Discrete Applied Mathematics 157 (2009) 2015–2030

Contents lists available at ScienceDirect

Discrete Applied Mathematics

journal homepage: www.elsevier.com/locate/dam

Optimal matrix-segmentation by rectangles

Konrad Engel *

Institut für Mathematik, Universität Rostock, Universitätsplatz 1, D-18051 Rostock, Germany

ARTICLE INFO

Article history:
Received 20 June 2006
Received in revised form 6 March 2008
Accepted 9 December 2008
Available online 15 January 2009

Keywords:
Network flow
Intensity-modulated radiation therapy (IMRT)
Segmentation algorithm
Linear programming
Rectangular partition
Matrix-decomposition
Gomory cut
Revised simplex method

ABSTRACT

We study the problem of decomposing a nonnegative matrix into a nonnegative combination of 0-1-matrices whose ones form a rectangle such that the sum of the coefficients is minimal. We present for the case of two rows an easy algorithm that provides an optimal solution which is integral if the given matrix is integral. An additional integrality constraint makes the problem more difficult if the matrix has more than two rows. An algorithm that is based on the revised simplex method and uses only very few Gomory cuts yields exact integral solutions for integral matrices of reasonable size in a short time. For matrices of large dimension we propose a special greedy algorithm that provides sufficiently good results in numerical experiments.

1. Introduction

In the following we study matrices of nonnegative real numbers of dimension $m \times n$. Some auxiliary matrices have other dimensions, but this will be clear from the context. Let $[n] = \{1, \ldots, n\}$. A nonempty subset R of $[m] \times [n]$ is called a *rectangle* if there are $l, r \in [n]$ and $b, t \in [m]$ such that $R = \{(i, j) : b \le i \le t, l \le j \le r\}$. A matrix $S = (s_{ij})$ is said to be a *rectangular segment* if there is a rectangle R such that

$$s_{ij} = \begin{cases} 1 & \text{if } (i, j) \in R, \\ 0 & \text{otherwise.} \end{cases}$$

We write S_R to indicate that the segment is given by R. A *segmentation* of a matrix A is a decomposition of A into a nonnegative combination of rectangular segments:

$$A = \sum_R u_R S_R$$

where $u_R \ge 0$ for all R. The *DT of the segmentation* is defined to be

$$U = \sum_R u_R$$

(DT is an abbreviation for *delivery time* — see below). The *rectangular DT-segmentation problem* is the following: For a given matrix A, find a rectangular segmentation of A such that its DT is minimal. Let $c(A)$ be this minimal DT. Moreover, let $c_I(A)$ be the minimal DT under integrality constraint, i.e. if the coefficients u_R are required to be integral.

* Fax: +49 381 4986553.
E-mail address: konrad.engel@uni-rostock.de.

0166-218X/$ – see front matter
doi:10.1016/j.dam.2008.12.008

The problem is motivated by radiation therapy planning. The radiation is produced by a linear accelerator and delivered through a rectangular region bounded by jaws. These jaws can be shifted so that the size of the rectangle can be changed (realized by a *collimator*). Usually the largest possible rectangle is discretized into bixels. Optimization algorithms provide values for the desired fluence through each bixel. Since the fluence varies from bixel to bixel one speaks of *intensity-modulated radiation therapy (IMRT)*. The fluence values for each bixel are given by a nonnegative fluence matrix that must be realized as a superposition of adjustable positions. The segments S correspond to the positions and the coefficients u correspond to the time of delivery, counted by monitor units. The total time of radiation is then the sum of the coefficients, i.e. the *delivery time — DT*. For the purpose that there are sufficiently many adjustable positions modern additional devices, called *multileaf collimators (MLCs)*, have been created, cf. [1]. But these MLCs are expensive and not used everywhere. Moreover, the practical verification of the positions is an important task. Hence the question of the much simpler realization by jaw movements using conventional collimators is plausible. There exist commercial systems using these simpler collimators for IMRT, e.g. the "Prowess's unique jaws-only IMRT system". But in that system the planning process is not split into the two steps "bixel-fluence-optimization" and "segmentation" — the process is combined and the fluence through the rectangles is directly optimized. A new algorithm for this *Direct Aperture Optimization* is contained in [2]. For the one-step approach, the number of rectangles or, more generally, segments cannot be very large, hence a restriction to a smaller set of allowed segments is indispensable. These allowed segments have to be chosen by heuristic means.

In this paper we focus on the two-step approach, where on principle all segments are allowed, and further discuss the segmentation step using jaws only, i.e. rectangles. An interesting special case of the rectangle segmentation, namely that A is a Boolean matrix, was considered by several authors in a geometrical setting, see [3–5]. These papers contain an algorithm for the determination of an optimal integral solution, i.e. of $c_I(A)$ in the case of a 0-1-matrix A.

The first results for general segmentation were obtained by Dai and Hu [6] using a simple heuristic, see Section 3.4. In a series of papers, Webb [7–9] studied the case where rectangular segments are combined with certain masks. This was extended by Webb et al. [10] for variable masks, called variable-aperture collimators.

The results show, that on the average the use of MLC-segments leads to much smaller DTs than the use of rectangular segments since with an MLC many more positions can be realized. This is partly compensated for by the fact that the transmission of radiation through the leaves of an MLC is significantly greater than the transmission through the jaws of a collimator. At least, as a mathematical problem, the rectangular DT-segmentation problem is very challenging and probably also interesting for other applications.

For brevity, we add two zero-columns and two zero-rows to A, i.e., we put

$$a_{i,0} = a_{i,n+1} = 0 \quad \text{for all } i \in \{0, \dots, m+1\},$$
$$a_{0,j} = a_{m+1,j} = 0 \quad \text{for all } j \in \{0, \dots, n+1\}.$$

Let M be the entry-rectangle incidence matrix, i.e. a 0-1-matrix of dimension $mn \times \binom{m+1}{2}\binom{n+1}{2}$ whose rows are indexed by the elements from $[m] \times [n]$ and whose columns are indexed by the rectangles, where – as usual – an element of M equals 1 iff the corresponding entry is contained in the rectangle. In order to make M well-defined we must fix a linear ordering of the elements of $[m] \times [n]$ and of the rectangles. With A we associate the vector $\boldsymbol{a}$ whose entries are the entries of A in the order that is given by the linear ordering of the elements of $[m] \times [n]$ (usually row by row in A). With these notations the rectangular DT-segmentation problem can be written in the form

$$\begin{array}{ll} \min & \mathbf{1}^{\mathbf{T}}\boldsymbol{u} \\ \text{s.t.} & M\boldsymbol{u} = \boldsymbol{a} \\ & \boldsymbol{u} \geq \mathbf{0}. \end{array} \tag{1}$$

The case $m = 1$ is well-known and well studied (cf. [11,12,1]). We have

$$c(A) = \sum_{j=1}^{n} \max\{a_{1,j} - a_{1,j-1}, 0\}. \tag{2}$$

Moreover, M is an interval matrix and hence totally unimodular. This immediately implies the existence of integral optimal solutions if A is integral, i.e. $c(A) = c_I(A)$ (cf. [13, pp. 540 ff]).

Simple inspection shows that for $m = n = 2$ the matrix M is still totally unimodular. But for $m \geq 2, n \geq 3$ (and similarly for $m \geq 3, n \geq 2$) the matrix M is no longer totally unimodular. Indeed, the submatrix of M induced by the entries $(1, 1), (1, 3), (2, 2)$ and the rectangles $\{(1, 1), (1, 2), (2, 1), (2, 2)\}$, $\{(1, 1), (1, 2), (1, 3)\}$, and $\{(1, 2), (1, 3), (2, 2), (2, 3)\}$ reads

$$\begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 1 & 0 & 1 \end{pmatrix}$$

and has determinant 2. Moreover, for $m \geq 2, n \geq 3$ (and similarly for $m \geq 3, n \geq 2$) the polyhedron of feasible solutions of (1) is not necessarily integral for integral A. One can easily check that

$$\frac{1}{2}\begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} + \frac{1}{2}\begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix} + \frac{3}{2}\begin{pmatrix} 0 & 1 & 0 \\ 0 & 1 & 0 \end{pmatrix} + \frac{7}{2}\begin{pmatrix} 0 & 0 & 0 \\ 1 & 1 & 0 \end{pmatrix} + \frac{5}{2}\begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & 1 \end{pmatrix} + \frac{1}{2}\begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix}$$