on the left are $\max_{\mathcal{K}} \mathcal{D}^o$ and pictures on the right are the union of the nondominated facets of $\mathcal{P}^i$. Pictures on the top are for $\epsilon = 0.01$ while pictures on the bottom are for $\epsilon = 0.1$.

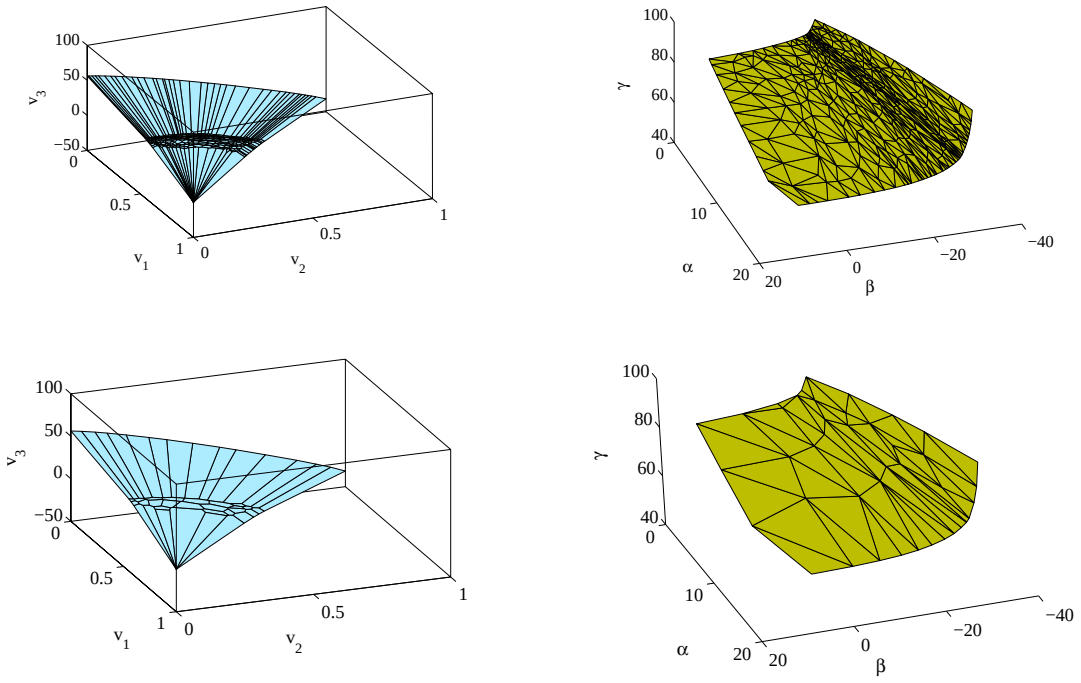

Figure 6.12: *The results for the PL case.*

Summarizing information comparing the number of vertices and number of cuts of the dual variant of Benson's algorithm and the approximate dual variant of Benson's algorithm with various values of ϵ is given in Table 6.1. The dual variant of Benson's algorithm can solve the dual problem of the first two cases exactly in one hour, but not the problem of the pancreatic lesion case. The approximate dual variant of Benson's algorithm can solve all three problems within 15 minutes with approximation error $\epsilon = 0.005$.

Table 6.1 and Figures 6.10, 6.11 and 6.12 clearly show the effect of the choice of ϵ. The smaller the approximation error, the more vertices and cuts that are generated

and the longer the computation time.

To make comparisons with the results obtained by Benson's algorithm and the approximation version of Benson's algorithm, we show $\mathcal{Y}'$ and $\mathcal{Y}'^{o}$ with different values of approximation error ϵ for the above three cases. Figure 6.13 shows $\mathcal{Y}'$ for the acoustic case and Figure 6.15 shows $\mathcal{Y}'$ for the prostate case. $\mathcal{Y}'^{o}$ with $\epsilon = 0.1$ for the acoustic case obtained by the approximation version of Benson's algorithm is shown in Figure 6.14, while $\mathcal{Y}'^{o}$ with $\epsilon = 0.1$ for the prostate case obtained by the approximation version of Benson's algorithm is shown in Figure 6.16. For the pancreatic lesion case, we show $\mathcal{Y}'^{o}$ with approximation error $\epsilon = 0.005$ in Figure 6.17 and $\epsilon = 0.1$ in Figure 6.18.

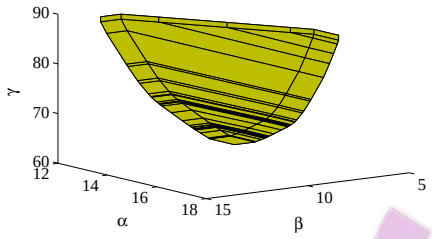

Figure 6.13: *AN: $\mathcal{Y}'$ solved by Benson's algorithm.*

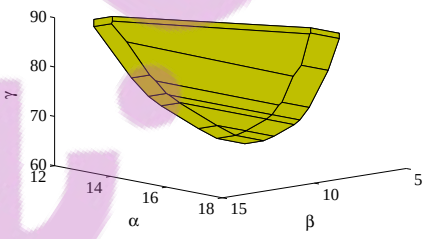

Figure 6.14: *AN: $\mathcal{Y}'^{o}$ with $\epsilon =$ 0.1.*

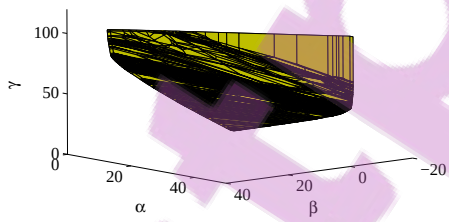

Figure 6.15: *PR: $\mathcal{Y}'$ solved by Benson's algorithm.*

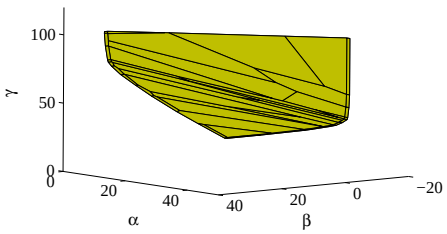

Figure 6.16: *PR: $\mathcal{Y}'^{o}$ with $\epsilon =$ 0.1.*

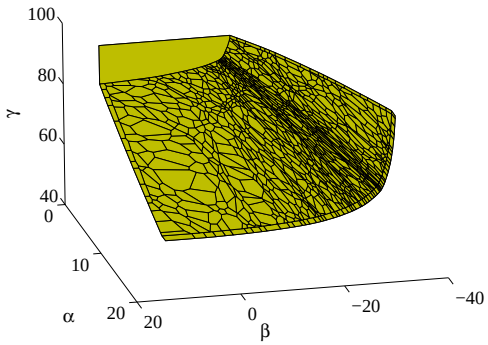

Figure 6.17: *PL:* $\mathcal{Y}'^{o}$ *with* $\epsilon =$ 0.005.

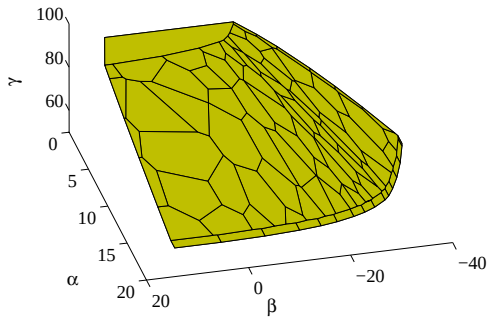

Figure 6.18: *PL:* $\mathcal{Y}'^{o}$ *with* $\epsilon =$ 0.1.

The number of vertices and number of cuts of $\mathcal{Y}'^{o}$ with various values of ϵ are also given in Table 6.1. Comparing the computation time of exactly solving the primal and exactly solving the dual problem for both the acoustic case and the prostate case, solving the primal problem exactly using Benson's algorithm needs more computation time than solving the dual problem exactly using the dual variant of Benson's algorithm.

If the approximation error ϵ in the approximation version of Benson's algorithm and the approximation error ϵ in the approximate dual variant of Benson's algorithm are the same, then both algorithms guarantee finding ε-nondominated set ($\varepsilon = \epsilon e$). Thus we can compare the computation time for solving $\mathcal{Y}'^{o}$ and $\mathcal{P}^{i}$ with the same ϵ. In Table 6.1, we observe that for the three clinical cases solving the dual approximately is always faster than solving the primal approximately with the same approximation error ϵ.

Solving the primal MOLP gives us $\mathcal{Y}'$. If we only need the nondominated facets, we need to get rid of the weakly nondominated facets. Solving the dual problem allows us directly calculate the nondominated facets. This can be taken as an advantage of solving the dual problem.

Table 6.1: *Running time and number of vertices and cutting planes to solve the dual problem and the primal problem for three cases with different approximation error ϵ ($\epsilon = 0$ means it is solved exactly).*

Case	ϵ	Solving the dual			Solving the primal		
		Time (seconds)	Vertices of $\mathcal{D}^o$	Cuts of $\mathcal{D}^o$	Time (seconds)	Nondominated vertices of $\mathcal{Y}'^o$	Cuts of $\mathcal{Y}'^o$
AN	0.1	1.484	17	8	5.938	27	21
	0.01	3.078	33	18	8.703	47	44
	0	8.864	85	55	13.984	55	85
PR	0.1	4.422	39	19	14.781	56	42
	0.01	18.454	157	78	64.954	296	184
	0.001	86.733	729	366	257.328	1157	692
	0	792.390	3280	3165	995.050	3165	3280
PL	0.3	29.110	40	21	70.796	57	37
	0.1	58.263	85	44	164.360	152	90
	0.05	102.761	151	78	303.630	278	159
	0.01	401.934	582	298	1184.950	1097	586
	0.005	734.784	1058	539	2147.530	1989	1041

6.6 Summary

In this chapter, we have developed an approximate dual variant of Benson's algorithm to solve MOLPs in objective space. We have shown the algorithm guarantees to find ε-nondominated points with a specified accuracy ϵ.

This algorithm was applied to the beam intensity optimization problem of radiation therapy treatment planning. Three clinical cases were used and the results were compared with those obtained by approximately solving the primal with the approximation version of Benson's algorithm. When both algorithms use the same approximation error ϵ, both guarantee producing ε-nondominated set ($\varepsilon = \epsilon e$). We

found that approximately solving the dual with the approximate dual variant of Benson's algorithm is faster than approximately solving the primal with the approximation version of Benson's algorithm for all the three clinical cases.

Chapter 7

Finding Representative Nondominated Points in Multiobjective Linear Programming

7.1 Introduction

For a decision maker, it is nearly impossible to study the infinite set of nondominated points to identify the most preferred solution. A discrete representation of the nondominated set by finitely many distinguishable points that cover the whole nondominated set simplifies this task. The decision maker can interactively navigate through the nondominated points to choose the most preferred solution. Therefore, it is of interest to find a good discrete representative subset of the nondominated set.

In Sections 7.2 and 7.3, we review quality attributes of discrete representations and summarize current methods for computing discrete representations of the nondominated set. In Section 7.4, we propose a method which combines the global shooting method (Benson and Sayin, 1997) and the normal boundary intersection (NBI) method (Das and Dennis, 1998). We analyze our proposed method and show

that the obtained points are evenly distributed and that the quality of the representation in terms of coverage and uniformity can be guaranteed. Neither the global shooting method, nor the normal boundary intersection method have this property. In Section 7.5, we apply the proposed method to a radiation therapy treatment planning problem. The results we obtain for some clinical cases illustrate the quality of our method.

7.2 Quality of Discrete Representations

In this section, let $\mathcal{Z} \subseteq \mathbb{R}^p$ be a set and let $\mathcal{R} \subseteq \mathcal{Z}$ be a finite subset. Sayin (2000) defines coverage, uniformity, and cardinality as the three quality attributes of a discrete representation. According to these three attributes, a good representation needs to contain a reasonable number of points, should not miss large portions of the nondominated set, and should not contain points that are very close to each other.

Moreover, Sayin (2000) proposes measures to quantify these attributes. The number of points contained in a representation is used to measure the cardinality. The coverage error ϵ and uniformity level δ are defined as follows.

Definition 7.2.1. *Let $\epsilon \geqq 0$ be a real number and d be a metric. $\mathcal{R}$ is called a d_ϵ-representation of $\mathcal{Z}$ if for any $z \in \mathcal{Z}$, there exists $r \in \mathcal{R}$ such that $d(z, r) \leqq \epsilon$.*

Definition 7.2.2. *Let $\mathcal{R}$ be a d_ϵ-representation of $\mathcal{Z}$. $\mathcal{R}$ is called a δ-uniform d_ϵ-representation if*

$$\min_{r^1, r^2 \in \mathcal{R}, r^1 \neq r^2} \{d(r^1, r^2)\} \geqq \delta.$$

The coverage error ϵ is a parameter that signifies how precisely the set $\mathcal{Z}$ is being represented by the discrete representative subset $\mathcal{R}$, it can be mathematical written as:

$$\epsilon = \max_{z \in \mathcal{Z}} \min_{r \in \mathcal{R}} d(z, r).$$

How well a fixed $z \in \mathcal{Z}$ is covered is determined by the closest point to z in the representation $\mathcal{R}$. For the entire set $\mathcal{Z}$, the coverage error depends on how well an

arbitrary element of $\mathcal{Z}$ is covered. Therefore, the coverage error ϵ is equal to the maximum of coverage error for individual points in $\mathcal{Z}$.

Similarly, the uniformity of a representation can be measured by the distance between a pair of closest points of $\mathcal{R}$. Thus it can be expressed as

$$\delta = \min_{r^1, r^2 \in \mathcal{R}} d(r^1, r^2).$$

For a discrete representation, a small number of points, low coverage error, and high uniformity level are desirable. This in its own right is a multiobjective problem, namely to find finite $\mathcal{R} \subseteq \mathcal{Z}$ to maximize δ and minimize ϵ and the number of points of $\mathcal{R}$.

7.3 Existing Methods

Survey of Existing Methods The nondominated set of an MOLP is the union of the (maximal) nondominated faces and these nondominated faces are polyhedral due to $\mathcal{Y}$ being a polytope. Therefore, finding discrete representations of $\mathcal{Y}_N$ is equivalent to finding discrete representations of a union of polyhedra.

There are two groups of methods for finding representations of the nondominated set, one is based on the knowledge of $\mathcal{X}_E$ and the other works without the knowledge of $\mathcal{X}_E$.

Based on the knowledge of $\mathcal{X}_E$, Sayin (2003) proposes a procedure to find discrete representations with specified coverage errors. The procedure also specifies the uniformity level of the representations. Knowledge of $\mathcal{X}_E$ can, however, not be assumed when solving an MOLP.

Most of the methods work without the knowledge of $\mathcal{X}_E$.

Benson and Sayin (1997) propose a global shooting method to find a representation of the nondominated set. This method has the coverage property, but it can not directly control the uniformity of the representations it generates.

Das and Dennis (1998) propose a normal boundary intersection method for finding several nondominated points for a general multiple objective nonlinear programming problem. It uses the convex hull of the individual minima (CHIM) as reference

plane. Equidistant reference points are placed on the CHIM and for each reference point a corresponding nondominated point is found by solving a scalar optimization problem. This method can produce evenly distributed nondominated points, however, some parts of the nondominated set may be missed, a problem caused by the use of the CHIM. We will illustrate this limitation later in this section.

Based on the NBI method, Messac *et al.* (2003) propose the normalized normal constraint (NC) method. NC works in a normalized objective space and uses an inequality constraint to reduce the feasible region in objective space. However, it has the same problem as the NBI method because it uses the CHIM as a reference plane. Realizing this limitation of using the CHIM, Messac and Mattson (2004) improve the NC method by using an extended CHIM instead of the CHIM as reference plane. They use examples to illustrate that their method provides an even representation of the entire nondominated set but they do not give any mathematical proof.

Analogously to Messac and Mattson (2004) we revise the NBI method in a way that guarantees coverage of the whole nondominated set and that allows us to prove a uniformity guarantee.

We emphasize that representation is different from approximation of $\mathcal{Y}_N$. While a representation $\mathcal{R}$ of $\mathcal{Y}_N$ is a finite set of points that must be nondominated, an approximation of $\mathcal{Y}_N$ may be an infinite or continuous set that has no intersection with $\mathcal{Y}_N$. Approximations can be feasible (inner approximation) or infeasible (outer approximation), as seen in Chapters 4 and 6. Quality measures for approximations are quite different from those of representations (Lotov *et al.*, 2004). While there are many approximation methods for multiobjective programming (the reader is referred to Ruzika and Wiecek (2005) for a survey) that compute some nondominated points, they do not aim at finding evenly distributed nondominated points and may yield bad representations in terms of coverage error and uniformity.

The Global Shooting Method For MOLP, define $\mathcal{Y}' = \{y \in \mathbb{R}^p : Cx \leqq y \leqq \hat{y} \text{ for some } x \in \mathcal{X}\}$, where $\hat{y}$ is chosen as a point so that for all $y \in \mathcal{Y}_N$ we have $y \leqq \hat{y}$ (the existence of $\hat{y}$ is an assumption for this method, see Benson and Sayin (1997)).

E.g., $\hat{y}$ can be chosen as the anti ideal point y^{AI}, $y_k^{AI} = \max\{y_k, y \in \mathcal{Y}\}, k = 1, \ldots, p$. $\mathcal{Y}'$ has dimension p and $\mathcal{Y}'$ and $\mathcal{Y}$ have the same nondominated set (see Section 3.4).

First, a big simplex $\mathcal{S}$ is constructed that contains $\mathcal{Y}'$ and a subsimplex $\hat{\mathcal{S}}$ of $\mathcal{S}$ is taken as the reference plane. Equidistant reference points are placed on $\hat{\mathcal{S}}$ and the method "shoots" from $\hat{y}$ towards each reference point as far as possible while remaining in $\mathcal{Y}'$. This is achieved by solving an LP. Thus a set of points on the boundary of $\mathcal{Y}'$ is calculated. Each reference point corresponds to a boundary point of $\mathcal{Y}'$, but not every such point is nondominated. Therefore it needs to be checked whether the intersection point is dominated or not by solving another LP.

Figure 7.1 illustrates the global shooting method. Three weakly nondominated points of $\mathcal{Y}'$ are found. Those are shown as crosses.

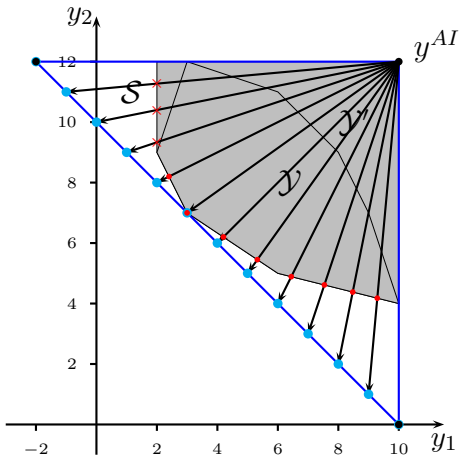

Figure 7.1: *Global shooting method.*

The global shooting method is simple and computationally tractable for the MOLP case. It guarantees coverage because it puts equidistant reference points on $\hat{\mathcal{S}}$ and $\mathcal{Y}_N \subseteq \hat{\mathcal{S}} + \mathbb{R}^p_{\leqq}$. However, the uniformity of the discrete representative set can not be controlled directly.

The Normal Boundary Intersection Method For MOLP (3.2), assume that individual minima of the functions $c^k x$ over $\mathcal{X}$ are attained at x^k for $k = 1, 2, \ldots, p$. Let $y^k = Cx^k$ and let $y^I = (c^1x^1, c^2x^2, \ldots, c^px^p)^T$ be the ideal point. The points

$y^1, \ldots, y^p$ define the convex hull of the individual minima (CHIM).

A set of equidistant reference points on the CHIM is generated and, for each of them, a NBI subproblem is solved to find the farthest point on the boundary of $\mathcal{Y}$ along the normal $\hat{n}$ of the CHIM pointing toward the ideal point (the normal direction should be negative). The NBI subproblem for a given reference point q is as follows:

$$\max\{t : q + t\hat{n} \in \mathcal{Y}, t \geqq 0\}. \tag{7.1}$$

Figure 7.2 shows how the NBI method works for the same MOLP example with two objectives as in Figure 7.1. For this example, all the points obtained are nondominated and no part of the nondominated set is overlooked. However, for problems with more than two objectives, even if the normal direction of the CHIM is negative, the solution method may still fail to find representatives in some facets of the nondominated set. These overlooked points are likely near the periphery of the nondominated set (Das and Dennis, 1998). If $\hat{n}$ is not negative, the NBI method may not find any nondominated points not contained in $CHIM - \mathbb{R}^p_{\geqq}$. Later we will show an example to see why CHIM based algorithms do not work in some cases.

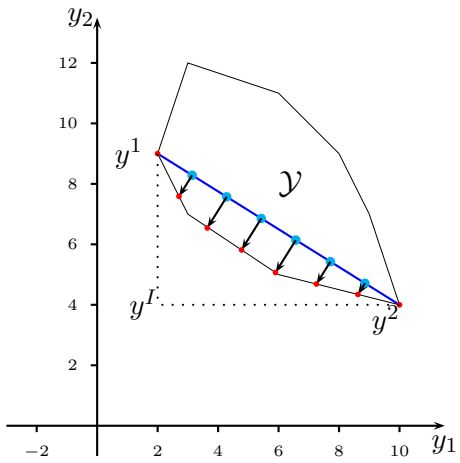

Figure 7.2: *NBI method.*

Although Das and Dennis (1998) claim that this method does compute evenly distributed nondominated points, they do not provide bounds on the spacing of the

resulting points, which means the uniformity of the discrete representative is not measured.

The Limitation of CHIM Based Algorithms Consider the linear relaxation of an assignment problem with three objectives. The cost matrices of the three objectives are

$$c^1 = \begin{pmatrix} 3 & 6 & 4 & 5 \\ 2 & 3 & 5 & 4 \\ 3 & 5 & 4 & 2 \\ 4 & 5 & 3 & 6 \end{pmatrix}, c^2 = \begin{pmatrix} 2 & 3 & 5 & 4 \\ 5 & 3 & 4 & 3 \\ 5 & 2 & 6 & 4 \\ 4 & 5 & 2 & 5 \end{pmatrix}, c^3 = \begin{pmatrix} 4 & 2 & 4 & 2 \\ 4 & 2 & 4 & 6 \\ 4 & 2 & 6 & 3 \\ 2 & 4 & 5 & 3 \end{pmatrix}.$$

Define $\mathcal{Y}' = \{y \in \mathbb{R}^3 : Cx \leqq y \leqq \hat{y} \text{ for some } x \in \mathcal{X}\}$ with $\hat{y} = (21, 21, 21)$, which is greater than the anti ideal point $(20, 20, 20)$.

In Figure 7.3, the four circles that represent points $(11, 11, 14)$, $(19, 14, 10)$, $(15, 9, 17)$ and $(13, 16, 11)$ are the nondominated extreme points of $\mathcal{Y}'$. The nondominated set consists of a line segment from point $(11, 11, 14)$ to point $(19, 14, 10)$ and a face which is the convex hull of $(11, 11, 14)$, $(19, 14, 10)$ and $(13, 16, 11)$.

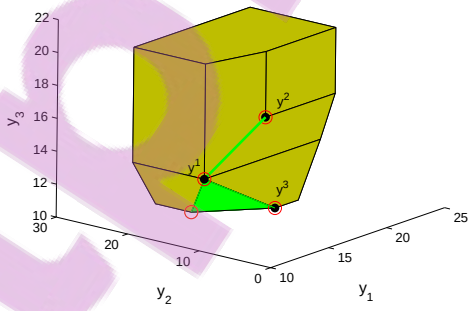

Figure 7.3: *$\mathcal{Y}'$ and the nondominated set.*

The three dots in Figure 7.3 represent the (unique) individual minima of the three objectives, $y^1 = (11, 11, 14)$, $y^2 = (15, 9, 17)$, $y^3 = (19, 14, 10)$. The normal of the CHIM is $\hat{n} = (1, -40, -28)$, which is not negative. Placing reference points on the CHIM, we can not find nondominated points on the face defined by $(11, 11, 14)$,

$(19, 14, 10)$ and $(13, 16, 11)$. Therefore, for this example, the CHIM based algorithms NBI and NC do not work very well.

7.4 Revised Normal Boundary Intersection

The global shooting method has the advantage of guaranteeing coverage, and the NBI method can produce evenly distributed nondominated points. Hence, we propose a revised NBI method that combines the two. This revised NBI method has the advantage of guaranteeing coverage and uniformity.

Instead of the CHIM, the revised NBI method uses the subsimplex $\hat{\mathcal{S}}$ of the simplex $\mathcal{S}$ that is used in the global shooting method as the reference plane (the same as global shooting method, we assume that $\hat{y}$ exists, thus the reference plane is bounded). By doing this, we overcome the limitations of the NBI method, i.e, we have the property of coverage. By solving subproblems similar to (7.1), we obtain evenly spaced nondominated points.

Thus, the revised normal boundary intersection method involves choosing a reference plane, placing equidistant reference points on the plane and computing the intersection point of the normal of the plane through reference points and the boundary of $\mathcal{Y}$. At last, we need to check if the intersection point is nondominated or not because not every intersection point is nondominated. In the following paragraphs, we explain the details of the revised NBI method.

Reference Plane. Here we use the subsimplex $\hat{\mathcal{S}}$ of the simplex $\mathcal{S}$ used in the global shooting method (Benson and Sayin, 1997) as the reference plane.

Let

$$\hat{\beta} = \min\{e^T y : y \in \mathcal{Y}\}, \tag{7.2}$$

where $e \in \mathbb{R}^p$ is a vector in which each entry is 1.

Define $p+1$ points $v^k \in \mathbb{R}^p$, $k = 0, 1, \ldots, p$. Let $v^0 = y^{AI}$ and, for $k = 1, 2, \ldots, p$,

let

$$v_l^k = \begin{cases} y_l^{AI}, & \text{if } l \neq k, \\ \hat{\beta} + y_k - \langle e, v^0 \rangle, & \text{if } l = k, \end{cases} \tag{7.3}$$

$l = 1, 2, \ldots, p$. Then the convex hull $\mathcal{S}$ of $\{v^k : k = 0, 1, \ldots, p\}$ is a p-dimensional simplex, and $\mathcal{S}$ contains $\mathcal{Y}$, as shown by Benson and Sayin (1997).

The subsimplex of $\mathcal{S}$ given by the convex hull $\hat{\mathcal{S}}$ of $\{v^k : k = 1, 2, \ldots, p\}$ is the reference plane. It is a supporting hyperplane of $\mathcal{Y}_N$ with normal e.

Equidistant Points on the Reference Plane. We place equidistant reference points on $\hat{\mathcal{S}}$. For $p = 2$, $\hat{\mathcal{S}}$ is a line segment. For $p = 3$, $\hat{\mathcal{S}}$ is an equilateral triangle in the three dimensional objective space. Therefore, we can use a triangular lattice to produce the equidistant points, see Figure 7.4.

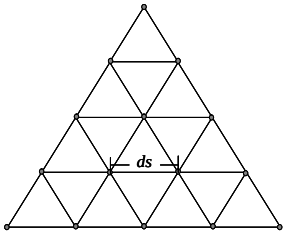

Figure 7.4: *Equidistant reference points on the reference plane.*

In the general case of p objectives, $\hat{\mathcal{S}}$ is a $p-1$ dimensional simplex with equal edge length and with the normal direction e according to the construction of $\mathcal{S}$. The ith reference point q^i is given by

$$q^i = \sum_{k=1}^{p} \rho_k^i v^k$$

where $0 \leq \rho_k^i \leq 1$ and $\sum_{k=1}^{p} \rho_k^i = 1$. By varying ρ_k from 0 to 1 with a fixed increment of η_k an evenly distributed set of points on the reference plane can be generated. For the three objective case in Figure 7.4 $\eta_k = 0.25$.

Computing the Intersection Points and Checking Nondominance. Given a reference point q on $\hat{\mathcal{S}}$, the revised NBI subproblem searches for the closest point

to the reference point on the boundary of $\mathcal{Y}$ along the normal direction e. The revised NBI subproblem is as follows:

$$\min\{t : q + te \in \mathcal{Y}, t \geqq 0\}. \tag{7.4}$$

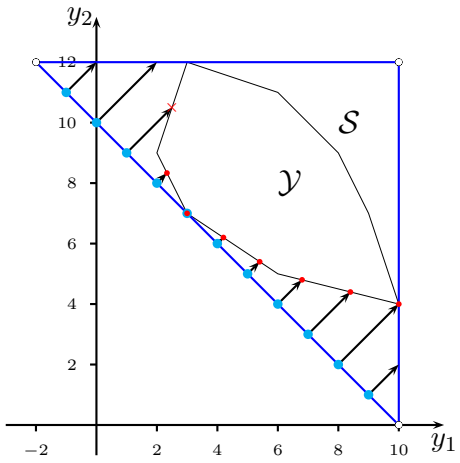

Figure 7.5: *Solutions obtained in the revised NBI method.*

There are three scenarios for the solution of (7.4), as we can see in Figure 7.5 (the same example as Figure 7.1 and Figure 7.2).

1. There is no intersection between the normal and the boundary of $\mathcal{Y}$.

2. The normal and the boundary of $\mathcal{Y}$ intersect, but the intersection point is dominated.

3. The intersection point is nondominated.

If LP (7.4) is infeasible, then there is no intersection between the normal and the boundary of $\mathcal{Y}$, else there is an intersection point. Not every intersection point is a nondominated point. Therefore, we need to check whether it is dominated or not.

A simple nondomination filter can be used to exclude some of the dominated points (Messac *et al.*, 2003). This method has the advantage of being fast, but it

may accept some of the dominated points which are near the boundary of $\mathcal{Y}_N$ as nondominated.

An exact way to check nondominance is according to the following theorem.

Theorem 7.4.1. *Assume that $\lambda \in \mathbb{R}^p_>$ and $\bar{y} \in \mathcal{Y}$. Then $\bar{y}$ belongs to $\mathcal{Y}_N$ if and only if $\bar{y}$ is an optimal solution to the the following problem*

$$\min\{\lambda^T y : y \leqq \bar{y}, y \in \mathcal{Y}\}. \tag{7.5}$$

The reader is referred to Ehrgott (2005) for a proof. By solving (7.5), we can get rid of all the dominated points that remain after filtering. In our implementation we have used $\lambda = e$.

Analysis of the Nondominated Points Given a nondominated facet, the angle between the reference plane and the plane of the nondominated facet is

$$\cos\theta = \frac{\hat{m}^T \hat{n}}{||\hat{m}||\,||\hat{n}||}. \tag{7.6}$$

Here, $\hat{m} \in \mathbb{R}^p$, $\hat{n} \in \mathbb{R}^p$ are the normal vector of the reference plane and the plane of the nondominated facet, respectively.

Because the normal $\hat{m}$ of the reference plane is $e \in \mathbb{R}^p$ (7.6) can be written as:

$$\cos\theta = \frac{\hat{n}_1 + \cdots + \hat{n}_p}{\sqrt{(\hat{n}_1)^2 + \cdots + (\hat{n}_p)^2}\sqrt{p}}. \tag{7.7}$$

According to Theorem 3.3.4 and Definition 3.3.5, a set $\mathcal{F} \in \mathbb{R}^p$ is a face of $\mathcal{Y}_N$ of the MOLP (3.2) if and only if $\mathcal{F}$ equals the optimal solution set $\mathcal{Y}^*(\lambda)$ of the problem

$$\min\{\lambda^T y : y \in \mathcal{Y}\} \tag{7.8}$$

for some $\lambda \in \mathbb{R}^p_>$. Therefore, we know $\hat{n} \in \mathbb{R}^p_>$ and we have

$$\begin{aligned} & \frac{\hat{n}_1 + \cdots + \hat{n}_p}{\sqrt{(\hat{n}_1)^2 + \cdots + (\hat{n}_p)^2}\sqrt{p}} \\ > \quad & \frac{\hat{n}_1 + \cdots + \hat{n}_p}{\sqrt{(\hat{n}_1 + \cdots + \hat{n}_p)^2}\sqrt{p}} = \frac{1}{\sqrt{p}}. \end{aligned} \tag{7.9}$$

When $\hat{m} = k\hat{n}, k \neq 0$, we have $\cos\theta = 1$. So the range of $\cos\theta$ is

$$\frac{1}{\sqrt{p}} < \cos\theta \leqq 1 \tag{7.10}$$

and θ is in the range of $0 \leqq \theta < \arccos \frac{1}{\sqrt{p}}$.

If $p = 2$, $0 \leqq \theta < \frac{\pi}{4}$. If $p = 3$, $0 \leqq \theta < \arccos \frac{\sqrt{3}}{3}$. We see that as p increases, the range of angles between the reference plane and the plane of a nondominated face can increase.

Suppose we have equidistant reference points with distance ds on the reference plane. This implies that the distance between the nondominated points can be calculated as $ds/\cos\theta$.

Figure 7.6 shows an example with two objectives ($p = 2$). The nondominated facets are line segments. $\mathcal{F}_1$ is a nondominated facet, while $\mathcal{F}_2$ is a weakly nondominated facet. The biggest possible angle between the nondominated face and the reference plane is $\frac{\pi}{4}$. The angle between the reference plane and the weakly nondominated facet is $\theta = \frac{\pi}{4}$. The distance between the nondominated points obtained by the revised NBI method is between ds and $\sqrt{2}ds$.

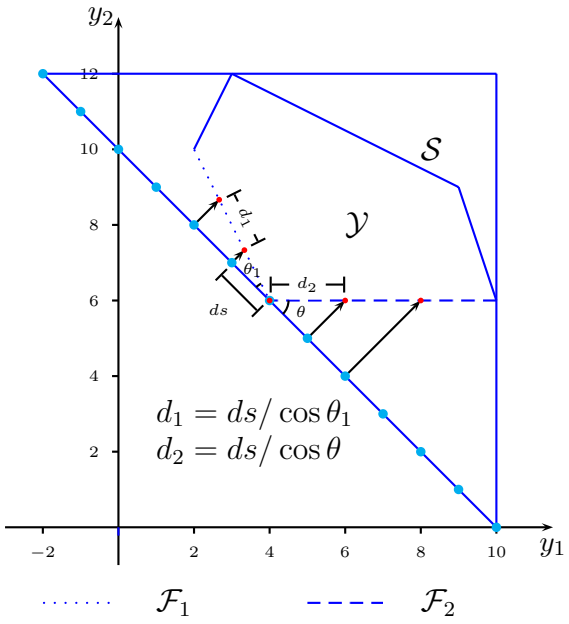

Figure 7.6: *Discrepancy analysis for a 2D example.*

For p objectives, the distance between the nondominated points is $ds \leqq d < \sqrt{p}ds$. As p increases, the range of the distance d between the representative nondominated points on $\mathcal{Y}_N$ increases.

Now, we have the following results.

Theorem 7.4.2. *Let $\mathcal{R}$ be the representative subset of $\mathcal{Y}_N$ obtained by the revised NBI method. $\mathcal{R}$ is a ds-uniform representation of $\mathcal{Y}_N$.*

Proof. We had shown that the distance of two neighbor nondominated points in $\mathcal{R}$ is in between ds and $\sqrt{p}ds$. Therefore, we have $\min_{r^1,r^2\in\mathcal{R},r^1\neq r^2}\{d(r^1,r^2) \geqq ds\}$. □

The width $w(\mathcal{A})$ of a convex set $\mathcal{A} \in \mathbb{R}^p$ is defined as the smallest Eudlidean distance between two supporting hyperplanes. Since this means that the width of any convex set of dimension less than p is 0 but we want to measure the width of projections of $\mathcal{Y}_N$, we also define the width of a convex set on a hyperplane $\hat{\mathcal{S}}$, $w_{\hat{\mathcal{S}}}(\mathcal{A})$ as the minimal distance between two parallel supporting hyperplanes perpendicular to $\hat{\mathcal{S}}$.

Let $\mathcal{Y}_N^p$ be the projection of $\mathcal{Y}_N$ to the reference plane $\hat{\mathcal{S}}$. $\mathcal{Y}_N^p$ can be represented by a union of K convex sets $\bigcup_{k=1}^K \mathcal{O}_k$ on $\hat{\mathcal{S}}$.

Theorem 7.4.3. *If each $\mathcal{O}_k$ of $\mathcal{Y}_N^p$ satisfies $w_{\hat{\mathcal{S}}}(\mathcal{O}_k) \geqq ds$, then the representative subset $\mathcal{R}$ obtained by the revised NBI method is a $d_{\sqrt{p}ds}$-representation of $\mathcal{Y}_N$.*

Proof. Let $\mathcal{T}$ be the set of reference points which corresponds to $\mathcal{R}$. Let the set of reference points both in $\mathcal{T}$ and $\mathcal{O}_k$ be $\mathcal{T}_{\mathcal{O}_k}$. Since the width of $\mathcal{O}_k, k = 1\ldots, K$ on the reference plane $\hat{\mathcal{S}}$ is greater than or equal to ds, then $\mathcal{T}_{\mathcal{O}_k} \neq \emptyset$.

For a point $o \in \mathcal{O}_k$, there must exist $t \in \mathcal{T}_{\mathcal{O}_k}$ such that $d(o,t) \leqq ds$ because the distance of reference points is ds. Project point o to $\mathcal{Y}_N$ and get z. Similarly, project t to $\mathcal{Y}_N$ and get $r \in \mathcal{R}$. Then $d(z,r) \leqq \sqrt{p}ds$ as the angle between the reference plane and the nondominated facets is at most $\arccos\frac{1}{\sqrt{p}}$. Therefore, for a nondominated point $z \in \mathcal{Y}_\mathcal{N}$, there must exist $r \in \mathcal{R}$ such that $d(z,r) \leqq \sqrt{p}ds$. □

The above results quantify the quality of representation in terms of coverage and uniformity. The parameter of the method is the distance ds between reference points. As ds decreases, the cardinality of $\mathcal{R}$ increases, the coverage error decreases, and the uniformity decreases.

If the number of objectives is not very big, then we think the revised NBI method finds quality representations. Moreover, to the best of our knowledge, this is the first

method that allows the computation of a discrete representative set with guaranteed coverage and uniformity measures.

7.5 Numerical Results

We apply the revised normal boundary intersection method to the beam intensity optimization problem of radiation therapy planning.

The algorithm was implemented in Matlab 7.1(R14) using CPLEX 10.0 as LP solver and the tests were run on a dual processor CPU with 1.8 GHz and 1 GB RAM.

In Table 7.1, we list the number of reference points (RFP), the number of intersection points between the normal and the boundary of $\mathcal{Y}$ (ITP), the number of nondominated points (NDP), the distance ds between reference points, and the computation time (CPU) in seconds for calculating the nondominated points for each case. For all three cases, more than half of the reference points do not produce intersection points. No intersection means that LP (7.4) is infeasible. Detecting infeasibility is simple, so the reference points that do not yield intersection points do not contribute much to the computation time. Moreover, we can see from the prostate and pancreatic lesion cases in Table 7.1, that not every intersection point corresponds to a nondominated point. Therefore, it is necessary to check nondominance even though it takes time.

The computation time is related to the number of reference points which corresponds to the number of LPs to be solved. Therefore, for the same case, more reference points need more computation time as we can see in Table 7.1.

We show the nondominated points of the three clinical cases in Figure 7.7 and 7.8. We can see from these pictures that the nondominated points are evenly distributed. The revised NBI method overcomes the deficiency of the NBI method, i.e., the calculated nondominated points cover the whole nondominated set. As long as we have enough equidistant points on the reference plane, the nondominated points produced will be a good representation of the nondominated set according to

Table 7.1: *Data for the application problems.*

	RFP	ITP	NDP	*ds*	CPU
AN	378	72	72	1.04	21.235
PR	378	144	112	4.79	27.163
PL	378	145	129	3.31	96.972
AN	153	29	29	1.59	10.265
PR	153	62	48	7.30	19.548
PL	153	59	54	5.06	42.129

coverage, uniformity and cardinality, the three attributes of discrete representation.

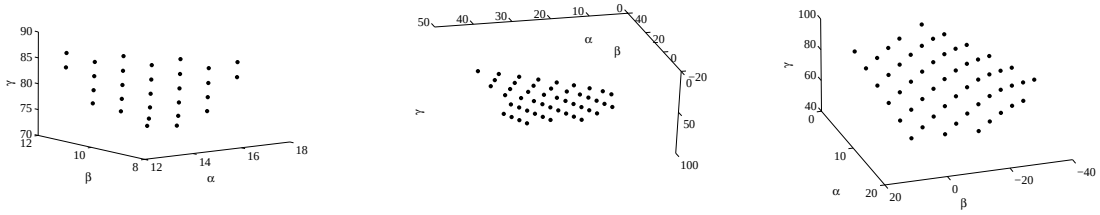

Figure 7.7: *Pictures from left to right are the nondominated points of acoustic, prostate and pancreatic lesion with 153 reference points.*

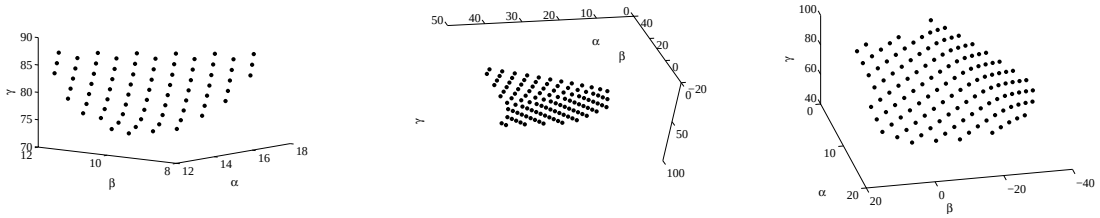

Figure 7.8: *Pictures from left to right are the nondominated points of acoustic, prostate and pancreatic lesion with 378 reference points.*

7.6 Summary

In this chapter, we have addressed the problem of finding well distributed nondominated points for an MOLP. A revised normal boundary intersection method was proposed. By combining features of the normal boundary intersection method and the global shooting methods it overcomes the limitation of CHIM based algorithms. This is the first method for which quality guarantees for coverage and uniformity have been proved. Moreover, numerical results on intensity optimization problems from radiotherapy treatment planning showed that the nondominated points were indeed evenly distributed in practice. The issue of choosing a final solution from amongst the discrete representation is an issue that deserves further study. It is amenable to the large variety of methods of multicriteria decision analysis (Figueira *et al.*, 2005). We show how to use representations for decision support in Chapter 8.

Chapter 8

Case Study

In Chapters 4 and 6 we developed an approximation version of Benson's algorithm and an approximate dual variant of Benson's algorithm to approximate the nondominated set of an MOLP. These algorithms have been used to solve the beam intensity optimization problem for simplified two dimensional clinical cases in previous chapters. In this chapter, we apply these algorithms to the beam intensity optimization problem for three dimensional clinical cases. To limit computation time, we reduce the problem size by reducing resolution in the normal tissue (Lim, 2002). Three different reduced versions of the problem have been tried and the solution quality has been measured by comparing their nondominated sets with the nondominated set of the full version of the problem. We tested the reductions on the 2D clinical cases. We found that normal tissue voxels some distance away from the PTV and OARs do not need to be considered in the optimization, thereby reducing the number of constraints without sacrificing solution quality. Then we use the best reduction in 3D clinical cases. Moreover, representative subsets of the nondominated set for each case are calculated with the revised normal boundary intersection method. We use an example to illustrate how a representative subset can help a decision maker select a treatment.

8.1 Reducing the Computation Time

IMRT aims at delivering a high uniform dose to the tumor while sparing organs at risk and normal tissues by using several intensity modulated beams. Therefore, normal tissues that are some distance away from the PTV get less radiation than those that are close to the PTV. Thus, for optimization, it may suffice to compute dose only on a subset of the normal tissue voxels that are significant for the optimization problem and use this subset to enforce constraints and to formulate their contribution to the objectives.

Lim (2002) includes only a subset of the normal tissue voxels in a beam intensity optimization problem. The subset of the normal tissue voxels includes the normal tissue voxels that are close to the target or lie in an OAR. Additionally, for the rest of the normal tissue voxels every eighth voxel (for 3D cases) is also included into the subset. He shows that the computation time is improved dramatically by reducing the number of voxels in the normal tissue without sacrificing solution quality for a pancreatic cancer case.

Küfer *et al.* (2003) exploit similarity of constraints and apply a clustering procedure to fuse the voxels which form only slightly different constraints to a voxel cluster, thereby reducing the number of constraints.

We have formulated the beam intensity optimization problem in (3.1) in Chapter 3. In this section, we carry out an experiment on how reducing resolution (i.e. increasing voxel size) in the normal tissue (including only a subset of the normal tissue voxels in the optimization) affects the result of (3.1). Four different versions of the problem, the full version N and reduced versions N_1, N_2 and N_3 have been used for the experiment. Reduced version N_1 only includes the normal tissue voxels that are close to the PTV or OARs or skin surface in the optimization. N_2 only includes normal tissue voxels that are close to the PTV or OARs while N_3 only includes normal tissue voxels that are close to the PTV. The three 2D clinical cases, acoustic neuroma (AN), prostate (PR) and pancreatic lesion (PL) used in previous chapters are used for the experiment. Figure 8.1 shows the voxels for the

four different versions of the problem. In the pictures, red represents PTV, grey represents normal tissues and cyan represents critical organs. The number of voxels in the normal tissue for the four versions of the problem is shown in Table 8.1.

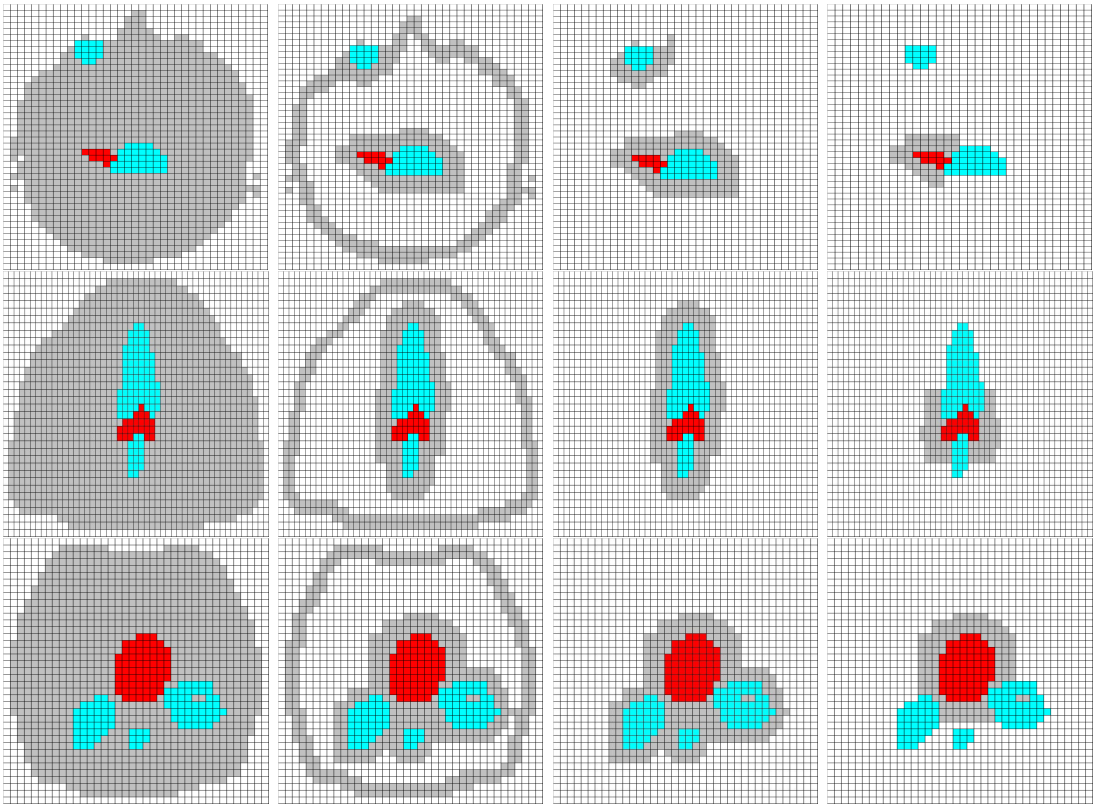

Figure 8.1: *Pictures from left to right show the voxels for the full version N, reduced versions N_1, N_2 and N_3 of the problem, pictures from top to bottom are for AN, PR and PL case, respectively.*

Table 8.1: *Number of voxels in the normal tissue for the full version N, the reduced versions N_1, N_2 and N_3 of the problem for AN, PR and PL cases.*

Case	N	N_1	N_2	N_3
AN	999	380	148	44
PR	1182	447	187	70
PL	986	470	225	112

To measure the solution quality of the three reduced versions of the problem, we

compare their nondominated sets with the nondominated set of the full version of the problem. However, it is hard to compare nondominated sets with one another as they include an infinite number of points. Therefore, we use the revised normal boundary intersection method to determine some representative nondominated points. For all four versions of the problem, the same reference plane and reference points are used. To compare N with N_i $(i = 1, 2, 3)$, we compute the average and maximal distance of the nondominated points of N and N_i corresponding to the same reference point. Note that a reference point may result in a nondominated point for both N and N_i, neither, or one of them. Only the cases where nondominated point for both is found are considered.

2D Acoustic Neuroma Case For the acoustic case, we solve the four versions of the problem with the revised normal boundary intersection method. 28 reference points which are uniformly distributed on the triangle simplex with vertices (-2.3243, 12, 87.55), (16.490, -6.814, 87.550) and (16.49002, 12, 68.73568) are used. The computation times are 6.528, 5.728, 5.210 and 5.064 seconds for the full version N, reduced versions N_1, N_2 and N_3 of the problem, respectively. We obtain 7 nondominated points for the full version N, 8 nondominated points for the reduced versions N_1 and N_2, and 11 nondominated points for the reduced version N_3 (see Figure 8.2). The nondominated points for the four versions of the problem are listed in Table 8.2. As the nondominated points for N_1 and N_2 are the same, we list them in the same column.

We calculate the Euclidean distances between the nondominated points of N and their corresponding nondominated points of N_1 and N_3, respectively. For N_1 these are 0.1060, 0.1308, 0.0590, 0.1004, 0.1067, 0.0674 and 0.0944, respectively. The maximum distance is 0.1308 and the average distance is 0.095. For N_3 they are 1.3358, 1.2599, 0.9972, 1.2565, 1.2616, 0.7252 and 1.1679, respectively. The maximum distance is 1.3358 and the average distance is 1.1434. The number of the nondominated points of the reduced versions N_1, N_2 and N_3 of the problem is bigger than that of the full version N. This may be due to the fact that the

Table 8.2: *AN: The nondominated points solved by the revised normal boundary intersection method for four different versions of the problem.*

	N			N_1, N_2			N_3		
	α	β	γ	α	β	γ	α	β	γ
1		-		12.268	11.959	85.418	11.550	11.241	84.701
2		-			-		12.100	11.791	78.979
3	13.588	11.189	82.558	13.527	11.127	82.496	12.817	10.417	81.786
4	14.462	9.972	85.522	14.387	9.897	85.447	13.735	9.245	84.795
5	14.046	11.646	76.744	14.012	11.612	76.710	13.470	11.070	76.168
6	14.854	10.364	79.643	14.796	10.306	79.585	14.129	9.639	78.918
7	15.704	9.123	82.583	15.642	9.062	82.521	14.975	8.395	81.854
8		-			-		15.928	7.257	84.897
9	15.360	10.870	73.877	15.321	10.831	73.838	14.941	10.451	73.459
10	16.121	9.540	76.729	16.066	9.486	76.674	15.447	8.866	76.054
11		-			-		16.225	7.554	78.923

feasible sets of the reduced versions N_1, N_2 and N_3 are larger than the feasible set of the full version N of the problem because removing the constraints for some of the normal tissue voxels enlarges the feasible set in decision space, thereby the feasible set in objective space enlarges, too. We see that the values of the nondominated points always get smaller with versions. According to the Euclidean distances being calculated, we can say that the solution quality of the reduced version N_1 is the same as the solution quality of N_2, it is better than the solution quality of the reduced version N_3. As discussed in Chapter 4, for clinical purposes a computation error of 0.1 Gy is acceptable, thus the solution quality of the reduced versions N_1 and N_2 of the problem is acceptable.

Next, we solve the four versions of the problem with both Benson's algorithm (primal algorithm) and the dual variant of Benson's algorithm (dual algorithm). The

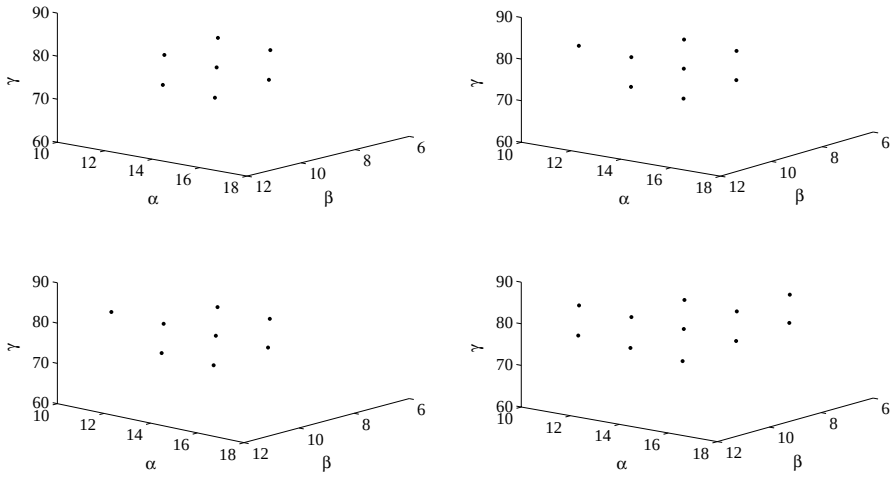

Figure 8.2: *AN: The nondominated points of the full version N (top left), reduced versions N_1 (top right), N_2 (bottom left) and N_3 (bottom right) of the problem.*

number of extreme points of $\mathcal{D}$ is 85 for the full version N, 73 for the reduced version N_1, 73 for the reduced version N_2 and 64 for the reduced version N_3. The number of $\mathcal{K}$-maximal facets of $\mathcal{D}$ is 55, 47, 47, 41 for N, N_1, N_2 and N_3, respectively. The union of the nondominated facets for the four versions of the problem are shown in Figure 8.3. As can be seen in the pictures, the union of the nondominated facets of N_3 is "larger" than that of N, N_1 and N_2. This is because the feasible set enlarges with the versions. Let $\mathcal{Y}_N^0$, $\mathcal{Y}_N^1$, $\mathcal{Y}_N^2$, $\mathcal{Y}_N^3$ be the nondominated set of the full version N and reduced versions N_1, N_2 and N_3. Then $\mathcal{Y}_N^i + \mathbb{R}^3_{\leqq} \subseteq \mathcal{Y}_N^{i-1} + \mathbb{R}^3_{\leqq}$ since we relax more constraints. For the four versions of the problem, Table 8.3 summarizes the computation times and the savings in comparison with the time in column N representing solving the full version of the problem. Comparing columns N and N_1, we see that more than a 60 percent reduction is obtained for the reduced version N_1. Column N_2 shows that more savings (a further 10 percent) are obtained, with essentially no degradation in the solution quality. Although more than 80 percent of the computation time can be saved by using N_3, the solution quality degrades considerably, which is not acceptable.

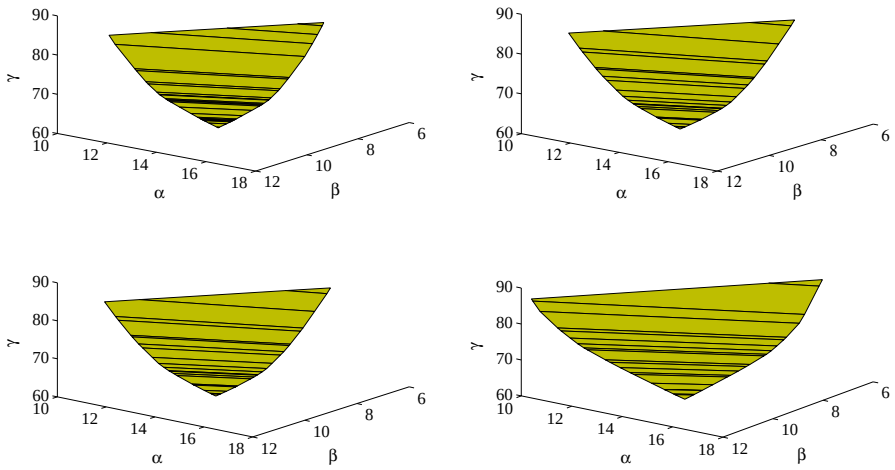

Figure 8.3: *AN: The union of the nondominated facets of $\mathcal{P}$ solved by the dual variant of Benson's algorithm for the full version (top left), reduced versions N_1 (top right), N_2 (bottom left) and N_3 (bottom right) of the problem.*

Table 8.3: *AN: Comparison of computation time for solving four different versions of the problem with different algorithms.*

Algorithm		N	N1	N2	N3
Primal	Time (seconds)	12.704	4.594	3.266	2.031
	Time saved (%)	–	63.84	74.29	84.01
Dual	Time (seconds)	8.188	2.906	2.312	1.422
	Time saved (%)	–	64.51	71.76	82.63

2D Prostate Case For the prostate case, we use 45 reference points and solve the four versions of the problem with the revised normal boundary intersection method. The computation times are 6.937, 6.14, 5.828 and 5.499 seconds for solving the full version N and reduced versions N_1, N_2 and N_3, respectively. 16 nondominated points are obtained for the full version N and reduced versions N_1 and N_2, while 15 nondominated points are obtained for the reduced version N_3 (see Figure 8.4). We list the nondominated points in Table 8.4. As the sets of nondominated points being calculated for the full version N and the reduced versions N_1 and N_2 are the

same, we list them in the same column.

Table 8.4: *PR: The nondominated points solved by the revised normal boundary intersection method for four different versions of the problem.*

	N, N_1, N_2			N_3		
	α	β	γ	α	β	γ
1	6.130	28.124	90.096	6.089	28.083	90.055
2	10.781	24.106	77.409	10.655	23.981	77.284
3	14.748	19.405	90.045	14.697	19.354	89.994
4	16.577	21.234	65.869	16.420	21.077	65.711
5	19.291	15.280	77.251	19.257	15.246	77.217
6	23.472	10.792	90.100		-	
7	24.095	20.083	56.049	23.847	19.836	55.802
8	23.895	11.215	64.518	23.826	11.146	64.449
9	27.957	6.608	77.248	27.925	6.576	77.216
10	32.362	19.682	46.980	31.914	19.234	46.532
11	29.656	8.308	52.942	29.533	8.185	52.819
12	32.482	2.465	64.437	32.460	2.443	64.414
13	41.265	19.917	38.546	41.034	19.686	38.315
14	37.294	7.276	43.242	37.079	7.062	43.028
15	37.070	-1.616	51.687	37.038	-1.648	51.655
16	41.253	-6.101	64.539	41.250	-6.104	64.536

The number of nondominated points being calculated for the reduced version N_3 is less than the number of the nondominated points being calculated for the full version N, reduced versions N_1 and N_2. Some nondominated points disappear with reductions because they are no longer nondominated. If we relax all normal voxels we would end up with only two objectives, i.e., a much smaller nondominated set. The Euclidean distances from the 15 nondominated points of N to their corresponding nondominated points of N_3 are 0.071, 0.218, 0.088, 0.273, 0.059, 0.429,

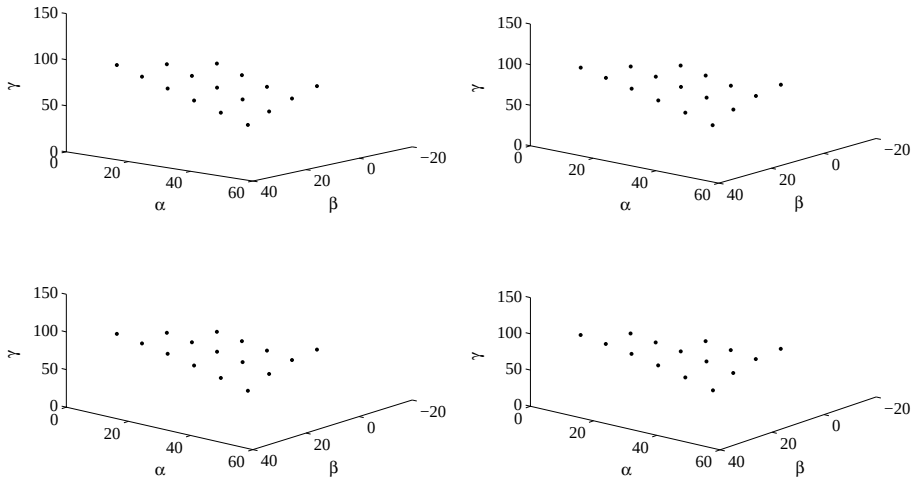

Figure 8.4: *PR: The nondominated points for the full version (top left), the reduced versions N_1 (top right), N_2 (bottom left) and N_3 (bottom right) of the problem.*

0.119, 0.056, 0.775, 0.213, 0.039, 0.400, 0.372, 0.056 and 0.005, respectively. The maximum distance is 0.775 and the average distance is 0.198. This means that the solution quality for different regions of the nondominated set varies a lot. For some regions the quality is good (for example there is a nondominated point of N_3 with Euclidean distance to its corresponding nondominated point of N equal to 0.005) but for others the solution quality is bad (for example there is a nondominated point of N_3 with the Euclidean distance to its corresponding nondominated point of N equal to 0.775). In general, the solution quality of N_3 is not acceptable.

We also solve the four versions of the problem using the approximation version of Benon's algorithm (approximate primal algorithm) and the approximate dual variant of Benson's algorithm (approximate dual algorithm), both with approximation error $\epsilon = 0.1$. Figure 8.5 shows the union of the nondominated facets of $\mathcal{P}^i$ solved by the approximate dual variant of Benson's algorithm. We show the computation time and the savings of the reduced versions of the problem in comparison with the full version of the problem in Table 8.5. Comparing columns N and N_1, we see that more than a 30 percent reduction is obtained by using the reduced version N_1. For the reduced version N_2, a 40 percent saving is obtained when approximately solving

the primal problem while a 54 percent saving is obtained when approximately solving the dual problem. As the reduced versions N_1 and N_2 have the same solution quality as the full version N, we achieve computational savings with N_1 and N_2 without degradation in quality. Around 60 percent of the computation time can be saved by solving the reduced version N_3 approximately, however the solution quality is not acceptable.

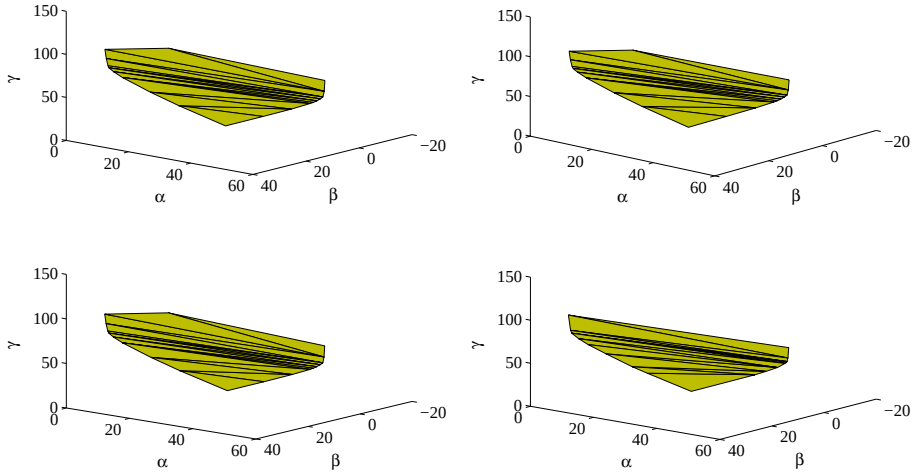

Figure 8.5: *PR: The union of the nondominated facets of $\mathcal{P}^i$ solved by the approximate dual variant of Benson's algorithm with $\epsilon = 0.1$ for the full version (top left), the reduced versions N_1 (top right), N_2 (bottom left) and N_3 (bottom right) of the problem.*

Table 8.5: *PR: Comparison of computation time for solving four different versions of the problem with different algorithms.*

Algorithm		N	N1	N2	N3
Approximate primal	Time (seconds)	4.375	2.969	2.593	1.656
	Time saved (%)	–	32.14	40.73	62.15
Approximate dual	Time (seconds)	14.750	9.265	8.000	6.125
	Time saved (%)	–	37.19	54.24	58.47

2D Pancreatic Lesion Case For the pancreatic lesion case, we use 45 reference points and solve the four different versions of the problem with the revised normal boundary intersection method. The computation time is 15.563 seconds for solving the full version N, 13.862 seconds for solving the reduced version N_1, 12.584 seconds for N_2 and 10.994 seconds for N_3. 20 nondominated points are obtained for the full version while 19 nondominated points are obtained for the reduced versions N_1, N_2 and N_3 of the problem (see Figure 8.6). The nondominated points of the reduced versions N_2 and N_3 are the same, while they are slightly different from the nondominated points of the full version N and the reduced version N_1. Except for nondominated point $(0, -10.98, 85.384)$ depicted by $\odot$ in Figure 8.6, all the nondominated points being calculated for the full version N have corresponding nondominated points in the nondominated set of the reduced versions N_1, N_2 and N_3. Out of the 19 nondominated points of N that have corresponding nondominated points in the nondominated sets of N_1, N_2 and N_3, 13 nondominated points are exactly the same as their corresponding nondominated points of the reduced versions N_1, N_2 and N_3. Those nondominated points are depicted by $\bullet$ in Figure 8.6 and are (4.553, 11.297, 72.214), (2.354, 3.190, 75.922), (0.621, -4.451, 80.097), (9.599, 10.435, 65.443), (7.222, 2.150, 68.974), (5.124, -5.856, 72.784), (3.843, -13.045, 77.411), (14.836, 9.764, 58.864), (12.341, 1.361,62.277), (10.094, -6.794, 65.938), (8.445, -14.350, 70.198), (15.244, -7.551, 59.273) and (13.333, -15.370, 63.270). The 6 nondominated points of N that are slightly different from their corresponding nondominated points of the reduced versions N_1, N_2 and N_3 are listed in Table 8.6. Since N_2 and N_3 have the same nondominated points, we list them in the same column.

The distances between the six nondominated points of N and their corresponding nondominated points of N_1 are 0.103, 0.068, 0.196, 0.001, 0.003 and 0.158, respectively. The maximum distance is 0.196 and the average distance is 0.026. The distances from the six nondominated points of N to their corresponding nondominated points of N_2 are 0.107, 0.084, 0.211, 0.001, 0.003 and 0.206, respectively. The maximum distance is 0.211 while the average distance is 0.031. In Figure 8.6 we use

Table 8.6: *PL: The nondominated points solved by the revised normal boundary intersection method for four different versions of the problem.*

	N			N_1			N_2, N_3		
	α	β	γ	α	β	γ	α	β	γ
1	4.717	-18.078	84.193	4.658	-18.138	84.134	4.655	-18.140	84.131
2	8.219	-20.485	75.879	8.180	-20.524	75.840	8.170	-20.533	75.831
3	10.176	-24.435	83.744	10.063	-24.549	83.631	10.054	-24.557	83.622
4	12.463	-22.149	68.307	12.461	-22.150	68.306	12.461	-22.150	68.306
5	13.650	-26.869	75.402	13.559	-26.960	75.311	13.531	-26.988	75.283
6	16.104	-30.324	83.764	15.997	-30.430	83.657	15.995	-30.432	83.655

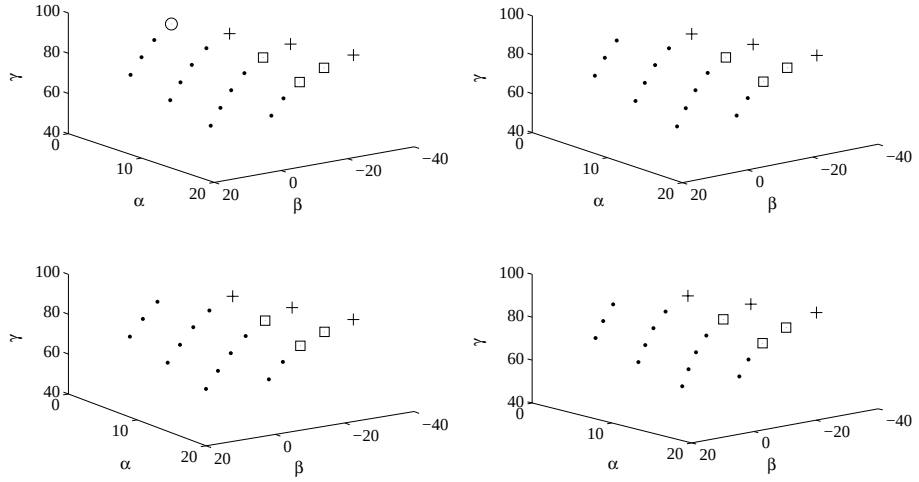

Figure 8.6: *PL: The nondominated points for the full version N (top left), the reduced versions N_1 (top right), N_2 (bottom left) and N_3 (bottom right) of the problem.*

$+$ to represent the nondominated points of N (or N_1, N_2 and N_3) with Euclidean distances to their corresponding nondominated points of N_1, N_2 and N_3 (or N) being greater than 0.1, $\boxdot$ to represent the nondominated points of N (or N_1, N_2 and N_3) with Euclidean distances to their corresponding nondominated points of N_1, N_2

and N_3 (or N) being greater than 0 but less than 0.1. It can be seen that reducing voxels in normal tissue only affects the solution quality for part of the nondominated set. All reduced versions have an acceptable solution quality compared to the full version N.

We solve the four different versions of the problem with both the approximation version of Benson's algorithm and the approximate dual variant of Benson's algorithm with approximation error $\epsilon = 0.1$. Figure 8.7 shows the union of the nondominated facets of $\mathcal{P}^i$ solved by the approximate dual variant of Benson's algorithm. A summary of the computation time saved by using the reduced versions N_1, N_2 and N_3 of the problem is shown in Table 8.7. Comparing columns N and N_1, we see that more than a 20 percent reduction is obtained by using the reduced version N_1. Column N_2 shows that more savings (more than 30 percent) are obtained by using the reduced version N_2. Column N_3 shows that further savings are obtained with no degradation in solution quality compared to N_2.

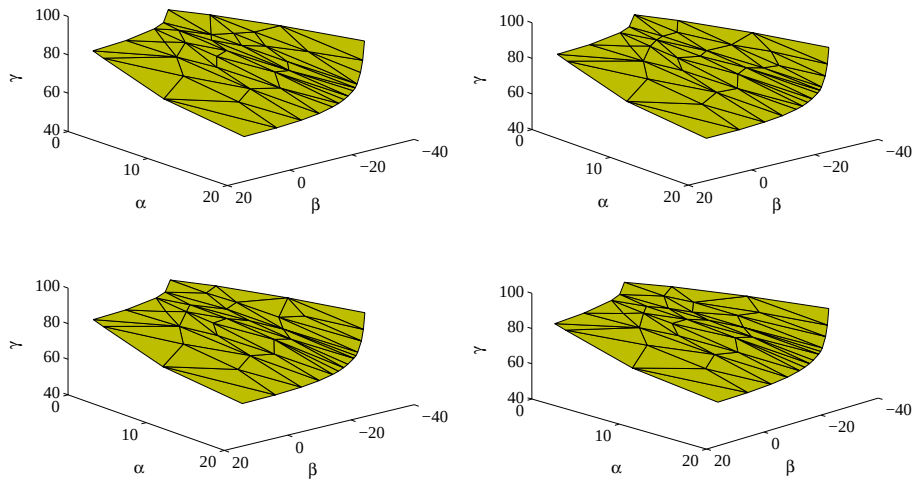

Figure 8.7: *PL: The union of the nondominated facets of $\mathcal{P}^i$ solved by the approximate dual variant of Benson's algorithm with $\epsilon = 0.1$ for the full version N (top left), reduced versions N_1 (top right), N_2 (bottom left) and N_3 (bottom right) of the problem.*

Table 8.7: *PL: Comparison of computation time for solving four different versions of the problem with different algorithms.*

Algorithm		N	N_1	N_2	N_3
Approximate primal	Time (seconds)	57.821	42.406	40.110	37.578
	Time saved (%)	–	26.66	30.63	35.01
Approximate dual	Time (seconds)	164.360	127.313	112.845	98.641
	Time saved (%)	–	22.54	31.34	39.98

Summary For all three 2D cases, the reduced version N_2 which includes the normal tissue voxels close to the PTV or OARs in the optimization has a considerable reduction in the computation time with small degradation in the solution quality. Reduction of computation time is related to the reduction in normal tissue voxels; the more reduction in normal tissue voxels, the more reduction in the computation time.

Including only a subset of the normal tissue voxels into the optimization enlarges the feasible set in objective space (see the AN case), but not necessarily the nondominated set (see the PR and PL cases).

Reducing voxels in normal tissue affects the solution quality although differently throughout the nondominated set. This is seen in the PR and PL cases.

8.2 3D Clinical Cases

We use the approximation version of Benson's algorithm and the approximate dual variant of Benson's algorithm to solve the beam intensity optimization problem of 3D clinical cases. Four 3D clinical cases are used, they are "3D acoustic neuroma (5mm) problem" (AN5), "3D acoustic neuroma (3mm) problem" (AN3), "3D abdomen (5mm) problem" (AD5) and "3D Lung (5mm) problem" (LG5). Here 3mm and 5mm means that each anatomy slice is 3mm or 5mm apart from each other and the voxel size is $3 \times 3 \times 3\text{mm}^3$ or $5 \times 5 \times 5\text{mm}^3$. Figure 8.8 shows one sim-

plified CT image indicating the outline of the tumor and critical organs at risk for each case. These examples are available at `http://lagrange.math.trinity.edu/tumath/research/reports/misc/report97`. For all examples, a total of 9 evenly spaced beams were used at angles $40^o n$, where $n = 0, \ldots, 8$. The number of slices, the number of voxels and bixels used for optimization of each case and the prescription information that defines parameters in (3.1) is shown in Table 8.8. All tests are run on a dual processor CPU with 1.8 GHz and 1 GB RAM. All code is written in MATLAB 7.1(R14) and links to the CPLEX solvers (CPLEX v. 10.0).

Table 8.8: *Number of slices, number of voxels and bixels, lower and upper bounds for tumor, critical organs, and normal tissue (in Gy).*

Case	AN5	AN3	AD5	LG5
Number of slices	2	3	3	3
Tumor voxels	15	78	443	426
Critical organ voxels	113	472	87	1939
Normal tissue voxels	1626	6778	7895	6623
Bixels	597	597	1776	1773
TUB	61.1408	61.1408	54.59	74.16
TLB	57.5793	57.5793	51.41	69.84
CUB	50/5	50/5	20	36/9
NUB	0.00	0.00	0.00	0.00
α_u	11.5159	11.5159	10.282	13.968
β_u	10	10	4	7.2
γ_u	61.1408	61.1408	54.59	74.16

Note: 50 and 5, 36 and 9 are the upper bounds for different critical organs.

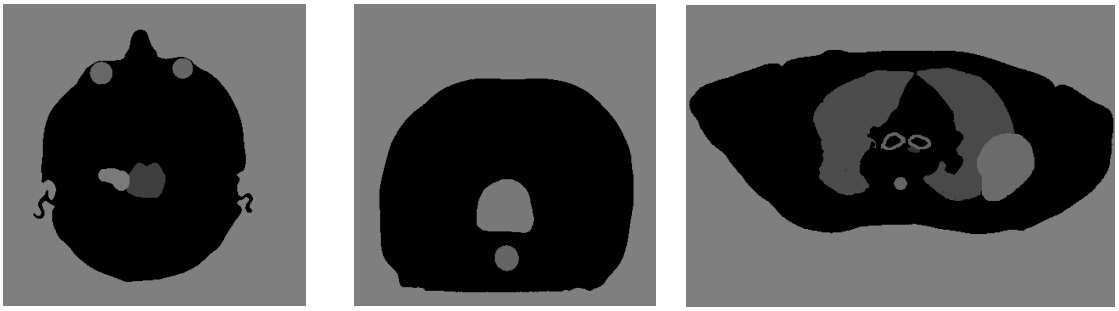

Figure 8.8: *Pictures from left to right show one simplified CT slice for 3D acoustic neuroma, abdomen and lung cases.*

3D Acoustic Neuroma (5mm) Case We solve the full version of the problem with the revised normal boundary intersection method. 3003, 1653 and 703 reference points which are uniformly distributed on the triangle simplex with the vertices (-49.92043, 10, 61.1408), (11.51586, -51.43629, 61.1408) and (11.51586, 10, -0.29549) are used. In Table 8.9, we list the number of reference points (RFP), the number of intersection points (ITP) between the normal and the boundary of $\mathcal{Y}$, the number of nondominated points (NDP), the distance ds between reference points, and the computation time (CPU) in seconds for calculating the nondominated points. As can be seen, with 3003 reference points, we only obtain 675 intersection points and 60 nondominated points. With 1653 reference points, we only obtain 387 intersection points and 33 nondominated points while with 703 reference points, we only obtain 168 intersection points and 15 nondominated points. This means around 23 percent of the reference points correspond to intersection points while only 2 percent of them correspond to nondominated points.

Moreover, we solve the full version problem with the approximation version of Benson's algorithm and the approximate dual variant of Benson's algorithm. Table 8.10 shows the number of vertices and cuts of $\mathcal{Y}'^o$ and $\mathcal{D}^o$, the computation time for solving the problem with different approximation error ϵ. As can be seen, with the same approximation error, solving the dual approximately is faster than solving the primal approximately.

Figure 8.9 shows the results of solving the full version of the problem with dif-

Table 8.9: *AN5: The results of solving the problem with the revised normal boundary intersection method.*

RFP	ITP	NDP	*ds*	CPU (seconds)
3003	675	60	1.1000	239.491
1653	387	33	1.4726	101.610
703	168	15	2.2278	47.218

Table 8.10: *AN5: The results of solving the primal and the dual approximately with different approximation error ϵ.*

AN5	ϵ	Vertices	Cuts	CPU (seconds)
$\mathcal{Y}'^o$	0.1	17	12	8.297
	0.01	79	48	32.328
	0.001	406	223	157.422
$\mathcal{D}^o$	0.1	13	8	5.735
	0.01	57	32	19.39
	0.001	316	170	108.312

ferent solution methods. The picture on the top left is a representative subset of the nondominated set. It is solved by the revised normal boundary intersection method with 1653 reference points and there are 33 nondominated points in total. The picture on the top right shows $\mathcal{Y}'^o$ solved by the approximation version of Benson's algorithm with approximation error 0.001. The pictures on the bottom are the results of solving the problem approximately with the approximate dual variant of Benson's algorithm with approximation error 0.001, the one on the left is $\max_{\mathcal{K}} \mathcal{D}^o$ while the one on the right is the union of the nondominated facets of $\mathcal{P}^i$.

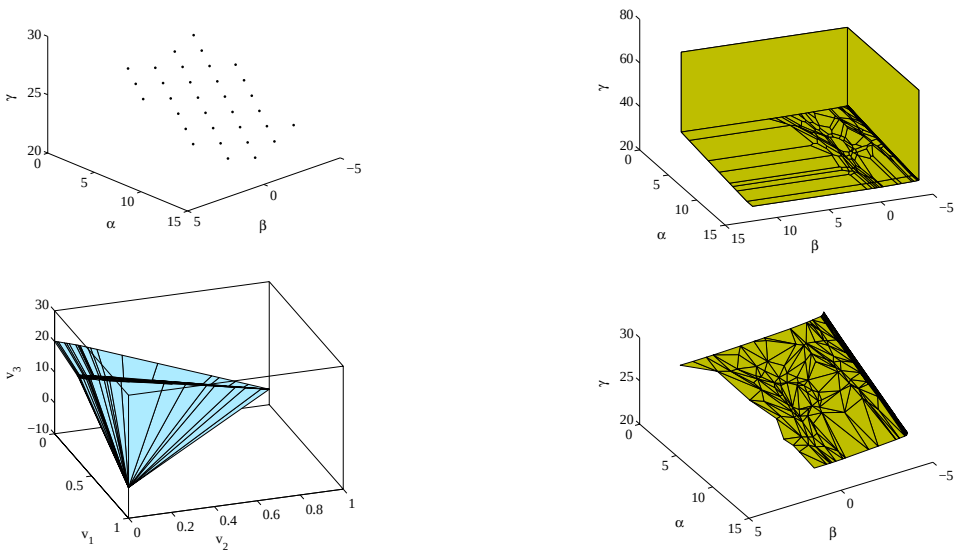

Figure 8.9: *AN5: A representative subset of the nondominated set (top left),* $\mathcal{Y}'^{o}$ *(top right),* $\max \mathcal{D}^{o}$ *(bottom left) and the union of the nondominated facets of* $\mathcal{P}^{i}$ *(bottom right).*

3D Acoustic Neuroma (3mm) Case For the 3D acoustic neuroma (3mm) case, the full version N of the problem has 6778 voxels in the normal tissue. We also consider to solve the reduced version N_2, which only includes the normal tissue voxels close to the PTV or OARs in the optimization. The reduced version N_2 has 906 normal tissue voxels.

We use 153 reference points and solve the full version N and the reduced version N_2 of the problem with the revised normal boundary intersection method. 22 nondominated points are obtained for both versions of the problem, see Figure 8.10. The maximum distance and the average distance between the nondominated points of N to their corresponding nondominated points of N_2 are 0.0137 and 0.002483, respectively. Therefore, the solution quality of the reduced version N_2 is acceptable.

Also we solve the two versions of the problem with the revised normal boundary intersection method using 528 and 78 reference points, respectively. Table 8.11 summarizes the number of reference points, the number of intersection points between the normal and the boundary of $\mathcal{Y}$, the number of nondominated points, the

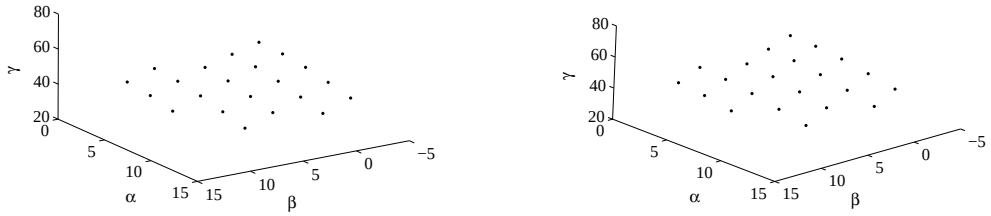

Figure 8.10: *AN3: The nondominated points for N (left) and N_2 (right).*

distance ds between reference points, the computation time for calculating the nondominated points and the time savings of solving N_2. As can be seen, solving the reduced version N_2 can save more than 60 percent of the computation time.

Table 8.11: *AN3: The results of solving the full version N and the reduced version N_2 of the problem with the revised NBI method.*

Version	RFP	ITP	NDP	ds	CPU (seconds)	Time saved (%)
N	528	163	71	1.9364	419.795	–
N_2	528	163	71	1.9366	148.316	64.67
N	153	54	22	3.4654	139.281	–
N_2	153	54	22	3.4654	47.359	66.00
N	78	28	10	4.7031	117.694	–
N_2	78	28	10	4.7031	27.441	76.68

Moreover, we solve the problem with the approximation version of Benson's algorithm and the approximate dual variant version of Benson's algorithm, with different approximation error ϵ. Table 8.12 shows the number of vertices and cuts of $\mathcal{Y}'^{o}$ and $\mathcal{D}^{o}$, the computation time and the time savings of solving the reduced version N_2 compared to solving the full version N of the problem. For solving the same version of the problem approximately with the same approximation error, the approximate dual variant of Benson's algorithm is more than 3 times faster than the approximation version of Benson's algorithm. For example, solving the full version N of the problem approximately with approximation error $\epsilon = 0.1$, the approximate

dual variant of Benson's algorithm needs 88 seconds while the approximation version of Benson's algorithm needs 283 seconds.

Table 8.12: *AN3: The results of solving the primal and dual approximately with different approximation error ϵ.*

AN3	ϵ	Version	Vertices	Cuts	CPU (seconds)	Time saved (%)
$\mathcal{D}^o$	0.1	N	41	22	88.010	–
		N_2	41	22	46.953	46.66
	0.02	N	141	77	288.619	–
		N_2	132	72	150.028	48.02
$\mathcal{Y}'^o$	0.1	N	59	36	283.404	–
		N_2	51	32	132.483	53.25
	0.02	N	228	125	934.323	–
		N_2	245	134	521.878	44.13

Some results of solving the two versions of the problem are shown in Figure 8.11. The pictures on the left are the results of the full version N of the problem while the pictures on the right are the results of the reduced version N_2 of the problem. From top to bottom, they are $\mathcal{Y}'^o$ solved by the approximation version of Benson's algorithm with approximation error 0.02, $\max_{\mathcal{K}} \mathcal{D}^o$ and the union of the nondominated facets of $\mathcal{P}^i$ solved with the approximate dual variant of Benson's algorithm with approximation error 0.02.

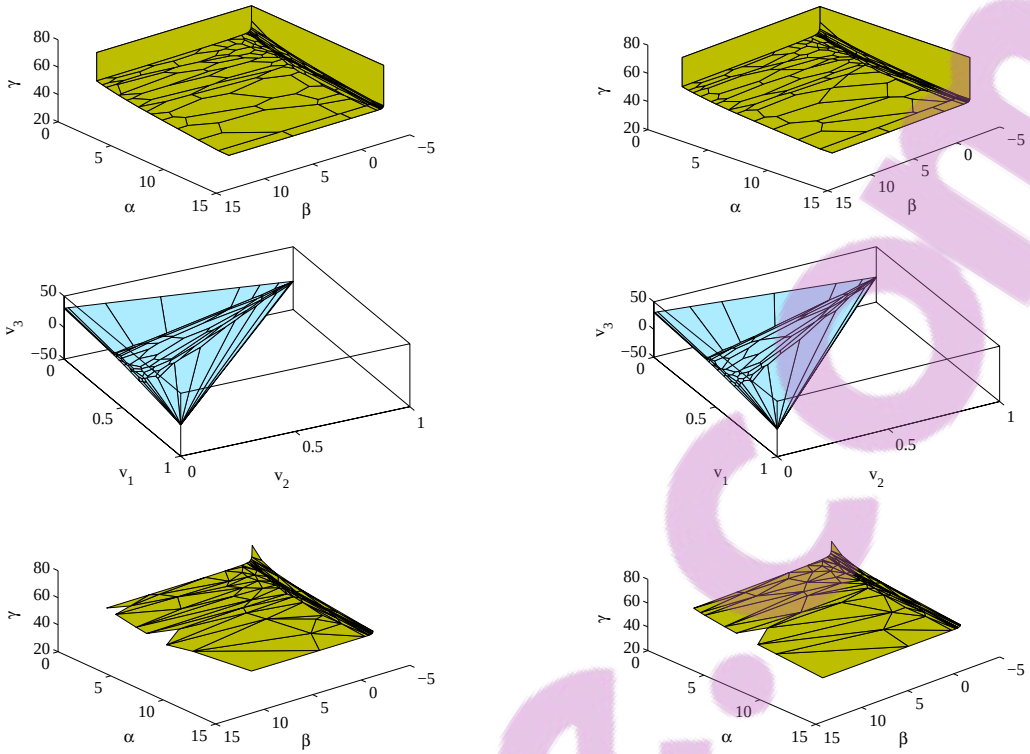

Figure 8.11: *AN3: Pictures from top to bottom are* $\mathcal{Y}'^{o}$, $\max \mathcal{D}^{o}$ *and the union of the nondominated facets of* $\mathcal{P}^{i}$. *The pictures on the left are for the full version* N *while the pictures on the right are for the reduced version* N_2.

3D Abdomen (5mm) Case For the abdomen (5mm) case, the full version of the problem has 7895 voxels in the normal tissue. To improve computation time, we solve the reduced version N_2 of the problem. It has 845 voxels in the normal tissue.

The revised normal boundary intersection method is used to analyze the solution quality of the reduced version N_2 of the problem. With 153 reference points, the full version obtains 28 nondominated points while the reduced version of the problem obtains 27 nondominated points, see Figure 8.12. The maximum distance between the nondominated points of the full version N and their corresponding nondominated points of the reduced version N_2 is 0.1371 and the average distance is 0.089. Therefore, we think that the solution quality of the reduced version N_2 is acceptable.

We also solve the two versions of the problem with the revised normal boundary

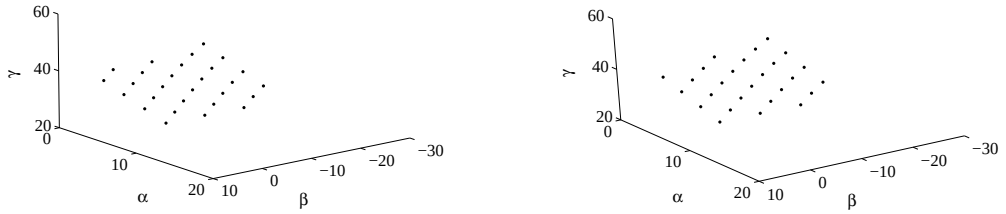

Figure 8.12: *AD5: The nondominated points for the full version (left), the reduced version N_2 (right) solved by the revised NBI method.*

intersection method using 528 and 78 reference points, respectively. Table 8.13 shows the number of reference points, the number of intersection points between the normal and the boundary of $\mathcal{Y}$, the number of nondominated points, the distance ds between reference points, the computation time and the time savings for solving N_2. As can be seen, compared to the full version, solving the reduced version N_2 can save more than 80 percent of the computation time.

Table 8.13: *AD5: The results of solving the full version N and the reduced version N_2 of the problem with the revised NBI method.*

Version	RFP	ITP	NDP	ds	CPU (seconds)	Time saved (%)
N	528	182	85	2.0605	23662.056	–
N_2	528	182	83	2.0605	4299.858	81.83
N	153	55	28	3.6872	6538.917	–
N_2	153	55	27	3.6872	1176.835	82.00
N	78	27	15	5.004	3796.794	–
N_2	78	26	13	5.004	696.668	81.65

Comparisons for the full version N of the problem versus the reduced version N_2 with the approximation version of Benson's algorithm and the approximate dual variant of Benson's algorithm are shown in Table 8.14. Solving both versions of the problem with the approximate dual variant of Benson's algorithm with approximation error 0.1, we see that the reduced version needs 2021 seconds while the

full version needs 5600 seconds, this means solving the reduced version can save 64 percent computation time. Solving both versions of the problem with the approximation version of Benson's algorithm with approximation error 0.1, the reduced version needs 6207 seconds while the full version needs 14800 seconds, thus a 58 percent saving is achieved by solving the reduced version N_2 of the problem.

Table 8.14: *AD5: The results of solving the primal and the dual approximately with different approximation error ϵ for the full version N and the reduced version N_2.*

AD5	ϵ	Version	Vertices	Cuts	CPU (seconds)	Time saved (%)
$\mathcal{D}^o$	0.1	N	51	27	5600	–
		N_2	51	27	2021	63.91
	0.01	N	341	179	38218	–
		N_2	350	183	14580	61.85
$\mathcal{Y}'^o$	0.1	N	73	43	14800	–
		N_2	77	45	6207	58.06
	0.01	N	595	316	115570	–
		N_2	624	330	48412	58.11

Some results of solving both versions of the problem with different algorithms are shown in Figure 8.13. The pictures on the left are the results of the full version N while the pictures on the right are for the reduced version N_2. From top to bottom, they are $\mathcal{Y}'^o$ solved by the approximation version of Benson's algorithm with approximation error 0.01, $\max_{\mathcal{K}} \mathcal{D}^o$ and the union of the nondominated facets of $\mathcal{P}^i$ solved by the approximate dual variant of Benson's algorithm with approximation error 0.01.

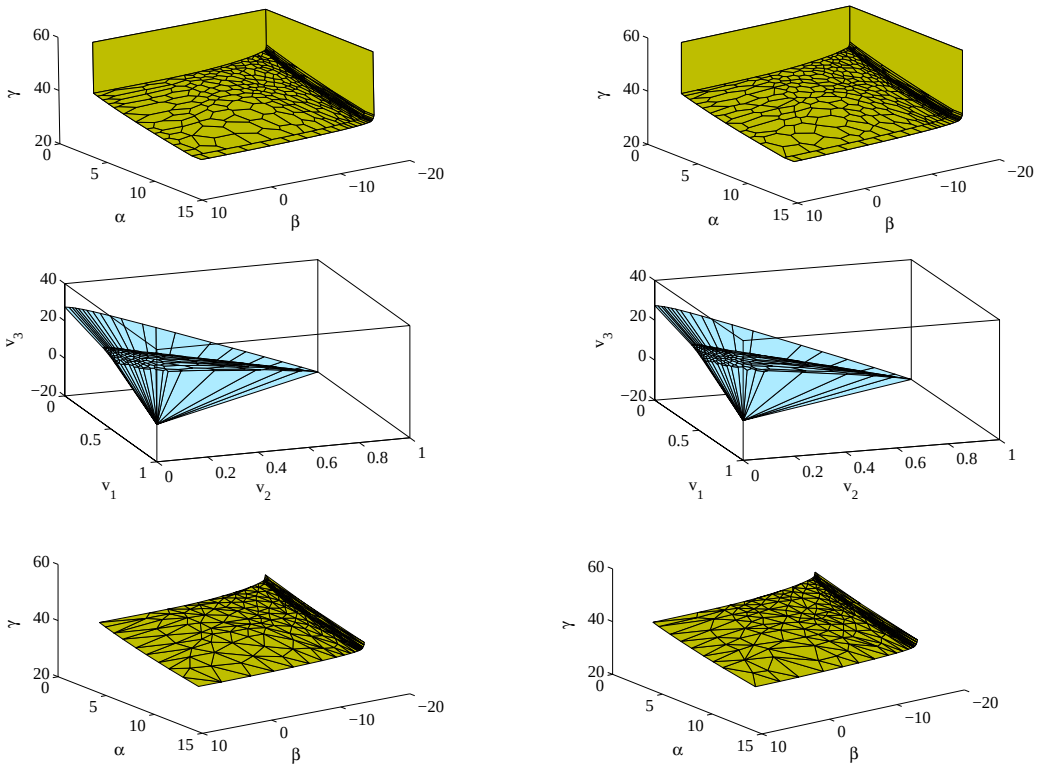

Figure 8.13: *AD5: Pictures from top to bottom are* $\mathcal{Y}'^{o}$, $\max \mathcal{D}^{o}$ *and the union of the nondominated facets of* $\mathcal{P}^{i}$. *The pictures on the left are for the full version* N *while the pictures on the right are for the reduced version* N_2.

3D Lung (5mm) Case For the 3D lung (5mm) case, the full version N of the problem has 6623 voxels in normal tissue. To reduce the computation time, we consider the reduced version N_2 of the problem. The number of normal tissue voxels of the reduced version N_2 is 1190.

To measure the solution quality of the reduced version N_2, we solve both versions of the problem with the revised normal boundary intersection method. The same reference plane is used for both versions and total 28 reference points are used for the calculation. 13 nondominated points are obtained for both N and N_2 (see Figure 8.14) and listed in Table 8.15.

The Euclidean distances between the 13 nondominated points of N_2 to their corresponding nondominated points of N are 0.0512, 0.0443, 0.0751, 0.0789, 0.0568,

Table 8.15: *LG5: The nonnominated points obtained with the revised NBI method for the full version N and the reduced version N_2 of the problem.*

	N			N_2		
	α	β	γ	α	β	γ
1	1.527	4.413	66.546	1.498	4.383	66.517
2	3.743	1.802	73.589	3.718	1.776	73.563
3	2.475	5.361	53.014	2.432	5.317	52.970
4	3.983	2.042	59.348	3.937	1.996	59.303
5	6.354	-0.414	66.546	6.321	-0.447	66.513
6	9.005	-2.589	74.024	8.970	-2.625	73.988
7	5.802	3.861	46.687	5.769	3.828	46.654
8	6.864	0.096	52.576	6.796	0.028	52.508
9	9.164	-2.431	59.703	9.123	-2.472	59.661
10	11.906	-4.515	67.271	11.862	-4.560	67.227
11	9.879	3.111	41.111	9.835	3.067	41.066
12	10.196	-1.399	46.254	10.131	-1.464	46.189
13	12.201	-4.220	53.086	12.153	-4.269	53.038

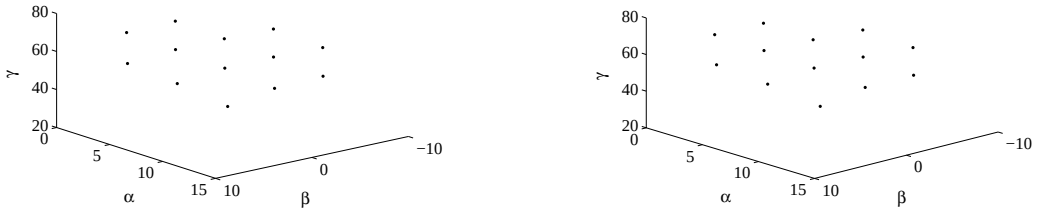

Figure 8.14: *LG5: The nondominated points for the full version (top left), the reduced versions N_2 (top right).*

0.0618, 0.0561, 0.1187, 0.0718, 0.0769, 0.0763, 0.1130 and 0.0834, respectively. The the maximum distance is 0.1187 and the average distance is 0.0742. This means that the reduced version N_2 is acceptable.

We also solve the two versions of the problem with the revised normal boundary

intersection method using 78 and 153 nondominated points. Table 8.16 summarizes the number of reference points, the number of intersection points and the number of nondominated points, the distance between the reference points, the computation time and the time savings of solving the reduced version N_2 compared to solving the full version N. Solving the reduced version N_2 saves more than 60 percent of the computation time compared to solving the full version N of the problem.

Table 8.16: *LG5: The results of solving the full version N and the reduced version N_2 of the problem with the revised NBI method.*

Version	RFP	ITP	NDP	*ds*	CPU (seconds)	Time saved (%)
N	28	13	13	6.826	1841.968	–
N_2	28	13	13	6.826	655.276	64.42
N	78	33	33	4.3882	5562.188	–
N_2	78	34	34	4.3882	1560.578	71.95
N	153	60	58	3.2334	9907.878	–
N_2	153	60	58	3.2334	2860.863	71.13

Moreover, we solve versions N and N_2 of the problem with the approximation version of Benson's algorithm and the approximate dual variant of Benson's algorithm. Table 8.17 shows the computation time, the number of vertices and cutting planes of $\mathcal{D}^o$ and $\mathcal{Y}'^o$. Solving the two versions of the problem with the approximate dual variant of Benson's algorithm, when approximation error $\epsilon = 0.1$, we find N_2 can save 51.9 percent of the computation time; when $\epsilon = 0.3$, the reduced version N_2 saves 52.06 percent of the computation time. Solving both versions of the problem with the approximation version of Benson's algorithm, with approximation error 0.1, the reduced version N_2 saves 29.23 percent of the computation time, and with approximation error 0.3, solving the reduced version N_2 saves 24.7 percent of the computation time. This means solving the reduced version N_2 of the problem achieves a modest improvement in the computation time with an acceptable solution quality.

Table 8.17: *LG5: The results of solving the primal and the dual approximately with different approximation error ϵ for the full version N and the reduced version N_2 of the problem.*

LG5	ϵ	Version	Vertices	Cuts	CPU (seconds)	Time saved (%)
$\mathcal{Y}'^o$	0.1	N	142	86	39864	-
		N_2	149	89	28211	29.23
	0.3	N	54	36	15608	-
		N_2	56	37	11751	24.70
$\mathcal{D}^o$	0.1	N	68	34	7458	-
		N_2	72	36	3598	51.90
	0.3	N	29	14	3302	-
		N_2	31	15	1583	52.06

Figure 8.15 shows the results of both versions of the problem solved by different solution approaches. The pictures on the left are for the full version N while the pictures on the right are for the reduced version N_2. From top to bottom, they are $\mathcal{Y}'^o$ solved by the approximation version of Benson's algorithm with approximation error 0.1, $\max_{\mathcal{K}} \mathcal{D}^o$ and the union of the nondominated facets of $\mathcal{P}^i$ solved by the approximate dual variant of Benson's algorithm with approximation error 0.1.

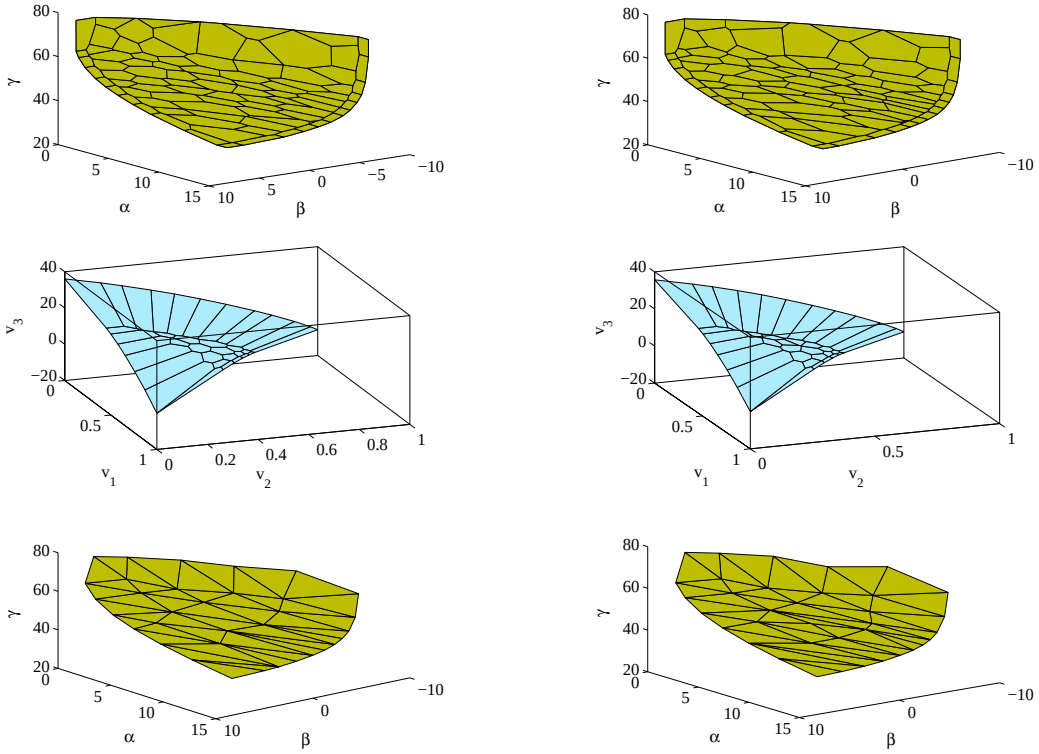

Figure 8.15: *LG5: Pictures from top to bottom are* $\mathcal{Y}'^{o}$, $\max \mathcal{D}^{o}$ *and the union of the nondominated facets of* $\mathcal{P}^{i}$. *The pictures on the left are for the full version* N *while the pictures on the right are for the reduced version* N_2.

8.3 Decision Support

Accessing a representative subset of the nondominated set allows treatment planners and radiation oncologists to evaluate how changes in the dose to some structure affects the doses to other structures.

Each nondominated point corresponds to a treatment plan. The planner evaluates the quality of a plan by looking at its isodose curves and DVHs. The isodose curves are the level curves of the radiation per anatomy slice, where high doses of radiation (red curves on the graphs) are focused on the tumor. The DVHs plot the dose (relative to TLB) versus the percentage of volume of the tumor or OAR.

We take the 3D acoustic neuroma (3mm) case (see the leftmost picture in Figure

8.8) as an example to show how a representative subset of the nondominated set can help a decision maker select a treatment. We recall that the lower bound and the upper bound for the target are 57.58 and 61.14 Gy, respectively. The upper bound for the brain stem is 50 Gy while the upper bound for the eye sockets is 5 Gy.

In Section 8.2, we obtained a representative subset of the nondominated set by solving the problem with the revised normal boundary intersection method using 153 reference points. The representative subset comprises 22 nondominated points (Figure 8.16). The Euclidean distance between the reference points is 3.4654. The representative subset shows the tradeoff information between the maximum deviation α from tumor lower bounds, the maximum deviation β from critical organ upper bounds and the maximum deviation γ from normal tissue upper bounds. These nondominated points and the corresponding treatment plans are stored in a database for a decision maker to analyze.

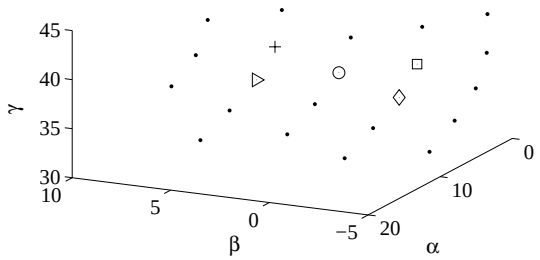

Figure 8.16: *AN3: A representative subset of the nondominated set.*

For each plan, we plot its isodose curves and its DVH curves for treatment plan evaluation. In DVH plots, there are five curves, which are for the PTV, normal tissue, left eye socket, right eye socket and brain stem, respectively.

The planner might look at the representative nondominated points in Figure 8.16, initially considering a plan that is in the middle of the figure such as $\odot$ with the objectives $(3.882, 2.366, 36.354)$, with dose distribution and DVHs shown in Figure 8.17, but then, considering an improvement of the objective value for the tumor, which is achievable with tolerable deteriorating to the objective values for the critical organs and normal tissue, the planner may decide on the plan that is depicted by $+$

in Figure 8.16 with the objectives $(2.663, 6.048, 37.585)$, the dose distribution and DVHs are shown Figure 8.18. Alternatively, the planner may also consider a plan with better objective value for the critical organs and the normal tissue, such as the plan depicted by $\diamondsuit$ in Figure 8.16 with objectives $(5.231, -1.185, 35.253)$. Its dose distribution and DVHs are shown in Figure 8.19.

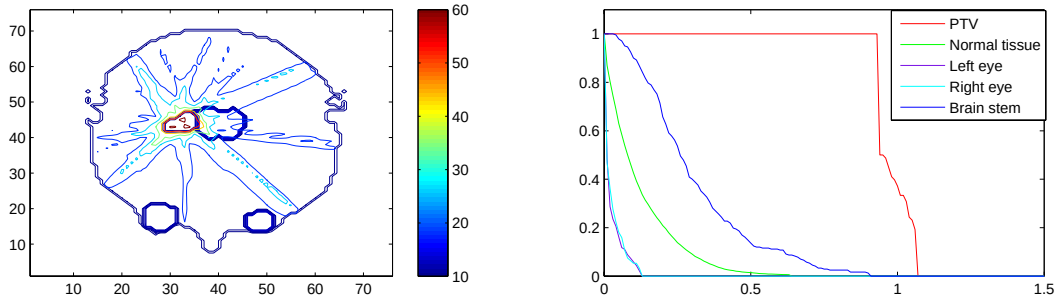

Figure 8.17: *AN3: Dose distribution and DVHs for the plan with objectives (3.882, 2.366, 36.354) depicted by $\odot$ in Figure 8.16.*

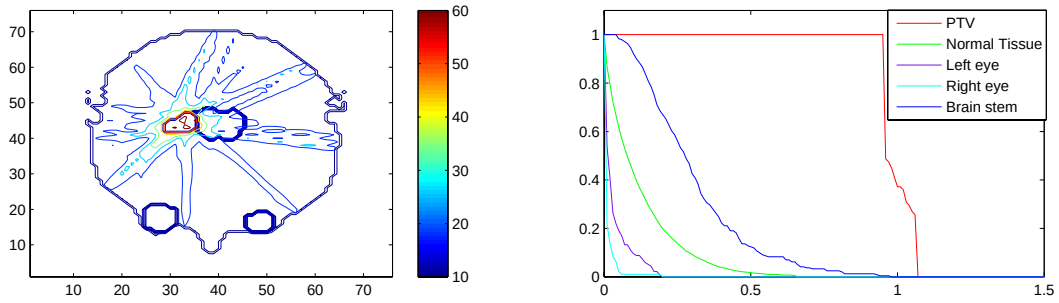

Figure 8.18: *AN3: Dose distribution and DVHs for the plan with objectives (2.663, 6.048, 37.585) depicted by $+$ in Figure 8.16.*

Additionally, if the planner wants a better objective value for the tumor by deteriorating the objective value for the normal tissue, he or she may consider the plan depicted by $\boxdot$ in Figure 8.16 with the objectives $(2.770, -1.196, 37.693)$, the dose distribution and the DVHs are shown in Figure 8.20. If he or she wants a better objective value for the normal tissue, he or she may also consider the plan depicted

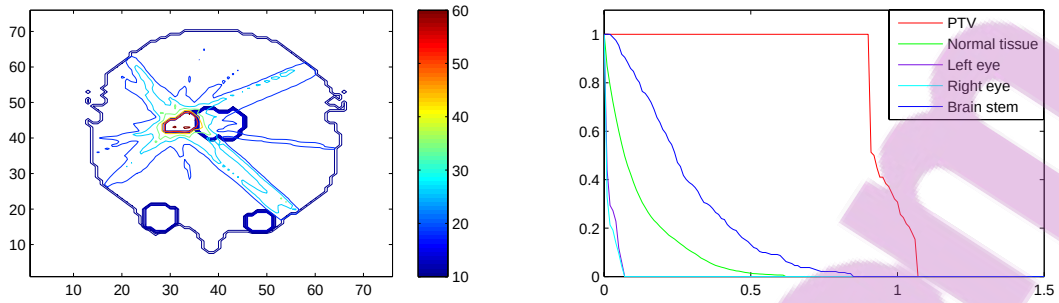

Figure 8.19: *AN3: Dose distribution and DVHs for the plan with objectives (5.231, -1.185, 35.253) depicted by ⋄ in Figure 8.16.*

by ▷ in Figure 8.16 with the objectives $(5.148, 6.083, 35.170)$. Its dose distribution and DVHs are shown in Figure 8.21.

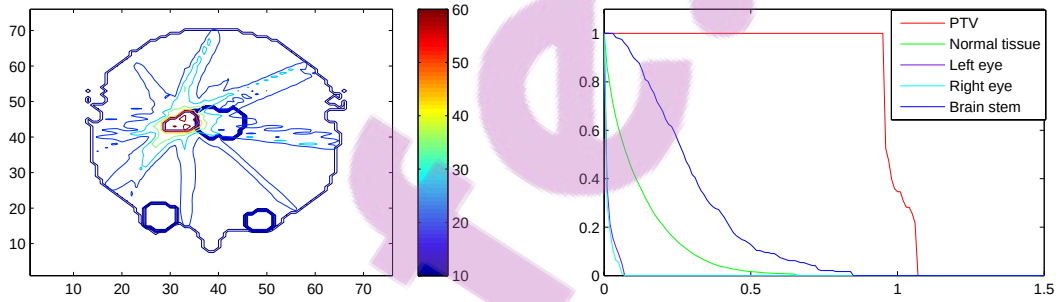

Figure 8.20: *AN3: Dose distribution and DVHs for the plan with objectives (2.770, -1.196, 37.693) depicted by ⊡ in Figure 8.16.*

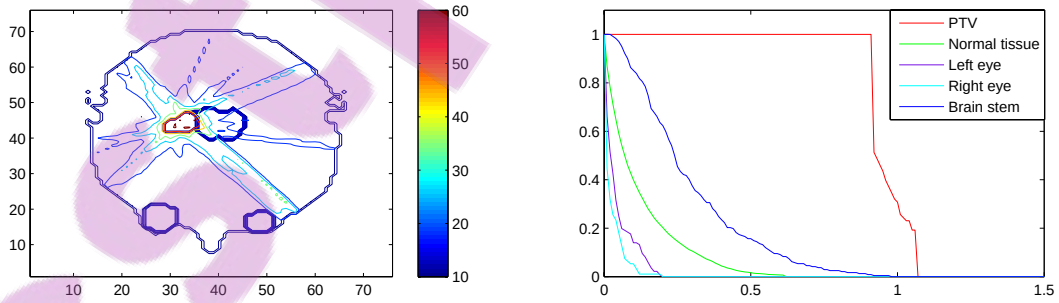

Figure 8.21: *AN3: Dose distribution and DVHs for the plan with objectives (5.148, 6.083, 35.170) depicted by ▷ in Figure 8.16.*

8.4 Summary

We summarize this chapter as follows.

- For all 2D cases and 3D cases we tested, the reduced version N_2 of the problem has an acceptable solution quality with a modest improvement of the computation time. We had showed that clinical cases with 3mm voxel size can be solved. Therefore, it is possible for us to solve a reduced version of the beam intensity optimization problem for three dimensional cases with clinically relevant 2mm spaced CT slices and a voxel size of 2mm using computational resources that are available at leading cancer treatment centres.

- The computation savings for a reduced version of the problem varies for different clinical cases. It seems related to the number of voxels being eliminated.

- Reducing the resolution in normal tissue enlarges the feasible region in objective space. This is verified by the results of solving the problem with the revised normal boundary intersection method. However, the nondominated set can become "larger" or "smaller".

- We observed that the nondominated sets of AN5, AN3 and AD5 are not as "curved" as the nondominated set of LG5 case. Using the revised normal boundary intersection method to find nondominated points for AN5, AN3 and AD5 is not very efficient because many reference points do not intersect with the feasible set in objective space and also because some of the intersection points are actually dominated by other points in objective space.

- Accessing a representative subset of the nondominated set allows treatment planners and oncologists to evaluate how changes in the dose to some structure affects the doses to other structures. Thus, it can help them to select a treatment plan.

Chapter 9

Conclusion

IMRT is a complicated process. Inverse planning of IMRT involves solving three optimization problems, the beam angle optimization problem, the beam intensity optimization problem and the realization problem. In this thesis, we focused on the beam intensity optimization problem and we have formulated the beam intensity optimization as a multiobjective linear programming problem as well as developed solution methods to solve MOLPs.

Two objective space methods for exactly finding the nondominated set of an MOLP have been studied. They are Benson's outer approximation algorithm and a dual variant of Benson's algorithm, which solve the primal MOLP and dual MOLP, respectively. Based on the two exact solution methods, we developed an approximation version of Benson's algorithm and an approximate dual variant of Benson's algorithm. We proved that they are guaranteed to find ε-nondominated sets. Moreover, a revised normal boundary intersection method is proposed for finding a representative subset of the nondominated set.

Application of these methods to the beam intensity optimization problem for 2D and 3D clinical cases shows that the approximation methods improve solution time with an acceptable solution quality. Comparing the primal method with the dual method, the dual method always shows a computational advantage in our experiments.

Throughout this work we have pointed to interesting questions for future re-

search.

- Since the solutions of the beam angle optimization, the beam intensity optimization and the realization problems affect each other, combining the three optimization problems should be considered.
- We notice that the inefficiency of the revised NBI method for some of the clinical cases is due to the reference plane being much bigger than the projection area of the nondominated set onto the plane. Therefore, some strategy which only uses the reference points in a small extension of the projection area should be developed.
- Due to the memory limitation of MATLAB, some other languages such as C++ should be used for implementations.

References

Agmon, S. (1954). The relaxation method for linear inequalities. *Canadian Journal of Mathematics*, **6**, 382–392.

Ahuja, R. and Hamacher, H. (2004). A network flow algorithm to minimize beam-on-time for unconstrained multileaf collimator problems in cancer radiation therapy. *Networks*, **45**(1), 36–41.

Alber, M. and Nüsslin, F. (2001). A representation of an NTCP function for local complication mechanisms. *Physics in Medicine and Biology*, **46**, 439–447.

Alber, M. and Reemtsen, R. (2007). Intensity modulated radiotherpay treatment planning by use of a barrier-penalty multiplier algorithm. *Optimization Methods and Software*, **22**(3), 391–411.

Bahr, G., Kereiakes, J., Horwitz, H., Finney, R., Galvin, J., and Goode, K. (1968). The method of linear programming applied to radiation treatment planning. *Radiology*, **91**, 686–693.

Bednarz, G., Michalski, D., Houser, C., Huq, M., Xiao, Y., Anne, P., and Galvin, J. (2002). The use of mixed-integer programming for inverse treatment planning with pre-defined field segments. *Physics in Medicine and Biology*, **47**, 2235–2245.

Bednarz, G., Michalski, D., Anne, P., and Valicenti, R. (2004). Inverse treatment planning using volume-based objective functions. *Physics in Medicine and Biology*, **49**, 2503–2514.

Benson, H. P. (1998a). Further analysis of an outcome set-based algorithm for multiple-objective linear programming. *Journal of Optimization Theory and Applications*, **97**(1), 1–10.

Benson, H. P. (1998b). Hybrid approach for solving multiple-objective linear programs in outcome space. *Journal of Optimization Theory and Applications*, **98**, 17–35.

Benson, H. P. (1998c). An outer approximation algorithm for generating all efficient extreme points in the outcome set of a multiple objective linear programming problem. *Journal of Global Optimization*, **13**, 1–24.

Benson, H. P. and Sayin, S. (1997). Towards finding global representations of the efficient set in multiple objective mathematical programming. *Naval Research Logistics*, **44**, 47–67.

Benson, H. P. and Sun, E. (2000). Outcome space partition of the weight set in multiobjective linear programming. *Journal of Optimization Theory and Applications*, **105**(1), 17–36.

Boland, N., Hamacher, H. W., and Lenzen, F. (2004). Minimizing beam-on time in cancer radiation treatment using multileaf collimators. *Networks*, **43**, 226–240.

Bortfeld, T. (1999). Optimized planning using physical objectives and constraints. *Seminars in Radiation Oncology*, **9**, 20–34.

Bortfeld, T., Burkelbach, J., Boesecke, R., and Schlegel, W. (1990). Method of image reconstructions from projections applied to confromation therapy. *Physics in Medicine and Biology*, **35**(10), 1423–1434.

Bortfeld, T., Boyer, A. L., Schlegel, W., Kahler, D. L., and Waldron, T. J. (1994). Realisation and verification of three-dimensional conformal radiotherapy with modulated fields. *International Journal of Radiation Oncology, Biology, Physics*, **30**, 899–908.

Bortfeld, T., Stein, J., and Preiser, K. (1997). Clinically relevant intensity modulated optimization using physical criteria. In D. Leavitt, editor, *XIIth International Conference on the Use of Computers in Radiation Therapy*, pages 1–4, Salt Lake City, Utah, U.S.A. Madison Medical Physics Publishing.

Brahme, A. (2001). Individualizing cancer treatment: biological optimization models in treatment planning and delivery. *International Journal of Radiation Oncology, Biology, Physics*, **49**(2), 327–337.

Brahme, A. and Agren, A. K. (1987). Optimal dose distribution for eradication of heterogeneous tumours. *Acta Oncologica.*, **26**, 377–85.

Burkard, R. E., Hamacher, H. W., and Rote, G. (1991). Sandwich approximation of univariate convex functions with an application to separable convex programming. *Naval Research Logistics*, **38**, 911–924.

Carlsson, F. and Forsgren, A. (2006). Iterative regularization in intensity-modulated radiation therapy optimization. *Medical Physis*, **33**(1), 225–234.

Carlsson, F., Forsgren, A., Rehbinder, H., and Eriksson, K. (2006). Using eigenstructure of the hessian to reduce the dimension of the intensity modulated radiation therapy optimization problem. *Annals of Operations Research*, **148**(1), 81–94.

Censor, Y., Altschuler, M., and Powlis, W. (1988a). A computational solution of the inverse problem in radiation therapy treatment planning. *Applied Mathematics and Computation*, **25**, 57–87.

Censor, Y., Altschuler, M., and Powlis, W. (1988b). On the use of Cimmino's simultaneous projections method for computing a solution of the inverse problem in radiation therapy treatment planning. *Inverse Problems*, **4**, 607–623.

Chen, P. C. and Hansen, P. (1991). On-line and off-line vertex enumeration by adjacency lists. *Operations Research Letters*, **10**, 403–409.

Chen, Y., Michalski, D., Houser, C., and Galvin, J. (2002). A deterministic iterative least-squares algorithm for beam weight optimization in conformal radiotherapy. *Physics in Medicine and Biology*, **47**, 1647–1658.

Cho, P., Lee, S., Marks, R., Oh, S., Sutlief, S., and Phillips, M. (1998). Optimization of intensity modulated beams with volume constraints using two methods: Cost function minimization and projections onto convex sets. *Medical Physics*, **25**, 435–443.

Choi, B. and Deasy, J. (2002). The generalized equivalent uniform dose function as a basis for intensity-modulated treatment planning. *Physics in Medicine and Biology*, **47**, 3579–3589.

Cotrutz, C. and Xing, L. (2002). Using voxel-dependent importance factors for interactive DVH-based dose optimisation. *Physics in Medicine and Biology*, **47**, 1659–1669.

Cotrutz, C. and Xing, L. (2003). Segment-based dose optimisation using a genetic algorithm. *Physics in Medicine and Biology*, **48**, 2987–2998.

Cotrutz, C., Lahanas, M., Kappas, C., and Baltas, D. (2001). A multiobjective gradient-based dose optimization algorithm for external beam conformal radiotherapy. *Physics in Medicine and Biology*, **46**(8), 2161–2175.

Craft, D., Halabi, T., and Bortfeld, T. (2005). Exploration of tradeoffs in intensity-modulated radiotherapy. *Physics in Medicine and Biology*, **50**, 5857–68.

Craft, D., Halabi, T., Shih, H., and Bortfeld, T. (2006). Approximating convex Pareto surfaces in multiobjective radiotherapy planning. *Medical Physics*, **33**, 3399–3407.

Crooks, S. M. and Xing, L. (2002). Application of constrained least-squares techniques to IMRT treatment planning. *International Journal of Radiation Oncology, Biology, Physics*, **54**(4), 1217–1224.

Dai, J. and Zhu, Y. (2003). Conversion of dose-volume constraints to dose limits. *Physics in Medicine and Biology*, **48**, 3927–3941.

Das, I. and Dennis, J. (1997). A closer look at drawbacks of minimizing weighted sums of objectives for pareto set generation in multicriteria optimization problems. *Structural and Multidisciplinary Optimization*, **14**, 63–69.

Das, I. and Dennis, J. E. (1998). Normal-boundary intersection: A new method for generating the pareto surface in nonlinear multicriteria optimization problems. *SIAM Journal on Optimization*, **8**(3), 631–657.

Deasy, J. O. (1997). Multiple local minima in radiotherapy optimization problems with dose-volume constraints. *Medical Physics*, **24**(7), 1157–1161.

Ehrgott, M. (2005). *Multicriteria Optimization*. Springer-Verlag, Berlin, 2nd edition.

Ehrgott, M. and Burjony, M. (2001). Radiation therapy planning by multicriteria optimisation. In *Proceedings of the 36th Annual Conference of the Operational Society of New Zealand*, pages 244–253.

Ehrgott, M. and Gandibleux, X. (2007). Bound sets for biobjective combinatorial optimization problems. *Computers and Operations Research*, **34**(9), 2674–2694.

Ehrgott, M. and Wiecek, M. (2005). Multiobjective programming. In J. Figueira, S. Greco, and M. Ehrgott, editors, *Multicriteria Decision Analysis: State of the Art Surveys*, pages 667–722. Springer Science + Business Media, New York.

Ehrgott, M., Holder, A., and Reese, J. (2005). Beam selection in radiotherapy design. Technical Report 95, Department of Mathematics, Trinity University , USA.

Ehrgott, M., Hamacher, H., and Nußbaum, M. (2007a). Decomposition of matrices and static multileaf collimators: A survey. In C. Alves, P. Pardalos, and L. Vicente, editors, *Optimization in Medicine*, pages 27–48. Springer Verlag, Berlin.

Ehrgott, M., Löhne, A., and Shao, L. (2007b). A dual variant of Benson's outer approximation algorithm. Report 654, Department of Engineering Science, The University of Auckland.

Emami, B., Lyman, J., and Brown, A. (1991). Tolerance of normal tissue to therapeutic irradiation. *International Journal of Radiation Oncology, Biology, Physics*, **21**, 109–22.

Ezzell, G. A. (1996). Genetic and geometric optimization of three-dimensional radiation therapy treatment planning. *Medical Physics*, **23**, 293–305.

Figueira, J., Greco, S., and Ehrgott, M., editors (2005). *Multiple Criteria Decision Analysis: State of the Art Surveys*, volume 78 of *Springer's International Series in Operations Research & Management Science*. Springer Science + Business Media, New York.

Gong, Y. (2007). *Integer Programming Methods for Beam Selection in Radiotherapy Treatment Planning*. M.E. Thesis, Department of Engineering Science, University of Auckland, New Zealand.

Grünbaum, B. (2003). *Convex Polytopes*. Springer, New York, 2nd edition. Graduate Texts in Mathematics 221.

Halabi, T., Craft, D., and T., B. (2006). Dose-volume objectives in multi-criteria optimization. *Physics in Medicine and Biology*, **51**, 3809–3818.

Hamacher, H. and Küfer, K.-H. (2002). Inverse radiation therapy planing – A multiple objective optimization approach. *Discrete Applied Mathematics*, **118**(1-2), 145–161.

Heyde, F. and Löhne, A. (2006). Geometric duality in multi-objective linear programming. Reports on Optimization and Stochastics 06-15, Department of Mathematics and Computer Science, Martin-Luther-University Halle-Wittenberg.

Heyde, F., Löhne, A., and Tammer, C. (2007). Set-valued duality theory for multiple objective linear programs and application to mathematical finance. Reports on

Optimization and Stochastics 07-04, Department of Mathematics and Computer Science, Martin-Luther-University Halle-Wittenberg.

Hodes, L. (1974). Semiautomatic optimization of external beam radiation treatment planning. *Radiology*, **110**, 191–196.

Holder, A. (2003). Designing radiotherapy plans with elastic constraints and interior point methods. *Health Care Management Science*, **6**(1), 5–16.

Holder, A. (2004). Radiotherapy treatment design and linear programming. In M. Brandeau, F. Sainfort, and W. Pierskalla, editors, *Operations Research and Health Care*, pages 741–774. Kluwer Academic Publishers, Norwell MA.

Holder, A. (2006). Partitioning multiple objective optimal solutions with applications in radiotherapy design. *Optimization and Engineering*, **7**(4), 501–526.

Holmes, T. and Mackie, T. R. (1994). A filtered backprojection dose caculation method for inverse treatment planning. *Medical Physics*, **21**, 303–313.

Horst, R., Thoai, N. V., and Devries, J. (1988). On finding the new vertices and redundant constraints in cutting plane algorithms for global optimization. *Operations Research Letters*, **7**, 85–90.

Hristov, D. and Fallone, B. (1997). An active set algorithm for treatment planning optimization. *Medical Physics*, **24**, 91–106.

Hristov, D. and Fallone, B. (1998). A continuous penalty function method for inverse treatment planning. *Medical Physics*, **25**(2), 208–223.

Isermann, H. (1974). Proper efficiency and the linear vector maximum problem. *Operations Research*, **22**, 189–191.

Jackson, A. and Kutcher, G. J. (1993). Probability of radiation-induced complications for normal tissues with parallel architecture subject to non-uniform irradiation. *Medical Physics*, **20**(3), 613–625.

Jeraj, R., Wu, C., and Mackie, T. (2003). Optimizer convergence and local minima errors and their clinical importance. *Physics in Medicine and Biology*, **48**, 2809–2827.

Källman, P., Agren, A., and Brahme, A. (1992). Tumor and normal tissue responses to factionated non-uniform dose delivery. *International Journal of Radiation Biology*, **62**(2), 249–262.

Küfer, K.-H. and Hamacher, H. (2000). A multicriteria optimization approach for inverse radiotherapy planning. In W. Schlegel and T. Bortfeld, editors, *XIIIth International Conference on the Use of Computers in Radiation Therapy*, pages 26–28, Heidelberg, Germany. Springer Verlag, Berlin.

Küfer, K.-H., Scherrer, A., Monz, M., Alonso, F., Trinkaus, H., Bortfeld, T., and Thieke, C. (2003). Intensity-modulated radiotherapy – A large scale multi-criteria programming problem. *OR Spectrum*, **25**, 223–249.

Kutcher, G. J., Burman, C., Brewster, L., Goitein, M., and Mohan, R. (1991). Histogram reduction method for calculating complication probabilities for three-dimensional treatment planning evalutions. *Internation Journal of Radiation Oncology, Biology, Physics*, **21**, 137–146.

Lahanas, M., Schreibmann, E., Milickovic, N., and Baltas, D. (2003a). Intensity modulated beam radiation therapy dose optimization with multi-objective evolutionary algorithms. Technical report, Department of Medical Physics and Engineering, Strahlenklinik, Klinikum Offenbach, Germany.

Lahanas, M., Schreibmann, E., and Baltas, D. (2003b). Multiobjective inverse planning for intensity modulated radiotherapy with constraint-free gradient-based optimization algorithms. *Physics in Medicine and Biology*, **48**(17), 2843–2871.

Langer, M. (1987). Optimization of beam weights under dose-volume restrictions. *International Journal of Radiation Oncology, Biology, Physics*, **13**, 1255–1260.

Langer, M. and Morrill, S. (1996). A comparison of mixed integer programming and fast simulated annealing for optimized beam weights in radiation therapy. *Medical Physics*, **23**, 957–964.

Langer, M., Brown, R., Urie, M., Leong, J., Stracher, M., and Shapiro, J. (1990). Large scale optimization of beam weights under dose-volume restrictions. *International Journal of Radiation Oncology, Biology, Physics*, **18**, 887–893.

Langer, M., Brown, R., Morill, S., Lane, R., and Lee, O. (1996). A generic genetic algorithm for generating beam weights. *Medical Physics*, **23**, 965–971.

Lee, E., Fox, T., and Crocker, I. (2003). Integer programming applied to intensity-modulated radiation therapy treatment planning. *Annals of Operations Research*, **119**, 165–181.

Lee, S., Cho, P., Marks, R., and Oh, S. (1997). Conformal radiotherapy computation by the method of alternating projections onto convex sets. *Physics in Medicine and Biology*, **42**, 1065–1086.

Legras, J., Legras, B., and Lambert, J. (1982). Software for linear and non-linear optimization in external radiotherapy. *Computer Programs in Biomedicine*, **15**, 233–242.

Li, Y., Yao, J., and Yao, D. (2003). Genetic algorithm-based deliverable segments optimization for static intensity-modulated radiotherapy. *Physics in Medicine and Biology*, **48**, 3353–3374.

Lim, G. J., Choi, J., and Mohan, R. (2007). Iterative solution methods for beam angle and fluence map optimization in intensity modulated radiation therapy planning. *OR Spectrum*. DOI 10.1007/s00291-007-0096-1.

Lim, J. (2002). *Optimisation in radiation treatment planning*. Ph.D. thesis, University of Wisconsin, Madison, USA.

Lim, J., Ferris, M. C., Wright, S. J., Shepard, D. M., and Earl, M. A. (2002). An optimization framework for conformal radiation treatment planning. Technical report,

Computer Sciences Department, University of Wisconsin-Madison. Available online at `http://pages.cs.wisc.edu/~ferris/papers/conformal.pdf`. Accepted for publication, INFORMS Journal on Computing.

Liu, D. C. and Nocedal, J. (1989). On the limited memory bfgs method for large scale optimization. *Mathematical Programming*, **45**, 503–528.

Llacer, G. (1997). Inverse radiation treatment planning using the dynamically penalized likelihood method. *Medical Physics*, **24**(11), 1751–1764.

Llacer, J., Deasy, J., Bortfeld, T., Solberg, T., and Promberger, C. (2003). Absence of multiple local minima effects in intensity modulated optimization with dose-volume constraints. *Physics in Medicine and Biology*, **48**, 183–210.

Löf, J. (2000). *Development of a general framework for optimization of radiation therapy*. Ph.D. thesis, Department of Medical Radiation Physics, Karolinska Institute, Stockholm, Sweden.

Loridan, P. (1984). ε-solutions in vector minimization problems. *Journal of Optimization Theory and Applications*, **43**, 265–276.

Lotov, A. V., Bushenkov, V. A., and Kamenev, G. K. (2004). *Interactive Decision Maps: Approximation and Visualization of Pareto Frontier*. Kluwer Academic Publishers.

Lyman, J. T. and Wolbrast, A. B. (1989). Optimization of radiation therapy, iv: a dose-voume histogram reduction algorithm. *Internation Journal of Radiation Oncology, Biology, Physics*, **17**, 433–436.

Mageras, G. S. and Mohan, R. (1993). Appliction of fast simulated annealing to optimization of conformal radiation treatments. *Medical Physics*, **20**(3).

McDonald, S. C. and Rubin, P. (1977). Optimization of external beam radiation therapy. *International Journal of Radiation Oncology, Biology, Physics*, **2**, 307–317.

Merritt, M. and Zhang, Y. (2002). A successive linear programming approach to IMRT optimization problem. Technical report, Department of Computational and Applied Mathematics, Rice University. `http://www.caam.rice.edu/~zhang/reports/tr0216.ps`.

Messac, A. and Mattson, C. A. (2004). Normal constraint method with guarantee of even representation of complete pareto frontier. *AIAA Journal*, **42**(10), 2101–2111.

Messac, A., Ismail-Yahaya, A., and Mattson, C. A. (2003). The normalized constraint method for generating the pareto frontier. *Structural Multidisciplinary Optimization*, **25**, 86–98.

Michalski, D., Xiao, Y., Censor, Y., and Galvin, J. (2004). The dose-volume constraint satisfaction problem for inverse treatment planning with field segments. *Physics in Medicine and Biology*, **49**, 601–616.

Morrill, S., Lane, R., and Rosen, I. (1990a). Constrained simulated annealing for optimized radiation therapy treatment planning. *Computer Methods and Programs in Biomedicine*, **33**, 135–44.

Morrill, S., Rosen, I., Lane, R., and Belli, J. (1990b). The influence of dose constraint point placement on optimized radiation therapy treatment planning. *International Journal of Radiation Oncology, Biology, Physics*, **19**, 129–141.

Morrill, S., Lane, R., Wong, J., and Rosen, I. (1991a). Dose-volume considerations with linear programming optimization. *Medcial Physics*, **18**(6), 1201–10.

Morrill, S., Lane, R., Jacobson, G., and Rosen, I. (1991b). Treatment planning optimization using constrained simulated annealing. *Physics in Medicine and Biology*, **36**(10), 1341–1361.

Morrill, S., Lam, K., Lane, R., Langer, M., and Rosen, I. (1995). Very fast simulated annealing in radiation therapy treatment plan optimization. *International Journal of Radiation Oncology, Biology, Physics*, **31**, 179–188.

Motzkin, T. S. and Schoenberg, I. J. (1954). The relaxation method for linear inequalities. *Canadian Journal of Mathematics*, **6**, 393–404.

Niemierko, A. (1992). Random search algoithm (RONSC) for optimization of radiation therapy with both physical and biological end points and constraints. *International Journal of Radiation Oncology, Biology, Physics*, **33**, 89–98.

Niemierko, A. (1997). Reporting and analysing dose distributions: A concept of equivalent uniform dose. *Medical Physics*, **24**, 103–110.

Niemierko, A. (1999). A generalized concept of equivalent uniform dose. *Medical Physics*, **26**, 1100.

Niemierko, A. and Goitein, M. (1991). Calculation of normal tissue complication probability and dose-volume histogram reduction schemes for tissues with a critical element architecture. *Radiotherapy and Oncology*, **20**, 166–176.

Nizin, P., Kania, A., and Ayyangar, K. (2001). Basic concepts of corvus dose model. *Medical Dosimetry*, **26**(1), 65–69.

Penagaricano, J. A., Papanikolaou, N., and Wu, C. Yan, Y. (2005). An assessment of biologically-based optimization (BORT) in the IMRT era. *Medical Dosimetry*, **30**(1), 12–19.

Powlis, W., Altschuler, M., Censor, Y., and Buhle, J. (1989). Semi-automatic radiotherapy treatment planning with a mathemathical model to satisfy treatment goals. *International Journal of Radiation Oncology, Biology, Physics*, **16**, 271–276.

Preciado-Walters, F., Rardin, R., Langer, M., and Thai, V. (2004). A coupled column generation, mixed integer approach to optimal planning of intensity modulated radiation therapy for cancer. *Mathematical Programming, Series B*, **101**, 319–338.

Przybylski, A., Gandibleux, X., and Ehrgott, M. (2007). Recursive algorithms for finding all nondominated extreme points in the outcome set of a multiobjective

integer program. Technical Report, Laboratoire d'Informatique de Nantes Atlantique, Université de Nantes.

Redpath, A. T., Vickery, B. L., and Wright, D. H. (1976). A new technique for radiotherapy planning using quadratic programming. *Physics in Medicine and Biology*, **21**, 781–91.

Rockafellar, R. T. (1970). *Convex Analysis*. Princeton University Press, Princeton, New Jersey.

Romeijn, H., Ahuja, R., Dempsey, J., Kumar, A., and Li, J. (2003). A novel linear programming approach to fluence map optimization for intensity modulated radiation therapy treatment planning. *Physics in Medicine and Biology*, **48**, 3521–3542.

Romeijn, H., Dempsey, J., and Li, J. (2004). A unifying framework for multi-criteria fluence map optimization models. *Physics in Medicine and Biology*, **49**, 1991–2013.

Romeijn, H., Ahuja, R., Dempsey, J., Kumar, A., and Li, J. (2006). A new linear programming approach to radiation therapy treatment planning problems. *Operations Research*, **54**(2), 201–216.

Rosen, I., Lane, R., Morrill, S., and Belli, J. (1991). Treatment planning optimisation using linear programming. *Medical Physics*, **18**(2), 141–152.

Rowbottom, C. and Webb, S. (2002). Configuration space analysis of common cost functions in radiotherapy beam-weight optimization algorithms. *Physics in Medicine and Biology*, **47**, 65–77.

Rowbottom, C., Khoo, V., and Webb, S. (2001). Simultaneous optimization of beam orientations and beam weights in conformal radiotherapy. *Medical Physics*, **28**(8), 1696–1702.

Ruzika, S. and Wiecek, M. M. (2005). Approximation methods in multiobjective programming. *Journal of Optimization Theory and Applications*, **126**(3), 473–501.

Sayin, S. (2000). Measuring the quality of discrete representations of efficient sets in multiple objective mathematical programming. *Mathematical Programming*, **87**, 543–560.

Sayin, S. (2003). A procedure to find discrete representations of the efficient set with specified coverage errors. *Operations Research*, **51**(3), 427–436.

Schlegel, W. and Mahr, A. (2001). 3D conformal radiation therapy: Multimedia introduction to methods and techniques. Springer Verlag, Berlin.

Schreibmann, E., Lahanas, M., Xing, L., and Baltas, D. (2004). Multi-objective evolutionary optimization of the number of beams, their orientations and weights for intensity-modulated radiation therapy. *Physics in Medicine and Biology*, **49**, 747–770.

Shao, L. (2005). A survey of beam intensity optimization in IMRT. In T. Halliburton, editor, *Proceedings of the 40th Annual Conference of the Operational Society of New Zealand*, pages 255–265. Available online at `https://secure.orsnz.org.nz/conf40/content/papers/Shao.pdf`.

Shao, L. (2006). Finding representative nondominated points in multiobjective radiotherapy planning. In S. Dye, editor, *Proceedings of the 41th Annual Conference of the Operational Society of New Zealand*, pages 249–259.

Shao, L. and Ehrgott, M. (2006). Approximately solving multiobjective linear programmes in objective space and an application in radiotherapy treatment planning. Report 646, Department of Engineering Science, The University of Auckland. Available online at `http://www.esc.auckland.ac.nz/research/tech/esc-tr-646.pdf`. Accepted for publication, Mathematical Methods of Operations Research.

Shao, L. and Ehrgott, M. (2007). Finding representative nondominated points in multiobjective linear programming. In D. Fogel, editor, *Proceedings of IEEE Symposium Series on Computational Intelligence (IEEE SSCI 2007)*, pages 245–252.

Shao, L. and Ehrgott, M. (2008). Approximating the nondominated set of an MOLP by approximately solving its dual problem. *Mathematical Methods of Operations Research*. http://dx.doi.org/10.1007/s00186-007-0194-5.

Shepard, D., Ferris, M., Olivera, G., and Mackie, T. (1999). Optimizing the delivery of radiation therapy to cancer patients. *SIAM Review*, **41**(4), 721–744.

SouthEastMissouriHospital (2008). `http://www.southeastmissourihospital.com/technology/linear.htm` [2008, 25 February].

Spirou, S. V. and Chui, C.-S. (1998). A gradient inverse planning algorithm with dose-volume constraints. *Medical Physics*, **25**(3), 321–333.

Starkschall, G. (1984). A constrained least-squares optimization method for external beam treatment planning. *Medical Physics*, **11**(5), 659–65.

Starkschall, G., Pollack, A., and Stevens, C. W. (2001). Treatment planning using a dose-volume feasibility search algorithm. *International Journal of Rdaiation Oncology, Biology, Physics*, **49**(5), 1419–1427.

Stavrev, P., Hristov, D., Warkentin, B., and Fallone, B. G. (2003). Inverse treatment planning by physically constrained minimization of a biological objective function. *Medical Physics*, **30**, 2948–58.

Steuer, R. (1989). ADBASE multiple objective linear programming package. Technical report, University of Georgia, Athens, Georgia.

Thieke, C. (2003). *Multicriteria optimisation in inverse radiotherapy planning*. Ph.D. thesis, Ruprecht–Karls–Universität Heidelberg, Germany.

Thieke, C., Bortfeld, T., and Küfer, K.-H. (2002). Characterization of dose distributions through the max and mean dose concept. *Acta Oncologica*, **41**, 158–161.

Tucker, S. L., Thames, H. D., and Taylor, J. M. G. (1990). How well is the probability of tumor cure after fractionated irradiation described by Poisson statistics? *Radiation Research*, **124**, 273–82.

Ulungu, E. and Teghem, J. (1995). The two phases method: An efficient procedure to solve bi-objective combinatorial optimization problems. *Foundations of Computing and Decision Sciences*, **20**, 149–165.

Varian (2008). `http://www.varian.com/us/oncology/treatments/treatment_techniques/IMRT/` [2008, 25 February].

Verhaegen, F. (2003). Monte carlo modelling of external radiotherapy photon beams. *Physics in Medicine and Biology*, **48**, 107–164.

Wang, C., Dai, J., and Hu, Y. (2003). Optimization of beam orientations and beam weights for conformal radiotherapy using mixed integer programming. *Physics in Medicine and Biology*, **48**, 4065–4076.

Wang, X., Mohan, R., Jackson, A., Leibel, S., Fuks, Z., and Ling, C. (1995). Optimization of intensity-modulated 3D conformal treatment plans based on biological indices. *Radiotherapy and Oncology*, **37**, 140–152.

Webb, S. (1989). Optimisation of conformal radiotherapy dose distribution by simulated annealing. *Physics in Medicine and Biology*, **34**, 1349–1370.

Webb, S. (1991). Optimization of conformal radiotherapy dose distributions by simulated annealing: 2. inclusion of scatter in the 2D technique. *Physics in Medicine and Biology*, **36**, 1227–1237.

Webb, S. (1992). Optimization by simulated annealing of three-dimensional, conformal treatment planning for radiation fields defined by a multileaf collimator: II. inclusion of two-dimensional modulation of the x-ray intensity. *Physics in Medicine and Biology*, **37**, 1689–1704.

Webb, S. (1994a). Optimizing the planning of intensity-modulated radiotherapy. *Physics in Medicine and Biology*, **39**, 2229–2246.

Webb, S. (1994b). Optimum parameters in a model for tumor control probability including interpatient heterogeneity. *Physics in Medicine and Biology*, **39**, 1895–1914.

Webb, S. (1997). *The Physics of Conformal Radiotherapy*. Institute of Physics Publishing.

Webb, S. and Nahum, A. E. (1993). A model for calculating tumour control probability in radiotherapy including the effects of inhomogeneous distributions of dose and clonogenic cell density. *Physics in Medicine and Biology*, **38**, 653–666.

Webster, R. (1994). *Convexity*. Oxford Science Publications, Oxford, Oxford University Press.

Wu, C., Jeraj, R., and Mackie, T. (2003a). The method of intercepts in parameter space for the analysis of local minima caused by dose-volume constraints. *Physics in Medicine and Biology*, **48**, N149–N157.

Wu, Q. and Mohan, R. (2000). Algorithms and functionality of an intensity modulated radiotherapy optimization system. *Medical Physics*, **27**(4), 701–711.

Wu, Q. and Mohan, R. (2002). Multiple local minima in IMRT optimization-based on dose-volume criteria. *Medical Physics*, **29**, 1514–1527.

Wu, Q., Mohan, R., Niemierko, A., and Schmidt-Ulirich, R. (2002). Optimization of intensity-modulated radiotherapy plans based on the equivalent uniform dose. *International Journal of Radiation Oncology, Biology, Physics*, **52**(1), 224–235.

Wu, Q., Djajaputra, D., Wu, Y., Zhou, J., Liu, H. H., and Mohan, R. (2003b). Intensity-modulated radiotherapy optimization with gEUD-guided dose-volume objectives. *Physics in Medicine and Biology*, **48**, 279–291.

Wu, X. and Zhu, Y. (2001). An optimization method for importance factors and beam weights based on genetic algorithms for radiotherapy treatment planning. *Physics in Medicine and Biology*, **46**, 1085–99.

Xiao, Y., Censor, Y., Michalski, D., and Galvin, J. (2003). The least-intensity feasible solution for aperture-based inverse planning in radiation therapy. *Annals of Operations Research*, **119**, 183–203.

Xing, L. and Chen, G. T. Y. (1996). Iterative algorithms for inverse treatment planning. *Physics in Medicine and Biology*, **41**, 2107–2123.

Xing, L., Hamilton, R., Spelbring, D., Pelizzari, C., Chen, G., and Boyer, A. (1998). Fast iterative algorithms for three-dimensional inverse treatment planning. *Medical Physics*, **25**, 1845–1849.

Xing, L., Li, J. G., Donaldson, S., Le, Q. T., and Boyer, A. L. (1999). Optimization of importance factors in inverse planning. *Physics in Medicine and Biology*, **44**, 2525–36.

Yu, Y. (1997). Multi-objective decision theory for computational optimisation in radiation therapy. *Medical Physics*, **24**(9), 1445–1454.

Zaider, M. and Minerbo, G. N. (2000). Tumour control probability: a formulation applicable to any temporal protocol of dose delivery. *Physics in Medicine and Biology*, **45**, 279–93.

List of Figures

List of Tables