Approximation Algorithms for Minimizing Segments in Radiation Therapy

Shuang Luan*   Jared Saia*   Maxwell Young*

Abstract

Intensity modulated radiation therapy (IMRT) is one of the most effective modalities for modern cancer treatment. The key to successful IMRT treatment hinges on the delivery of a two-dimensional discrete radiation intensity matrix using a device called a multileaf collimator (MLC). Mathematically, the delivery of an intensity matrix using an MLC can be viewed as the problem of representing a non-negative integral matrix (i.e. the intensity matrix) by a linear combination of certain special non-negative integral matrices called segments, where each such segment corresponds to one of the allowed states of the MLC. The problem of representing the intensity matrix with the minimum number of segments is known to be NP-complete. In this paper, we present two approximation algorithms for this matrix representation problem. To the best of our knowledge, these are the first algorithms to achieve non-trivial performance guarantees for multi-row intensity matrices.

Keywords: approximation algorithms; intensity modulated radiation therapy; multileaf collimator; segmentation

1   Introduction

Intensity modulated radiation therapy (IMRT) is an advanced modality of modern cancer treatment aiming to deliver a high radiation dose to a cancerous region while minimizing damage to the surrounding healthy tissue. Successful implementation of IMRT technology hinges on the delivery of a two-dimensional integral radiation intensity matrix T using a device called the multileaf collimator (MLC) [3, 8].

The MLC contains two banks of metal leaves and is able to position these leaves to prevent radiation from being delivered to certain portions of T. A pair of leaves is associated with each row of T. One leaf is located to the left of a row and may slide inwards to cover entries from left to right. Similarly, another leaf is located to the right of a row and may slide inwards to cover indices from right to left. The two leaves in any particular row cannot overlap. A treatment plan consists of discrete steps. The leaves of the collimator are first set into appropriate positions forming an aperture. An amount of radiation, d, is then passed through the apertures onto the exposed portions of T. All entries of T that are exposed by the MLC aperture decrease by d. The radiation is then turned off and the leaves are then repositioned to form different apertures through which another amount of radiation can be delivered. This procedure is repeated until all entries of T are 0 and the treatment is complete.

The positions of the leaves at a step in the treatment plan can be represented by a matrix with 0s in those entries covered by leaves and 1s in those entries left uncovered. We call such a matrix a *segment*. Setting leaf positions in each step of the treatment plan requires time. Minimizing the number of segments reduces treatment time which helps to avoid patient movement, which can adversely affect the treatment quality, as well as reduce the cost of the procedure.

*Department of Computer Science, University of New Mexico, NM, USA; email: sluan, saia, young@cs.unm.edu, contact author: young@cs.unm.edu

Here we concern ourselves with minimizing the number of segments required to complete the treatment specified by an intensity matrix.

1.1 Definitions

We formally define a segment as given in [9]:

Definition 1. *A segment is an $m \times n$ matrix S with integers l_i, r_i for $1 \leq i \leq m$ satisfying the following two criteria:*

$$
\begin{aligned}
l_i &\leq r_i + 1 & \text{for all } i,\ 1 \leq i \leq m \\
S[i][j] &= \begin{cases} 1 & \text{if } l_i \leq j \leq r_i \\ 0 & \text{otherwise} \end{cases} & \text{for all } i,\ 1 \leq i \leq m,\ \text{for all } j,\ 1 \leq j \leq n
\end{aligned}
$$

Thus, a segment is a 0-1 matrix where the values of 1 appear as a single block in any particular row. This definition corresponds to the setting of the leaves in the MLC. The first criteria ensures that the left and right leaves of row i do not overlap. The second criteria ensures that at each row i, values of 1 are listed consecutively in a single block.

Definition 2. *An MLC-aperture is a pair (S, c) where S is a segment and c is a positive integer which we call the beam-on time.*

Definition 3. *Consider an intensity matrix T. A segmentation of T is a multiset X of MLC-apertures (S_i, c_i) such that $T = \sum_{i=1}^{|X|} c_i S_i$. We say that T has segmentation of length $|X|$.*

Example 1. *Let $T = \begin{pmatrix} 2 & 3 & 1 \\ 4 & 4 & 2 \\ 2 & 1 & 7 \end{pmatrix}$ and let:*

$$
S_1 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}, S_2 = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 1 & 1 \\ 1 & 0 & 0 \end{pmatrix}, S_3 = \begin{pmatrix} 0 & 1 & 1 \\ 1 & 1 & 0 \\ 0 & 1 & 1 \end{pmatrix}, S_4 = \begin{pmatrix} 0 & 0 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}
$$

Then a segmentation of T is $X = \{(S_1, 5), (S_2, 2), (S_3, 1), (S_4, 1)\}$.

1.2 Related Work

The problem of minimizing the segmentation length for a $1 \times n$ intensity matrix is known to be NP-complete [4, 1]. An unpublished proof by Woeginger demonstrates that the $1 \times n$ case is NP-complete in the strong sense [9]. The literature contains several leaf-sequencing algorithms which attempt to reduce the number segments through the use of various heuristic techniques [11, 6, 7, 10]. Bansal *et al* [2] show that the single row version of the problem is APX-complete. They provide a 24/13-approximation algorithm for the single row problem and give some better approximations for more constrained versions. Recent work by Collins *et al* [5] shows the single column version of the problem is NP-complete and provides some non-trivial lower bounds given certain constraints; however, in the most general case, the best approximation currently known is still the trivial $\min\{\lceil \log h \rceil, n\}/\lceil \log n \rceil$.

1.3 Results

For the family of $m \times n$ intensity matrices, we provide two approximation algorithms. The first is a $(\lceil \log h \rceil + 1)$-approximation where $h > 0$ is the largest entry in the intensity matrix. 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. These two algorithms

have a running time complexity of $O(mn \log h)$ and $O(mn \log D)$, respectively. It is certainly possible for D to be significantly less than h for an intensity matrix, in which case our second algorithm guarantees an improved approximation ratio[1]. To the best of our knowledge, these are the first algorithms for the general $m \times n$ case with non-trivial performance guarantees.

2 Approximation Algorithms for an $m \times n$ Intensity Matrix

To simplify the following discussion, we introduce some notion. Let T_b denote the intensity matrix T where the entries are represented as binary numbers. Let P_i be the 0-1 matrix where the ij^{th} entry is the jth digit in the binary representation of the ij^{th} entry in T_b. A *marker* b_i is defined at a cell boundary of a row $\mathcal{R}$ of an intensity matrix if $\mathcal{R}[i] \neq \mathcal{R}[i+1]$ for $1 \leq i \leq n-1$. Also, a marker exists to the left of $\mathcal{R}[1]$ and to the right of $\mathcal{R}[n]$. Therefore, we have markers $b_1, ..., b_k$ for some k, $2 \leq k \leq n+1$, positioned from left to right in $\mathcal{R}$. Let T be an $m \times n$ intensity matrix indexed starting at 1. Define D_j, for $j = 1, ..., m$, to be the following:

$$D_j = \max \begin{cases} |T[j][i] - T[j][i-1]| & \text{for } 2 \leq i \leq n \\ T[j][1] \\ T[j][n] \end{cases}$$

and now define D to be the following:

$$D = \max_{1 \leq j \leq m} \{D_j, 1\}$$

Therefore, D is the maximum element of a set containing 1) all absolute differences between any two consecutive row entries over all rows j, 2) the first entry of each row, 3) the last entry of each row and 4) the value 1. Finally, for ease of notation, throughout the paper, unless explicitly stated otherwise, we will simply use $\log k$ to mean $\lceil \log k \rceil$ and all logarithms are base 2.

Before we outline the approximation algorithm for the segmentation of an $m \times n$ intensity matrix, we need the following result given in Lemma 1. To the best of our knowledge, this well-known result has never been published and we include it here only for completeness[2].

Lemma 1. *There exists an approximation algorithm for the segmentation of a single row intensity matrix $\mathcal{R}$ that provides a segmentation X of length no more than $2 \cdot OPT$. Furthermore, each MLC-aperture $d \in X$ has a beam-on time c_d such that $c_d \leq D$.*

Proof. Consider that we have markers $b_1, ..., b_k$ for some k, $2 \leq k \leq n+1$, positioned from left to right in $\mathcal{R}$ as described above. In each step, OPT can remove at most two such markers per MLC-aperture. Once all markers are removed, the collection of these MLC-apertures is a segmentation for $\mathcal{R}$. However, consider a simple algorithm which creates a segmentation X for $\mathcal{R}$ by creating an MLC-aperture for $\mathcal{R}$ in the following manner at each step i:

1. Let h be the largest value that appears in $\mathcal{R}$. Set the left leaf to cover all cells up to the marker b_l immediately left of h and set the right leaf to cover all cells up to the marker b_{l+1} immediately right of h (if h appears more than once between different pairs of markers, then select the first such instance going left to right along the row). Let S_i denote the $1 \times n$ segment corresponding to this leaf setting.

2. Let c_1 be the value of the cell immediately left of h and let c_2 be the value immediately right of h (both of which are currently covered by the leaves). Without loss of generality, let $c_1 \geq c_2$. Add the MLC-aperture $(S_i, h - c_1)$ to X.

---

[1]Note that $D \leq h$ always.

[2]An approximation algorithm given in [2] gives an identical approximation ratio for the single row problem; however, it does not say anything about D.

3

3. Remove marker b_l.

At each step, the above algorithm removes at least one marker. Therefore, this algorithm provides a segmentation of length no more than $2 \cdot OPT$ for a single row intensity matrix. As will be important later on for ALG 2, step (2) guarantees that the beam-on time during the step-and-shoot scheme never exceeds D. □

2.1 $(\log h + 1)$-Approximation Algorithm

We now present pseudocode for the $(\log h + 1)$-approximation algorithm and proceed to prove the approximation ratio.

ALG 1 : $(\log h + 1)$-*approximation*

1: Construct the P_i matrices for T for $i = 0, ..., \log h$.
2: **for** $(i = 0$ to $\log h)$ **do**
3:    Obtain an optimal segmentation, Y_i, for P_i.
4: Output $Z = \bigcup_{i=0}^{\log h} Y_i$

The beam-on time in the segmentation Y_i is 2^i. The MLC-apertures of Z clearly satisfy T and it is easy to see that ALG 2 has a run-time complexity of $O(mn \log h)$. We now prove the approximation ratio.

Lemma 2. *Over all rows in T, let ρ denote the maximum number of markers present in any one row. Then, for all i, the number of markers in any row of P_i is at most ρ.*

Proof. There is a marker in T whenever there is a difference in the binary representation between neighbors in T_b. Since the corresponding entry in P_i is equal to the ith digit in the corresponding entry in T_b, P_i can have at most the same number of differences between neighbors as the corresponding row in T and, therefore, at most the same number of markers. □

Lemma 3. *An optimal segmentation of P_i has length at most $\rho/2$, for all $i = 0, ..., \log h$ where ρ is defined as in Lemma 2.*

Proof. A segmentation of P_i is obtained by positioning the leaves of each row such that only the left-most block of consecutive 1s entries are exposed in each row at each step of the treatment. This block is satisfied by applying this segment with beam-on time 1. By moving from left to right in this fashion along each row concurrently, a segmentation of P_i is obtained. For some fixed i, over all rows of P_i, let ρ_i denote the maximum number of markers present in any row of P_i. There are markers on either side of each consecutive block of 1s in addition to a marker at the endpoints of a row; therefore, the number of blocks of 1s in any row is at most $\rho_i/2$. Finally, by Lemma 2, $\rho_i \leq \rho$. □

Lemma 4. *OPT requires a segmentation of length at least $\rho/2$ to satisfy T.*

Proof. Clearly, any segmentation of T must satisfy every row of T. The result then follows from Lemma 1 since OPT can remove at most 2 markers per MLC-aperture. □

Theorem 1. ALG 1 *is a $(\log h + 1)$-approximation algorithm.*

Proof. Lemma 3 and Lemma 4 imply that the size of the segmentation required to satisfy any P_i is no more than OPT. The approximation ratio follows immediately from noting that the number of P_i matrices is $(\log h) + 1$. □

2.2 $2(\log D + 1)$-**Approximation Algorithm**

The pseudocode for our second algorithm, ALG 2, is given below. ALG 2 obtains a segmentation of length no more than $2(\log D + 1)OPT$ and has a run-time complexity of $O(mn \log D)$. We now prove both correctness and the approximation ratio.

ALG 2: $2(\log D + 1)$-*approximation*

1: Let X be the set of MLC-apertures obtained as follows:

- For each row $j = 1, ..., m$, let X_j denote the segmentation of row j obtained from Lemma 1.

- Each segment of an MLC-aperture in X_j is a $1 \times n$ matrix. Convert each segment into an $m \times n$ matrix by doing the following. Make the $1 \times n$ matrix the jth row of an $m \times n$ matrix and set all other entries of the $m \times n$ matrix to 0.

- $X \leftarrow \bigcup_{j=1 \text{ to } m} X_j$.

2: Let X' be the set of MLC-apertures obtained from X as follows:

- For each MLC-aperture $(S, c) \in X$, add $(S, 2^k)$ to X' if and only if $b_k = 1$ where $b_{\lfloor \lg(c) \rfloor} \cdots b_0$ is the binary representation of c.

3: Let Z_{dj} be the sets of MLC-apertures obtained from X' as follows:

- Define 2^q be the largest beam-on time present over all MLC-apertures in X'. For all $d = 0, ..., q$ and all $j = 1, ..., m$, let Z_{dj} be the set of all MLC-apertures in X' of the form $(S, 2^d)$ where S is an $m \times n$ matrix with non-zero entries only in the jth row.

   for ($d = 0$ to q) **do**
     for ($g = 1$ to $\max_{1 \le j \le n}\{|Z_{dj}|\}$) **do**
           For $j = 0, ..., n$, remove one MLC-aperture d_j from each non-empty Z_{dj}.
           Denote the sum of the segments removed in the previous step
           by the $m \times n$ matrix S'_{gd}.
     if (S'_{gd} contains at least one non-zero entry) **then**
         $Z = Z \cup \{(S'_{gd}, d)\}$.
4: Output Z.

Theorem 2. *Z is a segmentation of T.*

Proof. Recall that X is the collection of MLC-apertures from the segmentations of each row where the $1 \times n$ segments of each MLC-aperture have been converted into an $m \times n$ segment. Clearly, X is a segmentation of T.

An MLC-aperture (S, c) in X may be decomposed into several MLC-apertures in X' that share the same segment S and possess levels that are a power of two and all sum to c. Since X is a segmentation of T, X' is a segmentation of T. Consequently, by construction, the set $\bigcup_{d,j} Z_{dj}$ is also a segmentation of T.

However, the segmentation $\bigcup_{d,j} Z_{dj}$ is wasteful. Each segment of an MLC-aperture in $\bigcup_{d,j} Z_{dj}$ represents a leaf setting where only one row is receiving radiation per discrete step. ALG 2 achieves a more efficient segmentation through Step 3 which we now describe. Let us fix d and pick at most one $m \times n$ S_{dj} segment per j value. Denote the sum of these segments by the segment S'_{dg}. We are able to create $\max_{1 \le j \le n}\{|Z_{dj}|\}$ non-zero segments (there exists at least

5

one non-zero entry in each segment) in this manner for a fixed d. Each pair (S'_{dg}, d) is a valid MLC-aperture. Doing this over all d then gives a segmentation of T where *at least* one row is receiving radiation per step. □

Theorem 3. ALG 2 *returns a segmentation for T of length at most $2(\lg D + 1)OPT$.*

Proof. The length of the segmentation created by ALG 2 is upper-bounded by $(q + 1)t$ where $t = \max\limits_{1 \le j \le n, 0 \le d \le q} \{|Z_{dj}|\}$. We now bound q and t:

Bounding q: Recall that 2^q is the largest beam-on time present over all MLC-apertures in X'. The row approximation algorithm of Lemma 1 guarantees that any beam-on time of an MLC-aperture in X is no greater than D. By the construction of X', we have $2^q \le D$ and so $q \le \lceil \lg D \rceil$.

Bounding t: In Step 2, an MLC-aperture $(S, c) \in X_j$ can contribute at most one MLC-aperture $(S, 2^k)$ to X' for some fixed k, $0 \le k \le \lfloor \lg(c) \rfloor$. Therefore, $|Z_{dj}| \le |X_j|$ for a fixed j and all $d = 0, ..., q$. Consequently, $t \le \max_j |X_j|$. The row approximation algorithm of Lemma 1 used to achieve the segmentation X_j in Step 1 guarantees that $|X_j| \le 2 \cdot OPT$ for all $j = 1, ..., n$. Therefore, for all d, $t \le 2 \cdot OPT$.

Therefore, the segmentation created by ALG 2 has length at most $2(\lg D + 1)OPT$. □

Example 2. *We illustrate how* ALG 2 *executes on the intensity matrix T from Example 1. If we let:*

$$S_1 = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad S_2 = \begin{pmatrix} 1 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \quad S_3 = \begin{pmatrix} 1 & 1 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \quad S_4 = \begin{pmatrix} 0 & 0 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

$$S_5 = \begin{pmatrix} 0 & 0 & 0 \\ 1 & 1 & 1 \\ 0 & 0 & 0 \end{pmatrix} \quad S_6 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad S_7 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix} \quad S_8 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 1 & 1 & 1 \end{pmatrix}$$

then we have that $X = X_1 \cup X_2 \cup X_3$ where:

$$X_1 = \{(S_1, 1), (S_2, 1), (S_3, 1)\}, \quad X_2 = \{(S_4, 2), (S_5, 2)\}, \quad X_3 = \{(S_6, 6), (S_7, 1), (S_8, 1)\}.$$

and now we obtain X':

$$X' = \{(S_1, 2^0), (S_2, 2^0), (S_3, 2^0), (S_4, 2^1), (S_5, 2^1), (S_6, 2^1), (S_6, 2^2), (S_7, 2^0), (S_8, 2^0)\}$$

We then have the following non-empty Z_{dj} sets:

$$Z_{0,1} = \{(S_1, 2^0), (S_2, 2^0), (S_3, 2^0)\}, \quad Z_{1,2} = \{(S_4, 2^1), (S_5, 2^1)\}$$

$$Z_{0,3} = \{(S_7, 2^0), (S_8, 2^0)\}, \quad Z_{1,3} = \{(S_6, 2^1)\}, \quad Z_{2,3} = \{(S_6, 2^2)\}$$

We now construct the segmentation created in the for-loop of ALG 2:

- *Step $d = 0$:*

  - *Step $g = 1$: We choose $(S_1, 2^0) \in Z_{0,1}$ and $(S_7, 2^0) \in Z_{0,3}$. $S'_{1,0} = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix}$.*

  - *Step $g = 2$: We choose $(S_2, 2^0) \in Z_{0,1}$ and $(S_8, 2^0) \in Z_{0,3}$. $S'_{2,0} = \begin{pmatrix} 1 & 1 & 0 \\ 0 & 0 & 0 \\ 1 & 1 & 1 \end{pmatrix}$.*

– *Step g = 3: We choose* $(S_3, 2^0) \in Z_{0,1}$. $S'_{3,0} = \begin{pmatrix} 1 & 1 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$.

- *Step d = 1:*

  – *Step g = 1: We choose* $(S_4, 2^1) \in Z_{1,2}$ *and* $(S_6, 2^1) \in Z_{1,3}$. $S'_{1,1} = \begin{pmatrix} 0 & 0 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$.

  – *Step g = 2: We choose* $(S_5, 2^1) \in Z_{1,2}$. $S'_{2,1} = \begin{pmatrix} 0 & 0 & 0 \\ 1 & 1 & 1 \\ 0 & 0 & 0 \end{pmatrix}$.

- *Step d = 2:*

  – *Step g = 1: We choose* $(S_6, 2^2) \in Z_{2,3}$. $S'_{1,2} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}$.

Therefore, $Z = \{(S'_{1,0}, 2^0), (S'_{2,0}, 2^0), (S_{3,0}, 2^0), (S'_{1,1}, 2^1), (S'_{2,1}, 2^1), (S'_{1,2}, 2^2)\}$.

References

[1] D. Baatar and H. W. Hamacher. New LP model for Multileaf Collimators in Radiation Therapy. Contribution to the Conference ORP3, University of Kaiserslautern, 2003.

[2] N. Bansal, D. Coppersmith, B. Schieber. "Minimizing Setup and Beam-On Times in Radiation Therapy", To appear in Proceedings of APPROX, 2006.

[3] N. Boland, H. Hamacher, and F. Lenzen, Minimizing Beam-On Time in Cancer Radiation Treatment Using Multileaf Collimators, Networks 43, 226-240, 2004.

[4] 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. In the Proceedings of the 15th International Symposium on Algorithms and Computation (ISAAC), 271-281, 2004.

[5] M. J. Collins, D. Kempe, J. Saia, and M. Young. "Non-Negative Integral Subset Representations of Integer Sets", accepted to Information Processing Letters, 2006.

[6] C. Cotrutz and L. Xing, Segment-Based Dose Optimization Using a Genetic Algorithm, Physics in Medicine and Biology, 48, 2987-2998, 2003.

[7] K. Engel. A New Algorithm for Optimal Multileaf Collimator Field Segmentation. To appear in Discrete Appl. Math.

[8] M. Ferris, 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.

[9] T. Kalinowski, The algorithmic complexity of the minimization of the number of segments in multileaf collimator field segmentation, preprint, 2004.

[10] S. Luan, C. Wang, D.Z. Chen, X. S. Hu, S.A. Naqvi, C.L. Lee, and C.X. Yu A new MLC segmentation algorithm/software for step-and-shoot IMRT delivery. Medical Physics, 31, 695-707, 2004.

[11] P. Xia and L. Verhey. Multileaf Collimator Leaf Sequencing Algorithm for Intensity Modulated Beams with Multiple Static Segments. Medical Physics, 1998.