KUOPION YLIOPISTON JULKAISUJA C. LUONNONTIETEET JA YMPÄRISTÖTIETEET 170
KUOPIO UNIVERSITY PUBLICATIONS C. NATURAL AND ENVIRONMENTAL SCIENCES 170

PEKKA KOLMONEN

The Direct Control of the Multi-Leaf Collimator in the Inverse Problem of Radiotherapy Treatment Planning

Doctoral dissertation

To be presented by permission of the Faculty of Natural and Environmental Sciences of the University of Kuopio for public examination in Auditorium E16/E17, CENTEK-building, Microteknia 2, on Friday 19th March 2004, at 12 noon

Department of Applied Physics
University of Kuopio

KUOPION YLIOPISTO
KUOPIO 2004

Distributor:
Kuopio University Library
P.O. Box 1627
FIN-70211 KUOPIO
FINLAND
Tel. +358 17 163 430
Fax +358 17 163 410
http://www.uku.fi/kirjasto/julkaisutoiminta/julkmyyn.html

Series Editors:
Professor Lauri Kärenlampi, Ph.D.
Department of Ecology and Environmental Science

Professor Jari Kaipio, Ph.D.
Department of Applied Physics

Author's address:
Department of Applied Physics
University of Kuopio
P.O. Box 1627
FIN-70211 KUOPIO
FINLAND
Email: kolmonen@venda.uku.fi

Supervisors:
Jouko Tervo, Ph.D.
Department of Mathematics and Statistics
University of Kuopio

Docent Tapani Lahtinen, Ph.D.
Department of Oncology
Kuopio University Hospital

Reviewers:
Professor Yair Censor, Ph.D.
Department of Mathematics
University of Haifa
Haifa
Israel

Docent Simo Hyödynmaa, Ph.D.
Department of Oncology
Tampere University Hospital

Opponent:
Docent Mikko Tenhunen, Ph.D.
Department of Oncology
Helsinki University Central Hospital

ISBN 951-781-308-2
ISBN 951-27-0003-4 (PDF)
ISSN 1235-0486

Kopijyvä
Kuopio 2004
Finland

Kolmonen, Pekka. The Direct Control of the Multi-Leaf Collimator in the Inverse Problem of Radiotherapy Treatment Planning. Kuopio University Publications C. Natural and Environmental Sciences 170. 2004. 81 p.
ISBN 951-781-308-2
ISBN 951-27-0003-4 (PDF)
ISSN 1235-0486

ABSTRACT

In the inverse problem of radiotherapy treatment planning, the setup for treatment delivery is solved so that clinical treatment criteria are satisfied. The treatment criteria can be consisted of dose constraints for the cancerous tissue and vulnerable tissues, or the criteria can be determined by using the estimated radiobiological outcome of a treatment.

When the solution of the inverse problem is the intensity modulation treatment fields, the planning/delivery process is called the intensity modulated radiotherapy (IMRT). The multi-leaf collimator (MLC) is a widely used field-shaping device that can be used to control the intensity modulation.

In this thesis, a complete system to solve the inverse problem is described. The novel approach in the system is that the field-shaping device, namely the MLC, is controlled directly. Thus, all technical limitations of the MLC can be taken into account, when the inverse problem is solved. The system employs the so-called multiple static control of the MLC.

The thesis describes the mathematical background of the inverse problem and connects the mathematical models and limitations of the MLC to the inverse problem. The inverse problem cannot be straightforwardly solved but mathematical optimization methods must be applied. The implementation of the optimization is described and discussed.

Treament planning examples, where the developed system has been applied, and preliminary dosimetric tests are presented and discussed.

AMS (MOS) Classification: 90C90, 90C53, 45B05, 92C50
National Library of Medicine Classification: WN 110, WN 250.5.R2
Universal Decimal Classification: 621.384.66
INSPEC Thesaurus: radiation therapy; intensity modulation; collimators; inverse problems; optimisation; nonlinear programming; dosimetry

To Stina, Petja, Pinja and “Esko”

Acknowledgements

This study was carried out in the Department of Applied Physics at the University of Kuopio.

I am most grateful to my supervisors Jouko Tervo Ph.D. for his seemingly endless stream of new and original ideas and Docent Tapani Lahtinen Ph.D. for his great ability to set up fruitful research projects. I especially thank them both for their trust and support.

I wish to thank the official reviewers Professor Yair Censor Ph.D. and Docent Simo Hyödynmaa Ph.D. for their constructive criticism, suggestions and fresh point of view.

I warmly thank my colleagues in the research project Kai Jaatinen M.Sc. and Marko Juntunen Ph.D. for their friendship and good humour. Without you the work would have been very dry indeed.

I thank the good hospital physicists at the Kuopio University Hospital Anssi Väänänen Ph.D. and Tiina Lyyra Ph.D. for their professional aid during the measurements. I thank also physicist Jan Seppälä M.Sc. for testing the developed system.

I express my gratitude to the Department of Applied Physics and especially to Professor Jari Kaipio Ph.D. for the encouraging research atmosphere.

The study was funded by Technology Development Centre Finland (TEKES) and Varian Medical Systems Finland. From Varian I want to thank Mr. Pekka Aalto for giving us the opportunity and Mr. Harri Puurunen and Mr. Mika Miettinen for supervising and supporting our project.

I thank my parents Vuokko and Kalle for supporting me and my family in our life.

I wish to thank generally everyone involved in my studies for their patience. I may not be quick but I am steady.

Finally, I thank my wife Stina for her love and support and my son Petja as well as my daughter Pinja for keeping my feet on the ground. There are more important things in life than research work (though they are not many).

Kuopio, March 2004

Pekka Kolmonen

Abbreviations and notations

BEAM	A Monte Carlo code to simulate treatment machines
DDC	Dose deposition coefficient
DVH	Dose volume histogram
GUI	Graphical user interface
$\mathbf{H}$	The discretized dose deposition kernel
$h(x,u)$	The dose deposition kernel
IMRT	Intensity modulated radiotherapy
L-BFGS-B	A limited memory quasi-Newton method for bound constrained non-linear optimization
MLC	Multi-leaf collimator
MU	Monitor unit
NTCP	Normal tissue complication probability
OAR	An organ at the risk of radiation damage
$\boldsymbol{\psi}$	Discretized intensity distribution
$\psi(u)$	Intensity distribution of a treatment field
PTV	The planning target volume
PUC	The probability of uncomplicated cure
TCP	Tumor control probability

Contents

CHAPTER I

Introduction

The basic dilemma of external beam radiotherapy is to deliver a high, and most often, homogeneous dose to cancerous tissue while keeping dose in healthy tissue at such a level that the amount of side effects is as low as possible or at least at an acceptable level. Since the discovery of x-rays' interaction with living tissue, the dilemma has managed to interest researchers in the field of radiotherapy.

The dilemma can not be overcome in its simplest physical form. It is impossible for radiation to pass through a medium without any interaction, except when the medium is vacuum. There has been, however, much advancement during the 20th century that has enhanced the quality of radiotherapy. The use of high energies enables deeper penetration, delivery from multiple directions focuses high dose to the cancerous tissue, rotations of the whole treatment unit (gantry), treatment head (collimator) and patient couch give more degrees of freedom.

Arguably, the greatest innovations during the latter part of the last century did not modify treatment delivery, but brought new ideas and possibilities to radiotherapy treatment planning. First, computerized x-ray tomography (CT) was applied to treatment planning [7, 80]. Second, the accurately estimated dose irradiated by a treatment unit could be calculated using CT-data and numerical computer algorithms (see *e.g.* [2, 65]). It was possible to plan radiotherapy treatments three-dimensionally. This meant that there was enough anatomical information to reconstruct 3D tumor shape as well as shapes for other tissues, treatment fields could be non-coplanar, a treatment plan could be viewed from beam's eye or from some other direction and three-dimensional dose distributions could be computed.

The new enhancements in treatment planning gave tools to apply field accessories and still get reliable dose estimations. With beam limiting and attenuating devices, the intensities of fields could be modified. For example, by blocking certain areas of a treatment field, dose is constrained in tissue which can not stand high dose. With the use of accurate dose estimations and field intensity modifications it was possible to produce conformal treatment plans. This was a big step towards solving the basic dilemma since the region of high dose could now be planned to conform to the shape of a tumor.

Some intensity modulation can be done in conformal radiotherapy. The modulation is, however, limited, with the exception of a compensator, to blocking wedges. During the last decades, despite the fact that there was not a device to produce arbitrary intensity modulation of a treatment field, numerous methods have been developed for the optimization of the intensity distributions of treatment fields. The principle of inverse treatment planning for radiotherapy was introduced [6]. In inverse planning the dose prescription is given and then the treatment setup is solved so that a dose prescription is satisfied. The solution of the inverse problem is the result of mathematical optimization. The dose prescription consists typically of dose constraints for planning target volumes (PTVs) and organs at risk (OARs) [28, 78].

If, in the inverse planning, the solved parameters are the intensities of treatment fields, the method is called the intensity modulated radiotherapy (IMRT). The other parameters that could be solved are *e.g.* radiation quality, the field angles (gantry, couch and collimator) and field weights.

At present, the device to produce complex intensity distributions exists. It is a beam blocking device consisting of narrow shields, called leaves, that can move during the irradiation of a patient. The device is called the multileaf collimator (MLC). The MLC can be used to reproduce the optimized intensity distributions. The intensity distributions cannot, however, be truly arbitrary since the MLC can move only under its mechanical limitations. For this reason, techniques that control the MLC in such a way that the optimized intensity distributions of treatment fields can be reproduced have been developed.

All the contemporary inverse planning systems follow a similar kind of approach. First, the intensity distributions of treatment fields are optimized. Then, the intensity distributions are reproduced with the MLC. The drawback of this dual approach is that during the optimization of the intensity distributions the mechanical limitations of the MLC are not taken into account or they are addressed only roughly.

The aims and contents of the thesis

The aim of this thesis is to report the development and evaluation of a complete system for inverse planning in radiotherapy. The system optimizes the control parameters of the MLC directly. The optimization of the intensity distributions of treatment fields is not needed. The technical limitations of the MLC are used as constraint in the mathematical optimization. The result of the optimization is a complete treatment plan.

A crucial part of introducing any new method, such as the inverse planning system described in this thesis, to radiotherapy is to make sure that the safety of a treated patient will not be compromised. The main aspect of patient safety is the dosimetric accuracy of the new method. Also, to gain knowledge of the performance of the new system several examples must be tested using real patient data. In this thesis, the evaluation of the inverse planning system includes dosimetric accuracy testing and inverse treatment planning for real patient examples. The

treatment planning examples demonstrate the feasibility of the developed system, they do not try to prove the superiority of the system.

Part of the work, mostly the mathematical background, that is presented in this thesis has been published earlier [35, 36, 69, 73, 70, 71, 72].

In chapter 2 the basics of the inverse and forward problems of radiotherapy are discussed together with clinical planning criteria. Chapter 3 introduces intensity modulation of treatment fields as a tool to achieve optimized dose distributions. Techniques for the delivery of intensity modulated fields are also discussed, especially the multileaf collimator (MLC). Chapter 4 gives an overview of the implemented inverse planning system. Treatment plans computed using the system are shown in chapter 6. The results of preliminary dosimetric tests are in chapter 5.

Chapter II

The inverse problem

2.1 Introduction

Conventionally, a radiation oncologist together with a physicist decides a configuration of treatment fields (field angles, field weights, etc) that results an acceptable dose distribution. In complex treatment situations, to reach the acceptable dose distribution can lead to a long trial-and-error iteration. For this reason the treatment parameters are varied over a limited range [56]. In addition, modern treatment planning systems enable three-dimensional planning and non-coplanar (patient couch rotated) field orientations. It is difficult (for a human mind) to find the best beam orientations if both gantry and couch angles are modified. Conventional treatment planning may also yield unacceptable dose to healthy tissues. For these reasons the inverse planning in radiotherapy has been studied intensively for the last decades.

In inverse planning, the treatment delivery of radiotherapy is modeled as an inverse problem. A dose prescription or radiobiological criteria are given and the treatment configuration is solved based on the problem modelling and the dose prescription. Usually, mathematical optimization methods are used to optimize an object or a penalty function. The phrase plan optimization is somewhat misleading since an optimal plan is very hard to specify. Optimal dose distributions can be produced, but they are only part of a treatment plan. One has to consider *e.g.* the overall condition of a patient, organ movement during beam delivery, dose fractionation schedules etc. Whenever treatment plan optimization is mentioned in this thesis, it refers to mathematical optimization of a mathematically defined object function.

A number of delivery techniques have been under research to be potential tools in inverse planning. These include the optimization of field weights [3, 40, 41], optimal compensator design [13, 82], use of blocks, optimization of beam orientation (gantry and/or couch angles) [26]. None of these have, however, found their way to clinical practice. Either beam delivery would be too complicated or results have not been satisfactory.

The most promising technique for inverse planning is to modify the intensities of treatment fields. This means that when the field orientations are given the intensity distributions over fields are optimized to fulfill some dose constraints or radiobiological criteria. It is clear that a computer algorithm is the most suitable means to solve the problem.

Where there is an inverse problem there must be a forward problem. In this chapter a mathematical model is introduced for the forward problem of treatment planning. The model enables to determine dose as a function of the intensity distributions of treatment fields. The basic solution of the inverse problem and the practical meaning of the solution are also described. Finally, different clinical criteria for inverse planning are examined.

2.2 Forward problem

To be able to apply inverse methods for radiotherapy treatment planning one must have a model for the forward problem of treatment planning. This problem can be defined shortly. When a treatment configuration is given, determine dose distribution in patient. As the inverse planning is often an iterative process, one must solve the forward problem several times. Consequently, fast techniques must be employed while simultaneously the dose model must be accurate enough.

Modern treatment planning softwares can use semiempirical models for dose calculation (*e.g.* [65]). Although these methods are accurate enough for clinical use they are too inefficient when applied to inverse planning. Time used for dose calculation is far too long. Other ways of describing the dose deposition must be sought.

2.2.1 Fredholm integral equation formulation

Since only the intensity distribution of treatment fields is solved, a model suited for this purpose is needed. All knowledge about field angles and modalities is included in the model. Only parameters that can alter dose distribution are the intensity distributions of treatment fields. A vastly used model for inverse planning is the Fredholm integral equation of the first kind. The basic model is

$$D(x) = \int_U h(x,u)\psi(u)\mathrm{d}u, \tag{2.1}$$

where $D(x)$ is dose at point $x = (x_1, x_2, x_3)$ in patient space, $\psi(u)$ is intensity at a point $u = (u_1, u_2)$ in treatment space and $h(x,u)$ is the dose deposition kernel which describes how much the value of intensity at point u contributes (or deposits) to dose at point x.

Looking at equation (2.1), it is clear that if the forward or inverse problem is to be solved, one must know the dose deposition kernel $h(x,u)$. Since the main interest here is in inverse planning, a thorough study of different kernels is excluded. Dose or energy deposition kernels are described *e.g.* in [1, 22].

2.2.2 The modeling of the dose deposition kernel

Since the kernel h is unknown, a model must be found that approximates the kernel accurately and enables fast computation.

A simple approximation for the dose deposition is convolution. Based on the equation (2.1), the convolution approach is

$$D(x) = \int_U I(x_3)h(\mathcal{P}(x) - u)\psi(u)\mathrm{d}u, \tag{2.2}$$

where $h(\mathcal{P}(x) - u)$ is the convolution kernel. $\mathcal{P}$ is a projection operator that projects the point x from the three-dimensional patient space V to the two-dimensional treatment space U which is essentially a plane. $I(x_3)$ is a function representing relative depth dose. The convolution approach is attractive since there are effective techniques (Laplace and Fourier transforms) to solve the convolution equation. The assumption that dose is dependent only on the depth x_3 and radial distance $\mathcal{P}(x) - u$ of a point x from a point u is, however, over-simplified. Although skin obliquity can be corrected using ray tracing, the effects of tissue inhomogeneities are not taken into account. Hence, here are presented two variants of the original and accurate Fredholm integral equation.

Discrete dose deposition matrix

In this model, the patient and treatment spaces are divided into voxels and bixels, respectively. A bixel is a shortened form for a beam pixel. It is usually a small rectangular area in the two-dimensional treatment space. A narrow radiation beam from the area of a bixel is usually called a pencil beam. A voxel is a cubic volume in the three-dimensional patient space. Now, equation (2.1) is discretized and can be formulated as a matrix equation

$$\mathbf{D} = \mathbf{H}\boldsymbol{\psi}, \tag{2.3}$$

where $\mathbf{D}$ is a dose vector of size $N_1 \times 1$, $\mathbf{H}$ is a matrix of size $N_1 \times N_2$ and $\boldsymbol{\psi}$ is an intensity vector of size $N_2 \times 1$. N_1 is the number of voxels and N_2 the number of bixels. An element of dose vector is the dose integrated over a voxel in patient space or dose at a point inside the voxel. Equation (2.3) can be called a pencil beam model since the total dose is a superposition of dose depositions from all the pencil beams [51]. Schematic figure 2.1 shows the spaces, a bixel and a voxel.

Equation (2.3) is a useful presentation for radiotherapy treatment planning since it is suitable for computer algorithms. The equation can be used to solve the inverse planning problem which will be discussed later. Here, the other inverse problem that is imbedded in the equation is solved, namely the determination of the matrix $\mathbf{H}$. Suppose that a treatment could be delivered where only one of the bixels has a non-zero intensity value, a value of unity. Then, the dose distribution from such irradiation would indeed be the dose deposition of that particular bixel and one column of $\mathbf{H}$ would present the deposition in a patient or phantom. Continuing the process, each of the bixels would be set to unity

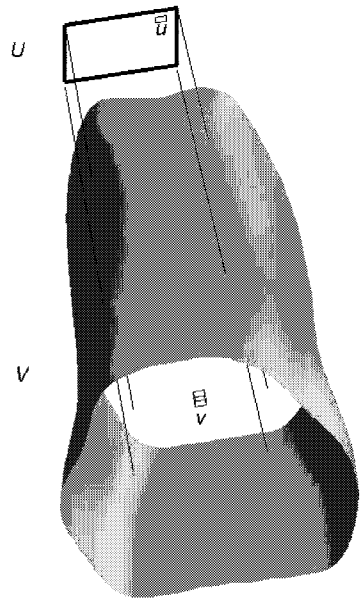

Figure 2.1: Treatment space U and patient space V together with a bixel u and voxel v.

while others would have zero value until dose depositions from all bixels would be determined. The matrix $\mathbf{H}$ would be now fully described.

Dose depositions of pencil beams cannot be measured, since a measurement setup where only a tiny part of a treatment field would be "open" is impossible because collimator scattering and leakage from the "blocked" area of the field would disturb the intensity of the open area. A practical approach is to use an accurate but slow dose calculation method for the determination of dose deposition from a bixel. In practice, the accuracy of a tested dose calculation method (semiempirical, Monte Carlo) and the fast speed of computation of the pencil beam model (2.3) are combined.

The size of matrix $\mathbf{H}$ is of significant importance. An example: the dose depositions of five 10×10 cm^2 treatment fields are determined at $N_1 = 30000$ voxels in a patient. If the bixel size is 10×2.5 mm^2, the number of bixels is $N_2 = 2000$. Thus there are 60×10^6 elements in the matrix which is a figure that cannot be handled by most computers. Fortunately, the matrix is sparse, especially when only voxels radially close enough (*e.g.* < 2.0 cm) to the central axis of a pencil beam are taken into account.

Non-homogeneous continuous approximation

The discrete dose deposition model introduced in the previous section is practical in the inverse problem of radiotherapy treatment planning but the model has certain disadvantages. Main drawback is that the model is tied locally to the points in the patient and treatment spaces where the kernel $h(x, u)$ has been discretized. If the original integral equation (2.1) is needed, a continuous approximation must be developed.

Based on the discrete dose deposition matrix $\mathbf{H}$ a continuous approximation can be constructed. The idea is to replace the kernel with suitable basis functions by formulating

$$h(x, u) = \sum_{n=1}^{N} c_n f_n(x, u),$$

where f_i are the chosen continuous basis functions. The unknown coefficients c_i can be determined using the discrete kernel $\mathbf{H}$. Let $x_i \in V$ $(i = 1 \dots N_1)$ and $u_j \in U$ $(j = 1 \dots N_2)$ be the discretization points. Denote $y_l = (x_l, u_l)$, $l = 1 \dots M$ $(M = N_1 \cdot N_2)$ to be some enumeration of the points (x_i, y_j). Then

$$\mathbf{H} = \begin{pmatrix} h(y_1) \\ \vdots \\ h(y_M) \end{pmatrix} \text{ and } \mathbf{C} = \begin{pmatrix} c_1 \\ \vdots \\ c_N \end{pmatrix}.$$

Furthermore, define a matrix $\mathbf{F}$ as

$$F_{ln} = f_n(y_l).$$

Then

$$\mathbf{H} = \mathbf{FC}. \tag{2.4}$$

Now, depending on the basis, the unknown coefficients can be computed from $\mathbf{C} = \mathbf{F}^{\dagger}\mathbf{H}$, where $(.)^{\dagger}$ is the pseudoinverse. One can use *e.g.* the Battle-Lemarié Spline Wavelet basis, which has spatially local support [69].

2.3 Inverse problem of radiotherapy treatment planning

Before the subject of optimizing dose delivery to a patient is discussed, the discrete inverse problem that was formulated in equation 2.3 is solved. The simplest form of the solution is

$$\boldsymbol{\psi} = \mathbf{H}^{\dagger}\mathbf{D}. \tag{2.5}$$

Two questions arise when the solution (2.5) is examined:

1. Is the intensity distribution $\boldsymbol{\psi}$ physically feasible?
2. What does contain the dose vector $\mathbf{D}$?

The answer to the first question is that the intensity distribution must be constrained. Otherwise it will contain negative intensities for some bixels. This means that a patient is irradiating the source of irradiation which would not be physically relevant. A constraint $\psi_i \geq 0,\ i = 1 \ldots N_2$ must be added.

The second question refers to an important aspect of inverse planning, namely the dose prescription. Here, in vector $\mathbf{D}$, one places all the requirements and clinical criteria that a treatment plan must meet as precisely as possible. A desired dose distribution is a rather simple way to assign a dose prescription. The rest of this chapter is devoted to the presentation of various kinds of dose prescriptions.

2.4 Clinical criteria

When one wants to find an optimal treatment plan to cure cancer, one must first decide what are the criteria that make a plan optimal. After the decision is made, the criteria must be put into mathematical context so that the applied optimization algorithm can use them.

The subject of defining an optimal treatment for an optimization algorithm is somewhat controversial and has rised some debate over years. The desired dose distribution $\mathbf{D}$ is clearly too simple formulation for dose prescription, although very understandable from the mathematical point of view. The prescription does not emerge from clinical demands. It is easy to assign dose in the PTV volume to a certain value, like 60 Gy. The dose prescription for OARs cannot be 0 Gy, since it is physically impossible to accomplish and all knowledge about the dose tolerance level of an organ would not be utilized. Better models must be sought.

More sophisticated formulations of clinical criteria can be divided into two distinct categories. One is based on physical and the the other on radiobiological criteria. It must be noted that fractionated radiotherapy is a process where uncertainties in treatment field output, patient set-up and internal motion of organs should also be taken into account or compensated [46].

Instead of formulating the criteria mathematically according to the linear equation (2.5), it will be assumed that dose depends nonlinearly on parameters that control the intensity *i.e.* $\psi = \psi(\mathbf{a})$, where $\mathbf{a}$ are the control parameters. Hence, $D(x) = D(x, \mathbf{a})$, where x is a point in the patient space. The linear problem has been solved in [35] and [70]. The subject of intensity control, or intensity modulation, is discussed in chapter 3.

2.4.1 Physical criteria

The physical criteria are a gathering of constraints, objectives and limitations that prescribe what the physical dose distribution inside a patient should be. By using the physical criteria, one tries to find mathematical equivalents for demands like "Inside the PTV volume, the prescribed dose should be 60 Gy or within 5 % to the prescribed dose", "dose in spinal cord can not exceed 47 Gy", "salivary glands should not get more than 10 Gy" or "no more than 10 % of the volume of a lung can be irradiated over 25 Gy". When a set of these demands are put together and their mathematical substitutes are inserted to an optimization algorithm, one should

obtain a treatment plan that can be physically "good". It can not be ensured that the plan is the optimal plan for a patient nor that it is the clinically best plan. There are mainly two reasons for this. The first is that the used optimization algorithm may not be capable of finding the best mathematical solution *i.e.* the algorithm can not find the global extremum (if an object function approach is used) but is trapped to a local one. The second reason are the clinical demands which are used to describe an acceptable treatment plan. The phrases that were introduced above are still too restrictive. Human perception can come up with far better definitions for a clinically good treatment plan. One of these definitions is visual inspection, something that an experienced dose planner does routinely and an algorithm simply cannot do.

Next, two mathematical approaches that use physical criteria to describe the treatment planning problem are discussed. First, discretized dose is defined, as the approaches will be described in their discrete forms. Patient space V is divided into voxels v_k, $k = 1, \ldots, N_1$. A voxel is centralized to a point $x_k \in v_k$. Dose in patient space is controlled by some parameters $\mathbf{a}$ which will be defined later: $D(x_k) = D(x_k, \mathbf{a})$. Different PTVs, OARs and other regions of interest are defined as disjoint index sets $J = J_{\mathrm{PTV}_1} \cup J_{\mathrm{PTV}_2} \cup \ldots \cup J_{\mathrm{PTV}_{\mathrm{N_p}}} \cup J_{\mathrm{OAR}_1} \cup J_{\mathrm{OAR}_2} \cup \ldots \cup J_{\mathrm{OAR}_{\mathrm{N_o}}}$, where $J = 1, \ldots, N_1$ and

$$\begin{aligned}
J_{\mathrm{PTV}_1} &= \{k \in J \mid x_k \in \mathrm{PTV}_1\}, \\
J_{\mathrm{PTV}_2} &= \{k \in J \mid x_k \in \mathrm{PTV}_2\}, \\
&\vdots \\
J_{\mathrm{PTV}_{\mathrm{N_p}}} &= \{k \in J \mid x_k \in \mathrm{PTV}_{\mathrm{N_p}}\}, \\
J_{\mathrm{OAR}_1} &= \{k \in J \mid x_k \in \mathrm{OAR}_1\}, \\
J_{\mathrm{OAR}_2} &= \{k \in J \mid x_k \in \mathrm{OAR}_2\}, \\
&\vdots \\
J_{\mathrm{OAR}_{\mathrm{N_o}}} &= \{k \in J \mid x_k \in \mathrm{OAR}_{\mathrm{N_o}}\},
\end{aligned} \tag{2.6}$$

with N_p being the number of PTVs and N_o the number of OARs.

Object function approach

Consider a treatment situation where 1 mm from the boundary of a PTV there is an OAR. The dose prescription dictates the minimum dose inside the PTV volume to be 70 Gy. The same prescription tells that the maximum dose inside the OAR volume must be under 47 Gy. Thus, the prescription demands that there must be a dose gradient of $\Delta \mathrm{D}/\Delta d = 23$ Gymm^{-1}. The gradient is impossible to obtain using the contemporary treatment delivery techniques. This implies that the optimization problem is infeasible. The infeasibility excludes many efficient formulations for object functions that would be based on the dose prescription. One can not use *e.g.* gradient projection [59] or augmented Lagrangian [48] methods, since they try to satisfy constraints exactly, they solve feasible problems.

Instead of exactly constrained techniques, penalty functions must be used, although they are not considered to be as effective as the gradient projection or augmented Lagrangian methods [24]. There is one additional demand that the penalty function must meet. It must be differentiable because a large-scaled optimization problem will be faced and if the derivate of the penalty/object function is needed, numerical differentiation (*e.g.* finite difference) is not recommended as computation times are to be kept at a reasonable level.

Here, an object function is presented which is initially based on discrete L_2 norm and is more specifically a sum of quadratic penalties [71]. Quadratic formulation is preferred over L_1 and L_∞ norms because of differentiability. Define an object function

$$\begin{aligned} F(\mathbf{a}) &= c_1 \sum_{k \in J_{\mathrm{PTV}}} |(D_{\mathrm{PTV}} - D(x_k, \mathbf{a}))_-|^2 \\ &+ c_2 \sum_{k \in J_{\mathrm{PTV}}} |(D(x_k, \mathbf{a}) - d_{\mathrm{PTV}})_-|^2 \\ &+ c_3 \sum_{k \in J_{\mathrm{PTV}}} \| \nabla_x D(x_k, \mathbf{a}) \|^2 \\ &+ c_4 \sum_{k \in J_{\mathrm{OAR}}} |(D_{\mathrm{OAR}} - D(x_k, \mathbf{a}))_-|^2 \\ &+ c_5 \left| \left(V_{\mathrm{dv,OAR}} - \frac{1}{|J_{\mathrm{OAR}}|} \sum_{k \in J_{\mathrm{OAR}}} \mathrm{erf}(D(x_k, \mathbf{a}) - D_{\mathrm{dv,OAR}}) \right)_- \right|^2 \end{aligned} \tag{2.7}$$

for a PTV and a single OAR. The scalar constants $c_1, \ldots, c_5$ are the relative weights for the penalties and the operator $(.)_-$ is

$$(y)_- = \begin{cases} 0, & y \geq 0 \\ y, & y < 0. \end{cases}$$

Since the physical size of a tissue must not bias the minimization of the object function (2.7), each weight is divided by the number of voxels in the volume of a tissue (PTV, OAR).

The different terms, whose sum the object function (2.7) is, are:

1. term: quadratic penalty for underdose in the PTV. This is the protagonist of the minimization of the object function (2.7) and must be heavily weighted against other terms.

2. term: quadratic penalty for overdose in the PTV. Here, different terms for underdose and overdose are used, since an applicable dose prescription for the PTV is sought. This and previous terms, however, could be replaced by $\sum_{k \in J_{\mathrm{PTV}}} |D_0 - D(x_k, \mathbf{a})|^2$ (weight excluded) where there is no penalty as such but a least-squares fit of dose $D(x_k, \mathbf{a})$ to prescribed dose D_0. This approach would be simpler but there would also be no means to handle severe underdose or overdose to small volumes *i.e.* "cold spots" or "hot

spots", respectively. When there are different weights for under/overdose, one can at least compromise between "cold" and "hot" volumes.

3. term: quadratic penalty for too large dose gradients in the PTV. This is an additional penalty that decreases the probability of "hot spots" and "cold spots" that are certain to emerge when the penalty of a constraint violation is quadratic.

4. term: quadratic penalty for overdose in the OAR. Similar to the penalty in term 2.

5. term: quadratic penalty for violation of dose-volume constraint in the OAR. $|J_{\text{OAR}}|$ is the number of voxels in the the OAR. The approximate formulation of the penalty in (2.7) is such that it enables analytic computation [60, 71]. The original constraint is

$$\frac{\left|\{m = 1, \ldots, M | D(x_m) \geq D_{\text{dv}}\}\right|}{M} \leq V_{\text{dv}},$$

where M is the number of voxels in an OAR and V_{dv} is the volume that should not be irradiated over D_{dv}. As the dose-volume constraint for an OAR must be physically relevant, all voxels must have equal volumes. There exist models for more complex volume divisions.

Using the object function (2.7) a generic extremum problem can be stated:
Find the global minimum of

$$\min_{\mathbf{a}} F(\mathbf{a})$$

under constraints for the control parameters $\mathbf{a}$.
The constraints depend on what control technique for the intensity of treatment fields is used. This subject will be discussed in chapter 3.

Feasible solution approach

With the definitions for object function, a feasible problem can also be formulated. The problem is associated with the physical criteria [71]. A generic feasible problem can be stated as:
Find control parameters $\mathbf{a}$ *for which the inequalities*

$$\begin{aligned} d_{\text{PTV}} &\leq D(x_k, \mathbf{a}), \ k \in J_{\text{PTV}}, \\ D(x_k, \mathbf{a}) &\leq D_{\text{PTV}}, \ k \in J_{\text{PTV}}, \\ D(x_k, \mathbf{a}) &\leq D_{\text{OAR}}, \ k \in J_{\text{OAR}}, \\ V_{\text{dv,OAR}} &\geq \frac{1}{|J_{\text{OAR}}|} \sum_{k \in J_{\text{OAR}}} \text{erf}(D(x_k, \mathbf{a}) - D_{\text{dv,OAR}}) \end{aligned} \tag{2.8}$$

are satisfied under constraints for $\mathbf{a}$.
The inequalities are, in order, the lower limit of PTV dose, higher limit of PTV

dose, higher limit of OAR dose and dose-volume constraint for OAR. Note that inequality for too large gradients in PTV is not included, although the inequality would be possible to formulate.

The feasible solution does seem ideal for treatment plans based on physical criteria. In reality, it is impossible to find the solution if the dose constraints are too tight. Fortunately, there are algorithms, *e.g.* the Cimmino's algorithm [10, 11] or its non-linear equivalent [12], that converge to a weighted least squares solution in case there is no feasible solution [21, 35].

2.4.2 Radiobiological criteria

While physical criteria enable dose prescription for a patient, they do not directly tell anything about the biological outcome of a treatment. Models have been developed that link physical dose to its biological effects. One of these models is the Poisson model of cell kill [55]. In the model, probability of tissue injury $P(D)$ as a function of dose is formulated as

$$P(D) = 2^{\exp[e\gamma(1-\frac{D}{D_{50}})]}, \tag{2.9}$$

where D_{50} is the dose causing the injury for 50 % of patients and γ is "close to the maximum normalized slope of the dose response relation", $\gamma = D(\mathrm{d}P/\mathrm{d}D)$ [42]. Another way to describe the probability is the logistic equation

$$P(D) = [1 + (D_{50}/D)^{\frac{\gamma}{4D_{50}}}]^{-1}. \tag{2.10}$$

It is assumed in the probabilities (2.9) and (2.10) that an entire organ is irradiated and the dose distribution is homogeneous. More sophisticated models have been developed where volume and the serial/parallel effects together with treatment fractionation have been taken into account (see *e.g.* [38]).

One possible way to define a biological object function is

$$\mathrm{PUC}(D(x_k, \mathbf{a})) = \mathrm{TCP}(D(x_k, \mathbf{a}))[1 - \mathrm{NTCP}(D(x_k, \mathbf{a}))], \tag{2.11}$$

where PUC is the probability of an uncomplicated cure, TCP is the tumor control probability and NTCP is the normal tissue complication probability [42, 39]. TCP and NTCP are based on the probabilities (2.9) or (2.10). Parameters $\mathbf{a}$ control dose and, thus, the probabilites in equation (2.11), for example via model (2.9) or (2.10). Using the biological object function, a generic extremum problem can be stated:
Find the global maximum of

$$\max_{\mathbf{a}} \mathrm{PUC}(D(x_k, \mathbf{a}))$$

under constraints for control parameters $\mathbf{a}$.

The biological response of radiotherapy has been studied widely but as Wang *et. al* have expressed: "Currently available models for computing the TCP and NTCP are simplistic, and the data they rely on are sparse and of questionable quality" [74].

Chapter III

IMRT, Intensity Modulated RadioTherapy

3.1 Introduction

Intensity modulated radiotherapy consists commonly of two main phases. First, the intensity distributions of treatment fields are optimized to produce a desired dose distribution. Then, the intensity distributions are converted to deliverable treatment plans by reproducing the distributions using a field accessory. In this chapter, a few devices that have been used to shape field intensity will be described. The multileaf collimator (MLC) is the most applied of these devices and, hence, it is discussed in detail.

The novel idea of the research and development described in this thesis is to have just one phase in the IMRT instead of the two discussed above. The parameters of the intensity shaping device, namely the MLC, are directly optimized. Thus, the intensity-to-delivery device conversion is avoided. To connect the MLC to optimization, an MLC head scatter model that can be used in the dose calculation formula (2.1) is needed. The head scatter model produces an intensity distribution that depends on the parameters of the MLC. The head scatter model and the complete dose calculation model will be presented in this chapter.

3.2 Dose delivery techniques

3.2.1 MultiLeaf Collimator, MLC

The MLC has movable leaves, narrow shields, and can act as a block that attenuates radiation (see figure 3.2). The leaves are arranged in pairs to a left and right leaf bank. The number of leaves ranges from 20 to 128. The leaves are controlled by a computer that loads the MLC leaf positions from a file that is written by a treatment planning system. Complex shaped fields can be generated but they are limited by the movement constraints of leaves (see section 3.2.2). The most important use of the MLC is to shape conformal fields *i.e.* the field is opened only for the tumor projection with a margin added. Previously, the treatment volume was conformed using blocks individually moulded for each treatment field.

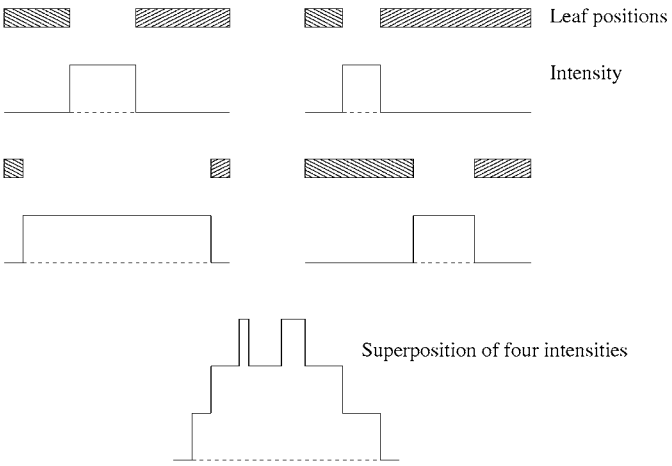

Figure 3.1: The illustration of the control of intensity using the delivery technique of multiple static collimation. Four segments construct a two-peaked intensity profile. Scattering and divergence have been excluded.

When blocks and the MLC are compared, it is obvious that time is saved as the individual blocks are not needed. The MLC might also be somewhat safer since accidents where a block is attached to its tray incorrectly are impossible with the MLC. In addition, a wrong block can be accidentally used for a treatment field or a block can drop from the tray. A drawback with MLCs is that they need a considerable amount of maintenance since they contain moving parts and electronic equipment.

First ideas about MLC were patented as early as 10 years after Wilhelm Röntgen's great discovery, but the concept was used first time in practise during 1960s in Japan [7]. Then, in the late 1970s and early 1980s, MLCs were developed that had an increased number of narrower leaves.

Next, the two most important techniques that are used when complex intensity distributions are created using the MLC are studied.

Multiple static MLC collimation

In the multiple static collimation intensity distributions are the superpositions of the intensities from a number of static, MLC-shaped treatment fields, the so-called subfields or segments. In figure 3.1 the superposition technique is illustrated. In figure 3.1, the amplitudes of the segments are equal which is not mandatory.

The use of the multiple static collimation is to reproduce the desired intensity distributions that are the solutions to the inverse problem of radiotherapy tretment planning (equation (2.5)). There are numerous approaches devoted to solve this problem (*e.g.* [4, 5, 73, 83]). One problem in the traditional use of the multiple static collimation is that smooth intensity profiles are reproduced approximately

by applying box-shaped functions (figure 3.1). This causes discrepancies between dose estimated using the continuous profiles and using the reproduced profiles. This is quite obvious consequence since the solution (2.5) is a linearized form of a basically non-linear problem, and the control device, the MLC, is not taken into account. For the reproduced profiles to be accurate, the number of segments would have to be very large.

One disadvantage of the multiple static collimation is that during treatment delivery x-rays are switched off, when the leaves are moved to their positions for the next segment. This can lenghten treatment times. In addition, the output of some treatment units is not instantly stable after the x-rays are swithced on. This can lead to errors in beam output. If there are a large number of segments having short beam-on times the instability error is bound to cumulate [78].

Dynamic MLC collimation

In the dynamic delivery of intensity modulated treatment fields, the leaves of the MLC are moved under computer control when the beam is on. The obvious advantage over the multiple static collimation is that intensity profiles are smoother, thus enabling more accurate reproduction of desired intensity distributions. Various control techniques for the dynamic leaf motion have been developed [17, 61, 62, 63, 67, 72].

3.2.2 Technical MLC constraints

The MLC, being a mechanical device, has a number of constraints for the movement of its leaves. Figure 3.2 shows examples of the different constraints. Since the main interest in this study is in the multiple static collimation, these constraints do not include the two constraints that are vital in dynamic MLC collimation, namely the maximum speed and acceleration of a leaf. The technical constraints are listed below. The first two constraints must be satisfied always while the rest are dependent on the MLC model.

1. The leaves may not collide:

$$a_{lsp} \leq b_{lsp},\ l = 1 \ldots L,\ s = 1 \ldots S_l,\ p = 1 \ldots P_l, \tag{3.1}$$

   where a_{lsp} is the left and b_{lsp} is the right edge of a leaf, L is the number of treatment fields, S_l is the number of segments and P_l is the number of the leaf pairs of the lth field.

2. The leaves can move only within certain limits:

$$a_{lsp} \geq W_l^{\text{left}},\ b_{lsp} \leq W_l^{\text{right}},\ l = 1 \ldots L,\ S = 1 \ldots S_l,\ p = 1 \ldots P_l, \tag{3.2}$$

   where W_l^{left} and W_l^{right} determine the width of a treatment field. If the field is symmetric in the leaf movement direction and the origin is at the central axis of the field, $W_l^{\text{left}} = -W_l^{\text{right}}$.

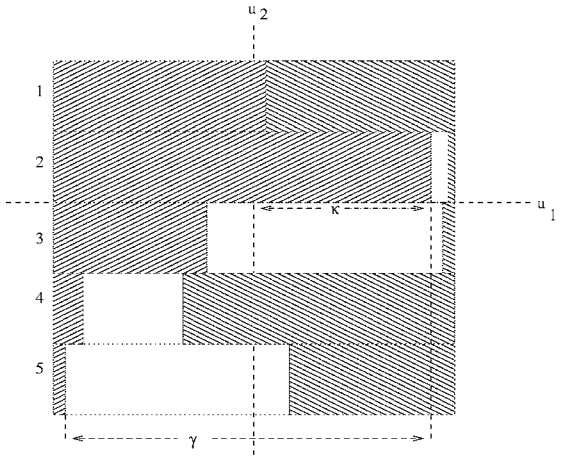

Figure 3.2: The schematic figure of some of the MLC leaf constraints. Leaf pair 1 shows the leaf overlapping condition, the left leaf of pair 2 shows restricted movement over field central axis (κ), pairs 3 and 4 (also pairs 1 and 2) show the violation of the interdigitation condition. Pairs 5 and 2 show the constrained distance between the leaves of the same bank (γ).

3. The leaf movements may have to satisfy the so-called interdigitation condition stating that the left and right (or the right and left) leaves of the adjacent leaf pairs cannot overlap:

$$b_{lsp} \leq a_{ls(p+1)},\ b_{ls(p+1)} \leq a_{lsp},\ l = 1 \ldots L,\ s = 1 \ldots S_l,\ p = 1 \ldots P_l - 1. \tag{3.3}$$

4. The travel of leaves over the central axis of a field can be restricted to

$$a_{lsp} \geq -\kappa,\ b_{lsp} \leq \kappa,\ l = 1 \ldots L,\ s = 1 \ldots S_l,\ p = 1 \ldots P_l. \tag{3.4}$$

   This is the maximum overtravel constraint.

5. The distance between the leftmost and rightmost left leaf, or between the rightmost and leftmost right leaf, can be limited to

$$a_{lsp} - a_{lsq} \leq \gamma,\ b_{lsp} - b_{lsq} \leq \gamma,\ l = 1 \ldots L,\ s = 1 \ldots S_l,\ p, q = 1 \ldots P_l. \tag{3.5}$$

   This is the leaf span constraint.

6. Some MLCs do not allow completely closed leaf pairs:

$$b_{lsp} - a_{lsp} \geq \varepsilon,\ l = 1 \ldots L,\ s = 1 \ldots S_l,\ p = 1 \ldots P_l. \tag{3.6}$$

Above κ, γ and ε are positive scalars.

3.2.3 Other techniques for beam delivery

A number of beam delivery techniques have been proposed for the IMRT. Besides the MLC, there are other suggested approaches like conventional blocks or dynamic wedges with the most futuristic of the suggestions being perhaps a small linac (linear accelerator) held by a robotic arm [81].

When a block of metal having a variable thickness is put in the shadow tray of a treatment unit, it is called a compensator. Because x-rays have to traverse through the compensator, they will experience location-dependent attenuation due to the variable thickness. With a compensator, it is relatively easy to modulate the intensity of a treatment field, but each modulated field must have an individual compensator and spectral changes in photon beam must be taken into account [49]. The fabrication of a compensator is time-consuming and requires special machinery. Also, compensators can get very thick, if high attenuation is desired [78].

Tomotherapy is a delivery technique that resembles the computerized tomography (CT). The analogy between tomotherapy and tomography is that in tomography, slices of a patient are imaged and 2D-figures reconstructed in series, whereas in tomotherapy patient is irradiated slice by slice. The treatment field is in principle one-dimensional, although the field has a finite height.

Because the treatment field in tomotherapy is narrow, a patient must be moved longitudinally during a treatment. Simultaneously to the movement of the patient, the gantry of the treatment unit is rotated. There are two techniques for the translation. In the first, the patient is translated between gantry locations (MIMiC, Multivane Intensity Modulating Collimator) [9]. In the second the patient is translated during the gantry rotation (Mackie's device) [47]. The field collimation is roughly similar in both techniques. In MIMiC, there is a slit aperture that can be blocked partly by vanes. Each vane is moved pneumatically and can be individually controlled. A vane can be moved in or out of the aperture in 40-60 ms. Intensity profiles can be constructed by blocking and opening parts of the aperture. The MIMiC collimator can be attached to an ordinary treatment unit whereas the Mackie's device is designed to be an independent CT-like treatment device.

The main disadvantage of tomotherapy is that only transaxial slices are irradiated at a time and thus, a patient must be translated during the treatment. This can cause difficulties in matching the individual slice-irradiations. Another concern is that there are a number of mechanical devices, like gantry, patient couch and the pneumatic vanes of the collimator, working at the same time. The treatment must be monitored carefully.

The scanning beam technique for treatment delivery has been proposed [43]. This can be done using dynamic jaw motion, but treatment times would be far too long. There is, however, a treatment unit (Racetrack Microtron) that can steer electron beam by bending magnets. The pencil beam that scans the area of a treatment field has too wide half-width (4 cm) at isocentre. For this reason the beam must be further collimated using *e.g.* the MLC [68] or a multi-hole

collimator [37].

3.3 Dose calculation model using MLC parameters

Here, the MLC is connected to the dose calculation model (2.1) and more attention is given to the parameters that affect the dose in the combined model. The main idea is to construct the intensity distribution $\Psi(u)$ using the positions of MLC leaves. The derivation of the model is for symmetric treatment fields. For asymmetric fields, modifications in the model are straightforward. The model is first described for one leaf pair of a field. Then the model is expanded to take into account multiple leaf pairs and several fields.

Let the maximal opening of a treatment field be a rectangle $U = [-W, W] \times [-K, K] \subset \mathbb{R}^2$. Denote the point of U by $u = (u_1, u_2)$. The leaves of an MLC are positioned orthogonally to u_2-axis and they have a positive width d. With the leaf positions, the positions of leaf edges in the leaf movement direction are meant. Assume that there are P leaf pairs (B_p, A_p), $p = 1, ..., P$. This means that the height of the field is $2K = Pd$. Let $U_p := [-W, W] \times [u_{2,p-1}, u_{2,p}]$, $p = 1, ..., P$ be the rectangular areas (along u_1-axis) determined by the leaf pairs (B_p, A_p). The areas U_p can be called channels [79].

Now, the intensity distribution for one channel is modelled. Here, the distribution is actually considered to be an intensity profile *i.e.* the effect of a leaf on the intensity distribution is assumed to be constant along u_2-axis. Another simplification is that intensity is 0 under a leaf and ψ_0 if a point is not under a leaf, where ψ_0 is the constant, non-modulated intensity distribution of a treatment field. The intensity profile for the pth leaf pair is

$$\psi_p(u_1) = \psi_0 \int_0^T \mathcal{H}(a_p(t) - u_1)\mathcal{H}(u_1 - b_p(t))\mathrm{d}t, \tag{3.7}$$

where $T \geq 0$ is the total irradiation time (beam-on time). Thus, the leaf positions $a_p : [0, T] \mapsto [-W, W]$ and $b_p : [0, T] \mapsto [-W, W]$ are functions of time. They are called the leaf trajectories. The trajectories are assumed to be sufficiently smooth, at least piecewise continuous and bounded functions. In addition, the trajectories must satisfy at least partially the MLC constraints that were formulated in section 3.2.2. The function $\mathcal{H}$ is the Heaviside function

$$\mathcal{H}(u) = \begin{cases} 1, & u \geq 0 \\ 0, & u < 0. \end{cases}$$

If the Fredholm integral equation (2.1) and the equation for MLC modulated intensity profile (3.7) are combined, dose deposited $D(x)$ by the channel U_p can be computed:

$$D(x) = \psi_0 \int_{U_p} h(x, u) \int_0^T \mathcal{H}(a_p(t) - u_1)\mathcal{H}(u_1 - b_p(t))\mathrm{d}t\mathrm{d}u_1\mathrm{d}u_2.$$

A treatment consists of L different treatment fields S_l. By different, it is meant that at least one of the rotations of a field (gantry, collimator or table)

is different from the other fields. Let T_l, $l = 1, ..., L$ be the beam-on times of the fields. Furthermore, for each segment S_l, let the end of leaf A_p be in a point $a_{lp}(t) \in [-W_l, W_l]$ at the moment $t \in [0, T_l]$ and let the end of leaf B_p be in a point $b_{lp}(t) \in [-W_l, W_l]$ at the moment $t \in [0, T_l]$. Also, let P_l be the number of leaf pairs of a field. Superposition is used to compute the dose deposition of an MLC-shaped field and then to compute the total dose from all the fields. Now, dose is formulated as

$$D(x) = \sum_{l=1}^{L} \psi_0^l \int_0^{T_l} \sum_{p=1}^{P_l} \int_{U_{lp}} h_l(x,u)\mathcal{H}(a_{lp}(t) - u_1)\mathcal{H}(u_1 - b_{lp}(t))\mathrm{d}u_1\mathrm{d}u_2\mathrm{d}t, \quad (3.8)$$

where U_{lp} is the pth channel of the lth treatment field and ψ_0^l is the constant, non-modulated intensity of the treatment field S_l. If the same treatment unit is used for a whole treatment, the intensity ψ_0^l can be divided into $\psi_0^l = \psi_0 w_l$, where ψ_0 is the output of the unit and w_l is the weight of the lth field. It is assumed that the radation quality is not changed during a treatment which enables the simple use of the weight w_l instead of $\psi_0 w_l$.

The dose calculation model (3.8) can be directly applied to the dynamic MLC delivery (section 3.2.1). It can be further enhanced by formulating the model to use leaf velocities as control parameters instead of leaf positions [72].

If the MLC is used traditionally as a block, the leaf trajectories would be constant functions. The obvious explanation is that leaves do not move during the delivery.

Unless there is a continuous kernel $h_l(x,u)$ available, a discrete version must be used. The discrete modification of the model (3.8) can be used in computation environment. Initially, the discrete kernel $\mathbf{H}$ (section 2.2, equation (2.3)) is assumed to have been measured or computed in such a way that the bixel size in u_2-direction is the same as the width of a leaf. Because of the discretization, intensity inside a bixel's area is simplified to be a constant. These assumptions lead to the discrete dose calculation model

$$D(x_k) = \sum_{l=1}^{L} w_l \sum_{p=1}^{P_l} \sum_{j=1}^{J_l} \int_0^{T_l} H_{lkpj}\mathcal{H}(a_{lp}(t) - u_1^j)\mathcal{H}(u_1^j - b_{lp}(t))\mathrm{d}t, \quad (3.9)$$

where J_l is the measure of width of treatment field l in bixels. This implies that the scaling of the parameters $a_{lp}(t)$ and $b_{lp}(t)$ must be changed accordingly. The integration over u_1 in equation (3.8) is replaced by the Riemann sum.

3.3.1 Dose calculation for multiple static MLC delivery

When treatment delivery consists of treatment fields that are further divided to segments (see section 3.2.1), the segments are superpositioned in a dose calculation model. Instead of one treatment field, there are a number of segments each having its own weight. The discrete model (3.9) is modified to take into account the segments of multiple static delivery. MLC leaves do not move when a beam is on in multiple static delivery. The constant leaf positions imply that the leaf

trajectories are piecewise constant functions. Using the Riemann sum to compute the integrals of the now piecewise constant trajectories, dose model becomes

$$D(x_k) = \sum_{l=1}^{L} \sum_{s=1}^{S_l} w_{ls} \sum_{p=1}^{P_l} \sum_{j=1}^{J_l} H_{lkpj} \mathcal{H}(a_{lsp} - u_1^j) \mathcal{H}(u_1^j - b_{lsp}), \tag{3.10}$$

where w_{ls} is the weight of sth segment of lth field. The position parameters of leaves a_{lsp} and b_{lsp} are not functions of time as in (3.9) but refer now to sth segment. The weight parameters w_{ls} control dose deposition together with leaf positions. The leaf positions and segment weights construct a complete set of control parameters that are needed to define dose $D(x_k)$. Integration over time is not needed since the trajectories of leaves are now piecewise constant. Instead, the integral is handled by the superposition of segments. In practice this means that leaves do not move during irradiation, when beam is "on".

3.4 The head scatter model for MLC

In the previous section, the model for dose calculation using the MLC parameters was formulated. However, a simplified model for the transport of radiation through the MLC was applied, namely the Heaviside function $\mathcal{H}$. The Heaviside function does not take into account the scattering from the edge and the sides of a leaf. It was also assumed that there is no radiation leakage through a leaf.

A relevant method to study radiation in a treatment unit is the Monte Carlo code BEAM to simulate radiotherapy treatment units [58]. The BEAM code was used to compute radiation flux at a plane under the MLC. Unfortunately, the BEAM model for the MLC was not accurate enough to fully describe the curvilinear geometry of the edge of an MLC leaf [1]. To overcome the problem of simulating the MLC scatter and leakage, a different approach, developed by Chen *et. al* [14], was used. The ray tracing and primary/extended source models were adopted from their work.

3.4.1 Ray tracing method

Here, the scattering from the leaf sides or the so-called tongue and groove effect [75] were not considered. The side scattering has been formulated in [73].

An intuitive way to determine fluence distribution at a scoring plane located under the MLC is to cast rays from a small focal source at the level of the target of a treatment unit to the plane. The attenuation of x-rays can be calculated using the pathlengths of the casted rays in the attenuating material (*i.e.* the MLC). Chen *et. al* [14] used a structure function for the calculation of the pathlengths since they wanted to model the MLC three-dimensionally. Here, instead, simple geometrical calculations are used to determine the pathlengths because only the effects of scatter at the edge of a leaf were studied. The cross-sectional geometry of a Varian (Palo Alto, USA) MLC was modelled.

[1] At the time of writing there is availabe a more sophisticated model for the MLC.

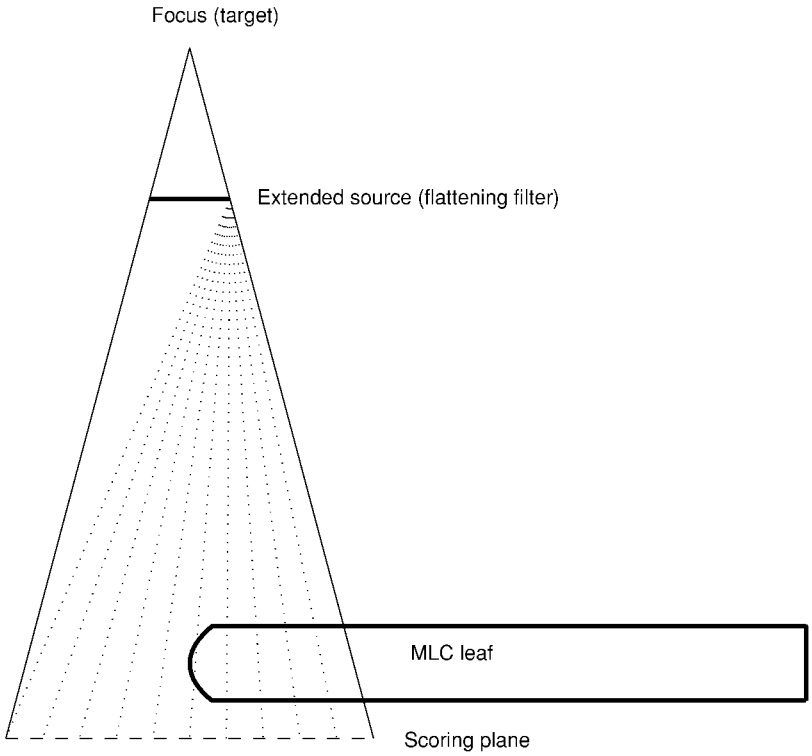

Figure 3.3: The example of the two-dimensional ray tracing for the determination of the MLC head scatter. There are two source of radiation: a focal and extended source. Rays casted from one point of the extended source are shown.

The determination of intensity using ray tracing from a tiny source (focal source) can be further enhanced by including scatter. Scatter is modelled by an extended source at the level of the flattening filter of a treatment unit. A dual source model can now be used: a small focal source at the target and a larger extended one at the flattening filter plane. Fluence distributions for the sources can be obtained from Monte Carlo simulations [44] and the source dimensions and material properties (tungsten, copper and lead) from the manufacturer of the treatment unit.

The convolution of the focal and extended sources, suggested in [14, 44], was not used. In convolution, a small angle approximation is used that causes small errors to the computed intensity. The full ray tracing is slow, when routine use is considered, but gives higher accuracy than convolution.

The intensity for the 4 MV Varian Clinac 600C (Varian, Palo Alto, USA) having a Varian MLC was computed. The fluence disitributions were estimated from the parameters represented by Liu *et. al* [44]. Figure 3.3 shows an example of

rays casted from the left edge of the extended source (the flattening filter). Since the intensity profile did not change when more than 1000 rays were traced, this number was used in computations.

3.4.2 Data fitting using an analytic basis function

As an efficient form of the MLC head scatter was needed for the dose calculation, an analytic function was fitted to the intensity profile that was determined by ray tracing. The first approximation, Heaviside function $\mathcal{H}$, was replaced by an analytic function

$$\tilde{\mathcal{H}}_{\text{right}}(y, a) = c_0 + \frac{c_1}{2}[1 - \tanh(c_2(b - y - c_3))], \tag{3.11}$$

where $c_0, \ldots, c_3$ are the fitting parameters, y is coordinate in leaf movement direction and b is the position of a leaf in the right leaf bank of the Varian MLC. For a leaf in the left bank, analogously to (3.11),

$$\tilde{\mathcal{H}}_{\text{left}}(y, a) = c_0 + \frac{c_1}{2}[1 - \tanh(c_2(y - c_3 - a))], \tag{3.12}$$

where a is the position of a left leaf.

Figure 3.4 shows the ray traced intensity profile of the right leaf of the Varian MLC and the fitted intensity using equation (3.11). For comparison, intensity that was computed without the contribution of the extended source is also shown.

It can be noticed in figure 3.4, that the analytic fit to the original, ray traced, intensity is not excellent. The original intensity is not "mirrored-symmetric" as was simplified, when the basis was chosen for $\tilde{\mathcal{H}}_{\text{right}}$ and $\tilde{\mathcal{H}}_{\text{left}}$. The fit could be improved for example by using a linear combination of two or more of the basis functions but, as will become clear in section 7.2, where the dosimetric testing is discussed, the goodness of the fit is of no great importance.

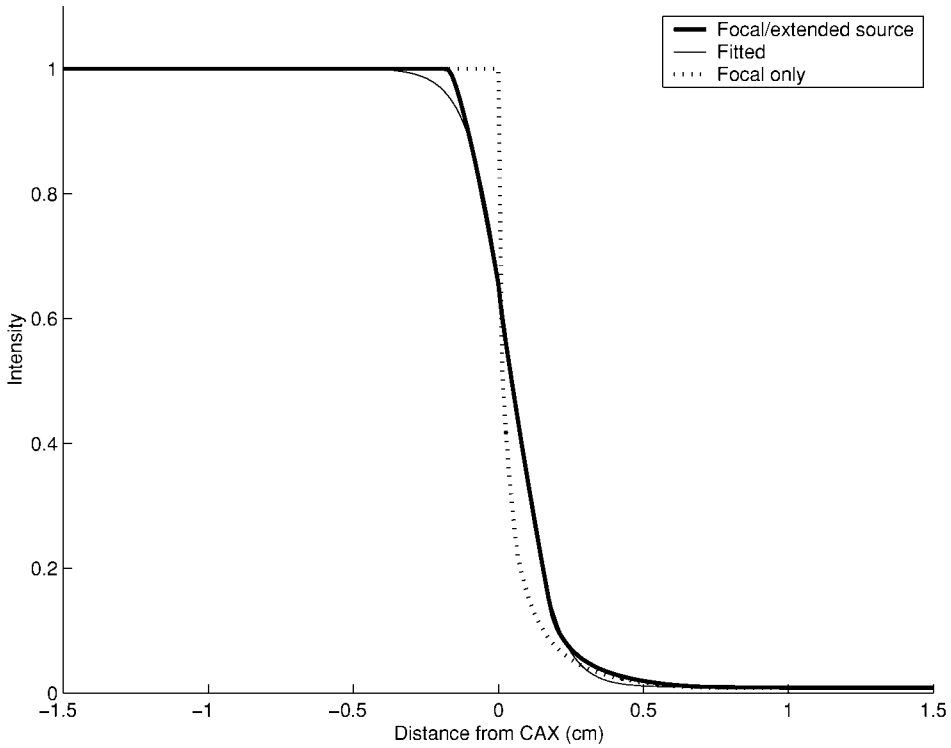

Figure 3.4: A comparison of the intensity profiles that were determined by ray tracing (focal source only, focal end extended sources) and using a fit to an analytic function. A right leaf of the MLC is at the central axis of the treatment field (CAX).

CHAPTER IV

Implementation

4.1 Introduction

Here, the inverse problem of radiotherapy treatment planning (chapter 2) and the IMRT (chapter 3) are combined to a complete inverse planning system that uses the multiple static collimation. The main goals of the design of the system were:

1. A system which optimizes MLC-parameters so that the resulting dose distribution is as close as possible to a dose prescription.
2. The MLC-parameters define treatment fields that use multiple static collimation (section 3.2.1).
3. The system should work for all MLCs *i.e.* the system must be flexible enough to accept the physical dimensions and the mechanical constraints (section 3.2.2) of any MLC.
4. The result of the computer optimization is a treatment plan, not a set of intensity distributions.
5. The estimated dose distribution of an IMRT treatment plan must be dosimetrically accurate.
6. The fast execution of the computer optimization is desirable.
7. The system must work together with the Cadplan treatment planning system (Varian Medical Systems Finland, Espoo, Finland).

The optimization system was written to be a stand-alone application. First, the system reads the dose deposition kernel ($\mathbf{H}$), the dose prescription and the characteristics of the treatment fields and the MLC. The kernel has been calculated by the dose calculation engine of Cadplan. After the computer optimization is finished, the system writes the positions of the leaves of the MLC, the weights of individual segments and the intensities of the treatment fields into separate files.

The intensities, produced by the multiple static collimation, can then be used to dose calculation purposes in Cadplan. From the leaf position files, MLC-files are generated that can be fed in to the MLC control of a treatment unit. Monitor units for the segments are calculated based on the weights of the segments.

In this chapter, the essential parts of the developed inverse planning system are described. The description follows the same sequence of actions as the process of inverse planning does.

4.2 The user interface

In the graphical user interface (GUI), the user gives information mainly concerning the treatment fields and the dose prescription. For the treatment fields the number of multiple static segments (3 – 20) must be given. The dose prescription includes:

- For a PTV, lower and higher dose constraints and relative weights.
- For an OAR, a possible higher dose constraint and a relative weight.
- For an OAR, possible dose volume constraints: doses, volumes and relative weights. An OAR can have more than one (up to ten) dose volume constraints.
- The densities of calculation points in the volume of an organ. Typical values range from 15 points/cm^3 to 35 points/cm^3.

In addition, the maximum number of iteration rounds (*e.g.* 100), the radial distance after a calculation point is discarded (sections 2.2.2 and 4.3) and the maximum optimization time must be given.

In addition to the interaction with the user, the GUI saves information about the treatment fields and the MLC including *e.g.* field sizes and MLC characteristics (leaf width etc.).

4.3 Calculation of the dose deposition kernel

The dose calculation engine of Cadplan was used to compute the discrete dose deposition kernel $\mathbf{H}$. The engine is a variant of the pencil beam algorithm introduced in [65]. Dose is determined at predetermined points in the patient space. The kernel is computed bixel by bixel *i.e.* dose contribution from an individual bixel is recorded to all points of interest in the patient space. These contributions are called dose deposition coefficients (DDCs).

An existing implementation, the pre-optimization routine `PrOp` of Cadplan, was exploited in the computation of the kernel. The routine is originally meant to be used with another optimization system which is based on [61]. Since the dose deposition kernel is universal for all models based on the Fredholm integral equation (2.1), the same pre-optimization routine can be used for all discrete, pencil beam based optimization methods.

For each treatment field and PTV/OAR `PrOp` saves the DDCs as a sparse array. Every point in patient space has a dose deposition matrix that defines how

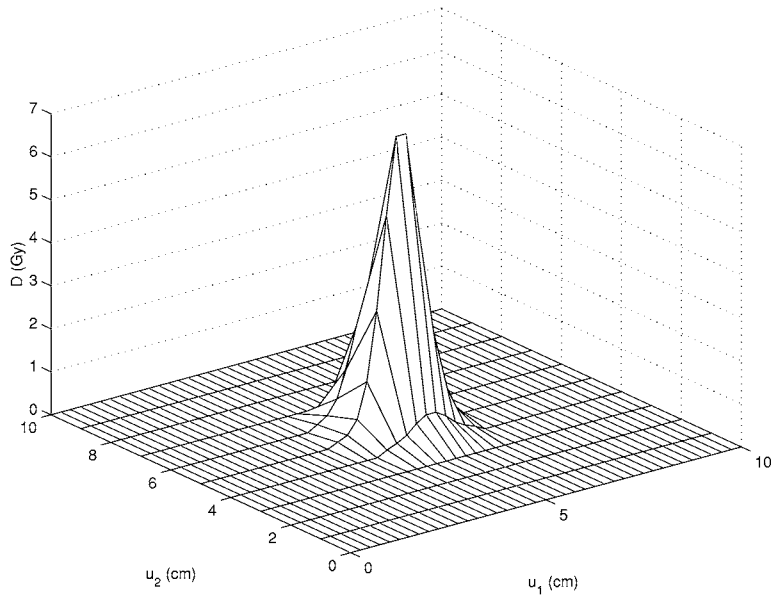

Figure 4.1: The dose deposition matrix in the treatment space (treatment field) for one point in the patient space. u_1 is the leaf movement direction. The total dose of the field can be computed by simply summing the dose depositions of bixels.

much each bixel of a treatment field contributes to the dose at the point (figure 4.1). Since the size of a bixel is small, $W_{\mathrm{MLC}} \times 2.5$ mm^2 (W_{MLC} is the width, *e.g.* 10.0 mm, of a leaf of the MLC at the distance where the field size is defined), the matrix is mostly filled with zeros because the intensity of the fluence of a small pencil beam decays radially fast. In fact, user can determine how close a point must be radially from the pencil beam axis to be included in DDCs. Typical value for the radial distance is 1.5 cm. The shorter the radial distance the fewer elements of the matrix get non-zero values. Naturally, the zero-valued bixels need not to be saved. This enables sparse matrix structure that can be used to save computer memory.

4.4 Specific extremum problem

Here, the generic extremum problem applying the 40 leaf-pair Varian MLC and the multiple static collimation is specified. The object function approach, equation (2.7) is used. The specifical, linear MLC constraints are as follows:

- The inhibition of the leaf collition – constraint (3.1).
- The leaves can move only within a treatment field – constraint (3.2).

- To avoid field shapes that have multiple openings, the interdigitation limitation was included, although it is not mandatory for the Varian MLC – constraint (3.3).
- The maximum overtravel is $\kappa = 16.0$ cm – constraint (3.4).
- The maximum leaf span is $\epsilon = 14.5$ cm – constraint (3.5).

The weight w_{ls} of a segment, introduced in section 3.3.1, needs special attention. In multiple static collimation, the beam is switched off during leaf movement. When the beam is then switched on, the flux of photons is not stable during the first few monitor units (<10 MUs). Thus, it is not advisable to use segments that have small weights corresponding to short irradiation times. A constraint can be used to either force the weight of a segment to be greater than a specified value or to force the weight to zero. The simple constraint is

$$w_{ls} \geq w_0 \ \vee \ w_{ls} = 0, \tag{4.1}$$

where w_0 is the lowest accepted value of a segment weight. The constraint (4.1) is not computationally practical. A better non-linear (quadratic) formulation is

$$w_{ls}(w_{ls} - w_0) \geq 0. \tag{4.2}$$

While the constraint (4.2) may increase the value of the object function, when the value of any of the w_{ls} is between zero and w_0, it does not ensure that the weights are non-negative. Hence, an additional constraint has to be added:

$$w_{ls} \geq 0. \tag{4.3}$$

Although it would be possible to link the monitor units and w_0, an arbitrary figure for the lowest accepted value w_0 was used.

The implemented object function is

$$\begin{aligned} F(\mathbf{a},\mathbf{b},\mathbf{w}) \quad = \quad & \sum_{i=1}^{N_{\mathrm{PTV}}} \Bigg[c^{\mathrm{high}}_{\mathrm{PTV},i} \sum_{k \in J_{\mathrm{PTV},i}} |(D_{\mathrm{PTV},i} - D(x_k,\mathbf{a},\mathbf{b},\mathbf{w}))_-|^2 \\ + \quad & c^{\mathrm{low}}_{\mathrm{PTV},i} \sum_{k \in J_{\mathrm{PTV},i}} |(D(x_k,\mathbf{a},\mathbf{b},\mathbf{w}) - d_{\mathrm{PTV},i})_-|^2 \Bigg] \\ + \quad & \sum_{i=1}^{N_{\mathrm{OAR}}} \Bigg[c_{\mathrm{OAR},i} \sum_{k \in J_{\mathrm{OAR},i}} |(D_{\mathrm{OAR},i} - D(x_k,\mathbf{a},\mathbf{b},\mathbf{w}))_-|^2 \\ + \quad & \sum_{j=1}^{N_{\mathrm{dv,OAR},i}} c_{\mathrm{dv,OAR},ij} \Bigg| \Bigg(V_{\mathrm{dv,OAR},ij} - \frac{1}{|J_{\mathrm{OAR},i}|} \\ \times \quad & \sum_{k \in J_{\mathrm{OAR},i}} \mathrm{erf}(D(x_k,\mathbf{a},\mathbf{b},\mathbf{w}) - D_{\mathrm{dv,OAR},ij}) \Bigg)_- \Bigg|^2 \Bigg] \end{aligned}$$

$$+ \quad c_{\text{weight}} \sum_{l=1}^{L} \sum_{s=1}^{S_l} |(w_{ls}(w_{ls} - w_0))_-|. \tag{4.4}$$

There are a number of modifications when the object function (4.4) is compared to the generic object function (2.7). The object is now a function of left and right leaf positions and segment weights, **a**, **b** and **w**, respectively. N_{PTV} is the number of PTVs and N_{OAR} the number of OARs. Each OAR can have $N_{\text{dv,OAR},i}$ dose volume constraints ($N_{\text{dv,OAR},i} \leq 10$). Also, the constraint for too small segment weights is added.

The computed dose is a combination of the dose calculation model for the multiple static collimation (3.10) and the analytic MLC head scatter models (3.12) and (3.11). At calculation point x_k,

$$D(x_k, \mathbf{a}, \mathbf{b}, \mathbf{w}) = \sum_{l=1}^{L} \sum_{s=1}^{S_l} w_{ls} \sum_{p=1}^{P_l} \sum_{j=1}^{J_l} H_{lkpj} \tilde{\mathcal{H}}_{\text{left}}(j, a_{lsp}) \tilde{\mathcal{H}}_{\text{right}}(j, b_{lsp}). \tag{4.5}$$

Now, the practical extremum problem can be stated:
Find the global minimum of

$$\min_{\mathbf{a},\mathbf{b},\mathbf{w}} F(\mathbf{a}, \mathbf{b}, \mathbf{w})$$

under the constraints (3.1),(3.2),(3.3), (3.4), (3.5) for leaf positions **a** *and* **b** *and under the constraint (4.3) for segment weights* **w**.

4.5 The optimization method

An optimization method is needed to find a minimum of the object function that was formulated in the previous section. A careful choice of the method is vital, when practical treatment plans are desired. There is no single optimization method that can solve all problems with equal efficiency and accuracy. The choice of the optimization method is always dependent on the capability of the person/persons who choose the method. Some understanding of the optimization problem is essential to categorize a problem optimizationwise. By categorizing, choices like differentiable/non-differentiable object function, linear/non-linear object function, no/box/linear/non-linear constraints etc. are meant. The deeper understanding of a problem and testing of different methods can lead to the choice of the method that solves the problem perfectly [24].

There were three main aspects that governed the choice of the optimization method: non-linear object function, linear constraints and large-dimensionality. The first two facets are obvious (section 4.4). The dimensionality of the problem needs some clarifying. Consider a five-field treatment plan, where each field has 15 leaf-pairs and 15 multiple static segments. The total number of segments is 75, which is also the number of the segment weights. For each segment, the field is shaped by 30 leaves, 15 left and 15 right leaves. Now, the total number of the MLC parameters is $N_{\text{tot}} = 2325$. For a non-linear optimization problem, N_{tot} is large for the present computer hardware and algorithms.

There are a few algorithms to choose for the non-linear, linearly constrained and large-dimensional extremum problem. Because the optimization problem is multiextremal, global optimization algorithms would be desirable [19]. Global optimization methods are not straightforwardly applicable. At present time, the generic global optimization algorithms, such as the simulated annealing [52, 57, 76] or the genetic programming [41], cannot solve large-dimensional, non-linear problems in a reasonably short time. On the other hand, the use of a local optimization algorithm forces one to determine an initial guess for the MLC parameters. This is a serious drawback because the final solution depends heavily on the initial guess. It is, however, possible to estimate the initial MLC leaf positions. This subject is discussed in section 4.6.

Local optimization methods have been studied in the inverse problem of radiotherapy, *e.g.* [18, 30, 31, 35, 45, 84]. In all these studies the linear model (2.3), or its variant, has been used for dose calculation, and the results of the optimization were the intensity distributions of treatment fields. Thus, the results of the studies are not applicable, since here the parameters in the dose calculation are the leaf positions and segment weights. To ensure the fast convergence of the object function (4.4), gradient of the object function must be used. Because the discrepancy between the optimized and prescribed dose distributions can be large, optimization methods that use only the gradient, or more precisely the Jacobian, of the object function will not be efficient [24]. Knowledge about the second derivate, namely the Hessian, of the object function is desirable.

The potential local optimization algorithms were LANCELOT [16], SNSOL [25], MINOS [53] and L-BFGS-B [85]. The algorithms use the same principal method. Let $f(x)$ be an object function that is to be minimized. A local quadratic approximation m_c (from Taylor series expansion) of f at point x_c is

$$m_c(x) = f(x_c) + \nabla f(x_c)^{\mathrm{T}}(x - x_c) + \frac{1}{2}(x - x_c)^{\mathrm{T}}\nabla^2 f(x_c)(x - x_c).$$

Let x_+ be the minimizer of m_c [32]. Now, x_+ is the approximation of the parameters that minimize f. Then, by setting $x_c = x_+$, a new approximation for the minimizer of f can be calculated. This iteration (or sequential quadratic approximation) can be continued until convergence to a minimum is achieved. The method is called the Newton's method, and when an approximation of the Hessian $\nabla^2 f(x_c)$ is used, the method is called the quasi-Newton or variable metric method. The amount of documentation about the Newton's method is vast [20, 23, 24, 32].

The main problem of the Newton's method, when it is used to solve large-dimensional problems, is the size of the Hessian matrix $\nabla^2 f(x_c)$. Let $x \in \mathbb{R}^{N\times 1}$. Then, there must be enough computer memory for N^2 double precision (8 bytes) floating point numbers. If N is large, computer memory must be used extensively. All four potential algorithms overcome this problem by using either a known sparsity pattern of the Hessian (LANCELOT) or a limited memory Hessian (SNSOL, MINOS and L-BFGS-B).

None of the algorithms is perfect for the treatment planning problem. LANCELOT was not suitable because there is no knowledge about the sparse-

ness or about the sparsity pattern of the Hessian in the problem. SNSOL and MINOS are effective only when the number of decision parameters roughly equals the number of active constraints. In inverse treatment planning, this would mean that almost all leaves should be under a constraint *i.e.* a leaf is about to collide the opposite leaf ($a_{lsp} = b_{lsp}$, constraint 3.1) or a leaf is at the edge of a treatment field (constraint 3.2) etc. Finally, L-BFGS-B accepts only box constraints, higher and lower limits, for parameters.

Despite the fact that L-BFGS-B is not capable of accepting the MLC constraints, it was considered to be sufficient for the purpose and was chosen to be the optimization algorithm. It is possible to determine the box constraints according to the MLC constraints. After the inital leaf positions are determined, for each leaf position, lower and higher limits are determined in such a way that a leaf will not violate the MLC constraints during the actual optimization. In other words, the leaf movement was restricted to feasible regions. Because of the rigid low/high limits, a leaf cannot move to position that would otherwise be feasible. For example, let the left edge of a treatment field, in leaf movement direction, be at $-W$ and the right edge at W. The central axis of the field is at 0. The initial values of the left and right leaf of a leaf-pair are at $-W/2$ and $W/2$, respectively. The intuitive choice for the higher limit of the left leaf's position is 0 which, at the same time, is the lower limit for the position of the right leaf. If, during the optimization, the right leaf is at the position $3W/4$ and left is at 0, the left leaf cannot move further right, *e.g.* to the position $W/5$, notwithstanding it will not collide the right leaf.

4.5.1 Optimization applying the L-BFGS-B

A description of the features specific for the L-BFGS-B, especially about the limited memory feature, are given in appendix A.

L-BFGS-B uses reverse communication *i.e.* it asks the user, or a user written driver routine, the necessary information instead of calling special subroutines, where the value of an object function or its gradient would be computed. The reverse communication property of the L-BFGS-B enables simple implementation since the optimization subroutine can be called from a driver routine and the only input parameters, besides the control parameters of the L-BFGS-B, are the value of the object function and its gradient, which can be computed in the driver. The reverse communication is a favourable feature since otherwise the large dose deposition kernel would have to be passed to optimization subroutine as an input argument or the kernel to be declared as a common (Fortran) parameter.

A simplified pseudo-code of the implemented driver routine:

```
PROGRAM main

  PROCEDURE do_optimization
  BEGIN
    read dose prescription
    read field related data and number of subfields
    read MLC related data
    FOR field DO
```

```
    BEGIN
      read the dose deposition coefficients for PTVs
      read the dose deposition coefficients for OARs
    END
    initialize leaf positions
    WHILE NOT last_round DO
      CALL L-BFGS-B algorithm
      compute the value of the object function
      compute the gradient of the object function
      IF NOT optimization_limit_reached THEN CONTINUE
      IF NOT time_limit_exceeded THEN CONTINUE
      write results (object function value, DVHs,
                     intensity matrices, solution)
      reread dose prescription
      IF NOT prescription_changed THEN CONTINUE
   ELSE restart optimization with the new prescription
    END
    write results (object function value, DVHs,
                   intensity matrices, solution)
  END

BEGIN
  read sizes for variable allocation
  CALL do_optimization
END
```

The description of the algorithm:
Before the actual optimization, the sizes of the arrays needed to define the dose deposition are read and memory is allocated for the arrays. This way the coding of the optimization algorithm is simple as dynamical memory allocation is avoided. First, in the actual optimization, the dose deposition kernels, the dose prescription and treatment field information are read form files and the MLC leaf positions are initialized. Then, the iteration is carried on as long as the time limit is not met, the value of the object function converges or the infinite norm of the gradient of the object function is above a predetermined threshold. The dose prescription can be modified from the GUI during the iteration. If modified dose constraints are detected, the iteration continues with the new prescription. At each iteration and at the end, results are written to files for evaluation and treatment planning purposes.

4.6 The initialization of leaf positions

As a local optimization algorithm was used for the computation of the MLC parameters, a method to determine an initial set of MLC parameters was developed. Two approaches were implemented and tested. In the first, geometrical one, the two-dimensional projections of PTVs and OARs to the treatment space (the plane of a treatment field) were used to determine field shapes [71]. Figures 4.2 and 4.3 explain the approach. One segment is used to open the field for a PTV while others open the field in such a way that an OAR is shielded from left or from right. The geometrical approach is simple to implement and to understand but no

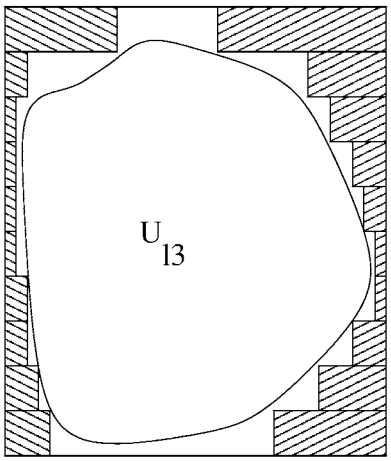

Figure 4.2: The conformation of the projection of a PTV (from Tervo and Kolmonen [71]).

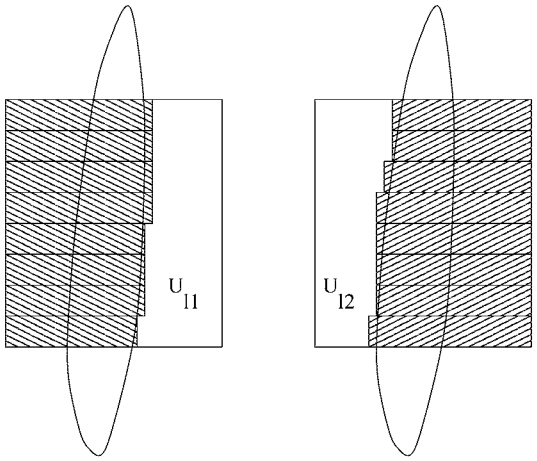

Figure 4.3: The avoidance of the projection of an OAR using two multiple static segments (from Tervo and Kolmonen [71]).

further work has been done to apply the approach for more than three segments and to more difficult OAR projections than the one in figure 4.3.

The other approach developed to initialize the leaf positions of the multiple static segments is based on the use of the discrete dose deposition kernel (section 2.2.2). This approach was used in the developed inverse planning system. The idea is to use the transpose of the kernel to mimick the backward (or adjoint)

transport of photons. The backward transport was used by Jeraj and Keall [29] to compute the initial guess of intensity distributions in their Monte Carlo based method for inverse treatment planning. They used Monte Carlo code to place a radiation source in PTV and scored particles exiting the patient geometry, thus obtaining an intensity distribution that was used as an initial guess for further optimization.

The initial intensity distribution $\hat{\Psi}_{lp}$ of the pth leaf-pair of the treatment field l is computed using equation

$$\hat{\Psi}_{lp} = \mathcal{S}\Big[\sum_{i=1}^{N_{\mathrm{PTV}}} c_{\mathrm{PTV},i} \mathcal{M}(H_{lpi}^{\mathrm{T}} I) - \sum_{i=1}^{N_{\mathrm{OAR}}} c_{\mathrm{OAR},i} H_{lpi}^{\mathrm{T}} I \Big], \tag{4.6}$$

where H_{lpi}^{T} is the transposed dose deposition kernel (a matrix) for lth field, pth leaf-pair and ith structure (a PTV or an OAR). The heuristic operator $\mathcal{M}$ modifies the intensity determined for PTVs by averaging and multiplying the intensity. The operator $\mathcal{S}$ smooths the intensity profile slightly. The weights for PTVs and OARs are $c_{\mathrm{PTV},i}$, $c_{\mathrm{OAR},i}$, respectively. I is a unit vector.

While equation (4.6) may not be physically relevant, when compared to the backward transport, it does have some heuristic analogy to the transport method. At a bixel in treatment space, all transposed dose depositions from a PTV are summed. It is clear that the unit of the summed quantity cannot be any of those that describe fluence, the quantity is simply the unitless weight of a bixel. This weight distribution must be heavily modified (operator $\mathcal{M}$) before a PTV is well conformed. From this weight distribution the transposed dose depositions from an OAR are subtracted, thus lowering the weights of the bixels that irradiate the OAR. A dose volume limit is too complex a constraint to be included in the simple model (4.6).

When intensity distributions are computed, the MLC leaf settings can be determined by methods described in *e.g.* [4, 5]. The implementation differs from the usual one because the leaf constraints must be taken into account and they must be transformed to box-constraints due to the choice of the optimization algorithm (see section 4.5).

4.7 Data output during optimization

During the iteration the value of object function is written to a file to observe the progress of the optimization. Dose volume histograms (DVHs) are also saved at each iteration, and they can be viewed during the optimization.

During and after the optimization, the intensities of treatment fields are written to files as intensity matrices. Each element of a matrix represents the intensity of a 2.5×2.5 mm^2 bixel. Reason for this kind of format is that the same bixel size is used by the dose calculation engine of Cadplan.

4.8 Dose calculation and MU determination

After optimization, the resulting intensity matrices are used to compute dose distribution employing the calculation algorithm of the Cadplan treatment planning system. Dose distribution is computed in transversal, two-dimensional slices. For plan evaluation, dose volume histrograms can then be determined from the computed dose distribution. The histrograms may differ from those that are computed during optimization. The reason for this is that the locations of the calculation points in patient space are not necessarily identical. When dose deposition coefficients are computed before the actual optimization, calculation points are spread to the entire volume of a PTV or an OAR, whereas during the dose computation after optimization, the dose distribution is computed in transversal slices.

To be able to deliver the optimized plan the number of Monitor Units (MUs) must be determined for each multiple static segment. The procedure is as follows:

- Compute the dose distribution applying optimized fluences.
- Based on the dose distribution and dose normalization, calculate MUs for treatment fields.
- Using the total amount of MUs for a treatment field and the weights of the multiple static segments of the field, compute MUs for the segments. This can be formulated as

$$N_{\mathrm{MU}}^{\mathrm{Seg}} = \mathrm{round}\left(N_{\mathrm{MU}}^{\mathrm{Tot}} \frac{w_{\mathrm{Seg}}}{w_{\mathrm{Tot}}}\right), \tag{4.7}$$

  where $N_{\mathrm{MU}}^{\mathrm{Seg}}$ is the amount of MUs for a segment, $N_{\mathrm{MU}}^{\mathrm{Tot}}$ is the total amount of MUs for the treatment field, w_{Tot} is the weight of the field and w_{Seg} is the weight of the segment. The weight of a field w_{Tot} is the maximum value of the intensity matrix of a field whose intensity has been modulated. The operator round(y) rounds y to the nearest integer since $N_{\mathrm{MU}}^{\mathrm{seg}} \in \mathbb{N}$.

Chapter V

Dosimetric testing

Patient safety demands extensive dosimetric testing when new methods are brought to radiation therapy. Agreement between the measured and the calculated dose must be within a narrow tolerance (set by *e.g.* the ICRU [27]).

The developed optimization system was tested applying film measurements in a plexiglass phantom. Film was chosen because it was readily available and the dose distribution of a plane could be measured at once. For more accurate point or profile dosimetry, other detectors, such as ionization chamber and detector array, can be used [54].

5.1 Measurement procedure

The measurement procedure consisted of:

1. Choose one field from an optimized treatment plan.
2. Calculate dose distribution from the chosen field to the phantom at a plane perpendicular to the field central axis.
3. Normalize the dose to 1 Gy at the central axis of the treatment field in the calculation plane.
4. Write an MLC-file for each multiple static segment.
5. Compute monitor units (MUs) for segments.
6. Set up the measurement (phantom, film and treatment unit).
7. Irradiate the film using the multiple static segments of the field and the computed MUs.
8. Scan the film with a film scanner (Hewlett-Packard, Palo Alto, CA, USA).
9. Convert the optical density of the measured film to dose using a measured calibration curve (the optical density as a function of absorbed dose).

After the measurements and film scans, the measured dose could be compared with the calculated dose. The calibration curve for film dosimetry was determined by irradiating several films with known values of dose.

5.2 The tested fields

Three intensity modulated fields were measured. The modulation of intensity was different for each field. The first field had a wedge type intensity distribution with a sharp peak at the thick end of the "wedge", the second one had an area of high intensity perpendicular to the direction of leaf motion and the intensity distribution of the third one was highly irregular. The intensity distributions of the tested fields featured mainly two different aspects: areas of flat intensity and steep intensity gradients. In addition, to test slow changes in the intensity, the intensity distribution of the second field decreases gradually across the width of the field. Because dose was measured in a homogeneous phantom, the intensity distributions, with a minor blurring caused by scattering in the phantom, can be seen in dose distributions that are showed in the figures in section 5.5.

The measured and computed dose distributions were determined using equal spatial resolution. The computed dose distributions were calculated using 1.25×1.25 mm^2 (the first and third fields) or 2.5×2.5 mm^2 (the second field) resolutions. The reason for the coarser resolution in the dose distribution of the second field is that the whole distribution would not have been calculated if the finer resolution had been used. The initial resolution of the measured dose distributions (150 dots per inch) was determined by the film scanner. To be able to compare the measured and computed distributions, the measured ones were averaged to decrease the amount of random noise over square areas whose dimensions were the same as the resolutions of the computed distributions (1.25×1.25 mm^2 or 2.5×2.5 mm^2).

5.3 Phantom

The phantom was cuboid in shape and was divided to four parts so that a radiographic film could be placed between the parts (figure 5.1). The material of the phantom was lexan (polycarbonate, $(C_{16}H_{14}O_3)_n$).

The computed dose distributions were determined in a synthetic water phantom. To be able to compare the measured and computed dose distributions, the depth from the surface of the phantom to the plane where the computed distributions were determined had to be corrected. The energy of the radiation (nominally 4 MV) indicates that the primary interaction between radiation and the phantom material is Compton scattering. Now, electron densities can be used to determine the depth correction [33]. The correction is

$$d_{\text{water}} = \frac{\rho_{\text{lexan}}}{\rho_{\text{water}}} \frac{(Z/A)_{\text{lexan}}}{(Z/A)_{\text{water}}} d_{\text{lexan}}, \tag{5.1}$$

where the density of lexan $\rho_{\text{lexan}} \approx 1.20$ gcm^{-3}, density of water $\rho_{\text{water}} \approx 1.00$ gcm^{-3}, $(Z/A)_{\text{lexan}} \approx 0.53$ and $(Z/A)_{\text{water}} \approx 0.56$ [64]. The result of the correction

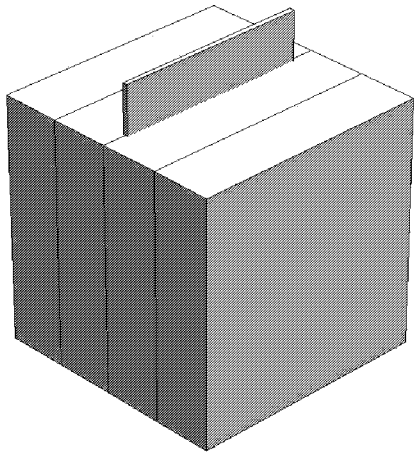

Figure 5.1: A schematic illustration of the phantom that was used in dosimetric film measurements. The phantom is constructed of four plexiglass parts. A radiographic film is visible between the two middle parts.

is that the calculations must be made in a plane that is deeper in the phantom than the plane were the measurements were made. The two depths are comparable by $d_{\text{water}} \approx 1.14 d_{\text{lexan}}$.

5.4 Radiographic film

Kodak X-Omat V film (Eastman Kodak Company, USA) was used . The film has a "nominal dose range of 0.25 – 1.75 Gy under normal viewing conditions" [34]. With the applied measurement equipment (the Kodak film and the HP film scanner) it was noticed, however, that the feasible dose range was 0.2 – 1.2 Gy. Outside this range the calibration curve (figure 5.2) became too steep (above 1.2 Gy) or too flat (below 0.2 Gy) for reliable results.

The reason for placing the film perpendicular to the central axis of a treatment field was that if the film is oriented along the central axis during irradiation, the optical density of the film is not only a function of dose but also a function of the distance that radiation has traversed in the film. This was discovered in preliminary measurements and is also reported in [66]. One possible explanation is that the film itself changes considerably the interaction conditions of radiation and matter. The substances in the film, such as silver, may have a small effect on scattering in the otherwise homogeneous lexan material. The phenomenon is further amplified if primary radiation traverses long distances in the film as is the situation when the film is oriented along the central axis of a treatment field. Thus, a logical orientation of the film is perpendicular to the central axis because then primary radiation traverses minimum distance in film resulting the radiation

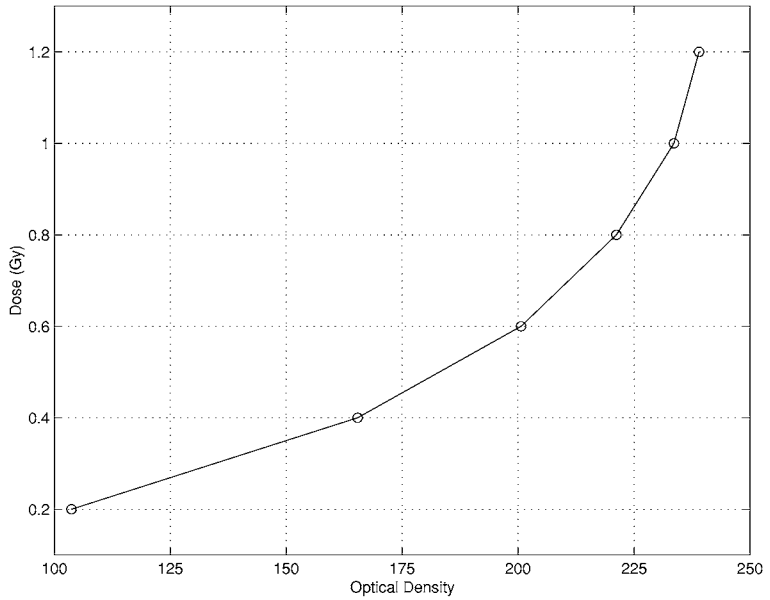

Figure 5.2: The calibration curve of the radiographic film. Dose is presented as a function of optical density of the film. The scanned reading is in range 0 – 255.

to cause minimum disturbance to the optical density of the film.

5.5 Results

Figures 5.3, 5.6 and 5.9 show comparisons between the measured and computed dose distributions. The measured dose distributions are derived from the scanned films by the optical density to dose calibration. The two-dimensional dose distributions are compared using difference maps. A statistical measure of the discrepancies between the measured and computed dose distributions is illustrated by histograms. The histograms show the discrepancy distributions when the measured dose distributions are subtracted from the computed ones. For more informative visual comparison, several dose profiles are shown in figures 5.4, 5.5, 5.7, 5.8, 5.10 and 5.11. The profiles are plotted either along the leaf movement direction or perpendicular to it.

The maximum and mean discrepancies between the computed and measured dose distributions were determined in dose range $D_{\max}/2$ – $D_{\max}$, where $D_{\max}$ is the maximum dose of a computed dose distribution. Results are in table 5.1.

Table 5.1: Maximum ($\Delta_{\max}$) and mean (Δ_{mean}) dose discrepancies between the computed and measured dose distributions in dose range $D_{\max}/2 - D_{\max}$.

Field	$\Delta_{\max}$ (Gy)	Δ_{mean} (Gy)
1	0.10	0.02
2	0.20	0.02
3	0.15	0.03

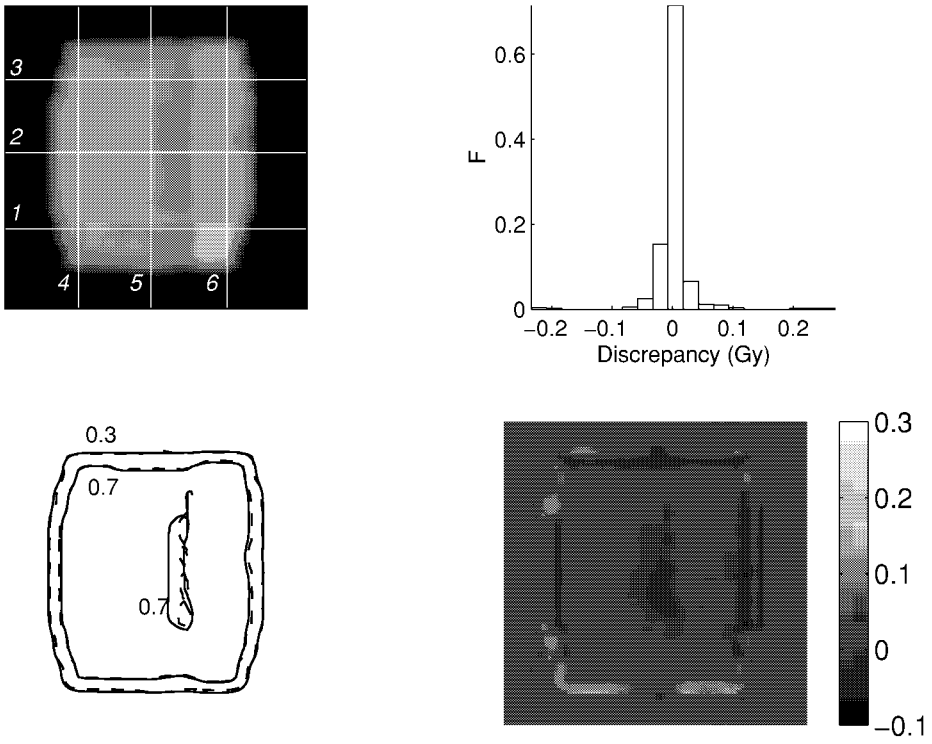

Figure 5.3: Overview of the first measurement. Upper row: the measured dose distribution and the histogram of the discrepancy between the computed and measured dose (F is the fraction of the total number of points in the dose distribution). Lower row: 0.3 and 0.7 Gy isodoses of the computed (solid line) and measured (dashed line) dose distributions and the difference map of the computed and measured dose distributions.

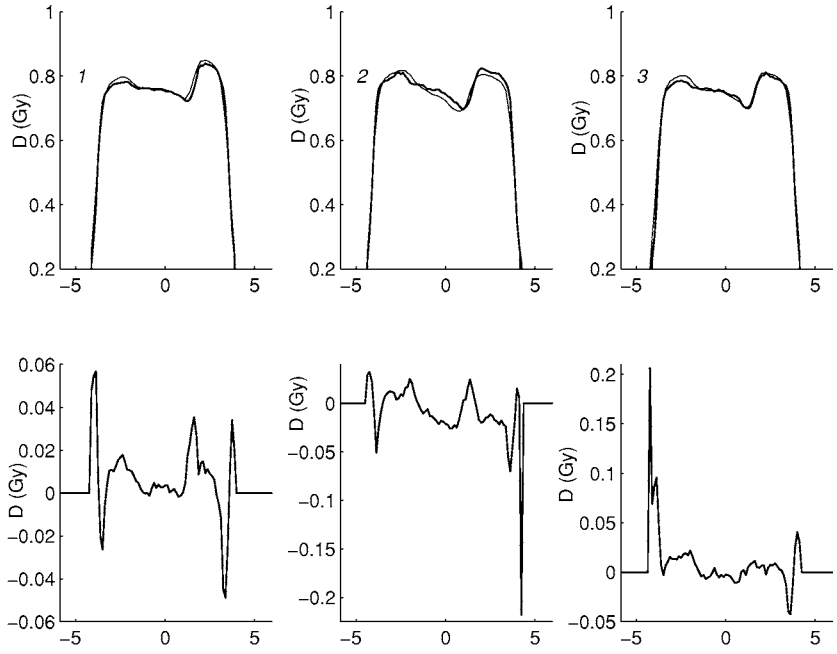

Figure 5.4: The profile comparisons of the first measurement. Upper row: Dose profiles along the direction of leaf movement. The measured profiles are drawn using thick lines and the computed ones using thin lines. Lower row: Absolute differences between the computed and measured profiles shown in the upper row. The numbers in the measured dose distribution, in figure 5.3, refer to the shown profiles.

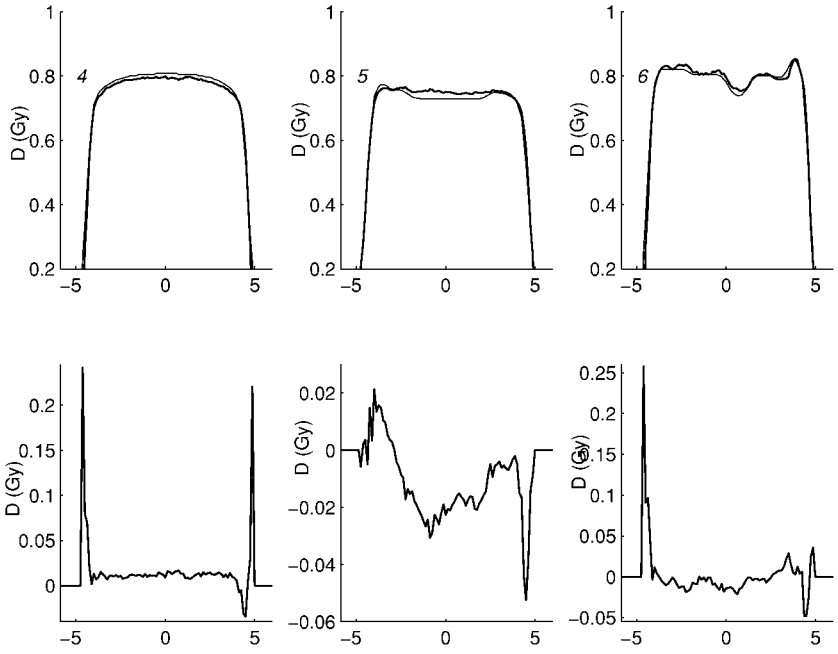

Figure 5.5: The profile comparisons of the first measurement. Upper row: Dose profiles perpendicular to the direction of leaf movement. The measured profiles are drawn using thick lines and the computed ones using thin lines. Lower row: Absolute differences between the computed and measured profiles shown in the upper row. The numbers in the measured dose distribution, in figure 5.3, refer to the shown profiles.

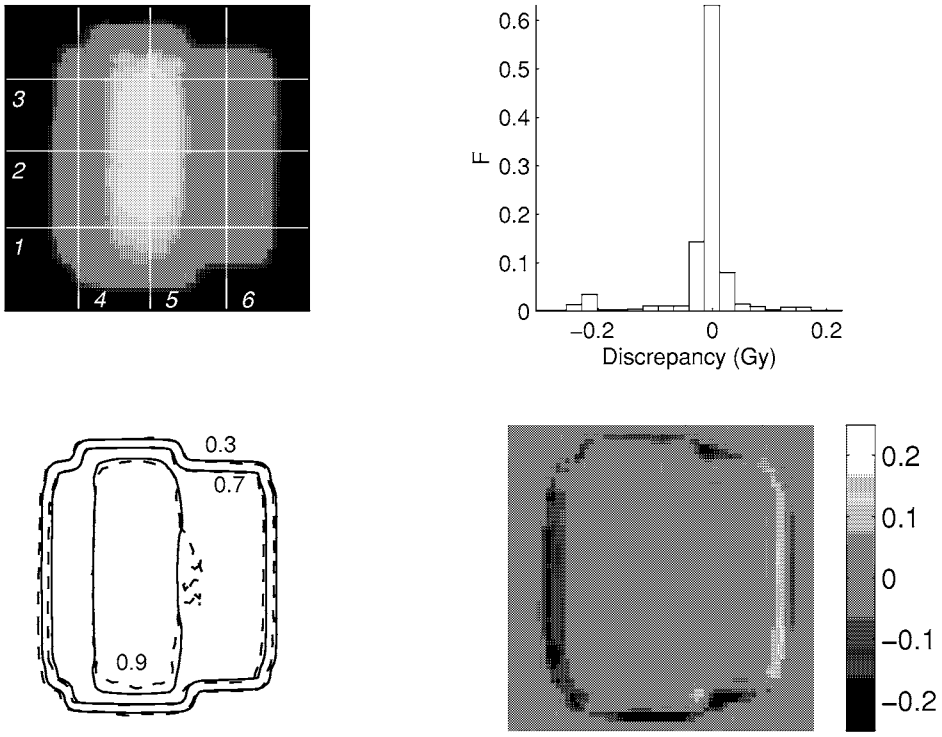

Figure 5.6: Overview of the second measurement. Upper row: the measured dose distribution and the histogram of the discrepancy between the computed and measured dose (F is the fraction of the total number of points in the dose distribution). Lower row: 0.3, 0.7 and 0.9 Gy isodoses of the computed (solid line) and measured (dashed line) dose distributions and the difference map of the computed and measured dose distributions.

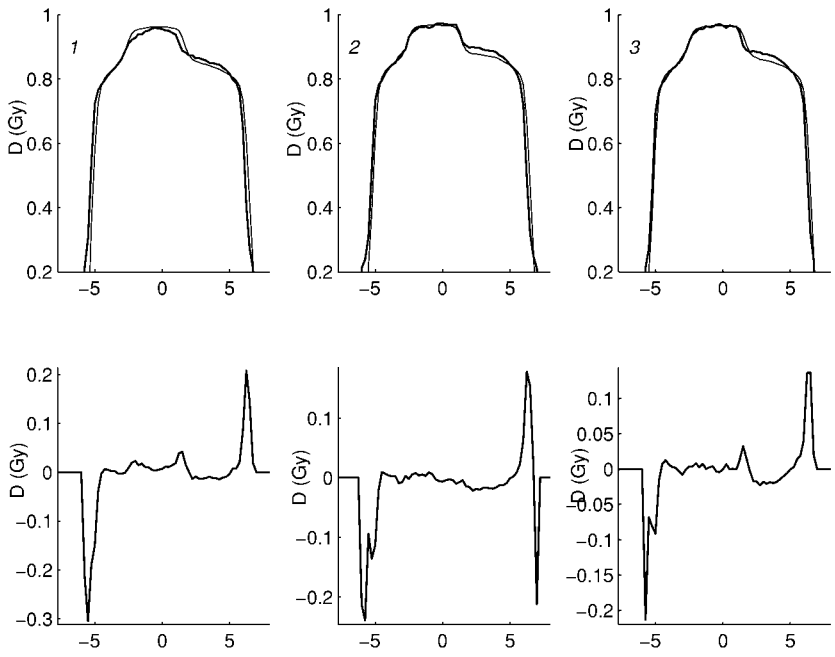

Figure 5.7: The profile comparisons of the second measurement. Upper row: Dose profiles along the direction of leaf movement. The measured profiles are drawn using thick lines and the computed ones using thin lines. Lower row: Absolute differences between the computed and measured profiles shown in the upper row. The numbers in the measured dose distribution, in figure 5.6, refer to the shown profiles.

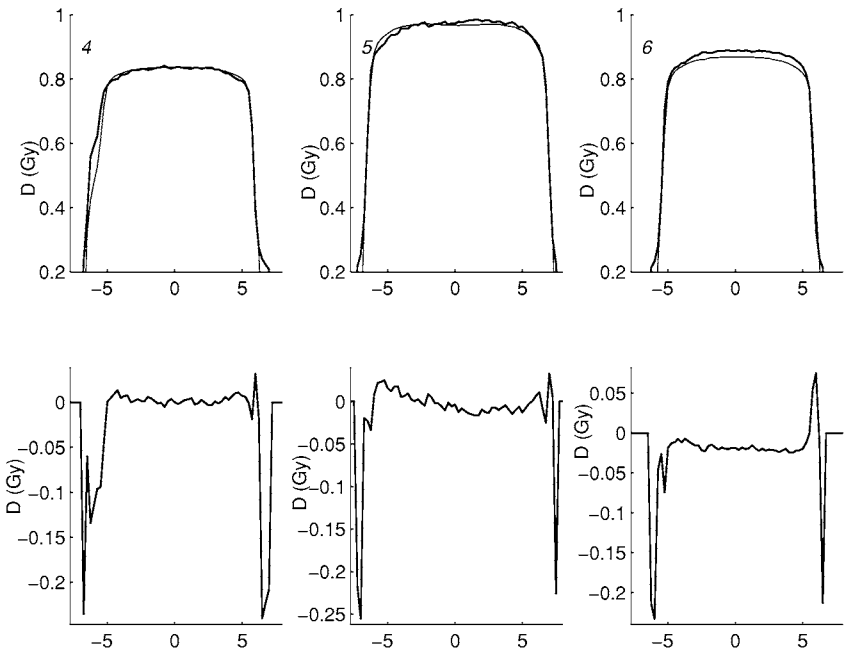

Figure 5.8: The profile comparisons of the second measurement. Upper row: Dose profiles perpendicular to the direction of leaf movement. The measured profiles are drawn using thick lines and the computed ones using thin lines. Lower row: Absolute differences between the computed and measured profiles shown in the upper row. The numbers in the measured dose distribution, in figure 5.6, refer to the shown profiles.

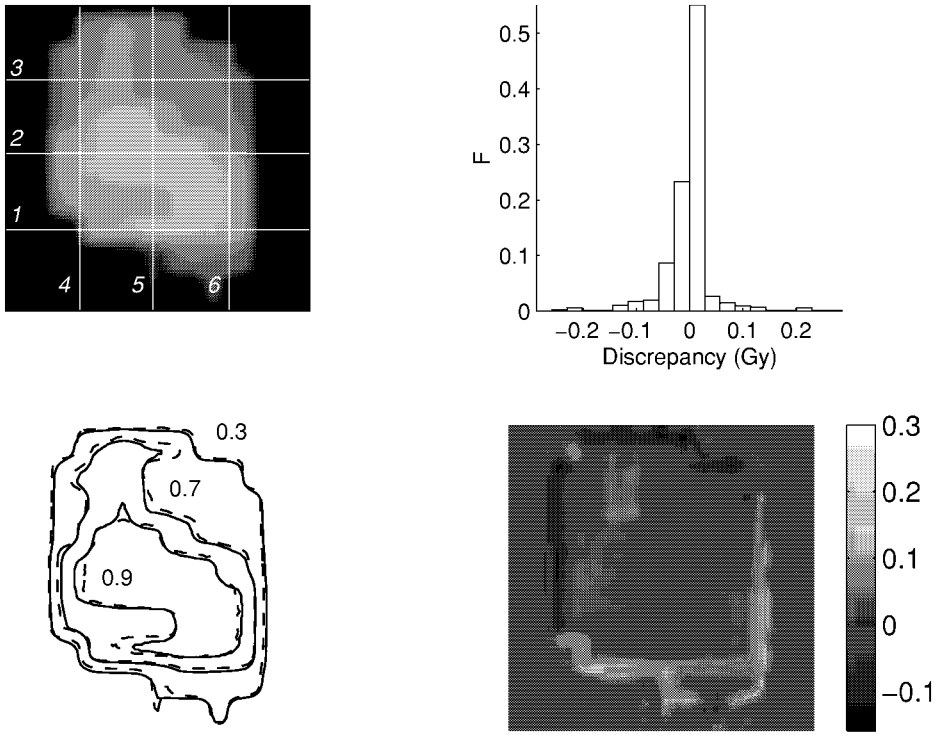

Figure 5.9: Overview of the third measurement. Upper row: the measured dose distribution and the histogram of the discrepancy between the computed and measured dose (F is the fraction of the total number of points in the dose distribution). Lower row: 0.3, 0.7 and 0.9 Gy isodoses of the computed (solid line) and measured (dashed line) dose distributions and the difference map of the computed and measured dose distributions.

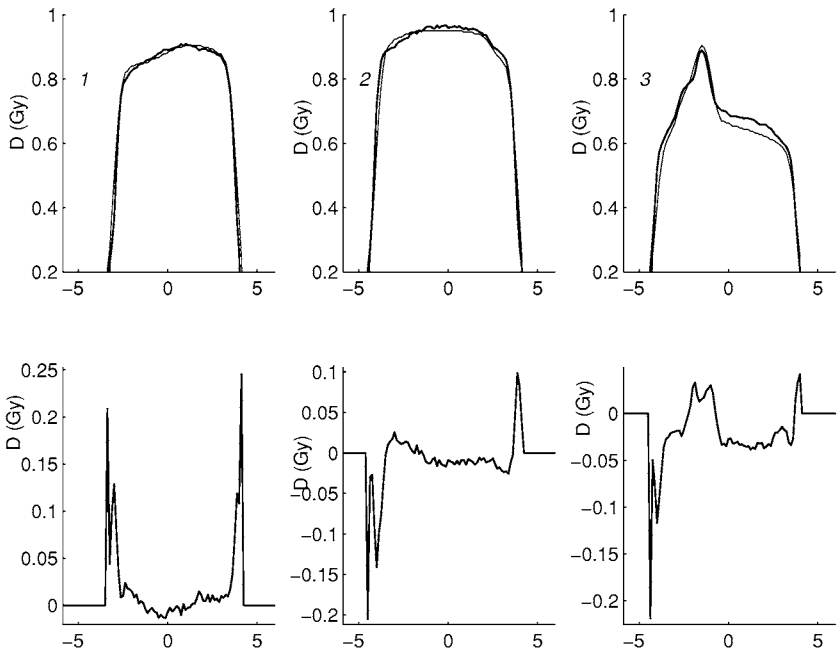

Figure 5.10: The profile comparisons of the third measurement. Upper row: Dose profiles along the direction of leaf movement. The measured profiles are drawn using thick lines and the computed ones using thin lines. Lower row: Absolute differences between the computed and measured profiles shown in the upper row. The numbers in the measured dose distribution, in figure 5.9, refer to the shown profiles.

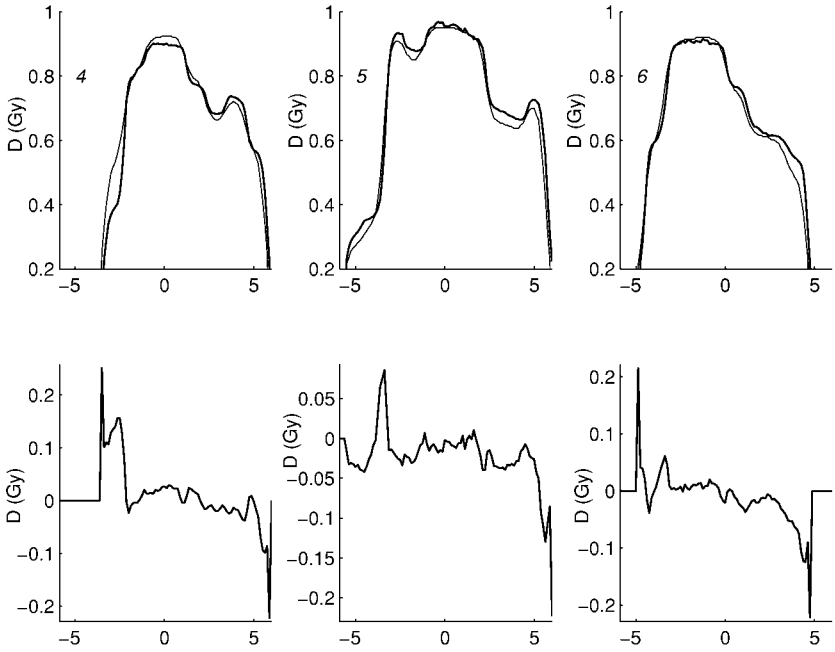

Figure 5.11: The profile comparisons of the third measurement. Upper row: Dose profiles perpendicular to the direction of leaf movement. The measured profiles are drawn using thick lines and the computed ones using thin lines. Lower row: Absolute differences between the computed and measured profiles shown in the upper row. The numbers in the measured dose distribution, in figure 5.9, refer to the shown profiles.

CHAPTER VI

Patient examples

To test the developed optimization system, several treatment plans were computed for artificial phantoms and anonymous patients. The test examples had a variety of PTV/OAR geometries and field settings. Three of the examples are represented and discussed. The plans are not intended to be clinically relevant but they show the basic features of IMRT with multiple static collimation. A more exhaustive testing and evaluation will be needed before the optimization system enters clinical practice.

The optimizations were done using a HP C180 XP (Hewlett-Packard, Palo Alto, USA) computer with a HP-UX 10.20 operating system.

The represented DVHs are based on the dose calculations using Cadplan treatment planning system. The intensities of the treatment fields were modulated according to the results from the developed optimization system.

6.1 A prostate example

The often existing problem, when external radiotherapy is used to cure a prostate cancer, is that while a homogeneous dose distribution is easy to create using a four-field "box" treatment plan, there are vulnerable OARs close to the PTV (prostate). One of the OARs is rectum, or more precisely the anterior rectal wall. The quality of life of a treated patient suffers significantly if the rectal wall receives a long-lasting radiation damage. If the anterior rectal wall is to be shielded after an accepted tolerance dose, but radiotherapy is continued to a high dose needed to destroy cancer in the PTV, a concave dose distribution is required because of the prostate-rectum anatomy (figure 6.1). Close to prostate is also bladder which should be shielded as effectively as possible.

Treatment configuration consisted of five isocentric coplanar treatment fields. Table 6.1 shows information about dose constraints and constraint priorities. Initially, 9 segments per treatment field were allowed. The density of calculation points in the patient space was 25 points/cm^3.

The optimization algorithm ran 95 iterations and evaluated the value of the

Table 6.1: Dose constraints in the prostate example. For each irradiated object the constrained volume is shown.

	Volume (%)	Dose (Gy)	Priority
PTV	100.0	≥ 69.3	100
PTV	100.0	≤ 70.7	75
Femur, dex	100.0	≤ 40.0	70
Femur, sin	100.0	≤ 40.0	70
Bladder	90.0	≤ 60.0	30
Rectum	100.0	≤ 60.0	80

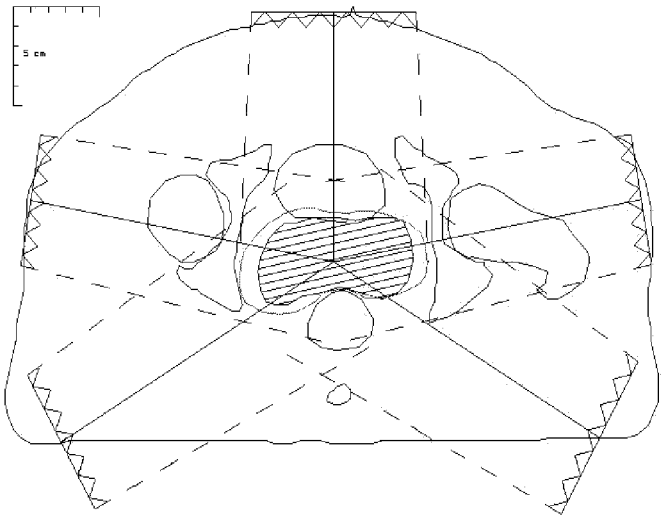

Figure 6.1: A transversal slice of the patient with prostate cancer. The shaded area shows the PTV. Anterior to the PTV is bladder and posterior the rectum. Five isocentric treatment fields irradiate the PTV. The 63 Gy isodose of the optimized treatment plan is also shown.

object function and its gradient 115 times. The whole optimization took 63 minutes. Finally, there were 3, 5, 4, 4 and 4 segments in the fields $1 \ldots 5$, respectively. Figure 6.2 shows the resulting dose volume histograms and table 6.2 dose statistics. The 63 Gy isodose that corresponds to 90 % of the prescribed dose in the PTV is shown in figure 6.1. There is an overdose in rectum and bladder, since the lower dose limit of the PTV has the highest priority. Especially, dose in the bladder

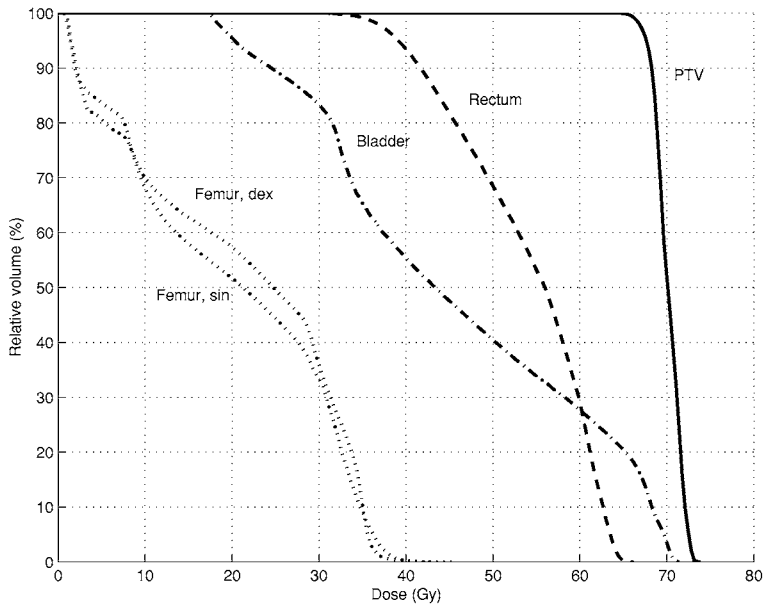

Figure 6.2: The cumulative dose volume histograms of the prostate example.

violates significantly the dose volume constraint because of the low priority. Mean doses in the femurs, on the other hand, are well below the dose limit, mainly due to their distant location from the PTV.

Table 6.2: The statistics of the optimized dose distribution in the prostate example. For each object the minimum dose (Min dose), maximum dose (Max dose), mean dose and standard deviation (S.D.) are shown.

	Min dose (Gy)	Max dose (Gy)
PTV	64.1	73.7
Femur, dex	0.8	45.7
Femur, sin	0.8	44.9
Bladder	17.0	71.4
Rectum	30.7	66.3
	Mean dose (Gy)	± 1 S.D. (%)
PTV	70.0	2.2
Femur, dex	20.7	18.1
Femur, sin	19.7	17.7
Bladder	45.5	23.2
Rectum	53.6	11.4

6.2 A nasopharynx example

The treatment configuration is non-symmetric. As opposed to the prostate example, where the patient was irradiated from all directions co-planarly, the fields are here mainly posterior. Hence, dose to the mouth and nose may be reduced. Dose to parotids, however, must be constrained since two of the lateral fields irradiate right through the parotids. Of other sensitive structures, the brain stem and spinal cord have the OAR status. Dose constraints are shown in table 6.3.

The treatment setting and patient anatomy are illustrated in figure 6.3. Each treatment field had initially 11 segments. Dose deposition was computed using 20 calculation points per cm^3.

Table 6.3: Dose constraints in the nasopharynx example. For irradiated objects the constrained volume is shown.

	Volume (%)	Dose (Gy)	Priority
PTV	100.0	≥ 60.0	100
PTV	100.0	≤ 60.6	90
Parotis, dex	100.0	≤ 25.0	90
Parotis, sin	100.0	≤ 25.0	90
Brain stem	100.0	≤ 40.0	90
Spinal cord	100.0	≤ 40.0	80

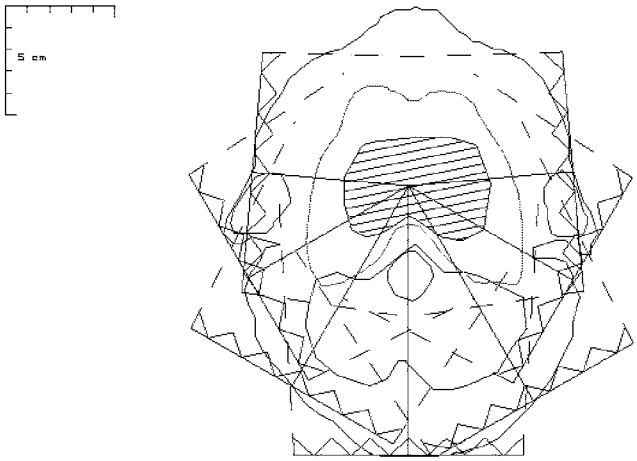

Figure 6.3: A transversal slice of the patient with a cancer of the nasopharynx. The shaded area shows the PTV. Laterally to the PTV are the parotid glands and posterior is the brain stem. The 54 Gy isodose of the optimized treatment plan is also shown.

The optimization algorithm ran 110 iterations and evaluated the value of the object function and its gradient 138 times. The whole optimization took 78 minutes. The final numbers of segments were 2, 5, 2, 2, 8, 10 and 10 for treatment fields 1...7, respectively. Dose statistics for the PTV and the OARs are in table 6.4 and dose volume histograms that were calculated from the optimized dose distribution are shown in figure 6.4. The 54 Gy isodose that corresponds to 90 % of the prescribed dose in the PTV is shown in figure 6.3.

Table 6.4: the statistics of the optimized dose distribution in the nasopharynx example. For each object the minimum dose (Min dose), maximum dose (Max dose), mean dose and standard deviation (S.D.) are shown.

	Min dose (Gy)	Max dose (Gy)
PTV	54.2	63.9
Parotis, dex	4.3	28.8
Parotis, sin	5.7	31.8
Brain stem	3.2	44.6
Spinal cord	4.0	39.7
	Mean dose (Gy)	±1 S.D. (%)
PTV	60.0	1.8
Parotis, dex	13.4	7.0
Parotis, sin	17.2	7.1
Brain stem	21.3	19.3
Spinal cord	23.1	15.6

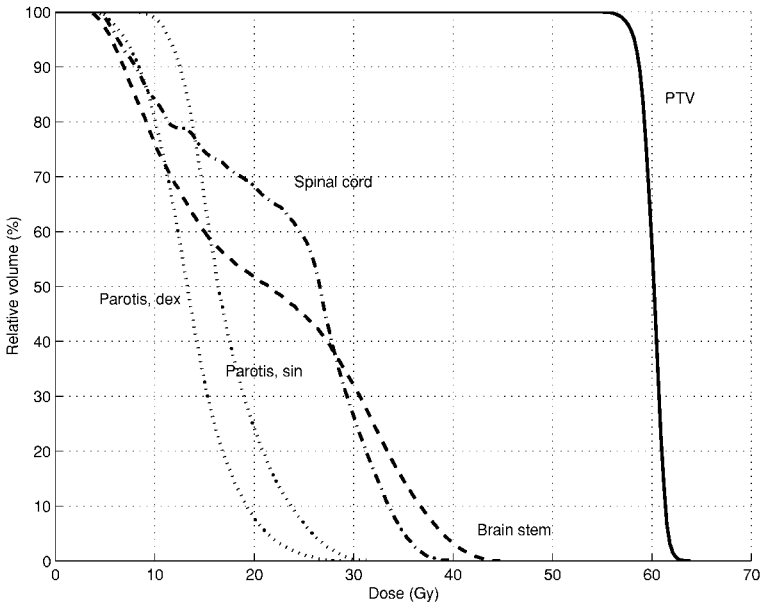

Figure 6.4: The cumulative dose volume histograms of the nasopharynx example.

6.3 A Mediastinal example

Due to the location of the (mediastinal) PTV, dose volume constraints must be assigned to both lungs if the irradiated parts of the lungs are to be kept at least

partially functional (table 6.5). The number of treatment fields, 3, was kept modest to not to raise integral dose (the overall dose level) in lungs (figure 6.5). In the beginning of the optimization, 11 segments were allowed for the treatment fields. The point densities in the patient space were not homogeneous. In the "small" volumes (PTV, spinal cord), a figure of 30 points/cm^3 was used, but in the lungs only 3 points/cm^3. Otherwise, the number of DDCs representing dose contribution in the lungs would have been too big for the computer that was used for the optimization.

Table 6.5: Dose constraints in the mediastinal example. For each object the constrained volume is shown.

	Volume (%)	Dose (Gy)	Priority
PTV	100.0	$\geq$ 59.4	100
PTV	100.0	$\leq$ 60.6	90
Lung, dex	90.0	$\leq$ 25.0	50
Lung, sin	79.5	$\leq$ 25.0	54
Lung, sin	92.4	$\leq$ 40.0	50
Spinal cord	100.0	$\leq$ 40.0	100

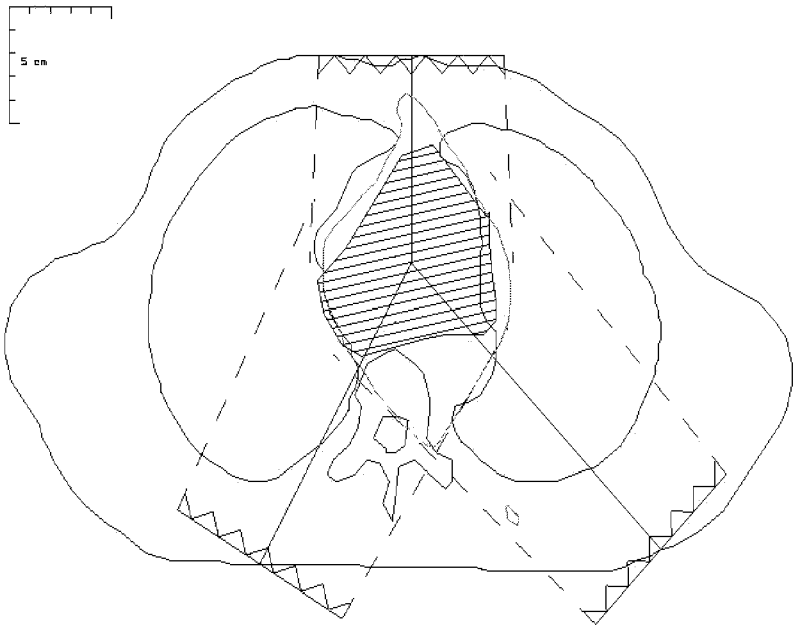

Figure 6.5: A transversal slice of the patient in the mediastinal case. The shaded area shows the PTV. Adjacent to the PTV are lungs and posterior is the spinal cord. The 48 Gy isodose of the optimized treatment plan is also shown.

Optimization was finished after 43 iteration rounds, and 56 function/gradient evaluations were needed. The whole optimization took 31 minutes. At the end there were 6, 11 and 10 segments of the fields 1, 2 and 3, respectively. Figure 6.6 shows the cumulative dose volume histograms of the optimized dose, while table 6.6 shows the dose statistics in the PTV and OARs. The 48 Gy isodose that corresponds to 80 % of the prescribed dose in the PTV is shown in figure 6.5.

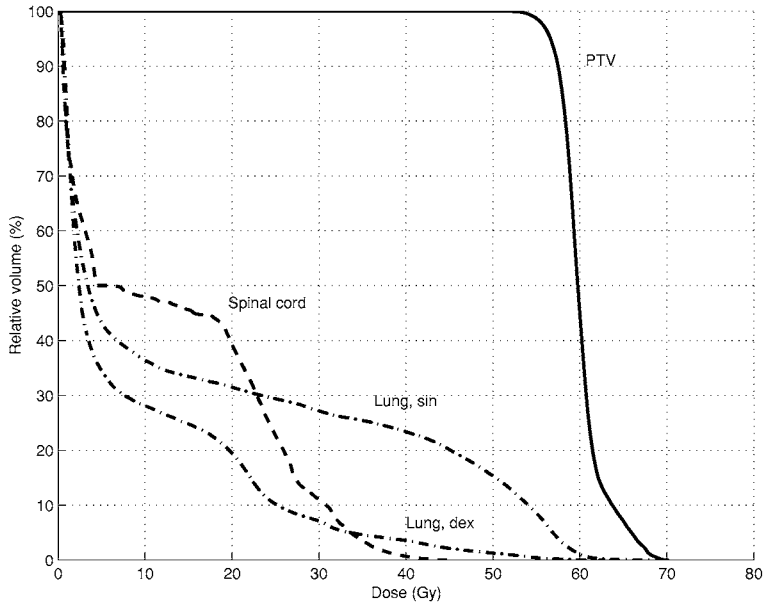

Figure 6.6: The cumulative dose volume histograms of the mediastinal example.

Table 6.6: The statistics of the optimized dose distribution in the mediastinal example. For each object the minimum dose (Min dose), maximum dose (Max dose), mean dose and standard deviation (S.D.) are shown.

	Min dose (Gy)	Max dose (Gy)
PTV	49.3	70.0
Lung, dex	0.1	62.6
Lung, sin	0.1	68.8
Spinal cord	0.1	45.3
	Mean dose (Gy)	± 1 S.D. (%)
PTV	60	4.3
Lung, dex	8.6	19.0
Lung, sin	16.6	35.5
Spinal cord	13.0	20.8

Chapter VII

Discussion

7.1 Implementation issues

A local optimization method was used. It was not, however, a method of choice. An efficient global optimization method would have been a more preferable approach. Global optimization has been used to optimize radiotherapy treatment plans. The simulated annealing algorithm, particularly, has been applied widely [52, 76] and is used in the commercially available inverse planning system CORVUS (NOMOS corporation, Sewickley, USA). Unfortunately, simulated annealing would be too inefficient for the developed inverse planning system. The main reason is that the intensity of a bixel is dependent on the leaf positions in several segments. For example, if there are 15 segments, then there are 30 parameters (15 left and 15 right leaves) that can potentially modify the intensity of a bixel. When the linear dose calculation model (2.3) is used, a global optimization method is applicable since only one parameter determines the intensity of a bixel.

To apply a global optimization algorithm for the developed inverse planning system, prior knowledge would have to be used to decrease the size of the parameter space where a global algorithm searches the global minimum. In addition, the use of initial values for parameters could shorten the optimization times. The use of prior knowledge and initial values must be used carefully because the algorithm can be guided accidentally to a region where the global minimum cannot be found. Thus, the benefits of global optimization may be lost.

A local optimization method always converges to the nearest local minimum. Consequently, the determination of the initial values of the MLC parameters has a large effect on the final solution. The equation (4.6) is the product of a thorough testing where the treatment plans of several patients were optimized. Because the patient anatomy and the location of the PTV varies substantially between different patients, equation (4.6) is a compromise that produces the best overall set of initial values for the MLC parameters.

7.2 Dosimetric tests

The results of the preliminary dosimetric measurements are satisfactory. They show that the positions of the leaves of the MLC are nearly identical in the computed and actual dose delivery. There are minor differences between the computed and irradiated dose distributions, especially when the dose profiles in the direction perpendicular to the movement of the leaves are studied (figures 5.4, 5.5, 5.7, 5.5, 5.10 and 5.11). The tinyest difference in leaf positions causes noticeable discrepancies in these perpendicular dose profiles. In addition, the round off of the monitor units to integers in equation (4.7) can cause small errors. This could be compensated, however, by recomputing segment weights using the rounded MUs.

Initial comparisons between the computed and measured dose showed unacceptable amount of discrepancy. After a thorough analysis of the dose distributions, it was understood that the head scatter model for the MLC (section 3.4) was not correct. The contribution of the dose calculation algorithm of Cadplan had not been taken into account in the applied ray tracing method. Cadplan already models the head scatter. To enhance the scattering model, a time consuming manual adjustment of the model parameters was done. The adjusted model was quite similar to the Heaviside function with an 8 mm offset in the leaf movement direction. All dose comparisons were made using the new model.

At the edges of the dose distributions, large discrepancies between the computed and measured dose can be seen in the difference profiles (lower portions of figures 5.5, 5.5 and 5.11). Presently, it is not clear what causes these discrepancies. Possible contributors are the head scatter model of the MLC, dose calculation system or the inaccuracy of the film measurements.

7.3 Patient tests

The discussion here is not meant to find out whether the optimized treatment plans are clinically acceptable or superb but to evaluate the solutions optimizationwise. Since the simulations were peliminary, the evaluation must answer questions about how error-free and reliable the optimization system is and how good it's performance was.

The optimization system is, at it's present condition, stable and does follow the dose prescriptions. Illogical solutions have not been encountered. The produced field shapes are simple and intuitive. All the applied technical constraints of the MLC are satisfied. As an example, figure 7.1 shows three segments that form the intensity modulation of the first field in the prostate example (section 6.1). It is easy to conclude from the shapes of the segments that the field faces an OAR and the modulation tries to protect it.

All optimizations terminated when the L-BFGS-B algorithm informed that the value of the object function was not converging anymore. The termination was controlled by a given tolerance. Another termination criterion of the L-BFGS-B is the infinite norm (the largest absolute value) of the gradient of the object function. The optimization never terminated because of the gradient criterion. Based on

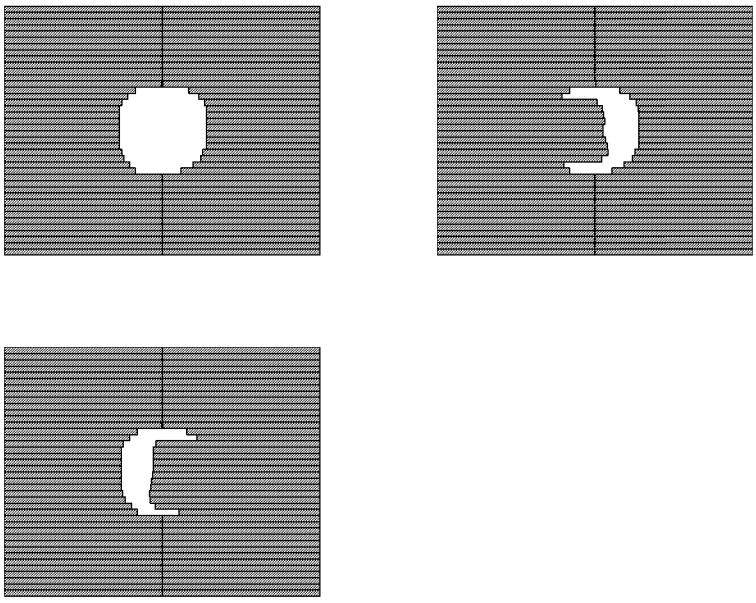

Figure 7.1: The MLC-shaped segments of the first treatment field in the prostate example.

the large gradient norms at the termination, it was considered that at least one parameter was having large values of gradient but the value of the parameter could not change since the parameter was already at a constraint boundary *i.e.* the parameter was active. The L-BFGS-B never terminated due to erroneous search direction. This implies that the gradient of the object function was correctly computed.

The optimized plans show the typical feature of a quadratic penalty. While the overall dose distributions satisfy the prescriptions, small regions of PTVs and OARs get too low or too high dose (tables 6.2, 6.4, 6.6). Another feature in the solutions is that the values of weighted penalties of PTVs and OARs are not equal after optimization. For example, in table 7.1 are shown the values of the non-weighted and weighted quadratic penalties for the PTV and the OARs in the prostate example. The values in table 7.1 are determined from the final dose distribution that was computed using Cadplan. This feature is self-explanatory as the weighted quadratic penalties are combined into one object function (4.4). The individual penalties are not separable. A more sophisticated approach for the treatment planning problem would be the multiobjective optimization [50]. In the multiobjective optimization, each dose constraint could be handled separately *i.e.* the object function is a vector containing the weighted dose constraints of PTVs and OARs. There is not, however, an implementation of the multiobjective optimization for large-scale problems.

Table 7.1: The non-weighted (W_1) and weighted (W_2) quadratic penalties of the PTV and the OARs in the prostate example.

	W_1	W_2
PTV, low	0.482	124.338
PTV, high	0.032	34.913
Femur, dex	0.241	143.650
Femur, sin	0.090	57.475
Bladder	0.011	549.250
Rectum	1.700	240.803

In each patient example, the number of segments decreased during the optimizations. This shows that the constraint for too small segment weights was working (section 4.4). Too small weights were forced either to zero or above the given minimum value. The smallest accepted segment weight corresponded approximately to 8 MU.

In the mediastinal example, the PTV dose distribution is not as good as in other examples. This is a consequence of the overlap of the PTV and lungs. The dose prescription in the region of intersection is not unique. In fact, the prescription is controversial because dose should be above the lower constraint of the PTV and, simultaneously, below the highest accepted dose of the lungs.

The computation times varied substantially between the examples. Time range was from half hour to almost 80 minutes. The calculation time is dependent on the number of treatment fields, the number of segments per a field and the point densities in patient space. The most influental of these dependencies is the number of fields since adding more fields increases the number of optimization parameters rapidly. As the computer hardware that was used was quite old even at the time of writing, the calculation times can be expected to decrese to just a few minutes in the near future. Remarkable in the mediastinal example is that only 43 optimization iterations were needed. This number is approximately half of the number of iterations in the other examples. There are mainly two reasons for this. First, the number of treatment fields was lower than in the other examples. Second, dose volume constraints were assigned for both lungs. There appears to be an unbalance between "ordinary" constraints and dose volume constraints that was wittnessed not only in the mediastinal example but in other examples that are not documented in this thesis. A dose volume constraint appears to be too strong when compared to other constraints. This leads to the premature termination of optimization because a strong dose volume constraint implies a strong local minimum of the object function. To balance the overly strong dose volume constraints their priorities must be set to low values. A careful balancing of different types of constraints needs to be addressed in the future.

Chapter VIII

Summary

A complete system for controlling the multileaf collimator (MLC) in the inverse problem of radiotherapy treatment planning has been described. The control of the MLC enables the modulation of the intensity of a treatment field. The control parameters of the MLC can be mathematically optimized according to a dose prescription. A dose description can contain weighted dose constraints and dose volume constraints. A correct head scatter model for the leaves of the MLC ensures that the resulting dose distribution is accurate. The technical limitations of MLCs are addressed during the mathematical optimization. To not to create segments with too short beam-on times, the minimum weight of a segment can be assigned.

Preliminary dosimetric phantom testing has been done using radiographic film. The performance of the inverse planning system has been tested by optimizing treatment plans of example patients.

The significant parts of the inverse planning system have been discussed and studied critically.

The aims of the work that was reported in this thesis were fulfilled. Based on the results of the dosimetric and patient example tests, the direct control of the MLC can be used in the inverse problem of radiotherapy treatment planning. Further testing will show whether the described system is clinically relevant.

REFERENCES

[1] A. Ahnesjö. *Dose Calculation Methods in Photon Beam Therapy Using Energy Deposition Kernels.* PhD Thesis, Stockholm University, 1991

[2] A. Ahnesjö and M.M. Aspradakis. Dose calculations for external photon beams in radiotherapy. *Phys. Med. Biol.* 44:R99-R155, 1999

[3] G.K. Bahr, J.G. Kereikes, H. Horwitz, R. Finney, J. Galvin and K. Goode. The method of linear programming applied to radiation treatment planning. *Radiology* 91:686-693 , 1968

[4] T.R. Bortfeld, D.L. Kahler, T.J. Waldron and A.L. Boyer. X-ray field compensation with multileaf collimators. *Int. J. Radiat. Oncol. Biol. Phys.* 30:899-908, 1994

[5] A.L. Boyer. Use of MLC for intensity modulation. *Med. Phys.* 21:1007-1021, 1994

[6] A. Brahme, J.-E. Roos and I. Lax. Solution of an integral equation encountered in rotation therapy. *Phys. Med. Biol.* 10:1221-1229, 1982

[7] A. Brahme. Optimization of radiation therapy and the development of multileaf collimation. *Int. J. Radiation Oncol. Biol. Phys.* 25:373-375, 1993

[8] R.H. Byrd, P. Lu, J. Nocedal and C. Zhu. A limited memory algorithm for bound constrained optimization. Technical report NAM-08, Northwestern University, Evanston, 1994

[9] M.P. Carol. Peacock: a system for planning and rotational delivery of intensity-modulated fields. *Int. J. Imaging Syst. Technol.* 6:56-61, 1995

[10] Y. Censor and T. Elfving. New methods for linear inequalities. *Linear Algebra Appl.* 42:199-221, 1982

[11] Y. Censor, M.D. Altschuler and W.D. Powlis. 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, 1988

[12] Y. Censor and S.A. Zenios *Parallel Optimization: Theory, Algorithms and Applications.* Oxford University Press, New York, 1997

[13] S.A. Chang, T.J. Cullip and K.M. Deschesne. Intensity modulation delivery techniques: "Step & shoot" MLC auto-sequence versus the use of a modulator. *Med. Phys* 27:948-959, 2000

[14] Y. Chen, A.L. Boyer and C.-M. Ma. Calculation of x-ray transmission through a multileaf collimator. *Med. Phys.* 27:1717-1726, 2000

[15] A.R. Conn, N.I.M. Gould and Ph.L. Toint. Global convergence of a class of trust region algorithms for optimization with simple bounds. *SIAM J. Numer. Anal.* 25:433-460, 1988

[16] A.R. Conn, N.I.M. Gould and Ph.L. Toint. *LANCELOT: a FORTRAN Package for Large-scale Nonlinear Optimization (Release A)*. Springer-Verlag, New York, 1992

[17] D.J. Convery and M.E. Rosenbloom. The generation of intensity-modulated fields for conformal radiotherapy by dynamic collimation. *Phys. Med. Biol.* 37:1359-1374, 1992

[18] S.M. Crooks and L. Xing. Linear algebraic methods applied to intensity modulated radiation therapy. *Phys. Med. Biol.* 46:2587-2606, 2001

[19] J.O. Deasy. Multiple local minima in radiotherapy optimization problems with dose-volume constraints. *Med. Phys.* 24:1157-1161, 1997

[20] J.E. Dennis and R.B. Schnabel. *Numerical Methods for Unconstrained Optimization and Nonlinear Equations.* SIAM, Philadelphia, 1996

[21] A.R. De Pierro and A.N. Iusem. A simultaneous projections method for linear inequalities. *Linear Algebra Appl.* 64:243-253, 1985

[22] A. Eklöf, A. Ahnesjö and A. Brahme. Photon beam energy deposition kernels for inverse radiotherapy planning. *Acta Oncol.* 29:447-454, 1990

[23] A.V. Fiacco and G.P. McCormick. *Nonlinear Programming: Sequential Unconstrained Minimization Techniques.* SIAM, Philadelphia, 1990

[24] P.E. Gill, W. Murray and M.H. Wright. *Practical Optimization.* Academic Press, London, 1981

[25] P.E. Gill, W.Murray and M.A. Saunders. SNOPT: An SQP algorithm for large-scale constrained optimization. Report NA 97-3, Department of Mathematics, University of California, San Diego, 1997

[26] P. Gokhale, E.M.A. Hussein and N. Kulkarni. Determination of beam orientation in radiotherapy planning. *Med. Phys.* 21:393-400, 1994

[27] ICRU Report 24. *Determination of Absorbed Dose in a Patient Irradiated by Beams of X or Gamma Rays in Radiotherapy Procedures.* International Comission on Radiation Units and Measurements, Bethesda, 1976

[28] ICRU Report 50. *Prescribing, Recording and Reporting Photon Beam Therapy.* International Comission on Radiation Units and Measurements, Bethesda, 1993

[29] R. Jeraj and P. Keall. Monte Carlo-based inverse treatment planning. *Phys. Med. Biol.* 44:1885-1896, 1999

[30] T. Holmes and T.R. Mackie. A comparison of three inverse treatment planning algorithms. *Phys. Med. Biol.* 39:91-106, 1993

[31] D.H. Hristov and B.G. Fallone. An active set algorithm for treatment planning optimization. *Med. Phys.* 24:1455-1464, 1997

[32] C.T. Kelley. *Iterative Methods for Optimization.* SIAM, Philadelphia, 1999

[33] F.M. Khan, K.P. Doppke, K.R. Hogstrom, G.J. Kutcher, R. Nath, S.C. Prasad, J.AA. Purdy, M. Rozenfeld and B.L. Werner. Report of the AAPM radiation therapy committee task group nr. 25. *Med. Phys*, 18:73-109, 1991

[34] Kodak X-Omat V film / 4508. A technical information data sheet. Eastman Kodak company, 1994

[35] P. Kolmonen, J. Tervo and T. Lahtinen. Use of the Cimmino algorithm and continuous approximation for the dose deposition kernel in the inverse problem of radiation treatment planning. *Phys. Med. Biol.* 43:2539-2554, 1998

[36] P.Kolmonen, J.Tervo, K.Jaatinen and T.Lahtinen. Direct computation of the 'step-and-shoot' IMRT plan. *XIIth International Conference on the Use of Computers in Radiotherapy*, Editors: T. Bortfeld and W. Schlegel, Heidelberg, Germany, May 22-25, 2000

[37] G. Küster and T. Bortfeld. Applicability of a multi-hole collimator for scanned photon beams: A Monte Carlo study. *XIIth International Conference on the Use of Computers in Radiotherapy*, Editors: T. Bortfeld and W. Schlegel, Heidelberg, Germany, May 22-25, 2000

[38] P. Källman, A. Ågren and A. Brahme. Tumor and normal tissue responses to fractionated non-uniform dose delivery. *Int. J. Radiat. Biol.* 62:249-262, 1992

[39] T. Lahtinen and L.R. Holsti (editors). *Kliininen säteilybiologia* (in Finnish). Duodecim, Helsinki, 1997

[40] M. Langer and J. Leong. Optimization of beam weights under dose-volume restrictions. *Int. J. Radiat. Oncol. Biol. Phys.* 13:1255-1260, 1987

[41] M. Langer, R. Brown, S. Morrill, R. Lane and O. Lee. A generic genetic algorithm for generating beam weights. *Med. Phys.* 23:965-971, 1996

[42] B.K. Lind. *Radiation Therapy Planning and Optimization Studied as Inverse Problems.* PhD Thesis, Stockholm University, 1991

[43] B. Lind and A. Brahme. Development of treatment techniques for radiotherapy optimisation. *Int. J. Imaging Syst. Technol.* 6:33-42, 1995

[44] H.H. Liu, T.R. Mackie and E.C. McCullough. A dual source photon beam model used in convolution/superposition dose calculation for clinical megavoltage x-ray beams. *Med. Phys.* 24:1960-1974, 1997

[45] J. Llacer. Inverse radiation treatment planning using the Dynamically Penalized Likelihood method. *Med. Phys.* 24:1751-1764, 1997

[46] J. Löf, B.K. Lind and A. Brahme. An adaptive control algorithm for optimization of intensity modulated radiotherapy considering uncertainties in beam profiles, patient set-up and internal organ motion. *Phys. med. Biol.* 43:1605-1628, 1998

[47] T.R. Mackie, T. Holmes, P. Reckwerdt, J.O. Deasy, J. Yang, B. Paliwal and T. Kinsella. Tomotherapy: a new concept for the delivery of dynamic conformal therapy. *Med. Phys.* 20:1709-1719 , 1993

[48] O.L. Mangasarian. Unconstrained Lagrangians in nonlinear programming. *SIAM J. Control and Optimization* 13:772-791, 1975

[49] Y. Mejaddem, S. Hyödynmaa, R. Svensson and A. Brahme. Photon scatter kernels for intensity modulating radiation therapy filters. *Phys. Med. Biol.* 46:3215-3228, 2001

[50] K. Miettinen. *Nonlinear Multiobjective Optimization.* Kluwer Academic Publishers, Boston, 1999

[51] R. Mohan, C. Chui and L. Lidofsky. Differential pencil beam dose computation model for photons. *Med. Phys.* 13:64-73, 1986

[52] S.M. Morrill, R.G. Lane, G. Jacobson and I.I. Rosen. Treatment planning optimization using constrained simulated annealing. *Phys. Med. Biol.* 36:1341-1361, 1991

[53] B.A. Murtaugh and M.A. Saunders. MINOS 5.4 USERS'S GUIDE (revised). Technical report SOL83-20R, Department of Operations Research, Stanford University, Stanford, 1993. Revised 1995.

[54] S. Papatheodorou, J-C. Rosenwald, S. Zefkili, M-C. Murillo, J. Drouard and G. Gaboriaud. Dose calculation and verification of intensity modulation generated by dynamic multileaf collimators. *Med. Phys.* 27:960-971, 2000.

[55] E.H. Porter. The statistics of dose/cure relationships for irradiated tumours Part I. *Br. J. Radiol.* 53:210-227, 1980

[56] A.T. Redpath, B.L. Vickery and D.H. Wright. A new technique for radiotherapy planning using quadratic programming. *Phys. Med. Biol.* 21:781-791, 1976

[57] A.T. Redpath. Planning of beam intensity modulation using an advanced 3D dose calculation algorithm and a simulated annealing mehtod. *Radiother. Oncol.* 49:295-304, 1998

[58] D.W.O. Rogers, B.A. Faddegon, G.X. Ding, C.-M. Ma and J. We. BEAM: A Monte Carlo code to simulate radiotherapy units. *Med. Phys.* 22:503-524, 1995

[59] J.B. Rosen The gradient projection method for nonlinear programming, Part II – nonlinear constraints. *SIAM J. Appl. Math.* 9:514-532, 1961

[60] D.M. Shepard, M.C. Ferris, G.H. Olivera and T.R. Mackie. Optimizing the delivery of radiation therapy to cancer patients. *SIAM Review* 41:721-744, 1999

[61] S.V. Spirou. *Design and Delivery of Intensity-modulated Profiles in Radiation Therapy.* PhD Thesis, Columbia University, 1996

[62] S.V. Spirou and C-S. Chui. Generation of arbitrary intensity profiles by combining the scanning beam with dynamic multileaf collimation. *Med. Phys.* 23:1-8, 1996

[63] J. Stein. Dynamic x-ray compensation for conformal radiotherapy by means of multileaf collimation. *Radiother. Oncol.* 32:163-173, 1994

[64] R.M. Sternheimer, M.J. Berger and S.M. Seltzer. *Atomic Data and Nuclear Data Tables* 30:261-271, 1984

[65] P. Storchi and E. Woudstra. Calculation models for determining the absorbed dose in water phantoms in off-axis planes of rectangular fields of open and wedged photon beams. *Phys. Med. Biol.* 40:511-527, 1995

[66] N. Suchowerska, P. Hoban, M. Butson, A. Davison and P. Metcalfe. Directional dependence in film dosimetry: radiographic and radiochromic film. *Phys. Med. Biol.* 46:1391-1397, 2001

[67] R. Svensson, P. Källman and A. Brahme. An analytical solution for the dynamic control of multileaf collimators. *Phys. Med. Biol.* 39:37-61, 1994

[68] R. Svensson, B. Lind and A. Brahme. Beam chracteristics and clinical possibilities of a new compact treatment unit design combining narrow pencil beam scanning and segmental multileaf collimation. *Med. Phys.* 25:2358-2369, 1998

[69] J. Tervo and P. Kolmonen. Data fitting model for the kernel of integral operator from radiation therapy. *Mathl. Comput. Modelling* 28:59-77, 1998

[70] J. Tervo, P. Kolmonen, M. Vauhkonen, L.M. Heikkinen and J.P. Kaipio. A finite element model of electron transport in radiation therapy and related inverse problem. *Inverse Problems* 15:1345-1361, 1999

[71] J. Tervo and P. Kolmonen. A model for the control of a multleaf collimator in radiation therapy treatment planning. *Inverse Problems* 16:1875-1895, 2000

[72] J.Tervo, T.Lyyra-Laitinen, P.Kolmonen and E.Boman. An inverse treatment planning model for intensity modulated radiation therapy with dynamic MLC. *Appl. Math. Comput.* 135:227-250, 2003

[73] J. Tervo, P. Kolmonen, T. Lyyra-Laitinen, J.D. Pintér and T. Lahtinen. An optimization-based approach to the multiple static delivery technique in radiation therapy. *Annals of Operational Research* 119:205-227, 2003

[74] X-H. Wang, R. Mohan, A.J. Jackson, S.A. Leibel, Z. Fuks and C.C. Ling. Optimization of intensity modulated 3D conformal treatment plans based on biological indices. *Radiother. Oncol.* 37:140-152, 1995

[75] X-H. Wang, S. Spirou, T. LoSasso, J. Stein, C-S. Chui and R. Mohan. Dosimetric verification of intensity-modulated fields. *Med. Phys.* 23:317-327, 1996

[76] S. Webb. Optimisation of conformal radiotherapy dose distributions by simulated annealing. *Phys. Med. Biol.* 34:1349-1369, 1989

[77] S. Webb. *The Physics of Three-Dimensional Radiation Therapy.* IOP, Bristol, 1997

[78] S. Webb. *The Physics of Conformal Radiotherapy.* IOP, Bristol, 1997

[79] S. Webb. Configuration options for intensity-modulated radiation therapy using multiple static fields shaped by a multileaf collimator. *Phys. Med. Biol.* 42:595-602, 1998

[80] S. Webb. The physics of radiation treatment. *Physics World* 11(11):39-42, 1998

[81] S. Webb. Conformal intensity-modulated radiotherapy (IMRT) delivered by robotic linac – testing IMRT to the limit. *Phys. Med. Biol.* 44:1639-1654, 1999

[82] K.J. Weeks and M.R. Sontag. 3-D dose-volume compensation using nonlinear least-squares regression technique. *Med. Phys.* 18:474-480, 1991

[83] P. Xia and L.J. Verhey. Multileaf collimator leaf sequencing algorithm for intensity modulated beams with multiple static segments. *Med. Phys.* 25:1424-1434, 1998

[84] L. Xing and G.T.Y Chen. Iterative methods for inverse treatment planning. *Phys. Med. Biol.* 41:2107-2123, 1996

[85] C. Zhu, R.H. Byrd, P. Lu and J. Nocedal. Algorithm 778: L-BFGS-B: Fortran subroutines for large-scale bound-constrained optimization. *ACM Transactions on Mathematical Software* 23:550-560, 1997

Appendix I

L-BFGS-B

L-BFGS-B is a quasi-Newton method [24] designed to solve large dimensional non-linear simple bounded (box constrained) optimization problems of the form

$$\begin{aligned} \min \quad & f(\mathbf{x}) \\ \text{subject to} \quad & l_i \leq x_i \leq u_i,\ i = 1, \ldots, n, \end{aligned} \tag{A.1}$$

where $f : \mathbb{R}^n \to \mathbb{R}$ is a nonlinear function [8, 85]. The gradient $\mathbf{g}$ of f must be provided by the user. The number n of variables is assumed to be large. The algorithm does not require the knowledge of the second derivate of f since it uses a quasi-Newton update of the Hessian $\mathbf{B}$ of f.

At iteration k a quadratic model

$$m_k(x) = f(x_k) + \mathbf{g}_k^{\mathrm{T}}(\mathbf{x} - \mathbf{x}_k) + \frac{1}{2}(\mathbf{x} - \mathbf{x}_k)^{\mathrm{T}}\mathbf{B}_k(\mathbf{x} - \mathbf{x}_k) \tag{A.2}$$

is approximately minimized subject to the bounds $\mathbf{l}$ and $\mathbf{u}$. Before the minimization, gradient projection method is used to determine a set of active constraints at each iteration *i.e.* the set $I = \{i \mid x(i) = l(i) \text{ or } x(i) = u(i),\ i = 1, \ldots, n\}$ (see *e.g.* [15]). Then, the constraints belonging to the active set are treated as equality constraints during the minimization of $m_k(x)$.

The ordinary BFGS update of the Hessian $\mathbf{B}_{k+1}$ is of the form

$$\mathbf{B}_{k+1} = \mathbf{B}_k + \frac{\mathbf{y}_{k+1}\mathbf{y}_{k+1}^{\mathrm{T}}}{\mathbf{y}_{k+1}^{\mathrm{T}}\mathbf{s}_{k+1}} - \frac{(\mathbf{B}_k\mathbf{s}_{k+1})(\mathbf{B}_k\mathbf{s}_{k+1})^{\mathrm{T}}}{\mathbf{s}_{k+1}\mathbf{B}_k\mathbf{s}_{k+1}}, \tag{A.3}$$

where $\mathbf{s}_{k+1} = \mathbf{x}_{k+1} - \mathbf{x}_k$ and $\mathbf{y}_{k+1} = \mathbf{g}_{k+1} - \mathbf{g}_k$ [32]. The update A.3 requires dense matrices $\mathbf{B}_{k+1}$ and $\mathbf{B}_k$ which is impossible if n is large (several thousands). One method to overcome this is to use a sparse representation of $\mathbf{B}$ with a known sparsity pattern. An example of this kind of approach is in the well known LANCELOT algorithm [16].

What makes the L-BFGS-B algorithm suitable for large-scale optimization, when $\mathbf{B}_k$ is not sparse or the sparsity pattern is not known, is the use of limited

memory BFGS matrices to approximate the updated Hessian $\mathbf{B}_k$ [8]. At every iteration k, the algorithm stores a small number m of correction pairs $\{\mathbf{s}_i, \mathbf{y}_i\}$, $i = k-1, \ldots, k-m$, where

$$\mathbf{s}_k = \mathbf{x}_{k+1} - \mathbf{x}_k \text{ and } \mathbf{y}_k = \mathbf{g}_{k+1} - \mathbf{g}_k. \tag{A.4}$$

The correction pairs contain information about the curvature of f and they are used instead of the full Hessian.

Now, the limited memory Hessian can be determined using the correction pairs (A.4). Define first $n \times m$ correction matrices

$$\mathbf{Y}_k = [\mathbf{y}_{k-m} \ \ldots \ \mathbf{y}_{k-1}], \ \mathbf{S}_k = [\mathbf{s}_{k-m} \ \ldots \ \mathbf{s}_{k-1}]. \tag{A.5}$$

When the BFGS update (A.3) is used, the Hessian is (from [8])

$$\mathbf{B}_k = \Theta\mathbf{I} - \mathbf{W}_k\mathbf{M}_k\mathbf{W}_k^{\mathrm{T}}, \tag{A.6}$$

where

$$\mathbf{W}_k = \begin{bmatrix} \mathbf{Y}_k & \Theta\mathbf{S}_k \end{bmatrix},$$

$$\mathbf{M}_k = \begin{bmatrix} -\mathbf{D}_k & \mathbf{L}_k^{\mathrm{T}} \\ \mathbf{L}_k & \Theta\mathbf{S}_k^{\mathrm{T}}\mathbf{S}_k \end{bmatrix}^{-1},$$

with

$$(L_k)_{i,j} = \begin{cases} (\mathbf{s}_{k-m-1})^{\mathrm{T}}(\mathbf{y}_{k-m-1+j}) & \text{if } i > j \\ 0 & \text{otherwise} \end{cases}$$

and

$$\mathbf{D}_k = \operatorname{diag}\left[\mathbf{s}_{k-m}^{\mathrm{T}}\mathbf{y}_{k-m} \ \ldots \ \mathbf{s}_{k-1}^{\mathrm{T}}\mathbf{y}_{k-1}\right].$$

The scalar Θ is a positive scaling parameter.

In the L-BFGS-B algorithm, the limited memory presentation (A.6) is used to perform efficiently computations involving the Hessian $\mathbf{B}_k$. The saving of memory is remarkable since, in principle, only $2m$ vectors of length n need to be saved and m can be as small as 3. Typical values are $3 \leq m \leq 20$ [85].