数理解析研究所講究録
第1641巻 2009年 45-67

Algorithms for Geometric Problems in Intensity-Modulated Radiation Therapy*

Danny Z. Chen and Chao Wang
Department of Computer Science and Engineering
University of Notre Dame
Notre Dame, IN 46556, USA
E-mail: {chen, cwang1}@cse.nd.edu

Abstract

Intensity-modulated radiation therapy (IMRT) is a modern cancer treatment technique aiming to deliver a prescribed conformal radiation dose to a target tumor while sparing the surrounding normal tissue and critical structures. In this paper, we consider a set of geometric and combinatorial problems that arise in IMRT planning and delivery, such as static leaf sequencing, static leaf sequencing with error control, field splitting, dose simplification, dynamic leaf sequencing, single-arc leaf sequencing, and and image segmentation. We discuss efficient algorithms for these problems. The algorithms exploit the underlying geometric and combinatorial properties of the problems and transform them into geometric or graph-theoretic problems such as shortest paths, optimal matchings, maximum flows, multicommodity demand flows, and linear programming. Some open problems and promising directions for future research are also given.

1 Introduction

Algorithmic studies of combinatorial and geometric optimization problems that arise in the field of medicine are emerging as a significant interdisciplinary research area. Many medical applications call for effective and efficient algorithmic solutions for many discrete or continuous computational problems. In this paper, we cover a number of geometric and combinatorial problems in the research and clinical practice of *intensity-modulated radiation therapy* for cancer treatment.

Intensity-modulated radiation therapy (IMRT) is a modern cancer treatment technique that aims to deliver highly conformal prescribed radiation distribution to a target volume (e.g., a tumor) in 3-D while sparing the surrounding normal tissue and critical structures. Good quality IMRT treatment is based on the ability to *accurately* and *efficiently* deliver prescribed dose distributions of radiation, called *intensity maps (IMs)*, to the target. An IM is specified by a set of nonnegative integers on a uniform 2-D grid (see Figure 1(a)). The value for each grid cell of the IM indicates the intensity level (in units) of radiation to be delivered to the body region corresponding to that grid cell. For common clinical prescriptions, the size of an IM cell is typically measured from $0.5cm \times 0.5cm$ to $1cm \times 1cm$ [29].

Most current IMRT delivery systems consist of two major components: (i) a medical linear accelerator (LINAC) as the radiation source, which generates photon beams used in IMRT treatment,

*This research was supported in part by NSF Grant CCF-0515203 and NIH Grant 1R01-CA117997-01A2.

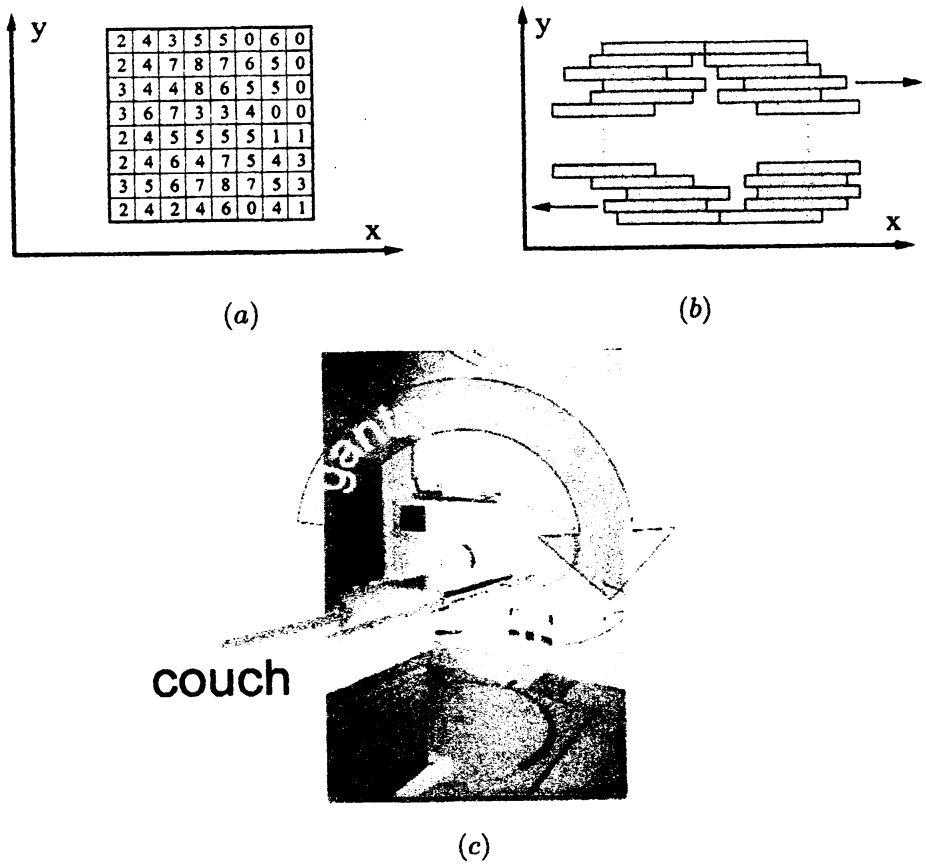


Figure 1: (a) An intensity map (IM). (b) A schematic drawing of a multileaf collimator (MLC), in which the shaded rectangles represent the MLC leaves that can move left and right (as indicated by the arrows) to form an MLC-aperture. (c) The patient couch and rotatable gantry.

and (ii) a *multileaf collimator* (MLC) [44, 72, 73], which defines or specifies the shapes of cylindrical radiation beams for delivering the prescribed intensity maps (IM). An MLC is made of many pairs of tungsten alloy leaves of the same rectangular shape and size (see Figure 1(b)). The opposite leaves of each pair are aligned to each other. These leaves, which are controlled by a computer, can move left and right to form a rectilinear polygonal region, called *MLC-aperture*, which defines the cross-section of a cylindrical radiation beam. Note that such a cross-section is of the shape of a y-monotone rectilinear polygon (Figure 1(b)). During IMRT treatment, the patient is positioned and secured to the treatment couch (see Figure 1(c)), and the LINAC delivers radiation beams to the target tumor from various directions. The direction of a radiation beam is controlled by rotating the gantry (which contains the radiation source) to the desired angle. The MLC is mounted on the gantry and the cross-section of the cylindrical radiation beam is shaped by an MLC-aperture to deliver a uniform dose to (a portion of) an IM [17, 33, 51, 52, 53, 63, 67, 80, 83].

Currently, there are three popular commercial MLC systems in use for clinical cancer treatment: Elekta, Siemens, and Varian [44]. Depending on the actual MLC system used, there are some differences among the particular geometric shapes of the rectilinear y-monotone polygonal regions that can be formed by the corresponding MLC. The details on the differences between these MLC

systems will be discussed in Section 2.1.

Using a computer-controlled MLC, the IMs are usually delivered either statically or dynamically. In the *static* approach [15, 16, 11, 12, 13, 14, 30, 35, 41, 40, 74, 79], the MLC leaves do not move during irradiation, and are repositioned to form another beam shape only when the radiation source is turned off. In the *dynamic* approach [28, 29, 31, 45, 70, 69, 71], the MLC leaves keep moving across an IM field while the radiation remains on. Note that MLC leaves are capable of changing their positions quickly during irradiation, and the actually delivered IMs can still remain to be of integer values in the dynamic settings. In both these delivery approaches, the gantry is fixed during irradiation. *Arc-modulated radiation therapy* (AMRT) is a newly emerging IMRT delivery technique [82]. AMRT differs from the commonly used static and dynamic IMRT approaches in that the gantry is kept rotating during delivery, typically along a circular arc path in 3-D around the patient. The radiation source remains on for irradiation during the gantry rotation. In the meantime, the MLC leaves also move continuously across the treatment field to produce the prescribed IM.

A *treatment plan* for delivering a given IM provides a precise description of how to control the MLC leaves and radiation source to deliver the dose distribution of the IM. For static IMRT, the treatment plan must also specify how to turn on/off the radiation beam source. Two key criteria are used to evaluate the quality of an IMRT treatment plan:

(i) **Delivery time (the efficiency):** Minimizing the delivery time is important because it not only lowers the treatment costs but also increases the patient throughput. Short delivery time also reduces the possible recurrence of tumor cells. For static IMRT, the delivery time consists of two parts: (1) the *beam-on time* and (2) the *machine setup time*. Here, the *beam-on time* refers to the total amount of time when a patient is exposed to actual irradiation, and the *machine setup time* refers to the time associated with turning on/off the radiation source and repositioning MLC leaves. Since the machine setup time is the dominating portion of the total delivery time in static IMRT, minimizing the machine setup time normally can significantly shorten the delivery time. On the other hand, minimizing the beam-on time is considered as an effective way to enhance the radiation source efficiency of the machine (referred to as the *monitor unit efficiency* or *MU efficiency* in medical literature) as well as to reduce the patient's risk and treatment uncertainties under radiation [10]. For dynamic IMRT or arc-modulated radiation therapy, since the radiation source is always on when delivering an IM, the delivery time is almost equal to the beam-on time.

(ii) **Delivery error (the accuracy):** For various reasons, there are discrepancies between the prescribed IM and the actually delivered IM. We distinguish two types of delivery error: (1) *approximation error* and (2) *tongue-and-groove error*. The *approximation error* refers to the discrepancy between the prescribed IM and actually delivered IM that resides in the interior of the IM grid cells; the *tongue-and-groove error* [79, 81, 68] refers to the discrepancy between the prescribed IM and actually delivered IM that appears on the boundaries of the IM cells, which is caused by the special geometric shape design, called *tongue-and-groove* design, of the MLC leaves [81] (Section 2.2 discusses more on its nature).

There are many interesting and important geometric optimization problems arising in the planning and delivery of radiation cancer treatment. In this paper, we discuss the following problems in IMRT, such as: (1) static leaf sequencing (SLS), (2) static leaf sequencing with tongue-and-groove

error control, (3) field splitting, (4) dose simplification (DLS), (5) dynamic leaf sequencing, (6) single-arc leaf sequencing, and (7) image segmentation. The definitions of the above problems will be given in later sections.

The rest of this paper is organized as follows. Section 2 discusses the MLC constraints and its tongue-and-groove feature. Sections 3 to 7 discuss the aforementioned problems and give efficient algorithms. Section 8 discusses some open problems and promising directions for future research.

2 Preliminaries

In this section, we first discuss the three popular MLC systems and their constraints, then characterize the tongue-and-groove error associated with IMRT treatments using such MLC systems.

2.1 Constraints of Multileaf Collimators

As mentioned in Section 1, there are three popular MLC systems currently used in clinical treatments [44]: the Elekta, Siemens, and Varian MLC systems. The mechanical structure of these MLCs, although is quite flexible, is not perfect in that it still precludes certain aperture shapes from being used [44, 29, 73] for treatment. In the following, we summarize the common constraints that appear in these MLC systems:

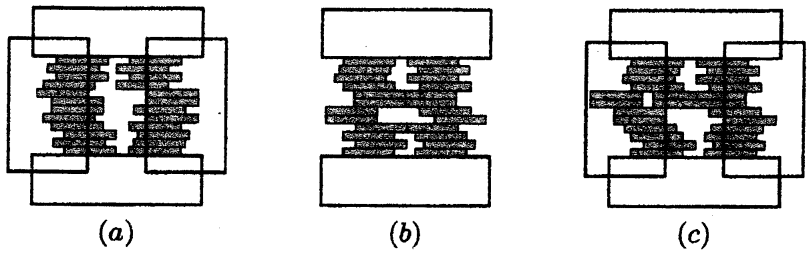
(1) The *minimum leaf separation* constraint. This requires the distance between the opposite leaves of any MLC leaf pair (e.g., on the Elekta or Varian MLC) to be no smaller than a given value δ (e.g., $\delta = 1$cm).

(2) The *interleaf motion* constraint. On the Elekta or Siemens MLC, the tip of each MLC leaf is not allowed to surpass those of its neighboring leaves on the opposite leaf bank.

(3) The *maximum leaf spread* constraint. This requires the maximum distance between the tip of leftmost left leaf and the tip of the rightmost right leaf of the MLC is no more than a threshold distance (e.g., 25 *cm* for Elekta MLCs). This constraint applies to all existing MLCs.

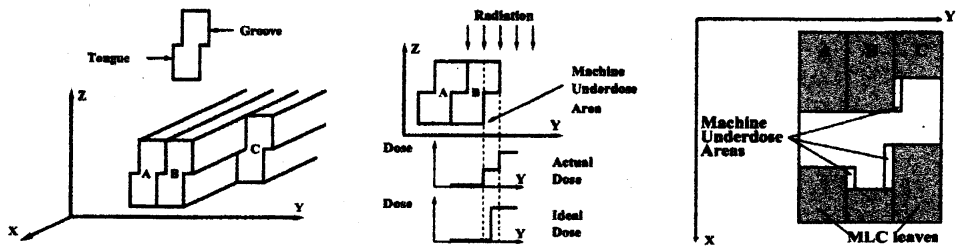
(4) The *maximum leaf motion speed* constraint. This requires the MLC leaves cannot move faster than a threshold value (e.g., 3cm/s for Varian MLCs). This constraint applies to all existing MLCs.

Figure 2 shows the constraints (1) and (2) of these MLC systems. The Elekta MLC is subject to both the minimum leaf separation and interleaf motion constraints. The Siemens MLC is subject to only the interleaf motion constraint. Hence a degenerate rectilinear y-monotone polygon can be formed by the Siemens MLC, by closing some leaf pairs (see Figure 2(b)). (A polygon P in the plane is called *y-monotone* if every line orthogonal to the y-axis intersects P at most twice. A y-monotone polygon is *degenerate* if its interior is not a single connected piece.) The Varian MLC is subject to the minimum leaf separation constraint, but allows interleaf motion. Thus, to "close" a leaf pair in the Varian system, we can move the leaf opening under the backup diaphragms (see Figure 2(c)). But, the Elekta MLC cannot "close" its leaf pairs in a similar manner due to its interleaf motion constraint.



(a) (b) (c)

Figure 2: Illustrating the constraints of three different MLC systems. The shaded rectangles represent the MLC leaves, and the unshaded rectangles represent the backup diaphragms which form a bounding box of an MLC-aperture. (a) The Elekta MLC. (b) The Siemens MLC. (Notice that unlike the Elekta and Varian MLCs, the Siemens MLC only has a single pair of backup metal diaphragms.) (c) The Varian MLC.



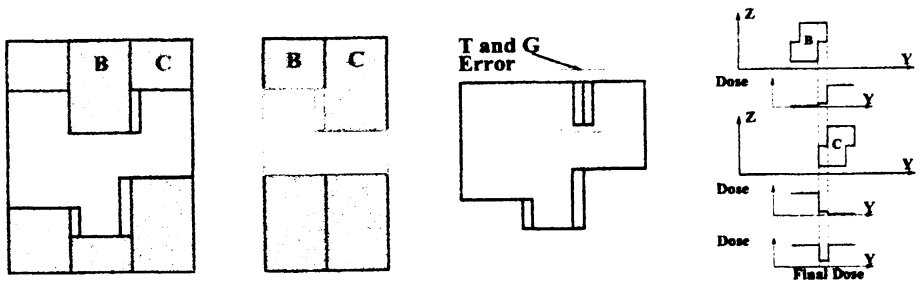
(a) (b) (c)

Figure 3: (a) Illustrating the tongue-and-groove interlock feature of the MLC in 3-D, where leaf B is used for blocking radiation. (b) When leaf B is used for blocking radiation, there is an underdose and leakage in the tongue or groove area. (c) The underdose and leakage areas of the tongue-and-groove feature on an MLC-aperture region.

Geometrically, on the Elekta, each MLC-aperture is a rectilinear y-monotone *simple* polygon whose minimum vertical "width" is $\geq$ the minimum separation value δ, while on the Siemens or Varian, an MLC-aperture can be a degenerate y-monotone polygon (i.e., with several connected components).

2.2 Tongue-and-Groove Design of the MLC Leaves

On most current MLCs, the sides of the leaves are designed to have a "tongue-and-groove" interlock feature (see Figure 3(a)). This design reduces the radiation leakage through the gap between two neighboring MLC leaves and minimizes the friction during leaf movement [68, 73, 79, 81]. But, it also causes an unwanted underdose and leakage situation when an MLC leaf is used for blocking radiation (see Figures 3(b) and 3(c)). Geometrically, the underdose and leakage error caused by the tongue-and-groove feature associated with an MLC-aperture is a set of 3-D axis-parallel boxes $w \cdot l_i \cdot h$, where w is the (fixed) width of the tongue or groove side of an MLC leaf, l_i is the length of the portion of the i-th leaf that is actually involved in blocking radiation, and $h = \alpha \cdot r$ is the amount of radiation leakage with α being the (fixed) leakage ratio and r being the amount of radiation delivered



(a) (b) (c) (d)

Figure 4: Illustrating the tongue-and-groove error. (a) and (b): Two MLC-apertures (the shaded rectangles represent MLC leaves). (c) When delivering the two MLC-apertures in (a) and (b) (one by one), the groove side of leaf B and tongue side of leaf C are both used for blocking radiation, causing a tongue-and-groove error in the area between the leaves B and C. (d) Illustrating a dose "dip" in the final dose distribution where a tongue-and-groove error occurs.

by that MLC-aperture. Figure 3(b) illustrates the height of the underdose and leakage error, and Figure 3(c) illustrates the width and length of the underdose and leakage error.

The **tongue-or-groove error** of an MLC-aperture is defined as the amount of underdose and leakage error occurred whenever the tongue side or groove side of an MLC leaf is used for blocking radiation. The tongue-or-groove error of an IMRT plan (i.e., a set of MLC-apertures) is the sum of the tongue-or-groove errors of all its MLC-apertures. The **tongue-and-groove error** occurs whenever the tongue side of an MLC leaf and the corresponding groove side of its neighboring leaf are both used for blocking radiation in any two different MLC-apertures of an IMRT plan (see Figure 4). Note that the tongue-or-groove error is defined on each individual MLC-aperture, while the tongue-and-groove error is defined by the relations between different MLC-apertures. Clearly, the tongue-and-groove error is a subset of the tongue-or-groove error. In medical physics, the tongue-and-groove error has received more attention than tongue-or-groove error because it usually occurs in the middle of the delivered intensity maps, causing insufficient dose coverage to the tumor [79]. According to a recent study [30], tongue-and-groove error may cause a point dose error up to 10%, well beyond the allowed 3-5% limit.

Chen *et al.* [24] introduced the notion of *error-norm*, which is closely related to the tongue-or-groove error. The error-norm $||A||_E$ of an IM $A = (a_{i,j})$ of size $m \times n$ is defined as

$$||A||_E = \sum_{i=1}^{m} \left(|a_{i,1}| + \sum_{j=1}^{n-1} |a_{i,j} - a_{i,j+1}| + |a_{i,n}| \right). \quad (1)$$

Chen *et al.* [23] proved that for any treatment plan S that exactly delivers IM A, the difference between the associated tongue-or-groove error $TorG(S)$ and the tongue-and-groove error $TandG(S)$ always equals to $||A||_E$. Thus minimizing tongue-or-groove error is equivalent to minimization tongue-and-groove error. Chen *et al.* [23] further showed that $||A||_E$ is the strict lower bound for $TorG(S)$, which implies that the tongue-and-groove error has a strict lower bound 0 for any input IM.

3 Static Leaf Sequencing

3.1 Problem Definition

The static leaf sequencing problem arises in the static IMRT delivery. In this delivery approach, an IM is delivered as follows:

1. Form an MLC-aperture.
2. Turn on the beam source and deliver radiation to the area of the IM exposed by the MLC-aperture.
3. Turn off the beam source.
4. Reposition the MLC leaves to form another MLC-aperture, and repeat steps 2-4 until the entire IM is done.

In this setting, the boundary of each MLC-aperture does not intersect the interior of any IM cell, i.e., any IM cell is either completely inside or completely outside the polygonal region formed by the MLC-aperture. In delivering a beam shaped by an MLC-aperture, all the cells inside the region of the MLC-aperture receive the same integral amount of radiation dose (say, one unit), i.e., the numbers in all such cells are decreased by the same integer value. The IM is done when each cell has a value zero.

The *static leaf sequencing (SLS)* problem, in general, seeks a treatment plan S (i.e., a set of MLC-apertures together with their weights) for exactly delivering a given IM A such that the delivery time is minimized. Note that the weight of an MLC-aperture corresponds to the amount of radiation dose delivered to the area of the IM exposed by the MLC-aperture, and by "exactly delivering" we mean there is no approximation error but may have tongue-or-groove error. We will discuss how to minimize the tongue-or-groove error in static IMRT delivery in the next section.

Several variations of the SLS problem have been studied in the literature: (1) the *SLS problem with minimum beam-on time* [3, 7, 10, 13, 46, 54], (2) the *SLS problem with minimum machine setup time* [7, 23, 25, 30, 79], and (3) the *SLS problem with minimum beam-on time plus machine setup time* [7, 54]. Recall that for static IMRT, the delivery time equals the sum of beam-on time and machine setup time.

The SLS problem with minimum beam-on time is polynomial time solvable [10]. Ahuja and Hamacher [3] formulated this problem as a minimum cost flow problem in a directed network and gave an optimal linear time algorithm for the case when the MLC is not subject to the minimum separation constraint and interleaf motion constraint, e.g., the Varian MLC. Kamath *et al.* [46] proposed a different approach by sweeping the IM field from left to right and generating MLC-apertures accordingly. They showed how to handle the minimum separation constraint and interleaf motion constraint during the sweeping and achieve minimum beam-on time.

The SLS problem with minimum machine setup time is, in general, NP-hard. Burkard [18] proved that for IMs with at least two rows, the problem can be reduced from the subset sum problem [38]. Baatar *et al.* [7] further showed that NP-hardness holds even for input IMs with a single row, using a reduction from the 3-partition problem [38]. Chen *et al.* [22], independently, gave an NP-hardness proof of this problem based on a reduction from the 0-1 knapsack problem [38]. The SLS problem

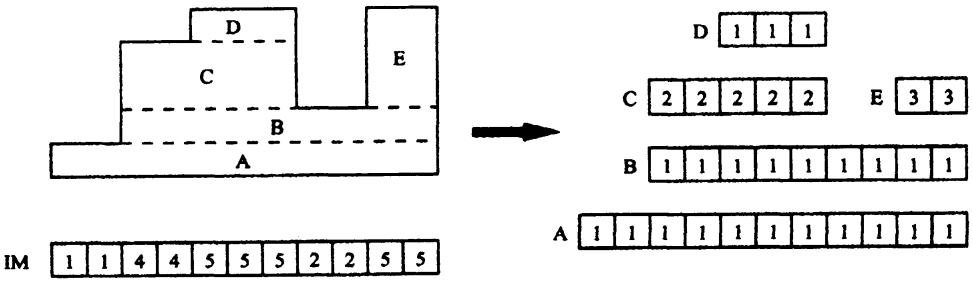


Figure 5: The leaf openings A, B, C, D, and E for delivering the dose profile curve of an IM of a single row.

with minimum beam-on time plus machine setup time is also NP-hard [10]. We are not aware of any efficient algorithms for the MLC problem with minimum beam-on time plus machine setup time.

In this section, we will study the following **3-D static leaf sequencing (SLS)** problem [79]: Given an IM, find a minimum set S of MLC-apertures (together with their weights) for delivering the IM (i.e., the size $|S|$ is minimized). Note that our goal is to minimize the *number of MLC-apertures* used for delivering an IM, which is equivalent to minimizing the machine setup time under the assumption that the setup time needed to go from one MLC-aperture shape to another shape is constant. While the MLC problem with minimum beam-on time plus machine setup time seems more general, our formulation of the 3-D SLS problem also well captures the total delivery time. This is because the machine setup time for the MLC-apertures dominates the total delivery time [30, 79], and algorithms that minimize the number of MLC-apertures used are desired to reduce the delivery time [79]. It should be pointed out that even though the SLS problem with the minimum beam-on time is polynomial time solvable, the resulting plan can use a very large number of MLC-apertures[36], and thus a prolonged total delivery time.

A key special case of the SLS problem, called the **basic 3-D SLS** problem, is also of clinical value [11, 12, 13, 14, 15, 16]: Given an IM, find a minimum set of MLC-apertures for the IM, such that each MLC-aperture has a unit height. Note that in the general SLS problem, the weight of each MLC-aperture can be any integer ≥ 1. Studying the basic case is important because the maximum heights of the majority of IMs used in current clinical treatments are around 5, and an optimal solution for the basic case on such an IM is often very close to an optimal solution for the general case.

The rest of this section is organized as follows. In Section 3.2, we consider the special SLS case that the input IM has a single row. In Section 3.3, we consider the special SLS case that the input IM has only 0 or 1 in its cells. In Section 3.4, we study the 3-D SLS problem.

3.2 The 1-D SLS Problem

When the input IM is of a single row, the 3-D SLS problem reduces to the 1-D case. The input IM becomes a one-dimensional intensity profile and the output MLC-apertures become a set of leaf openings with weighted intensity levels (see Figure 5). Such openings are obtained by setting the left and right leaves of an MLC leaf pair to specific positions. The goal is to minimize the number of leaf openings.

The **basic 1-D SLS problem**, in which the output leaf openings are required to have unit

weights, can be easily solved using a sweeping method proposed by Boyer and Yu [17]. The method first distinguishes the left and right ends along the intensity profile curve, and then uses a greedy approach to generate a *delivery option*, i.e., a valid pairing in which a left end is always paired with a right end on its right. An interesting question is how to generate all possible delivery options for a basic 1-D SLS instance. Yu [80] pointed out that for an IM row with N left ends and N right ends, there are in the worst case $N!$ delivery options. Webb [74] presented a general formula for determining the total number of delivery options for an arbitrary IM row, and showed that this number in general tends to be considerably smaller than $N!$. Luan *et al.* [56] gave an efficient algorithm for producing all possible delivery options which runs in optimal time, linearly proportional to the number of output delivery options. The key idea of the algorithm is to impose an order onto how the left and right ends should be paired: it recursively pairs the rightmost unpaired left end with an unpaired right end such that the pair does not define an illegal opening.

The **general 1-D SLS problem**, in which the output leaf openings are not required to have unit weights, is NP-hard [7, 22]. It is quite easy to observe that a horizontal trapezoidal decomposition of the input intensity profile curve yields a 2-approximation solution for the general 1-D SLS problem (e.g., see Figure 5).

Bansal *et al.* [8] modeled the general 1-D SLS problem as a maximum partition problem of prefix positive zero sum vectors, and proposed several approximation algorithms with approximation ratio less than 2. For the unimodal input intensity profile curves, i.e., of only one peak, they gave a $\frac{9}{7}$ approximation algorithm by reducing the general 1-D SLS problem to the set packing problem [43]. For arbitrary input intensity profile curves, they gave a $\frac{24}{13}$ approximation based on rounding the solution of a certain linear programming problem.

Independently, Chen *et al.* [21] studied the *shape rectangularizarion (SR)* problem of finding minimum set of rectangles to build a given functional curve, which turns out to be equivalent to the general 1-D SLS problem. They presented a $(\frac{3}{2}+\epsilon)$-approximation algorithm for the SR problem. They pointed out two interesting geometric observations: (1) For the SR problem, it is sufficient to consider only a special type of rectangle sets, called *canonical rectangle sets*, in which the left and right ends of the rectangles coincide with the left and right ends of the input intensity profiles; (2) An optimal rectangle set which is also canonical is isomorphic to a forest of weighted trees. Based on the above observations, they proposed a combinatorial optimization problem, called the *primary block set* (PBS) problem[21], which is related to the SR problem in the sense that for any $\mu \geq 2$, a μ-approximation PBS algorithm immediately implies a $(2-\frac{1}{\mu})$-approximation SR algorithm. Further, they showed that the PBS problem can be reformulated, in polynomial time, as a multicommodity demand flow (MDF) problem [19] on a path. Chekuri *et al.* [19] gave a polynomial time $(2+\epsilon)$-approximation algorithm for the MDF problem on a path when the maximum demand d_{max} is $\leq$ the minimum capacity c_{min}. Chen *et al.* [21] extended Chekuri *et al.*'s result and gave a $(2+\epsilon)$-approximation algorithm for the MDF problem on a path when $d_{max} \leq \lambda \cdot c_{min}$, where $\lambda > 0$ is any constant. This leads to a $(\frac{3}{2}+\epsilon)$-approximation algorithm for the SR problem when $M_f \leq \lambda \cdot m_f$, where m_f (or M_f) is the global positive minimum (or maximum) of the input intensity profile curve f.

3.3 The 2-D SLS Problem

When the input IM has only 0 or 1 in its cells, the 3-D SLS problem becomes a 2-D one (called the **2-D SLS** problem), and an optimal set of MLC-apertures is just a minimum set of rectilinear y-monotone simple polygons. For MLCs with the minimum separation constraint, each rectilinear y-monotone simple polygon must be have a minimum width $\geq \delta$ (the parameter δ represents the minimum separation constraint discussed in Section 2.1). As shown in Section 3.4, a good algorithm for partitioning a rectilinear polygonal region (possibly with holes) into a minimum set of rectilinear y-monotone simple polygons can become a key procedure for solving the general 3-D SLS problem.

Chen *et al.* [23] presented a unified approach for solving the minimum y-monotone partition problem on an n-vertex polygonal domain (possibly with holes) in various settings. This problem can be reduced to computing a maximum bipartite matching. But, due to the specific geometric setting, an explicit construction of the underlying bipartite graph for a maximum matching is too costly. They overcome this difficulty by reducing the matching problem to a maximum flow problem on a geometric graph. Since the matching information is implicitly stored in the flow graph, they use a special depth-first search to find an actual optimal matching. Thus, an $O(n^2)$ time algorithm is obtained for partitioning a polygonal domain (possibly with holes) into a minimum set of y-monotone parts, improving the previous $O(n^3)$ time algorithm [55] (which works only for the *simple* polygon case). The ideas can be extended to handling the minimum separation constraint by modifying the edge capacities of the flow graph.

3.4 The 3-D SLS Problem

In this subsection, we study the 3-D SLS problem, which seeks a minimum set S of MLC-apertures (together with their weights) for delivering the a given IM.

3.4.1 The Basic 3-D SLS Problem

The **basic 3-D SLS** problem, in which the all weights of the MLC-apertures are required to be 1, has been studied under different MLC constraints. For MLCs that are not subject to the minimum separation constraint nor the interleaf motion constraint, the problem can be easily reduced to the basic 1-D SLS problem (see Section 3.4), since each leaf pair is fully independent. For MLCs subjective to the minimum separation constraint but not the interleaf motion constraint, Kamath *et al.* [46] gave an algorithm that generates an optimal MLC-aperture set. The algorithm first generates an optimal MLC-aperture set without considering the minimum separation constraint, and then performs a sweeping of the IM field; during the sweeping, it repeatedly detects instances of violation and eliminates them accordingly.

For MLCs subjective to both the minimum separation and interleaf motion constraints, Chen *et al.* [25] modeled the corresponding 3-D SLS problem as an *optimal terrain construction* problem. To compute an optimal set of MLC-apertures for the IM, the algorithm builds a graph G: (1) Generate all distinct delivery options for each IM row, and let every vertex of G correspond to exactly one such delivery option; (2) for any two delivery options for two consecutive IM rows, put left-to-right directed edges whose weights are determined by the optimal bipartite matchings between the leaf openings of the two delivery options. The basic 3-D SLS problem is then solved by finding a shortest path in the graph G[25].

This algorithm, however, it does not handle two crucial issues well. The first issue is on the worst case time bound of the algorithm has a multiplicative factor of $N!$. Recall that there can be $N!$ distinct delivery options in the worst case for every IM row (see Section 3.2). The second issue is on the optimality of the MLC-aperture set produced. Since the algorithm uses only the *minimal* delivery options for each IM row to compute a set of MLC-apertures (i.e., each such delivery option has the minimum number of leaf opening for building the corresponding IM row), the output may not be truly optimal. It is possible that a truly optimal MLC-aperture set for an IM need not use a minimal delivery option for each row; instead, some of the delivery options used may be defined by *Steiner points* [23].

Chen *et al.*[23] proposed a polynomial time basic 3-D SLS algorithm that can handle the above two issues. The key idea is based on new geometric observations which imply that it is sufficient to use only very few special delivery options, which they called *canonical delivery options*, for each IM row without sacrificing the optimality of the output MLC-aperture set. This produces guaranteed optimal quality solutions for the case when a constant number of Steiner points is used on each IM row.

3.4.2 The General 3-D SLS Problem

We now study the **general 3-D SLS** problem, in which the weights of the MLC-apertures can be arbitrary positive integers.

Several heuristic methods for the general 3-D SLS problem have been proposed [11, 12, 13, 14, 15, 16, 29, 30, 35, 41, 74, 79]. They all have several main steps: (1) Choose the upper monotone boundary of an MLC-aperture (by using some simple criteria), (2) choose the lower monotone boundary, (3) check whether the two monotone boundaries enclose an MLC-aperture, and (4) output the MLC-aperture thus obtained. These steps are repeated until the entire IM is built. Depending on where the upper monotone boundary is placed, these algorithms are further classified as either the "sliding window" or "reducing level" methods. The sliding window methods always place the upper monotone boundary at the boundary of the planar projection of the remaining IM to be built, while the reducing level methods normally set the upper monotone boundary at the place with the maximum height of the remaining IM.

Chen *et al.* [23] presented a heuristic algorithm for the general 3-D SLS problem based on their solutions for the 2-D SLS problem (see Section 3.3 and basic 3-D SLS problem (see Section 3.4.1). To make use of the 2-D SLS algorithm and basic 3-D SLS algorithm, they partition the input IM into a "good" set of sub-IMs, such that each sub-IM can be handled optimally by one of these two algorithms (i.e., each resulting sub-IM must be of a uniform height, or the maximum height of each such sub-IM should be reasonably small, say, ≤ 5). The partition is carried out in a recursive fashion and for each recursive step, several partition schemes are tested and compared, and the algorithm finally chooses the best result using dynamic programming.

Luan *et al.* [57] gave two approximation general 3-D SLS algorithms for MLCs without the minimum separation constraint nor the interleaf motion constraint (e.g., the Varian MLC). The first is a $(\lceil \log h \rceil + 1)$-approximation where $h > 0$ is the largest entry in the IM. The second is a $2(\lceil \log D \rceil + 1)$-approximation algorithm where D is the maximum element of a set containing 1) all absolute differences between any two consecutive row entries over all rows, 2) the first entry of each row, 3) the last entry of each row, and 4) the value 1. The main ideas include (1) 2-approximation

algorithm for the single-row case, and (2) an IM partition scheme that partitions the input IM into logarithmic number of sub-IMs with weights that are all powers of 2. These two algorithms have a running time complexity of $O(mn \log h)$ and $O(mn \log D)$, respectively.

4 Static Leaf Sequencing with Tongue-and-Groove Error Control

In this section, we study the static leaf sequencing (SLS) problem with tongue-and-groove error control, i.e., we will seek to minimize both the delivery time and the tongue-and-groove error. As we will show later in this section, there is actually a trade-off between these two criteria. Thus this problem has been studied in two variations: (1) minimizing the delivery time subject to the constraint that the tongue-and-groove error is completely eliminated [24, 49, 65]. (2) minimizing the delivery time subject to the constraint that the tongue-and-groove error is no more than a given error bound [22].

Que *et al.* [65] gave a heuristic SLS algorithm that eliminates the tongue-and-groove error based on the "sliding-windows" method proposed by Bortfeld *et al.* [13]. The algorithm, however, does not guarantee optimal delivery time, either in terms of beam-on time or in terms of the number of MLC-apertures.

Kamath *et al.* [49] gave an SLS algorithm that minimizes beam-on time of the output plan subjective to the constraint that the tongue-and-groove error is completely eliminated. Their algorithm consists of a scanning scheme they proposed earlier [46] for solving the SLS problem with minimum beam-on time (and without tongue-and-groove error control), and a modification scheme that rectifies that possible tongue-and-groove error violation.

Chen *et al.* [22] presented a graph modeling of the basic SLS problem (i.e., each MLC-aperture is of unit weight) with a tradeoff between the tongue-and-groove error and the number of MLC apertures, and an efficient algorithm for the problem on the Elekta MLC model, which has both the minimum separation and interleaf motion constraints. In their solution, the problem is formulated as a k-weight shortest path problem on a directed graph, in which each edge is defined by a minimum weight g-cardinality matching. Every such k-weight path specifies a set S of k MLC-apertures for delivering the given IM, and the cost of the path indicates the tongue-and-groove error of the set S of MLC-apertures. Chen *et al.* [22] also extended the above algorithm to other MLC models, such as Siemens and Varian, based on computing a minimum g-path cover of a directed acyclic graph. They used a partition scheme developed in paper [23] for handling the general case, i.e., each MLC-aperture can be of an arbitrary weight.

Chen *et al.* [24] presented an algorithm for SLS problem with minimum number of MLC-apertures subjective to the constraint that the tongue-and-groove error is completely eliminated. The main ideas of their algorithm include:

- A novel IM partition scheme, called *mountain reduction*, for reducing a "tall" 3-D IM (mountain) to a small set of "low" sub-IM mountains that introduces no tongue-and-groove error. (In contrast, the partition scheme used in paper [22] may introduce tongue-and-groove error.) The key to their mountain reduction scheme is the *profile-preserving mountain cutting (PPMC)* problem, which seeks to cut an IM A into two IMs qB and C (i.e., $A = q \cdot B + C$, where $q \geq 2$ is a chosen integer), such that qB and C have the same profile as A. Two IMs M' and M'' are said to have the *same profile* if for all i, j, $M'_{i,j} \geq M'_{i,j+1} \Leftrightarrow M''_{i,j} \geq M''_{i,j+1}$ and

$M'_{i,j} \leq M'_{i,j+1} \Leftrightarrow M''_{i,j} \leq M''_{i,j+1}$. They showed that the profile-preserving cutting introduces no tongue-and-groove error. They formulated this PPMC problem as a bottleneck shortest path (BSP) problem on a DAG G' of a pseudo-polynomial size. By exploiting interesting properties of this DAG, they compute the BSP by searching only a small (linear size) portion of G' and achieve an optimal linear time PPMC algorithm.

- An efficient graph based algorithm for partitioning a sub-IM into a minimum number of MLC-apertures without tongue-and-groove error. They directly incorporated the zero tongue-and-groove error constraint with the previous graph-based algorithm proposed in paper [22], and by exploiting geometric properties, they show that the size of the graph can be significantly reduced and computing the weights of edges can be done in a much faster manner using a new matching algorithm.

5 Field Splitting

In this section, we study a few geometric partition problems, called *field splitting*, which arise in intensity-modulated radiation therapy. Due to the maximum leaf spread constraint (see Section 2.1), an MLC cannot enclose an IM of a too large width. For example, on one of the most popular MLC systems called Varian, the maximum leaf spread constraint limits the maximum allowed field width to about 14.5 *cm*. Hence, this necessitates a large-width IM field to be split into two or more adjacent subfields, each of which can be delivered separately by the MLC subject to the maximum leaf spread constraint [32, 42, 75]. But, such IM splitting may result in a prolonged beam-on time and thus affect the treatment quality. The **field splitting problem**, roughly speaking, is to split an IM of a large width into several subfields whose widths are all no bigger than a threshold value, such that the total beam-on time of these subfields is minimized.

In this section, we will focus our discussion on the Varian MLCs, whose maximum spread threshold value (i.e., 14.5cm) is the smallest among existing popular MLCs and where the field splitting is often needed for medium and large size tumor cases. Henceforth, we always assume the MLC under discussion do not have the minimum separation constraint nor the interleaf motion constraint.

Engel [34] showed that for an IM M of size $m \times n$, when n is no larger than the maximum allowed field width w, the **minimum beam-on time (MBT)** of M is captured by the following formula

$$MBT(M) = \max_{i=1}^{m}\{M_{i,1} + \sum_{j=2}^{n} \max\{0, M_{i,j} - M_{i,j-1}\}\} \quad (2)$$

Engel also described a class of algorithms achieving this minimum value.

Geometrically, we distinguish three versions of the field splitting problem based on how an IM is split (see Figures 6(c)-6(e)): (1) splitting using vertical lines; (2) splitting using y-monotone paths; (3) splitting with overlapping. Note that in versions (1) and (2), an IM cell belongs to exactly one subfield; but in version (3), a cell can belong to two adjacent subfields, with a nonnegative value in both subfields, and in the resulting sequence of subfields, each subfield is allowed to overlap only with the subfield immediately before or after it.

Kamath *et al.* [50] studied the *field splitting using vertical lines (FSVL)* problem with minimum *total MBT*, i.e., the sum of the MBTs of the resulting subfields. Their algorithm worked in a brute-force manner and split a size $m \times n$ IM using vertical lines into at most three subfields (i.e., $n \leq 3w$

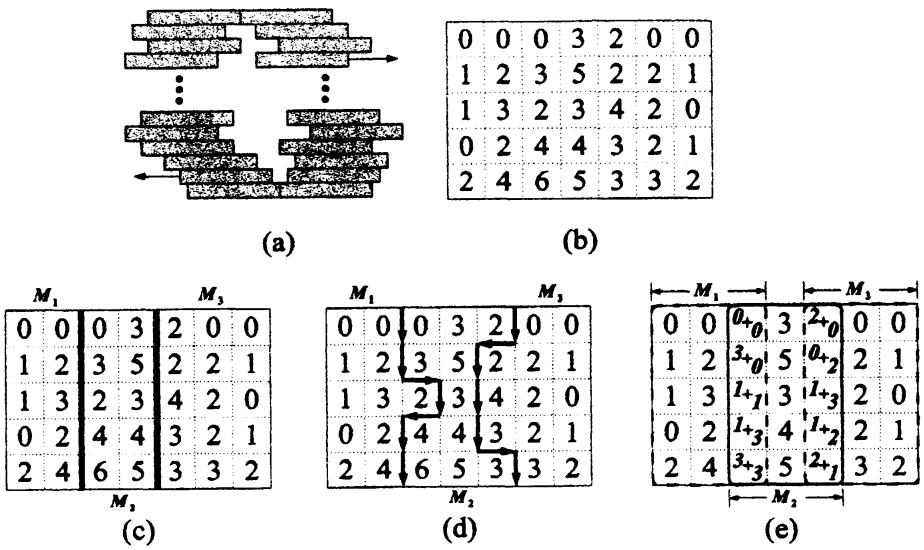


Figure 6: (a) An MLC. (b) An IM. (c)-(e) Examples of splitting an IM into three subfields, M_1, M_2, and M_3, using vertical lines, y-monotone paths, and with overlapping, respectively. The dark cells in (e) show the overlapping regions of the subfields; the prescribed dose in each dark cell is divided into two parts and allocated to two adjacent subfields.

for their algorithm, where w is the maximum allowed field width), and took $O(mn^2)$ time. Wu [77] formulated the FSVL problem that splits IMs of arbitrary widths into $k \geq 3$ subfields as a k-link shortest path problem in a directed acyclic graph. Each vertex in the graph represents a possible splitting line, and each edge represents the subfield enclosed by the two lines and is assigned a weight which equals the MBT for its associated subfield. With a carefully characterization of the intrinsic structures of the graph, an $O(mnw)$ time algorithm was achieved.

Chen and Wang [27] studied the *field splitting using y-monotone paths (FSMP)* problem. Their key observation is that there are at most a number of $mw + 1$, instead of $O((w+1)^m)$, candidates for the leftmost y-monotone path used in the splitting need to be considered. They then showed that all these candidate paths can be enumerated efficiently by using a heap data structure and an interesting new method called *MBT-sweeping*[27]. Further, they exploited the geometric properties of the underlying field splitting problem to speed up the computation of the total MBT of the induced subfields. The resulting FSMP algorithm took polynomial time as long as the number of subfields is a constant.

The *field splitting with overlapping (FSO)* problem has also been studied. Kamath *et al.* [47] studied a special FSO case in which an IM is split into at most three overlapping subfields and the overlapping regions of the subfields are *fixed*; a greedy scanning algorithm is proposed that produces optimal solutions in $O(mn)$ time for this special case. In fact, it was pointed out in paper [27] that by considering all possible combinations of overlapping regions, the algorithm in [47] can be extended to solving the general FSO problem, i.e., the overlapping regions of the subfields are not fixed and the number d of resulting subfields is any integer ≥ 2, in $O(mnw^{d-2})$ time.

Chen and Wang [27] presented an $O(mn + mw^{d-2})$ time FSO algorithm, improving the time

bound of Kamath *et al.*'s algorithm[47] by a factor of $\min\{w^{d-2}, n\}$. The algorithm hinges on a non-trivial integer linear programming (ILP) formulation of the *field splitting with fixed overlapping (FSFO) problem.* Basically, this is the FSO problem subject to the constraint that the sizes and positions of the d sought subfields are all fixed. They showed that the constraint matrix of the induced ILP problem is totally unimodular [64], and thus the ILP problem can be solved optimally by linear programming (LP). Further, they showed that the dual of this LP is a shortest path problem on a DAG, and the FSFO problem is solvable in totally $O(mn)$ time. The algorithm then reduced the original FSO problem to a set of $O(w^{d-2})$ FSFO problem instances. They pointed out that under the ILP formulation of the FSFO problem, with an $O(mn)$ time preprocess, each FSFO instance is solvable in only $O(m)$ time. This finally gave an $O(mn + mw^{d-2})$ time FSO algorithm.

6 Dose Simplification

In this section, we study the dose simplification problem that arises in IMRT. During the IMRT treatment planning process, the shapes, sizes, and relative positions of a tumor volume and other surrounding tissue are determined by 3-D image data, and an "ideal" radiation distribution is computed by a computer system. Without loss of generality, let the z-axis be the beam orientation. Then this "ideal" radiation distribution is a function defined on the xy-plane (geometrically, it is a 3-D functional surface above the xy-plane), which is usually of a complicated shape and not deliverable by the MLC. Thus, the "ideal" radiation distribution must be "simplified" to a discrete IM (see Section 1), i.e., a discrete IM approximates the "ideal" distribution under certain criteria. In some literature, a "ideal" radiation distribution is also referred to as a *continuous IM.*

The dose simplification problem has been studied under two major variants, depending on how the complexity of a (discrete) IM is defined:

- (Clustering) If the complexity of an IM is defined by the number of distinct intensity levels in the IM, then the dose simplification problem can be modeled as a **constrained 1-D K-means clustering** problem [9, 20, 61, 78]: Given an "ideal" radiation distribution of size $m \times n$, the mn intensity values in the distribution need to be grouped into K clusters for the smallest possible number K, such that the maximum difference between any two intensity values in each cluster is no bigger than a given bandwidth parameter δ and the total sum of variances of the K clusters is minimized. The resulting clusters are then used to specify a discrete IM.

- (Shape Approximation) If the complexity of an IM is defined by the number of MLC-apertures required to deliver the IM, then the dose simplification problem can be modeled as a shape approximation problem: Given an "ideal" radiation distribution, find K MLC-apertures for the smallest possible number K, such that the approximation error between the "ideal" radiation distribution and the discrete IM (defined by the sum of the K MLC-apertures) is within a given error bound $\mathcal{E}$.

Several algorithms for the constrained 1-D K-means clustering problems have been given in medical literature and used in clinical IMRT planning systems [9, 61, 78]. These algorithms use heuristic numerical methods to determine the clusters iteratively [9, 61, 78], and can be trapped in a local minimal. To the best of our knowledge, no theoretical analysis has been given on their convergence speed.

Chen *et al.* [20] presented an efficient geometric algorithm for computing optimal solutions to the constrained 1-D K-means problem. They modeled the problem as a K-link shortest path problem on a weighted directed acyclic graph (DAG). By exploiting the Monge property [1, 2, 62, 76] of the DAG, the algorithm runs in $O(\min\{Kn, n2^{\sqrt{\log K \log\log n}}\})$ time.

The shape approximation variant of the dose simplification problem appears to be much harder. Several papers [13, 21, 35, 73] discussed a simple version of the problem: Given a 1-D dose profile functional curve f, find a minimum set of rectangles whose sum approximates f within a given error bound $\mathcal{E}$. Due to its relation with the *shape rectangularization* problem (See Section 3.2), Chen *et al.* [21] called this special case the **generalized shape rectangularization (GSR)** problem and pointed out its NP-hardness. They also showed the relation between the GSR problem and the **forest of block towers (FBT)** problem: an optimal FBT solution immediately implies a 2-approximation GSR solution. Here, the FBT problem is a special GSR case in which the output rectangles are required to satisfy the *inclusion-exclusion constraint* [13, 35], meaning that for any two rectangles, either their projected intervals on the x-axis do not intersect each other (except possibly at an endpoint), or one interval is fully contained in the other interval.

Some work has been done on the FBT problem in the medical field [13, 35, 73]. An approach based on Newton's method and calculus techniques was used in [35]; but, it works well only when the input curve f has few "peaks", and must handle exponentially many cases as the number of "peaks" of f increases. Chen *et al.* [21] presented an FBT algorithm that produces optimal solutions for arbitrary input curves. They made a set of geometric observations, which imply that only a finite set of rectangles needs to be considered as candidates for the sought rectangles, and a graph G on such rectangles can be built. They then modeled the FBT problem as a k-MST [4, 5, 6, 37, 39, 59, 60, 66, 84] problem in G. By exploiting the geometry of G, they gave an efficient dynamic programming based FBT algorithm.

7 Leaf Sequencing for Dynamic IMRT and Arc Modulated Radiation Therapy

In this section, we study the dynamic leaf sequencing problem and the single-arc leaf sequencing problem, i.e., the leaf sequencing problems for dynamic IMRT delivery and arc-modulated radiation therapy (AMRT) (see Section 1), respectively.

7.1 Dynamic Leaf Sequencing (DLS)

Dynamic IMRT delivery is different from static delivery in that the radiation beam is always on and the leaf positions change with respect to time. The **dynamic leaf sequencing (DLS)** problem seeks a set of trajectories of the left and right leaf tips of the MLC leaf pairs, i.e., the leaf tip position at any time, to deliver a given IM with minimum total delivery time. Unlike the SLS problem (see Section 3), we need to consider an additional MLC constraint, namely, the maximum leaf motion speed constraint (see Section 2.1), in the DLS problem. Note that the actual delivered radiation dose to an IM cell $A(i,j)$ in the IM A, is proportional to the amount of time the cell is exposed to the radiation beam, (i.e., the total time the cell is inside the leaf opening formed by the i-th MLC leaf pair) during the dynamic delivery.

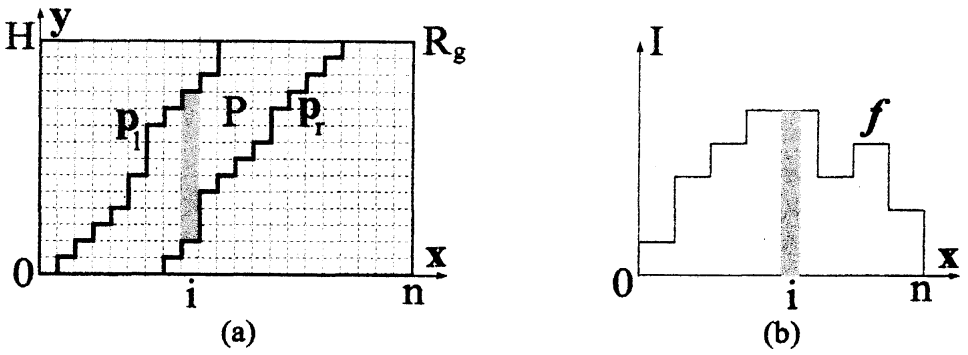


Figure 7: (a) Illustrating the CPP problem: p_l and p_r are two non-crossing c-steep paths with $c = 1$; the darkened area shows the vertical section on the i-th column of the region enclosed by the two paths, whose length corresponds to the amount of actually delivered radiation dose. (b) The input (intensity) function f for the CPP problem, which is defined on $\{1, 2, \ldots, n\}$; the darkened area shows the value of f at the i-th cell, which specifies the amount of prescribed dose at that cell.

Dynamic leaf sequencing (DLS) algorithms [13, 28, 45, 58, 70, 69] were given for *exactly* delivering an input IM with a short delivery time, under the maximum leaf motion speed constraint. Spirou and Chui's algorithm [69] computes the MLC leaf trajectories for exactly delivering the input IM. Their algorithm scans all IM cells from left to right, and when entering a new cell, produces the corresponding leaf positions. They proved that the algorithm optimizes the delivery time under the assumption that on each IM row, the corresponding MLC leaf tips always move from the left boundary of the leftmost non-zero cell to the right boundary of the rightmost non-zero cell. Kamath *et al.* [48] presented a DLS algorithm that can handle the interleaf motion MLC constraint, which occurs in some MLC systems such as Elekta and Siemens. The algorithm interleaves a scanning procedure similar to Spirou and Chui's [69] with detecting and rectifying possible violations of the interleaf motion constraint. The algorithm makes the same assumption on the starting and ending leaf tip positions as the algorithm in paper [69].

Chen *et al.* [26] pointed out that the assumption on the starting and ending leaf tip positions made in papers [48, 69] may sacrifice the optimality of the output DLS solution. For MLCs without the interleaf motion constraint, they modeled the DLS problem as the following **coupled path planning (CPP)** problem [26]. The problem is defined on a uniform grid R_g of size $n \times H$ for some integers n and H such that the length of each grid edge is one unit. A path on the plane is said to be *xy-monotone* if it is monotone with respect to both the x-axis and the y-axis. For an integer $c > 0$, an xy-monotone (rectilinear) path p along the edges of R_g is said to be *c-steep* if every vertical segment of p is of a length at least c (i.e., formed by c or more grid edges) and every horizontal segment has a unit length. The CPP problem is defined as follows: Given a non-negative function f defined on the integer set $\{1, 2, \ldots, n\}$ and positive integers c and Δ ($\Delta \leq H$), find two non-crossing c-steep paths on R_g, each starting at the bottom boundary and ending at the top boundary of R_g, such that the two paths, possibly with the bottom and top boundaries of R_g, enclose a (rectilinear) region P in R_g such that (1) for any column C_i of R_g, the vertical length of the intersection between C_i and P approximates the function value $f(i)$ (i.e., the value of f at i) within the given error bound Δ (see Figures 7(a)-7(b)), and (2) the total sum of errors on P is minimized. Note that in the CPP problem, f corresponds to the intensity profile function specifying one row of an IM (see Figure 7(b)), and the

two output paths specify the moving trajectories of the two MLC leaf tips, i.e., the leaf tip position (the x-coordinate) at any unit time (the y-coordinate). The maximum motion speed constraint of MLC is reflected in the c-steepness constraint on the paths. The CPP problem basically seeks to minimize the total approximation error of delivery within a given amount H units of delivery time.

Chen *et al.* [26] presented a novel approach based on interesting geometric observations for the CPP problem. The key idea is to formulate the problem as computing a shortest path in a weighted directed acyclic graph of $O(nH\Delta)$ vertices and $O(nH^2\Delta^2)$ edges. They exploited a set of geometric properties, such as certain domination relations among the vertices, to speed up the shortest path computation, resulting in an $O(nH\Delta)$ time CPP algorithm. One unique feature of their algorithm is it computes a tradeoff between the delivery time and the approximation error.

7.2 Arc Modulated Radiation Therapy

Arc modulated radiation therapy (AMRT) is a newly patented IMRT delivery technique [82]. In an AMRT delivery, the beam source rotates along a single arc path in 3-D, and for every θ degrees (usually $\theta = 10$), a prescribed intensity map (IM) of size $m \times n$ is delivered towards the target regions. A key problem in AMRT delivery is the so-called **single-arc leaf sequencing (SALS)** problem, which seeks to optimally convert a given set of K IMs (with $K = \frac{360}{\theta}$) into MLC leaf trajectories. The MLC is assumed to be free of the interleaf motion constraint, and thus each MLC leaf pair is treated independently.

Chen *et al* [26] pointed out the close relation between the SALS problem and the *constrained* CPP problem, which is a special version the CPP problem (see Section 7.1) where the starting and ending points of the sought paths are given as part of the input. Based on a graph modeling, it was shown [26] that the SALS problem can be computed by solving the set of all constrained CPP problem instances. More specifically, for all possible combinations of starting and ending leaf pair positions, the optimal solutions are sought for the corresponding constrained CPP instances (in total, there are $O(n^4)$ problems instances). They proposed a non-trivial graph transformation scheme that allows a *batch fashion* computation of the instances. Further, the shortest path computation is accelerated by exploiting the Monge property of the transformed graphs. Consequently, an $O(n^4\Delta + n^2H\Delta^2)$ time algorithm was given for solving the set of all constrained CPP problem instances. The final SALS algorithm takes $O(Kmn^2\Delta(n^2 + H\Delta))$ time.

8 Concluding Remarks

As pointed out by Webb [73], two of the most important developments in improving the quality of radiation therapy are (1) the introduction of intensity maps in the treatment planning stage to approximate the geometric shape of the planning target volume, and (2) the usage of multileaf collimators in the treatment delivery stage to geometrically shape the radiation beams. These developments are at the leading edge of the field and call for effective and efficient methods for treatment planning and delivery. In this paper, we have discussed various geometric and combinatorial problems that arise in the treatment planning and the treatment delivery phases of intensity-modulated radiation therapy.

There are many exciting and important computational problems in IMRT that are yet to be solved. We would like to discuss some of the open problems and research directions that are likely

to receive considerable attention in the future:

- For the general 3-D SLS problem discussed in Section 3.4.2, the existing approximation algorithm by Luan *et al.* [57] has an approximation ratio related to the complexity of the input IM and applies to MLCs without the interleaf motion constraint and the minimum separation constraint. However, what about the MLCs (e.g., Elekta and Siemens MLCs) that have one or both of these constraints? And is it possible to achieve a polynomial time algorithm with better approximation ratio, say a constant?

- Current field splitting algorithms (see Section 5) all target at the Varian MLCs which are not subject to the interleaf motion constraint and seek to minimize the total beam-on time. Can one extend the those field splitting algorithms to other types MLCs? Also, is it possible to develop efficient field splitting algorithms that minimize the total number of MLC-apertures?

- The shape approximation variant of the dose simplification problem still remains to be intriguing. The FBT algorithm (see Section 6) presented is a weakly polynomial time algorithm. Is it possible to develop strongly polynomial time algorithms for the FBT problem? Also, can we develop approximation algorithms for the shape approximation variant of the dose simplification problem, which is known to be NP-hard?

- For the dynamic leaf sequencing problem, the coupled path planning based algorithm (see Section 7.1) produces truly optimal solutions for MLCs without the interleaf motion constraint. Is is possible to extend this algorithm so that it can handle MLCs with the interleaf motion constraint?

- Recent development of radiation therapy calls for the combination of IMRT with *Image-guided radiation therapy* (IGRT). IGRT is a process of using various imaging technologies to locate a tumor target prior to a radiation therapy treatment. This process is aimed to improve the treatment accuracy by eliminating the need for large target margins which have traditionally been used to compensate for errors in localization. Advanced imaging techniques using CT, MRI, and ultrasound are applied to accurately delineate treatment target. One key problem arising in IGRT is the so-called **medical image registration** problem, which aims to transform one set of medical image data into others (e.g. data of the same patient taken at different points in time). The main difficulty here is to cope with elastic deformations of the body parts imaged. Existing medical image registration algorithms generally require lots of human interaction in order to find a satisfactory transformation, and do not work well for deformable cases. We believe the underlying problem could be formulated by some special matching and flow problems in graphs.

References

[1] A. Aggarwal, M.M. Klawe, S. Moran, P. Shor, and R. Wilber. Geometric Applications of a Matrix-Searching Algorithm. *Algorithmica*, 2:195–208, 1987.

[2] A. Aggarwal and J. Park. Notes on Searching in Multidimensional Monotone Arrays. In *Proc. 29th Annual IEEE Symp. on Foundations of Computer Science*, pages 497–512, 1988.

[3] R.K. Ahuja and H.W. Hamacher. A Network Flow Algorithm to Minimize Beam-on Time for Unconstrained Multileaf Collimator Problems in Cancer Radiation Therapy. *Networks*, 45:36–41, 2005.

[4] E.M. Arkin, J.S.B. Mitchell, and G. Narasimhan. Resource-Constrained Geometric Network Optimization. In *Proc. 14th ACM Symp. on Computational Geometry*, pages 307–316, 1998.

[5] S. Arora. Polynomial-Time Approximation Schemes for Euclidean TSP and Other Geometric Problems. *J. of the ACM*, 45(5):753–782, 1998.

[6] A. Arya and H. Ramesh. A 2.5 Factor Approximation Algorithm for the k-MST Problem. *Information Processing Letters*, 65:117–118, 1998.

[7] D. Baatar, M. Ehrgott, H.W. Hamacher, and G.J. Woeginger. Decomposition of Integer Matrices and Multileaf Collimator Sequencing. *Discrete Applied Mathematics*, 152:6–34, 2005.

[8] N. Bansal, D. Coppersmith, and B. Schieber. Minimizing Setup and Beam-on Times in Radiation Therapy. In *Proc. 9th International Workshop on Approximation Algorithms for Combinatorial Optimization Problems*, pages 27–38, 2006.

[9] W. Bär, M. Alber, and F. Nsslin. A Variable Fluence Step Clustering and Segmentation Algorithm for Step and Shoot IMRT. *Phys. Med. Biol.*, 46:1997–2007, 2001.

[10] N. Boland, H.W. Hamacher, and F. Lenzen. Minimizing Beam-on Time in Cancer Radiation Treatment Using Multileaf Collimators. *Networks*, 43(4):226–240, 2004.

[11] T.R. Bortfeld, A.L. Boyer, W. Schlegel, D.L. Kahler, and T.L. Waldron. Experimental Verification of Multileaf Conformal Radiotherapy. In A.R. Hounsell, J.M. Wilkinson, and P.C. Williams, editors, *Proc. 11th Int. Conf. on the Use of Computers in Radiation Therapy*, pages 180–181, 1994.

[12] T.R. Bortfeld, A.L. Boyer, W. Schlegel, D.L. Kahler, and T.L. Waldron. Realization and Verification of Three-Dimensional Conformal Radiotherapy with Modulated Fields. *Int. J. Radiat. Oncol. Biol. Phys.*, 30:899–908, 1994.

[13] 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.*, 28:723–730, 1994.

[14] T.R. Bortfeld, J. Stein, K. Preiser, and K. Hartwig. Intensity Modulation for Optimized Conformal Therapy. In *Proc. Symp. Principles and Practice of 3-D Radiation Treatment Planning*, 1996.

[15] A.L. Boyer. Use of MLC for Intensity Modulation. *Med. Phys.*, 21:1007, 1994.

[16] A.L. Boyer, T.R. Bortfeld, D.L. Kahler, and T.J. Waldron. MLC Modulation of X-ray Beams in Discrete Steps. In *Proc. 11th Int. Conf. on the Use of Computers in Radiation Therapy*, pages 178–179, 1994.

[17] A.L. Boyer and C.X. Yu. Delivery of Intensity-Modulated Radiation Therapy Using Dynamic Multileaf Collimator. In *Seminar in Radiat. Oncol.*, volume 9(2), 1999.

[18] R. Burkard. Open Problem Session. In *Oberwolfach Conference on Combinatorial Optimization*, November 24-29, 2002.

[19] C. Chekuri, M. Mydlarz, and F.B. Shepherd. Multicommodity Demand Flow in a Tree and Packing Integer Problem. In *Proc. of 30th Int. Colloquium on Automata, Languages and Programming*, pages 410–425, 2003.

[20] D.Z. Chen, M.A. Healy, C. Wang, and B. Xu. Geometric Algorithms for the Constrained 1-D K-Means Clustering Problems and IMRT Applications. In *Proc. of 1st Int. Frontiers of Algorithmics Workshop (FAW)*, pages 1–13, 2007.

[21] D.Z. Chen, X.S. Hu, S. Luan, E. Misiolek, and C. Wang. Shape Rectangularization Problems in Intensity-Modulated Radiation Therapy. In *Proc. 12th Annual Int. Symp. on Algorithms and Computation*, pages 701–711, 2006.

[22] D.Z. Chen, X.S. Hu, S. Luan, S.A. Naqvi, C. Wang, and C.X. Yu. Generalized Geometric Approaches for Leaf Sequencing Problems in Radiation Therapy. *International Journal of Computational Geometry and Applications*, 16(2-3):175–204, 2006.

[23] D.Z. Chen, X.S. Hu, S. Luan, C. Wang, and X. Wu. Geometric Algorithms for Static Leaf Sequencing Problems in Radiation Therapy. *International Journal of Computational Geometry and Applications*, 14(5):311–339, 2004.

[24] D.Z. Chen, X.S. Hu, S. Luan, C. Wang, and X. Wu. Mountain Reduction, Block Matching, and Applications in Intensity-Modulated Radiation Therapy. *International Journal of Computational Geometry and Applications*, 18(1-2):63–106, 2008.

[25] D.Z. Chen, X.S. Hu, S. Luan, X. Wu, and C.X. Yu. Optimal Terrain Construction Problems and Applications in Intensity-Modulated Radiation Therapy. *Algorithmica*, 42:265–288, 2005.

[26] D.Z. Chen, S. Luan, and C. Wang. Coupled Path Planning, Region Optimization, and Applications in Intensity-Modulated Radiation Therapy. In *Proc. 16th Annual European Symp. on Algorithms*, pages 271–283, 2008.

[27] D.Z. Chen and C. Wang. Field Splitting Problems in Intensity-Modulated Radiation Therapy. In *Proc. 12th Annual Int. Symp. on Algorithms and Computation*, pages 690–700, 2006.

[28] 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.

[29] D.J. Convery and S. Webb. Generation of Discrete Beam-Intensity Modulation by Dynamic Multileaf Collimation under Minimum Leaf Separation Constraints. *Phys. Med. Biol.*, 43:2521–2538, 1998.

[30] J. Dai and Y. Zhu. Minimizing the Number of Segments in a Delivery Sequence for Intensity-Modulated Radiation Therapy with Multileaf Collimator. *Med. Phys.*, 28(10):2113–2120, 2001.

[31] M.L.P. Dirkx, B.J.M. Heijmen, and J.P.C. van Santvoort. Trajectory Calculation for Dynamic Multileaf Collimation to Realize Optimized Fluence Profiles. *Phys: Med. Biol.*, 43:1171–1184, 1998.

[32] N. Dogan, L.B. Leybovich, A. Sethi, and B. Emami. Automatic Feathering of Split Fields for Step-and-Shoot Intensity Modulated Radiation Therapy. *Phys. Med. Biol.*, 48:1133–1140, 2003.

[33] M.N. Du, C.X. Yu, J.W. Wong, M. Symons, D. Yan, R.C. Matter, and A. Martinez. A Multi-Leaf Collimator Prescription Preparation System for Conventional Radiotherapy. *Int. J. Radiat. Oncol. Biol. Phys.*, 30(3):707–714, 1994.

[34] K. Engel. A New Algorithm for Optimal Multileaf Collimator Field Segmentation. *Discrete Applied Mathematics*, 152(1-3):35–51, 2005.

[35] P.M. Evans, V.N. Hansen, and W. Swindell. The Optimum Intensities for Multiple Static Collimator Field Compensation. *Med. Phys.*, 24(7):1147–1156, 1997.

[36] M.C. Ferris, R.R. Meyer, and W. D'Souza. Radiation Treatment Planning: Mixed Integer Programming Formulations and Approaches. Optimization Technical Report 02-08, Computer Sciences Department, University of Wisconsin, Madison, Wisconsin, 2002.

[37] M. Fischetti, H.W. Hamacher, K. Jørnsten, and F. Maffioli. Weighted k-Cardinality Trees: Complexity and Polyhedral Structure. *Networks*, 24:11–21, 1994.

[38] M.R. Garey and D.S. Johnson. *Computers and Intractability: A Guide to the Theory of NP-Completeness.* Freeman, San Francisco, CA, 1979.

[39] N. Garg. A 3-Approximation for the Minimum Tree Spanning k Vertices. In *Proc. 37th Annual IEEE Symp. on Foundations of Comp. Sci.*, pages 302–309, 1996.

[40] R.W. Hill, B.H. Curran, J.P. Strait, and M.P. Carol. Delivery of Intensity Modulated Radiation Therapy Using Computer Controlled Multileaf Collimators with the CORVUS Inverse Treatment Planning System. In *Proc. 12th Int. Conf. on Computers and Radiotherapy*, pages 394–397, 1997.

[41] T.W. Holmes, A.R. Bleier, M.P. Carol, B.H. Curran, A.A. Kania, R.J. Lalonde, L.S. Larson, and E.S. Sternick. The Effect of MLC Leakage on the Calculation and Delivery of Intensity Modulated Radiation Therapy. In *Proc. 12th Int. Conf. on the Use of Computers in Radiation Therapy*, pages 346–349, 1997.

[42] L. Hong, A. Kaled, C. Chui, T. Losasso, M. Hunt, S. Spirou, J. Yang, H. Amols, C. Ling, Z. Fuks, and S. Leibel. IMRT of Large Fields: Whole-Abdomen Irradiation. *Int. J. Radiat. Oncol. Biol. Phys.*, 54:278–289, 2002.

[43] C.A.J. Hurkens and A. Schrijver. On the Size of Systems of Sets Every t of Which Have an SDR, with an Application to the Worst-case Ratio of Heuristics for Packing Problem. *SIAM Journal on Discrete Mathematics*, 2:68–72, 1989.

[44] T.J. Jordan and P.C. Williams. The Design and Performance Characteristics of a Multileaf Collimator. *Phys. Med. Biol.*, 39:231–251, 1994.

[45] P. Kallman, B. Lind, and A. Brahme. Shaping of Arbitrary Dose Distribution by Dynamic Multileaf Collimation. *Phys. Med. Biol.*, 33:1291–1300, 1988.

[46] S. Kamath, S. Sahni, J. Li, J. Palta, and S. Ranka. Leaf Sequencing Algorithms for Segmented Multileaf Collimation. *Phys. Med. Biol.*, 48(3):307–324, 2003.

[47] S. Kamath, S. Sahni, J. Li, J. Palta, and S. Ranka. A Generalized Field Splitting Algorithm for Optimal IMRT Delivery Efficiency. *The 47th Annual Meeting and Technical Exhibition of the American Association of Physicists in Medicine (AAPM)*, 2005. Also, *Med. Phys.*, 32(6):1890, 2005.

[48] S. Kamath, S. Sahni, J. Palta, and S. Ranka. Algorithms for Optimal Sequencing of Dynamic Multileaf Collimators. *Phys. Med. Biol.*, 49(1):33–54, 2004.

[49] S. Kamath, S. Sahni, J. Palta, S. Ranka, and J. Li. Optimal Leaf Sequencing with Elimination of Tongue-and-Groove Underdosage. *Phys. Med. Biol.*, 49(3):N7–N19, 2004.

[50] S. Kamath, S. Sahni, S. Ranka, J. Li, and J. Palta. Optimal Field Splitting for Large Intensity-Modulated Fields. *Med. Phys.*, 31(12):3314–3323, 2004.

[51] H. Kobayashi, S. Sakuma, O. Kaii, and H. Yogo. Computer-Assisted Conformation Radiotherapy with a Variable Thickness Multi-Leaf Filter. *Int. J. Radiat. Oncol. Biol. Phys.*, 16(6):1631–1635, 1989.

[52] D.D. Leavitt, F.A. Gibbs, M.P. Heilbrun, J.H. Moeller, and G.A. Takach. Dynamic Field Shaping to Optimize Stereotaxic Radiosurgery. *Int. J. Radiat. Oncol. Biol. Phys.*, 21(5):1247–1255, 1991.

[53] D.D. Leavitt, M. Martin, J.H. Moeller, and W.L. Lee. Dynamic Wedge Field Techniques through Computer-Controlled Collimator or Motion and Dose Delivery. *Med. Phys.*, 17(1):87–91, 1990.

[54] F. Lenzen. An Integer Programming Approach to the Multileaf Collimator Problem. Master's thesis, University of Kaiserslautern, June 2000.

[55] R. Liu and S. Ntafos. On Decomposing Polygons into Uniformly Monotone Parts. *Information Processing Letters*, 27:85–89, 1988.

[56] S. Luan, D.Z. Chen, L. Zhang, X. Wu, and C.X. Yu. An Optimal Algorithm for Computing Configuration Options of One-dimensional Intensity Modulated Beams. *Phys. Med. Biol.*, 48(15):2321–2338, 2003.

[57] S. Luan, J. Saia, and M. Young. Approximation Algorithms for Minimizing Segments in Radiation Therapy. *Information Processing Letters*, 101:239–244, 2007.

[58] L. Ma, A. Boyer, L. Xing, and C.M. Ma. An Optimized Leaf-Setting Algorithm for Beam Intensity Modulation Using Dynamic Multileaf Collimators. *Phys. Med. Biol.*, 43:1629–1643, 2004.

[59] J.S.B. Mitchell. Guillotine Subdivisions Approximate Polygonal Subdivisions: A Simple New Method for the Geometric k-MST Problem. In *Proc. 7th Annual ACM-SIAM Symp. on Discrete Algorithms*, pages 402–408, 1996.

[60] J.S.B. Mitchell. Guillotine Subdivisions Approximate Polygonal Subdivisions: Part II – A Simple Polynomial-Time Approximation Scheme for Geometric TSP, k-MST, and Related Problems. *SIAM J. Comput.*, 28(4):1298–1309, 1999.

[61] R. Mohan, H. Liu, I. Turesson, T. Mackie, and M. Parliament. Intensity-Modulated Radiotherapy: Current Status and Issues of Interest. *Int. J. Radiat. Oncol. Biol. Phys.*, 53:1088–1089, 2002.

[62] G. Monge. Déblai et Remblai. In *Mémories de l'Académie des Sciences*, Paris, 1781.

[63] L.A. Nedzi, H.M. Kooy, E. Alexander, G.K. Svensson, and J.S. Loeffler. Dynamic Field Shaping for Stereotaxic Radiosurgery: A Modeling Study. *Int. J. Radiat. Oncol. Biol. Phys.*, 25(5):859–869, 1993.

[64] G.L. Nemhauser and L.A. Wolsey. *Integer and Combinatorial Optimization.* John Wiley, 1988.

[65] W. Que, J. Kung, and J. Dai. "tongue-and-Groove" Effect in Intensity Modulated Radiotherapy with Static Multileaf Collimator Fields. *Phys. Med. Biol.*, 49:399–405, 2004.

[66] R. Ravi, R. Sundaram, M.V. Marathe, D.J. Rosenkrantz, and S.S. Ravi. Spanning Trees Short and Small. In *Proc. 5th Annual ACM-SIAM Symp. on Discrete Algorithms*, pages 546–555, 1994.

[67] A. Schweikard, R.Z. Tombropoulos, and J.R. Adler. Robotic Radiosurgery with Beams of Adaptable Shapes. In Springer, editor, *Proc. 1st Int. Conf. on Computer Vision, Virtual Reality and Robotics in Medicine, Lecture Notes in Computer Science*, volume 905, pages 138–149, 1995.

[68] R.A.C. Siochi. Minimizing Static Intensity Modulation Delivery Time Using an Intensity Solid Paradigm. *Int J. Radiat. Oncol. Biol. Phys.*, 43(3):671–680, 1999.

[69] S.V. Spirou and C.S. Chui. Generation of Arbitrary Intensity Profiles by Dynamic Jaws or Multileaf Collimators. *Med. Phys.*, 21:1031–1041, 1994.

[70] J. Stein, T. Bortfeld, B. Dorschel, and W. Schlegel. Dynamic X-ray Compensation for Conformal Radiotherapy by Means of Multileaf Collimations. *Radiother. Oncol.*, 32:163–173, 1994.

[71] R. Svensson, P. Kallman, and A. Brahme. An Analytical Solution for the Dynamic Control of Multileaf Collimation. *Phys. Med. Biol.*, 39:37–61, 1994.

[72] S. Webb. *The Physics of Three-Dimensional Radiation Therapy.* Bristol, Institute of Physics Publishing, 1993.

[73] S. Webb. *The Physics of Conformal Radiotherapy — Advances in Technology.* Bristol, Institute of Physics Publishing, 1997.

[74] S. Webb. Configuration Options for Intensity-Modulated Radiation Therapy Using Multiple Static Fields Shaped by a Multileaf Collimator. *Phys. Med. Biol.*, 43:241–260, 1998.

[75] Q. Wu, M. Arnfield, S. Tong, Y. Wu, and R. Mohan. Dynamic Splitting of Large Intensity-Modulated Fields. *Phys. Med. Biol.*, 45:1731–1740, 2000.

[76] X. Wu. Optimal Quantization by Matrix Searching. *Journal of Algorithms*, 12:663–673, 1991.

[77] X. Wu. Efficient Algorithms for Intensity Map Splitting Problems in Radiation Therapy. In *Proc. 11th Annual Int. Computing and Combinatorics Conference*, pages 504–513, 2005.

[78] Y. Wu, D. Yan, M.B. Sharpe, B. Miller, and J.W. Wong. Implementing Multiple Static Field Delivery for Intensity Modulated Beams. *Med. Phys.*, 28:2188–2197, 2001.

[79] P. Xia and L.J. Verhey. MLC Leaf Sequencing Algorithm for Intensity Modulated Beams with Multiple Static Segments. *Med. Phys.*, 25:1424–1434, 1998.

[80] C.X. Yu. Intensity-Modulated Arc Therapy with Dynamic Multileaf Collimation: An Alternative to Tomotherapy. *Phys. Med. Biol.*, 40:1435–1449, 1995.

[81] C.X. Yu. Design Considerations of the Sides of the Multileaf Collimator. *Phys. Med. Biol.*, 43(5):1335–1342, 1998.

[82] C.X. Yu, S. Luan, C. Wang, D.Z. Chen, and M. Earl. Single-arc Dose Painting for Precision Radiation Therapy. US patent application, University of Maryland, 2008.

[83] C.X. Yu, D. Yan, M.N. Du, S. Zhou, and L.J. Verhey. Optimization of Leaf Positioning When Shaping a Radiation Field with a Multileaf Collimator. *Phys. Med. Biol.*, 40(2):305–308, 1995.

[84] A. Zelikovsky and D. Lozevanu. Minimal and Bounded Trees. In *Tezele Cong. XVIII Acad. Romano-Americane*, pages 25–26. Kishinev, 1993.