Near-optimal asymmetric binary matrix partitions*

Fidaa Abed† Ioannis Caragiannis‡ Alexandros A. Voudouris§

Abstract

We study the asymmetric binary matrix partition problem that was recently introduced by Alon et al. (WINE 2013). Instances of the problem consist of an $n \times m$ binary matrix A and a probability distribution over its columns. A *partition scheme* $B = (B_1, ..., B_n)$ consists of a partition B_i for each row i of A. The partition B_i acts as a smoothing operator on row i that distributes the expected value of each partition subset proportionally to all its entries. Given a scheme B that induces a smooth matrix A^B, the partition value is the expected maximum column entry of A^B. The objective is to find a partition scheme such that the resulting partition value is maximized. We present a $9/10$-approximation algorithm for the case where the probability distribution is uniform and a $(1 - 1/e)$-approximation algorithm for non-uniform distributions, significantly improving results of Alon et al. Although our first algorithm is combinatorial (and very simple), the analysis is based on linear programming and duality arguments. In our second result we exploit a nice relation of the problem to submodular welfare maximization.

1 Introduction

We study the *asymmetric matrix partition problem*, recently proposed by Alon et al. [2]. Consider an $n \times m$ matrix A with non-negative entries and a probability distribution p over its columns; p_j denotes the probability associated with column j. We distinguish between two cases for the probability distribution over the columns of the given matrix, depending on whether it is uniform or non-uniform. A partition scheme $B = (B_1, ..., B_n)$ for matrix A consists of a partition B_i of $[m]$ for each row i of A. More specifically, B_i is a collection of k_i pairwise disjoint subsets $B_{ik} \subseteq [m]$ (with $1 \le k \le k_i$) such that $\bigcup_{k=1}^{k_i} B_{ik} = [m]$. We can think of each partition B_i as a smoothing operator, which acts on the entries of row i and changes their value to the expected value of the partition subset they belong to. Formally, the *smooth value* of an entry (i, j) such that $j \in B_{ik}$ is defined as

$$A^B_{ij} = \frac{\sum_{\ell \in B_{ik}} p_\ell \cdot A_{i\ell}}{\sum_{\ell \in B_{ik}} p_\ell}. \tag{1}$$

Notice that all entries (i, j) such that $j \in B_{ik}$ have the same smooth value. Given a partition scheme B that induces the *smooth matrix* A^B, the resulting *partition value* is the expected maximum column entry of A^B, namely,

$$v^B(A, p) = \sum_{j \in [m]} p_j \cdot \max_i A^B_{ij}. \tag{2}$$

*A preliminary version of the results in this paper appeared in the 40th International Symposium on Mathematical Foundations of Computer Science (MFCS '15). This work was partially supported by the European Social Fund and Greek national funds through the research funding program Thales on "Algorithmic Game Theory", and by the Caratheodory research grant E.114 from the University of Patras.

†Faculty of Computing and Information Technology, University of Jeddah, Saudi Arabia. Email: fabed@uj.edu.sa

‡Computer Technology Institute and Press "Diophantus" & Department of Computer Engineering and Informatics, University of Patras, 26504 Rion, Greece. Email: caragian@ceid.upatras.gr

§Department of Computer Engineering and Informatics, University of Patras, 26504 Rion, Greece. Email: voudouris@ceid.upatras.gr

The objective of the asymmetric matrix partition problem is to find a partition scheme B such that the resulting partition value $v^B(A, p)$ is maximized.

The problem was introduced by Alon et al. [2]. They distinguish between two different cases depending on whether the matrix entries are binary (zero or one) or non-binary, and two different cases depending on whether the probability distribution over the matrix columns is uniform or not. For the simplest case of binary values and a uniform distribution, they prove that the problem is APX-hard and provide a 0.563-approximation algorithm. The partition scheme that achieves this approximation guarantee is selected as the one with the highest partition value among the partition schemes produced by three different algorithms. These algorithms use several interesting phases; we exploit two of them, namely, a "covering" and a "greedy completion" phase, which we put together in an intuitive greedy algorithm that we analyze. Alon et al. [2] also present a $1/13$-approximation bound for binary matrices and non-uniform probability distributions. Again, this bound follows by three different algorithms. For matrices with non-binary entries, they present a $1/2$- and an $\Omega(1/\log m)$-approximation algorithm for uniform and non-uniform distributions, respectively. A common idea underlying these results is that they try to identify a set of high-value entries that can be bundled together with other entries in order to increase the total contribution.

This interesting combinatorial optimization problem is strongly related to revenue maximization in *take-it-or-leave-it sales*. For example, consider the following setting. There are m items and n potential buyers. Each buyer has a value for each item; in general, she is not aware of the values of other buyers, or even of their existence. Nature selects at random (according to some probability distribution) an item for sale and, then, the seller approaches the highest value buyer and offers the item to her at a price equal to her valuation. A specific instantiation of this setting could be the following: the items correspond to keywords and the potential buyers correspond to advertisers. Every advertiser has a value for each keyword which represents the maximum amount of money she is willing to pay in order to occupy the advertising space that is allocated when the particular keyword is queried. The role of nature is played by users who submit queries and the role of the seller is played by the search engine, which allocates the advertising space according to the keyword queried each time, and in such a way that its revenue is maximized.

Can the seller exploit the fact that she has much more accurate information about the items for sale compared to the potential buyers? In particular, information asymmetry arises since the seller knows the realization of the randomly selected item whereas the buyers do not. The approach that is discussed in [2] is to let the seller define a buyer-specific signalling scheme. That is, for each buyer, the seller can partition the set of items into disjoint subsets (bundles) and report this partition to the buyer. For example, the search engine could bundle together keywords that are closely related to each other. After nature's random choice, the seller can reveal to each buyer the bundle that contains the realization, thus enabling her to re-evaluate her beliefs for the particular bundle (i.e., compute her expected value for the whole bundle and each item therein). The relation of this problem to asymmetric matrix partition should now be clear: the columns of the input matrix correspond to items, the rows correspond to potential buyers, and the value of the entry (i, j) corresponds to the value that buyer i has for item j. After the bundling of the items for a specific buyer, the smooth value of a bundle corresponds to the expected value the buyer has for each item included in the bundle. Finally, the partition value corresponds to the expected revenue of the seller. Interestingly, we will see that the seller can achieve revenue from items for which no buyer has any value.

This scenario falls within the line of research that studies the impact of information asymmetry to the quality of markets. Akerlof [1] was the first to introduce a formal analysis of "markets for lemons", where the seller has more information than the buyers regarding the quality of the products. Crawford and Sobel [7] study how, in such markets, the seller can exploit her advantage in order to maximize revenue. In [21], Milgrom and Weber provide the "linkage principle" which states that the expected revenue is enhanced when bidders are provided with more information. This principle seems to suggest full transparency but, in [18] and [20] the authors suggest that careful bundling of the items is the best way to exploit information asymmetry. Many different frameworks that reveal information to the bidders have been proposed in the literature.

More recently, Ghosh et al. [13] consider full information and propose a clustering scheme according to which, the items are partitioned into bundles and, then, for each such bundle, a separate second-price auction is performed. In this way, the potential buyers cannot bid only for the items that they actually want; they also have to compete for items that they do not have any value for. Hence, the demand for each item is increased and the revenue of the seller is higher. Emek et al. [10] present complexity results in similar settings and Miltersen and Sheffet [23] consider fractional bundling schemes for signaling.

In this work, we focus on the simplest binary case of asymmetric matrix partition. Of course, this case is very limited compared to the general one motivated above but poses interesting challenges in algorithm design and analysis; asymmetric binary matrix partition has been proved to be APX-hard and, still, the approximation ratios of the known algorithms are rather low. So, we design near-optimal approximation algorithms. In particular, we present a $9/10$-approximation algorithm for the uniform case and a $(1-1/e)$-approximation algorithm for non-uniform distributions. Both results significantly improve the previous bounds of Alon et al. [2]. The analysis of our first algorithm is quite interesting because, despite its purely combinatorial nature, it exploits linear programming techniques. Similar techniques have been used for the analysis of purely combinatorial algorithms in many different settings such as facility location [15], variants of set cover [3, 4, 6], online matching [19], maximum directed cut [11], and wavelength routing [5]; however, the application of the technique in the current context requires a quite involved reasoning about the structure of the solutions computed by the algorithm.

In our second result, we exploit a nice relation of the problem to submodular welfare maximization and use well-known algorithms from the literature. First, we discuss the application of a simple greedy $1/2$-approximation algorithm that has been studied by Lehmann et al. [17] and then apply Vondrák's smooth greedy algorithm [24] to achieve a $(1-1/e)$-approximation for our problem. Vondrák's algorithm is optimal in the value query model as Khot et al. [16] have proved. In a more powerful model where it is assumed that demand queries can be answered efficiently, Feige and Vondrák [12] have proved that $(1-1/e+\epsilon)$-approximation algorithms — where ϵ is a small positive constant — are possible. We briefly discuss the possibility/difficulty of applying such algorithms to asymmetric binary matrix partition and observe that the corresponding demand query problems are, in general, NP-hard.

The rest of the paper is structured as follows. We begin with preliminary definitions and examples in Section 2. Then, we present our $9/10$-approximation algorithm for the uniform case in Section 3 and our $(1-1/e)$-approximation algorithm for the non-uniform case in Section 4. We conclude with a short discussion on our model and results and present open problems in Section 5.

2 Preliminaries

Let $A^+ = \{j \in [m] :$ there exists a row i such that $A_{ij} = 1\}$ denote the set of columns of A that contain at least one 1-value entry, and $A^0 = [m] \backslash A^+$ denote the set of columns of A that contain only 0-value entries. In the next sections, we usually refer to the sets A^+ and A^0 as the sets of *one-columns* and *zero-columns*, respectively. Furthermore, let $A_i^+ = \{j \in [m] : A_{ij} = 1\}$ and $A_i^0 = \{j \in [m] : A_{ij} = 0\}$ denote the sets of columns that intersect with row i at a 1- and 0-value entry, respectively. All columns in A_i^+ are one-columns and, furthermore, $A^+ = \cup_{i=1}^n A_i^+$. The columns of A_i^0 can be either one- or zero-columns and, thus, $A^0 \subseteq \cup_{i=1}^n A_i^0$. Also, denote by $r = \sum_{j \in A^+} p_j$ the total probability of the one-columns. As an example, consider the 3×6 matrix

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

and a uniform probability distribution over its columns. We have $A^+ = \{2, 3, 5\}$ and $A^0 = \{1, 4, 6\}$. In the first two rows, the sets A_i^+ and A_i^0 are identical to A^+ and A^0, respectively. In the third row, the sets A_3^+ and A_3^0 are $\{2, 3\}$ and $\{1, 4, 5, 6\}$. Finally, the total probability of the one-columns r is $1/2$.

A partition scheme B can be thought of as consisting of n partitions B_1, B_2, ..., B_n of the set of columns $[m]$. We use the term *bundle* to refer to the elements of a partition B_i; a bundle is just a non-empty set of columns. For a bundle b of partition B_i corresponding to row i, we say that b is an *all-zero* bundle if $b \subseteq A_i^0$ and an *all-one* bundle if $b \subseteq A_i^+$. A singleton all-one bundle of partition B_i is called *column-covering bundle in row* i. A bundle that is neither all-zero nor all-one is called *mixed*. A mixed bundle corresponds to a set of columns that intersects with row i at both 1- and 0-value entries.

Let us examine the following partition scheme B for matrix A that defines the smooth matrix A^B according to equation (1).

B_1	$\{1, 2, 3, 4\}, \{5, 6\}$
B_2	$\{1, 2\}, \{3\}, \{4, 6\}, \{5\}$
B_3	$\{1, 4, 6\}, \{2, 3, 5\}$

A^B	1/2 1/2 0	1/2 1/2 2/3	1/2 1 2/3	1/2 0 0	1/2 1 2/3	1/2 0 0
$\max_i A_{ij}^B$	1/2	2/3	1	1/2	1	1/2

Here, the bundle $\{1, 2, 3, 4\}$ of (the partition B_1 of) the first row is a mixed one. The bundle $\{3\}$ of B_2 is all-one and, in particular, column-covering in row 2. The bundle $\{1, 4, 6\}$ of B_3 is all-zero.

By equation (2), the partition value is $25/36$ and it can be further improved. First, observe that the leftmost zero-column is included in two mixed bundles (in the first two rows). Also, the mixed bundle in the third row contains a one-column that has been covered through a column-covering bundle in the second row and intersects with the third row at a 0-value entry. Let us modify these two bundles.

B_1'	$\{1\}, \{2, 3, 4\}, \{5, 6\}$
B_2'	$\{1, 2\}, \{3\}, \{4, 6\}, \{5\}$
B_3'	$\{1, 4, 5, 6\}, \{2, 3\}$

$A^{B'}$	0 1/2 0	2/3 1/2 1	2/3 1 1	2/3 0 0	1/2 1 0	1/2 0 0
$\max_i A_{ij}^{B'}$	1/2	1	1	2/3	1	1/2

The partition value becomes $7/9 > 25/36$. Now, by merging the two mixed bundles $\{2, 3, 4\}$ and $\{5, 6\}$ in the first row, we obtain the smooth matrix below with partition value $47/60 > 7/9$. Observe that the contribution of column 4 to the partition value decreases but, overall, we have an increase in the partition value due to the increase in the contribution of column 6. Actually, such merges never decrease the partition value (see Lemma 2.1 below).

B_1''	$\{1\}, \{2, 3, 4, 5, 6\}$
B_2''	$\{1, 2\}, \{3\}, \{4, 6\}, \{5\}$
B_3''	$\{1, 4, 5, 6\}, \{2, 3\}$

$A^{B''}$	0 1/2 0	3/5 1/2 1	3/5 1 1	3/5 0 0	3/5 1 0	3/5 0 0
$\max_i A_{ij}^{B''}$	1/2	1	1	3/5	1	3/5

Finally, by merging the bundles $\{1, 2\}$ and $\{3\}$ in the second row and decomposing the bundle $\{2, 3\}$ in the last row into two singletons, the partition value becomes $73/90 > 47/60$ which can be verified to be optimal.

B_1'''	$\{1\}, \{2,3,4,5,6\}$
B_2'''	$\{1,2,3\}, \{4,6\}, \{5\}$
B_3'''	$\{1,4,5,6\}, \{2\}, \{3\}$

$A^{B'''}$	0 2/3 0	3/5 2/3 1	3/5 2/3 1	3/5 0 0	3/5 1 0	3/5 0 0
$\max_i A_{ij}^{B'''}$	2/3	1	1	3/5	1	3/5

We will now give some more definitions that will be useful in the following. We say that a one-column j is *covered* by a partition scheme B if there is at least one row i in which $\{j\}$ is column-covering. For example, in B''', the singleton $\{5\}$ is a column-covering bundle in the second row and the singletons $\{2\}$ and $\{3\}$ are column-covering in the third row. We say that a partition scheme *fully covers* the set A^+ of one-columns if all of them are covered. In this case, we use the term *full cover* to refer to the pairs of indices (i, j) of the 1-value entries A_{ij} such that $\{j\}$ is a column-covering bundle in row i. For example, the partition scheme B''' has the full cover $(2,5), (3,2), (3,3)$.

It turns out that optimal partition schemes always have a special structure like the one of B'''. Alon et al. [2] have formalized observations like the above into the following statement.

Lemma 2.1 (Alon et al. [2]). *Given a uniform instance of the asymmetric binary matrix partition problem with a matrix A, there is an optimal partition scheme B with the following properties:*

P1. B fully covers the set A^+ of one-columns.

P2. For each row i, B_i has at most one bundle containing all columns of A_i^+ that are not included in column-covering bundles in row i (if any). This bundle can be either all-one (if it does not contain zero-columns) or the unique mixed bundle of row i.

P3. For each zero-column j, there exists at most one row i such that j is contained in the mixed bundle of B_i (and j is contained in the all-zero bundles of the remaining rows).

P4. For each row i, the zero-columns that are not contained in the mixed bundle of B_i form an all-zero bundle.

Properties P1 and P3 imply that we can think of the partition value as the sum of the contributions of the column-covering bundles and the contributions of the zero-columns in mixed bundles. Property P2 comes from the following more general statement that has been proved in [2]; we give an alternative more direct proof here using Milne inequality [14, page 61]. Lemma 2.2 will be very useful several times in our analysis in both the uniform and the non-uniform case.

Lemma 2.2 (Alon et al. [2]). *Consider $t \geq 2$ mixed bundles. For $i = 1, ..., t$, bundle i contains 1-value entries of total probability x_i and zero-columns of probability y_i. The total contribution of the zero-columns in these mixed bundles to the partition value is upper bounded by the contribution of zero-columns of probability $\sum_{i=1}^t y_i$ that form a single mixed bundle together with 1-value entries of probability $\sum_{i=1}^t x_i$.*

Proof. By the definitions, the smooth value of the i-th bundle is $\frac{x_i}{x_i+y_i}$ and the contribution of its zero-columns to the the partition value is $\frac{x_i y_i}{x_i+y_i}$. The proof follows by Milne inequality which states that

$$\sum_{i=1}^t \frac{x_i y_i}{x_i + y_i} \leq \frac{\sum_{i=1}^t x_i \cdot \sum_{i=1}^t y_i}{\sum_{i=1}^t x_i + \sum_{i=1}^t y_i},$$

where the right-hand side expression is the contribution of the zero-columns in the partition value of the single mixed bundle. □

Now, property P2 should be apparent; the columns of A_i^+ that do not form column-covering bundles in row i are bundled together with zero-columns (if possible) in order to increase the contribution of the latter to the partition value. Property P4 makes B consistent to the definition of a partition scheme where the disjoint union of all the partition subsets in a row should be $[m]$. Clearly, the contribution of the all-zero bundles to the partition value is 0. Also, the non-column-covering all-one bundles do not contribute to the partition value either.

Unfortunately, as we will see later in Section 4, Lemma 2.1 does not hold for non-uniform instances. This is due only to property P1 which requires a uniform probability distribution over columns. Luckily, it turns out that non-uniform instances also exhibit some structure (recall that the crucial Lemma 2.2 applies to the non-uniform case as well), which allows us to consider the problem of computing an optimal partition scheme as a *welfare maximization problem*. In welfare maximization, there are m items and n agents; agent i has a valuation function $v_i : 2^{[m]} \to \mathbb{R}^+$ that specifies her value for each subset of the items. I.e., for a set S of items, $v_i(S)$ represents the value of agent i for S. Given a disjoint partition (or allocation) $S = (S_1, S_2, ..., S_n)$ of the items to the agents, where S_i denotes the set of items allocated to agent i, the social welfare is the sum of values of the agents for the sets of items allocated to them, i.e., $\mathrm{SW}(S) = \sum_i v_i(S_i)$. The term welfare maximization refers to the problem of computing an allocation of maximum social welfare. We will discuss only the variant of the problem where the valuations are monotone and submodular; following the literature, we use the term submodular welfare maximization to refer to it.

Definition 2.1. *A valuation function v is monotone if $v(S) \leq v(T)$ for any pair of sets S, T such that $S \subseteq T$. A valuation function v is submodular if $v(S \cup \{x\}) - v(S) \geq v(T \cup \{x\}) - v(T)$ for any pair of sets S, T such that $S \subseteq T$ and for any item x.*

An important issue in (submodular) welfare maximization arises with the representation of valuation functions. A valuation function can be described in detail by listing explicitly the values for each of the possible subsets of items. Unfortunately, this is clearly inefficient due to the necessity for exponential input size. A solution that has been proposed in the literature is to assume access to these functions by queries of a particular form. The simplest such form of queries reads as "what is the value of agent i for the set of items S?" These are known as *value queries*. Another type of queries, known as *demand queries*, are phrased as follows: "Given a non-negative price for each item, compute a set S of items for which the difference of the valuation of agent i minus the sum of prices for the items in S is maximized." Approximation algorithms that use a polynomial number of valuation or demand queries and obtain solutions to submodular welfare maximization with a constant approximation ratio are well-known in the literature [12, 17, 24]. Our improved approximation algorithm for the non-uniform case of asymmetric binary matrix partition exploits such algorithms.

3 The uniform case

In this section, we present the analysis of a greedy approximation algorithm when the probability distribution p over the columns of the given matrix is uniform. Our algorithm uses a *greedy completion procedure* that was also considered by Alon et al. [2]. This procedure starts from a full cover of the matrix, i.e., from column-covering bundles in some rows so that all one-columns are covered (by exactly one column-covering bundle). Once this initial full cover is given, the set of columns from A_i^+ that are not included in column-covering bundles in row i can form a mixed bundle together with some zero-columns in order to increase the contribution of the latter to the partition value. Greedy completion proceeds as follows. It goes over the zero-columns, one by one, and adds a zero-column to the mixed bundle of the row that maximizes the marginal contribution of the zero-column. The marginal contribution of a zero-column to the partition value when it is added to a mixed bundle that consists of x zero-columns and y one-columns is proportional (due to the uniform distribution over columns) to the quantity

$$\Delta(x,y) \quad = \quad (x+1)\frac{y}{x+y+1} - x\frac{y}{x+y} = \frac{y^2}{(x+y)(x+y+1)}.$$

The right-hand side of the first equality is simply the difference between the contribution of $x+1$ and x zero-columns to the partition value when they form a mixed bundle with y one-columns. Note that $\Delta(0,y)$ indicates the marginal contribution of a zero-column when put together with y one-columns to form a (new) mixed bundle. Alon et al. [2] made the following important observation for the uniform case. We extensively use it below, as well as the fact that $\Delta(x,y)$ is non-decreasing with respect to y.

Lemma 3.1 (Alon et al. [2]). *Among all partition schemes that include a given full cover, the greedy completion procedure yields the maximum contribution from the zero-columns to the partition value.*

So, our algorithm consists of two phases. In the first phase, called the *cover phase*, the algorithm computes an arbitrary full cover for set A^+. In the second phase, called the *greedy phase*, it simply runs the greedy completion procedure mentioned above. Note that, intentionally, we have not used much detail in the description of the algorithm and there are three issues that might seem to cause ambiguity at first glance. First, we have not described any particular way the full cover is constructed. Second, we have not defined some particular order in which the zero-columns are examined during the greedy phase. And, third, we have not discussed how ties are broken when there are multiple rows that maximize the marginal contribution of a zero-column. So, our description essentially defines a family of greedy algorithms; a different greedy algorithm is defined, depending on how the above three issues are implemented. In the rest of this section, we will show that *any* greedy algorithm has an approximation ratio of at least $9/10$; actually, the three issues do not affect the analysis at all. We will also show that our analysis is tight by presenting a simple instance for which some greedy algorithm is at most $9/10$-approximate. Even though greedy algorithms are purely combinatorial, our analysis exploits linear programming duality. In the following, unless otherwise specified, the term greedy algorithm refers to any member of the family of greedy algorithms.

Overall, the partition value obtained by the algorithm can be thought of as the sum of contributions from column-covering bundles (this is exactly r) plus the contribution from the mixed bundles created during the greedy phase (i.e., the contribution from the zero-columns). Denote by ρ the ratio between the total number of appearances of one-columns in the mixed bundles of the optimal partition scheme (so, the number of times each one-column is counted equals the number of mixed bundles that contain it) and the number of zero-columns. For example, in the partition scheme B''' in the example of the previous section, the two mixed bundles are $\{2,3,4,5,6\}$ in the first row and $\{1,2,3\}$ in the second row. So, the one-columns 2 and 3 appear twice while the one-column 5 appears once in these mixed bundles. Since we have three zero-columns, the value of ρ is $5/3$. We can use the quantity ρ to upper-bound the optimal partition value as follows.

Lemma 3.2. *The optimal partition value is at most* $r + (1-r)\frac{\rho}{\rho+1}$.

Proof. The first term in the above expression represents the contribution of the one-columns in the full cover of the optimal partition scheme. To reason about the second term, recall that our definitions imply that the total probability of one-columns in the mixed bundles of an optimal partition scheme is $\rho(1-r)$, while the total probability of zero-columns in these mixed bundles is $1-r$. By Lemma 2.2, the second term upper-bounds the total contribution of the zero-columns to the optimal partition value. □

In our analysis, we distinguish between two main cases depending on the value of ρ. The first case is when $\rho < 1$; in this case, we show that the additional partition value which is obtained during the greedy phase of the algorithm (i.e., the contribution of the zero-columns; recall that the greedy algorithm maximizes this quantity) is lower-bounded by the additional partition value we would have by creating bundles containing exactly one one-column and an almost equal number of zero-columns each.

Lemma 3.3. *If $\rho < 1$, then the partition value obtained by the algorithm is at least 0.97 times the optimal one.*

Proof. Using the definition of ρ, we can lower-bound the number of 1-value entries in the input matrix A by the sum of the mr column-covering bundles that form the full cover of the optimal partition scheme and the at least $\rho m(1-r)$ appearances of one-columns in the mixed bundles.

Now, consider a selection of the full cover during the cover phase of the greedy algorithm (this can, of course be different than the full cover of the optimal partition scheme) and let X be a set of (exactly) $\rho m(1-r)$ 1-value entries in the matrix A among those that are not included in the cover.

Using Lemma 3.1, we will lower-bound the partition value returned by the algorithm by considering the following formation of mixed bundles as an alternative to the greedy completion procedure used in the greedy phase. If $1/\rho$ is an integer, for each 1-value entry of X, we create a mixed bundle that contains the corresponding one-column together with $1/\rho$ distinct zero-columns. Hence, the smooth value of each zero-column is $\frac{1}{1+1/\rho}$ and the total partition value of this scheme is $r+(1-r)\frac{\rho}{\rho+1}$; by Lemma 3.2, this is optimal.

If instead $1/\rho$ is not an integer, let $k = \lfloor 1/\rho \rfloor$. For each 1-value entry of X, we create a mixed bundle that contains the corresponding one-column together with k or $k+1$ distinct zero-columns. In particular, $m(1-r)(1-\rho k)$ of these mixed bundles contain one one-column and $k+1$ zero-columns and the remaining $m(1-r)(\rho(k+1)-1)$ mixed bundles contain one one-column and k zero-columns. Observe that the smooth value of a zero-column is $\frac{1}{k+2}$ in the first case and $\frac{1}{k+1}$ in the second case. Hence, we can bound the partition value obtained by the algorithm as follows:

$$\begin{aligned}\text{ALG} &\geq r+(1-r)(1-\rho k)\frac{k+1}{k+2}+(1-r)(\rho(k+1)-1)\frac{k}{k+1}\\ &= r+(1-r)\frac{1+\rho k(k+1)}{(k+1)(k+2)}.\end{aligned}$$

Using Lemma 3.2, we have

$$\frac{\text{ALG}}{\text{OPT}} \geq \frac{r+(1-r)\frac{1+\rho k(k+1)}{(k+1)(k+2)}}{r+(1-r)\frac{\rho}{\rho+1}} \geq \frac{\frac{1+\rho k(k+1)}{(k+1)(k+2)}}{\frac{\rho}{\rho+1}} = \frac{(1+1/\rho)(1+\rho k(k+1))}{(k+1)(k+2)}.$$

This last expression is minimized (with respect to ρ) for $1/\rho = \sqrt{k(k+1)}$. Hence,

$$\frac{\text{ALG}}{\text{OPT}} \geq \frac{\left(1+\sqrt{k(k+1)}\right)^2}{(k+1)(k+2)},$$

which is minimized for $k=1$ to approximately 0.97. □

For the case $\rho \geq 1$, we use completely different arguments. Of course, we assume that $r < 1$, i.e., the input matrix contains some zero-columns since, otherwise, any full cover computed during the cover phase of the greedy algorithm would give an optimal partition value. We will reason about the partition value of the solution produced by the algorithm by considering a particular decomposition of the set of mixed bundles computed in the greedy phase. Then, using Lemmas 2.2 and 3.1, the contribution of the zero-columns to the partition value in the solution computed by the algorithm is lower-bounded by their contribution to the partition value when they are part of the mixed bundles obtained after the decomposition. To justify the correctness of the decomposition, we will use the following observation.

Lemma 3.4. *If $\rho \geq 1$, no mixed bundle computed by the greedy algorithm has more zero-columns than one-columns.*

Proof. First observe that the total number of appearances of one-columns in mixed and column-covering bundles in the optimal partition scheme is at least $rm + (1-r)\rho m$, which includes rm appearances of one-columns in column-covering bundles and $(1-r)\rho m$ appearances of one-columns in mixed bundles (there may be additional 1-value entries included in all-one bundles). So, after the end of the cover phase, there are at least $(1-r)\rho m \geq (1-r)m$ 1-value entries that can be included in mixed bundles together with the $(1-r)m$ zero-columns.

Assume, for the sake of contradiction, that some zero-column Z is included as the $(x+1)$-th zero-column in a mixed bundle b together with x 1-value entries for $x \geq 1$ at some step of the greedy phase. Prior to that step, there is either some 1-value entry not included in any mixed bundle which could be used to form a mixed bundle together with Z for a marginal contribution of $\Delta(0,1) = 1/2$ or some mixed bundle with $y \geq 1$ zero-columns and $y+\alpha$ 1-value entries (with $\alpha \geq 1$) in which case the marginal contribution would be $\Delta(y, y+\alpha) > 1/4$. This contradicts the definition of the greedy algorithm since the marginal contribution of Z was $\Delta(x,x) < 1/4$ when included in b. ☐

Now, the decomposition is defined as follows. For every mixed bundle with y zero-columns and x one-columns (by Lemma 3.4, $x \geq y$) and decomposes it into y bundles as follows. If x/y is an integer, each bundle has one zero-column and x/y one-columns. Otherwise, $x - y\lfloor x/y \rfloor$ bundles have one zero-column and $\lceil x/y \rceil$ one-columns and $y\lceil x/y \rceil - x$ bundles have one zero-column and $\lfloor x/y \rfloor$ one-columns. Clearly, this process does not alter bundles with a single zero-column. The solution obtained after the decomposition of the solution returned by the algorithm has a very special structure as our next lemma suggests.

Lemma 3.5. *There exists an integer $s \geq 1$ such that each bundle in the decomposition has at least s and at most $3s$ one-columns.*

Proof. Consider the application of the decomposition step to the mixed bundles that are computed by the algorithm and let s be the minimum number of one-columns among the decomposed mixed bundles. This implies that one of the mixed bundles, say b_1, computed by the algorithm has μ zero-columns and at most $(s+1)\mu - 1$ one-columns. Denoting by ν the number of one-columns in this bundle, we have that the marginal partition value when the last zero-column Z is included in b_1 is exactly

$$\Delta(\mu,\nu) = \frac{\nu^2}{(\nu+\mu)(\nu+\mu-1)} \leq \frac{((s+1)\mu-1)^2}{((s+2)\mu-1)((s+2)\mu-2)}$$

since $\Delta(\mu,\nu)$ is increasing in ν and $\nu \leq (s+1)\mu - 1$. The rightmost expression is decreasing in μ and $\mu \geq 1$; hence, the marginal partition value of Z is at most $\frac{s}{s+1}$.

Now assume for the sake of contradiction that one of the mixed bundles obtained after the decomposition has at least $3s+1$ one-columns. Clearly, this must have been obtained by the decomposition of a mixed bundle b_2 (returned by the algorithm) with λ zero-columns and at least $(3s+1)\lambda$ one-columns. Denote by ν' the number of one-columns in this bundle and let us compute the marginal partition value if the zero-column Z would be included in b_2. This would be

$$\Delta(\lambda+1,\nu') = \frac{\nu'^2}{(\nu'+\lambda+1)(\nu'+\lambda)} \geq \frac{(3s+1)^2\lambda}{((3s+2)\lambda+1)(3s+2)} \geq \frac{(3s+1)^2}{(3s+3)(3s+2)}.$$

The first inequality follows since the marginal partition value function is increasing in ν' and $\nu' \geq (3s+1)\lambda$, and the second one follows since $\lambda \geq 1$. Now, the last quantity can be easily verified to be strictly higher that $\frac{s}{s+1}$ and the algorithm should have included Z in b_2 instead. We have reached the desired contradiction that proves the lemma. ☐

Now, our analysis proceeds as follows. For every triplet $r \in [0, 1]$, $\rho \geq 1$ and integer $s \geq 1$, we will prove that any solution consisting of an arbitrary cover of the rm one-columns and the decomposed set of bundles containing at least s and at most $3s$ one-columns yields a $9/10$-approximation of the optimal partition value. By the discussion above (in particular, by Lemmas 2.2 and 3.1), this will also be the case for the solution returned by the algorithm. In order to account for the worst-case contribution of zero-columns to the partition value for a given triplet of parameters, we will use the following linear program, which we denote by LP(r, ρ, s):

$$\begin{aligned}
\text{minimize} \quad & \sum_{k=s}^{3s} \frac{k}{k+1}\theta_k \\
\text{subject to:} \quad & \sum_{k=s}^{3s} \theta_k = 1 - r \\
& \sum_{k=s}^{3s} k\theta_k \geq \rho(1-r) - r \\
& \theta_k \geq 0, k = s, ..., 3s
\end{aligned}$$

The variable θ_k denotes the total probability of the zero-columns that participate in decomposed mixed bundles with k one-columns. The objective is to minimize the contribution of the zero-columns to the partition value. The equality constraint means that all zero-columns have to participate in bundles. The inequality constraint requires that the total number of appearances of one-columns in bundles used by the algorithm is at least the total number of appearances of one-columns in mixed bundles of the optimal partition scheme minus one appearance for each one-column, since for every selection of the cover, the algorithm will have the same number of (appearances of) one-columns available to form mixed bundles. Informally, the linear program answers (rather pessimistically) the question of how inefficient the algorithm can be. In particular, given an instance with parameters r and ρ, the quantity $\min_{\text{int } s \geq 1} \text{LP}(r, \rho, s)$ yields a lower bound on the contribution of the zero-columns to the partition value and $r + \min_{\text{int } s \geq 1} \text{LP}(r, \rho, s)$ is a lower bound on the partition value. The next lemma completes the analysis of the greedy algorithm for the case $\rho \geq 1$.

Lemma 3.6. *For every $r \in [0, 1]$ and $\rho \geq 1$,*

$$r + \min_{\text{int } s \geq 1} \text{LP}(r, \rho, s) \geq \frac{9}{10}\text{OPT}.$$

Proof. We will prove the lemma using LP-duality. The dual of LP(r, ρ, s) is:

$$\begin{aligned}
\text{maximize} \quad & (1-r)\alpha + ((1-r)\rho - r))\beta \\
\text{subject to:} \quad & k\beta + \alpha \leq \frac{k}{k+1}, k = s, ..., 3s \\
& \beta \geq 0
\end{aligned}$$

Using Lemma 3.2, we bound the optimal partition value as

$$\text{OPT} \leq r + (1-r)\frac{\rho}{\rho+1} = \frac{\rho + r}{\rho + 1}.$$

Hence, it suffices to show that, for every triplet of parameters (r, ρ, s), there is a feasible dual solution of objective value $D(r, \rho, s)$ that satisfies

$$r + D(r, \rho, s) - \frac{9}{10}\frac{\rho + r}{\rho + 1} \geq 0. \tag{3}$$

The feasible region of the dual is defined by the lines $\beta = 0$, $\alpha = \frac{s}{s+1} - s\beta$ and $\alpha = \frac{3s}{3s+1} - 3s\beta$; the remaining constraints can be easily seen to be redundant. The two important intersections of those lines are the points

$$(\alpha, \beta) = \left(\frac{s}{s+1}, 0\right) \text{ and } (\alpha, \beta) = \left(\frac{3s^2}{(s+1)(3s+1)}, \frac{1}{(s+1)(3s+1)}\right)$$

with objective values

$$D_1(r, \rho, s) = \frac{s}{s+1}(1-r) \text{ and } D_2(r, \rho, s) = \frac{3s^2}{(s+1)(3s+1)}(1-r) + \frac{\rho(1-r) - r}{(s+1)(3s+1)},$$

respectively. We will show that one of these two points can always be used as a feasible dual solution in order to prove inequality (3). We distinguish between two cases.

Case I: $r \geq \frac{\rho-1}{\rho}$. We will show that the point with dual objective value $D_1(r, \rho, s)$ satisfies inequality (3), i.e.,

$$r + \frac{s}{s+1}(1-r) - \frac{9}{10}\frac{\rho + r}{\rho + 1} \geq 0. \tag{4}$$

Since $s \geq 1$, we have that the left hand side of inequality (4) is at least

$$\frac{1+r}{2} - \frac{9}{10}\frac{\rho + r}{\rho + 1} = \frac{1}{2} - \frac{9\rho}{10(\rho+1)} + r\left(\frac{1}{2} - \frac{9}{10(\rho+1)}\right).$$

Since $\rho \geq 1$, we have that $\frac{1}{2} - \frac{9}{10(\rho+1)} \geq 0$, and we can lower-bound the above quantity using the assumption $r \geq \frac{\rho-1}{\rho}$, as follows:

$$\frac{1+r}{2} - \frac{9}{10}\frac{\rho + r}{\rho + 1} \geq \frac{1}{2} - \frac{9\rho}{10(\rho+1)} + \frac{\rho - 1}{\rho}\left(\frac{1}{2} - \frac{9}{10(\rho+1)}\right) = \frac{(\rho-2)^2}{10\rho(\rho+1)} \geq 0,$$

and inequality (4) follows.

Case II: $r < \frac{\rho-1}{\rho}$. We will now show that the point with dual objective value $D_2(r, \rho, s)$ satisfies inequality (3), i.e.,

$$r + \frac{3s^2}{(s+1)(3s+1)}(1-r) + \frac{\rho(1-r) - r}{(s+1)(3s+1)} - \frac{9}{10}\frac{\rho + r}{\rho + 1} \geq 0. \tag{5}$$

Let us denote by F the left hand side of inequality (5). With simple calculations, we obtain

$$F = \frac{10\rho^2 - (-3s^2 + 36s - 1)\rho + 30s^2}{10(3s+1)(s+1)(\rho+1)} - r \cdot \frac{10\rho^2 - (40s - 10)\rho + 27s^2 - 4s + 9}{10(3s+1)(s+1)(\rho+1)}. \tag{6}$$

Observe that the numerator of the left fraction in (6) is a quadratic function with respect to ρ with positive coefficient in the leading term. Its discriminant is $-1191s^4 - 216s^3 + 1296s^2 - 72s + 7$ which is clearly negative for every integer $s \geq 1$. Hence, the numerator of the left fraction is always positive. Now, if the numerator of the rightmost fraction is negative, then inequality (5) is obviously satisfied. Otherwise, using the assumption $r < \frac{\rho-1}{\rho}$, we have

$$F \geq \frac{10\rho^2 - (-3s^2 + 36s - 1)\rho + 30s^2}{10(3s+1)(s+1)(\rho+1)} - \frac{\rho - 1}{\rho} \cdot \frac{10\rho^2 - (40s - 10)\rho + 27s^2 - 4s + 9}{10(3s+1)(s+1)(\rho+1)}$$

$$= \frac{(3s^2+4s+1)\rho^2+(3s^2-36s+1)\rho+27s^2-4s+9}{10\rho(3s+1)(s+1)(\rho+1)}.$$

Now, the numerator of the last fraction is again a quadratic function in terms of ρ with positive coefficient in the leading term and discriminant equal to

$$-315s^4-600s^3+1150s^2-200s-35 = (-315s^3-915s^2+235s-35)(s-1) \leq 0,$$

for every integer $s \geq 1$. Hence, $F \geq 0$ and the proof is complete. □

The next statement summarizes the discussion above.

Theorem 3.7. *The greedy algorithm always yields a* $9/10$*-approximation of the optimal partition value in the uniform case.*

Our analysis is tight as our next counter-example suggests.

Theorem 3.8. *There exists an instance of the uniform asymmetric binary matrix partition problem for which a greedy algorithm computes a partition scheme with value (at most)* $9/10$ *of the optimal one.*

Proof. Consider the instance of the asymmetric binary matrix partition problem that consists of the matrix

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

with $p_i = 1/4$ for $i = 1, 2, 3, 4$. The optimal partition value is obtained by covering the one-columns in the first two rows and then bundling each of the two zero-columns with a pair of one-columns in the third and fourth row, respectively. This yields a partition value of $5/6$. A greedy algorithm may select to cover the one-columns using the 1-value entries A_{31} and A_{42}. This is possible since the greedy algorithm has no particular criterion for breaking ties when selecting the full cover. Given this full cover, the greedy completion procedure will assign each of the two zero-columns with only one one-column. The partition value is then $3/4$, i.e., $9/10$ times the optimal partition value. □

4 Asymmetric binary matrix partition as welfare maximization

We now consider the more general non-uniform case. Interestingly, property P1 of Lemma 2.1 does not hold any more as the following statement shows.

Lemma 4.1. *For every* $\epsilon > 0$*, there exists an instance of the asymmetric binary matrix partition problem in which any partition scheme containing a full cover of the columns in* A^+ *yields a partition value that is at most* $8/9 + \epsilon$ *times the optimal one.*

Proof. Consider the instance of the asymmetric binary matrix partition problem consisting of the matrix

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

with column probabilities $p_j = \frac{1}{\beta+3}$ for $j = 1, 2, 3$ and $p_4 = \frac{\beta}{\beta+3}$ for $\beta > 2$. We will first prove an upper bound on the partition value of any partition scheme containing a full cover. Then, we will present a partition scheme without a full cover, which has a strictly higher partition value. The desired ratio of $8/9+\epsilon$ will then follow by setting the parameter β appropriately.

Observe that there are four partition schemes containing a full cover (depending on the rows that contain the column-covering bundle of the first two columns). In each of them, there are two 1-value entries in different rows that are not included in the full cover, and only one of them can be bundled together with the zero-column. By making calculations, we obtain that the partition value in these cases is $\frac{4\beta+3}{(\beta+1)(\beta+3)}$. Here is one of these partition schemes:

B_1	$\{1\},\{2,3,4\}$
B_2	$\{2\},\{1,3,4\}$
B_3	$\{1,3\},\{2,4\}$
B_4	$\{1\},\{3\},\{2,4\}$

A^B	1	0	0	0
	0	1	0	0
	0	$\frac{1}{\beta+1}$	0	$\frac{1}{\beta+1}$
	1	0	1	0
$p_j \cdot \max_i A^B_{ij}$	$\frac{1}{\beta+3}$	$\frac{1}{\beta+3}$	$\frac{1}{\beta+3}$	$\frac{\beta}{(\beta+1)(\beta+3)}$

In contrast, consider the partition scheme B' in which the 1-value entries A_{11} and A_{22} form column-covering bundles in rows 1 and 2, the entries A_{32} and A_{33} are bundled together in row 3 and the entries A_{41}, A_{43}, and A_{44} are bundled together in row 4. As it can be seen from the tables below (recall that $\beta>2$), the partition value now becomes $\frac{4.5\beta+5}{(\beta+2)(\beta+3)}$.

B'_1	$\{1\},\{2,3,4\}$
B'_2	$\{2\},\{1,3,4\}$
B'_3	$\{1,4\},\{2,3\}$
B'_4	$\{2\},\{1,3,4\}$

$A^{B'}$	1	0	0	0
	0	1	0	0
	0	1/2	1/2	0
	$\frac{2}{\beta+2}$	0	$\frac{2}{\beta+2}$	$\frac{2}{\beta+2}$
$p_j \cdot \max_i A^{B'}_{ij}$	$\frac{1}{\beta+3}$	$\frac{1}{\beta+3}$	$\frac{1}{2(\beta+3)}$	$\frac{2\beta}{(\beta+2)(\beta+3)}$

Clearly, the ratio of the two partition values approaches $8/9$ from above as β tends to infinity. Hence, the theorem follows by selecting β sufficiently large for any given $\epsilon>0$. $\square$

Still, as the next statement indicates, the optimal partition scheme has some structure which we will exploit later.

Lemma 4.2. *Consider an instance of the asymmetric binary matrix partition problem consisting of a matrix A and a probability distribution p over its columns. There is an optimal partition scheme B that satisfies properties P2, P3, P4 (from Lemma 2.1) as well as the new property P5:*

- *P2. For each row i, B_i has at most one bundle containing all columns of A_i^+ that are not included in column-covering bundles in row i (if any). This bundle can be either all-one (if it does not contain zero-columns) or the unique mixed bundle of row i.*
- *P3. For each zero-column j, there exists at most one row i such that j is contained in the mixed bundle of B_i (and j is contained in the all-zero bundles of the remaining rows).*
- *P4. For each row i, the zero-columns that are not contained in the mixed bundle of B_i form an all-zero bundle.*
- *P5. Given any column j, denote by $H_j=\arg\max_i A^B_{ij}$ the set of rows through which column j contributes to the partition value $v^B(A,p)$. For every $i\in H_j$ such that $A_{ij}=1$, the bundle of partition B_i that contains column j is not mixed.*

Proof. We first focus on property P5. Consider an optimal partition scheme B that does not satisfy property P5, and let j^* be a column such that $A_{i^*j^*}=1$ for some $i^*\in H_{j^*}$. Furthermore, assume that the mixed bundle b of partition B_{i^*} that contains column j^*, also contains the columns of a (possibly empty) set $b_1 \subseteq A^+_{i^*} \setminus \{j^*\}$ and the columns of a non-empty set $b_0 \subseteq A^0_{i^*}$. Let $p^+ \geq 0$ and $p^0 > 0$ be the sum of probabilities of the columns in b_1 and b_0, respectively.

Let B' be the partition scheme that is obtained from B when splitting bundle b into two bundles $\{j^*\}$ and $b \setminus \{j^*\}$; we will show that B' must be optimal as well. Observe that $A^B_{i^*j} = \frac{p_{j^*}+p^+}{p_{j^*}+p^++p^0}$ and $A^{B'}_{i^*j} = \frac{p^+}{p^++p^0}$ for every $j \in b \setminus \{j^*\}$; hence, $A^B_{i^*j} > A^{B'}_{i^*j}$. Since, this is the only difference between B and B', the difference $\max_i A^B_{ij} - \max_i A^{B'}_{ij}$ is at most $A^B_{i^*j} - A^{B'}_{i^*j}$ for every $j \in b \setminus \{j^*\}$, and for j^*, $\max_i A^B_{ij^*} - \max_i A^{B'}_{ij^*} = A^B_{i^*j^*} - A^{B'}_{i^*j^*} = \frac{p_{j^*}+p^+}{p_{j^*}+p^++p^0} - 1$. Hence, we have

$$\begin{aligned}
v^B(A,p) - v^{B'}(A,p) &= \sum_{j\in[m]} p_j \cdot \max_i A^B_{ij} - \sum_{j\in[m]} p_j \cdot \max_i A^{B'}_{ij} \\
&= \sum_{j\in b} p_j \left(\max_i A^B_{ij} - \max_i A^{B'}_{ij} \right) \\
&\leq \sum_{j\in b} p_j \left(A^B_{i^*j} - A^{B'}_{i^*j} \right) \\
&= p_{j^*} \left(\frac{p_{j^*}+p^+}{p_{j^*}+p^++p^0} - 1 \right) + \sum_{j\in b\setminus\{j^*\}} p_j \left(\frac{p_{j^*}+p^+}{p_{j^*}+p^++p^0} - \frac{p^+}{p^++p^0} \right) \\
&= \frac{p_{j^*}+p^+}{p_{j^*}+p^++p^0} \left(p_{j^*} + \sum_{j\in b\setminus\{j^*\}} p_j \right) - p_{j^*} - \frac{p^+}{p^++p^0} \sum_{j\in b\setminus\{j^*\}} p_j \\
&= 0,
\end{aligned}$$

where the second last equality is just a rearrangement of terms and the last one follows from the fact that $\sum_{j\in b\setminus\{j^*\}} p_j = p^+ + p^0$. Hence, the partition value does not decrease. By repeating this argument, we will reach an optimal partition scheme that satisfies property P5. Then, using arguments similar to the ones used in the proof of Alon et al. [2] for Lemma 2.1[1] is we can prove that the resulting partition scheme can be transformed in such a way so that it satisfies properties P2, P3, and P4. □

What Lemma 4.2 says is that the contribution of column $j \in A^+$ to the partition value comes from a row i such that either $j \in A^+_i$ and $\{j\}$ forms a column-covering bundle (and, hence, its smooth value is 1) or $j \in A^0_i$ and j belongs to the mixed bundle of row i (and the smooth value of its entries is strictly smaller than 1). A non-zero contribution of a column $j \in A^0$ to the partition value always comes from a row i where j belongs to the mixed bundle. A column $j \in A^0$ can have a contribution of zero to the optimal partition value when no mixed bundle exists[2]. Hence, the problem of computing the partition scheme of optimal partition value is equivalent to deciding the row from which each column contributes to the partition value, either as a one-column that is part of a (not necessarily full) cover or as a zero-column that is part of a mixed bundle.

Let B be a partition scheme and S be a set of columns whose contribution to the partition value of B comes from row i (i.e., i is a row that maximizes the smooth value A^B_{ij} for each column j in S). Denoting the sum of these contributions by $R_i(S) = \sum_{j\in S} p_j \cdot A^B_{ij}$, we can equivalently express

[1] Invoking Lemma 2.2 in order to prove property P2 is crucial here; verifying properties P3 and P4 is much easier.

[2] As an example of such an extreme case, consider an instance with a $k \times (k+1)$ matrix that consists of the identity $k \times k$ matrix and an extra zero-column, and has a uniform probability distribution over the columns. The optimal partition scheme contains a full cover and all-zero bundles only, and the zero-column has no contribution to the partition value.

$R_i(S)$ as

$$R_i(S) = \sum_{j \in S \cap A_i^+} p_j + \frac{\sum_{j \in S \cap A_i^0} p_j \sum_{j \in A_i^+ \setminus S} p_j}{\sum_{j \in S \cap A_i^0} p_j + \sum_{j \in A_i^+ \setminus S} p_j}.$$

The first sum represents the contribution of columns of $S \cap A_i^+$ to the partition value (through column-covering bundles) while the second sum represents the contribution of the columns in $S \cap A_i^0$ which are bundled together with all 1-value entries in $A_i^+ \setminus S$ in the mixed bundle of row i. Then, the partition scheme B can be thought of as a collection of disjoint sets S_i (with one set per row) such that S_i contains those columns whose entries achieve their maximum smooth value in row i. Hence, the partition value of B is $v^B(A, p) = \sum_{i \in [n]} R_i(S_i)$ and the problem is essentially equivalent to welfare maximization where the rows act as the agents who will be allocated bundles of items (corresponding to columns).

Lemma 4.3. *For every row i, the function R_i is non-decreasing and submodular.*

Proof. We will show that the function R_i is non-decreasing and has decreasing marginal utilities, i.e.,

- (monotonicity) for every set S and item $x \notin S$, it holds that $R_i(S) \leq R_i(S \cup \{x\})$;
- (decreasing marginal utilities) for every pair of sets S, T such that $S \subseteq T$ and every item $x \notin T$, it holds that $R_i(S \cup \{x\}) - R_i(S) \geq R_i(T \cup \{x\}) - R_i(T)$.

In order to simplify notation, let us define the functions $\alpha(S) = \sum_{j \in S \cap A_i^+} p_j$, $\beta(S) = \sum_{j \in S \cap A_i^0} p_j$ and $\gamma(S) = \sum_{j \in A_i^+ \setminus S} p_j$. We can rewrite the function R_i as

$$R_i(S) = \alpha(S) + \frac{\beta(S) \cdot \gamma(S)}{\beta(S) + \gamma(S)}.$$

Let $S, T \subseteq [m]$ be two sets of columns such that $S \subseteq T$ and let x be a column that does not belong to set T. We distinguish between two cases depending on x. If $x \in A_i^+$, observe that

- $\alpha(S \cup \{x\}) = \alpha(S) + p_x$ and $\alpha(T \cup \{x\}) = \alpha(T) + p_x$;
- $\beta(S \cup \{x\}) = \beta(S)$ and $\beta(T \cup \{x\}) = \beta(T)$;
- $\gamma(S \cup \{x\}) = \gamma(S) - p_x$ and $\gamma(T \cup \{x\}) = \gamma(T) - p_x$.

Using the definition of function R_i, we have

$$\begin{aligned}
R_i(S \cup \{x\}) - R_i(S) &= p_x + \beta(S) \left(\frac{\gamma(S) - p_x}{\beta(S) + \gamma(S) - p_x} - \frac{\gamma(S)}{\beta(S) + \gamma(S)} \right) \\
&= p_x - \frac{p_x \beta(S)^2}{(\beta(S) + \gamma(S))(\beta(S) + \gamma(S) - p_x)} \\
&\geq p_x - \frac{p_x \beta(S)^2}{(\beta(S) + \gamma(T))(\beta(S) + \gamma(T) - p_x)} \\
&\geq p_x - \frac{p_x \beta(T)^2}{(\beta(T) + \gamma(T))(\beta(T) + \gamma(T) - p_x)} \\
&= R_i(T \cup \{x\}) - R_i(T).
\end{aligned}$$

The first inequality follows since γ is clearly non-increasing and $S \subseteq T$ and the second inequality follows by applying twice (with $a = \gamma(T)$ and $a = \gamma(T) - p_x$, respectively) the fact that the function $f(z) = \frac{z}{z+a}$ for $a \geq 0$ is non-decreasing.

If instead $x \in A_i^0$, observe that

- $\alpha(S \cup \{x\}) = \alpha(S)$ and $\alpha(T \cup \{x\}) = \alpha(T)$;
- $\beta(S \cup \{x\}) = \beta(S) + p_x$ and $\beta(T \cup \{x\}) = \beta(T) + p_x$;
- $\gamma(S \cup \{x\}) = \gamma(S)$ and $\gamma(T \cup \{x\}) = \gamma(T)$.

Hence, we have

$$\begin{aligned} R_i(S \cup \{x\}) - R_i(S) &= \gamma(S) \left(\frac{\beta(S) + p_x}{\beta(S) + \gamma(S) + p_x} - \frac{\beta(S)}{\beta(S) + \gamma(S)} \right) \\ &= \frac{p_x \gamma(S)^2}{(\beta(S) + \gamma(S))(\beta(S) + \gamma(S) + p_x)} \\ &\geq \frac{p_x \gamma(S)^2}{(\beta(T) + \gamma(S))(\beta(T) + \gamma(S) + p_x)} \\ &\geq \frac{p_x \gamma(T)^2}{(\beta(T) + \gamma(T))(\beta(T) + \gamma(T) + p_x)} \\ &= R_i(T \cup \{x\}) - R_i(T). \end{aligned}$$

Again, the first inequality follows since β is clearly non-decreasing and $S \subseteq T$ and the second inequality follows by applying twice (with $a = \beta(T)$ and $a = \beta(T) + p_x$, respectively) the fact that the function $f(z) = \frac{z}{z+a}$ with $a \geq 0$ is non-decreasing.

We have completed the proof that R_i has decreasing marginal utilities. In order to establish monotonicity, it suffices to observe that the quantity at the right-hand side of the second equality in each of the above two derivations starting with $R_i(S \cup \{x\}) - R_i(S)$ is non-negative. □

Lehmann et al. [17] studied the submodular welfare maximization problem and provided a simple algorithm that uses value queries and yields a $1/2$-approximation of the optimal welfare. Their algorithm considers the items one by one in arbitrary order and assigns item j to an agent that maximizes the marginal valuation (the additional value from the allocation of item j). In our setting, this algorithm can be implemented as follows. It considers the one-columns first and the zero-columns afterwards. Whenever considering a one-column j, a column-covering bundle $\{j\}$ is formed at an arbitrary row i with $j \in A_i^+$ (such a decision definitely maximizes the increase in the partition value). Once all one-columns have been processed, the remaining 1-value entries (that did not form column-covering bundles) in each row are grouped into a bundle. All these bundles are available to host zero-columns (that will be processed next) and evolve into mixed ones. Afterwards, whenever considering a zero-column, the algorithm includes it to a mixed bundle that maximizes the increase in the partition value. Using the terminology of Alon et al. [2] (or the terminology we used in Section 3), the algorithm essentially starts with an arbitrary cover of the one-columns and then it runs the greedy completion procedure. Again, we will use the term "greedy algorithm" to refer to the whole family of algorithms that are defined by different implementations of the several missing details in the above description, such as the order in which the one-columns are processed, the particular way the column-covering bundles are selected, the order in which the zero-columns are processed, and the way ties are broken between different mixed bundles to which a zero-column can be added. Our analysis below holds for any member of this family.

Theorem 4.4. *The greedy algorithm for the asymmetric binary matrix partition problem has approximation ratio at least* $1/2$. *This bound is tight.*

Proof. The lower bound holds by the equivalence of the greedy algorithm with the algorithm studied by Lehmann et al. [17]. Below, we prove the upper bound. In particular, we show that for every $\epsilon > 0$, there exists an instance of the problem in which the greedy algorithm obtains a partition scheme whose value is at most $1/2+\epsilon$ of the optimal one.

Let $k>0$ be a positive integer and α significantly higher than k. Consider the instance of the asymmetric binary matrix partition that consists of the following $(k+1)\times(k+1)$ matrix

$$A=\begin{pmatrix} 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 \\ 1 & 1 & \cdots & 1 & 0 \end{pmatrix}$$

where $p_j=\frac{1}{k+\alpha}$ for $j\in[k]$ and $p_{k+1}=\frac{\alpha}{k+\alpha}$. So, the first k columns and rows of A form an identity matrix, the last column has only 0-value entries and the last row consists of k 1-value entries in the first k columns. In order to lower-bound the optimal partition value, consider the partition scheme consisting of a full cover that contains the 1-value entries (i,i) for $i\leq k$, and a bundle containing the whole $(k+1)$-th row. The optimal partition value is lower-bounded by the value of this partition scheme. By simple calculations, we obtain

$$\mathrm{OPT}\geq\frac{k^2+2\alpha k}{(k+\alpha)^2}.$$

On the other hand, the greedy algorithm may select first to cover the k one-columns using the 1-value entries $(k+1,j)$ for $j\leq k$ and, then, bundle the zero-column together with only one 1-value entry in some of the first k rows. The partition value of the greedy algorithm is then

$$\mathrm{GREEDY}=\frac{k+(k+1)\alpha}{(k+\alpha)(\alpha+1)}.$$

Hence, the ratio between the two partition values is

$$\frac{\mathrm{GREEDY}}{\mathrm{OPT}}\leq\frac{(k+\alpha)(k+(k+1)\alpha)}{(k^2+2\alpha k)(\alpha+1)}.$$

Pick an arbitrarily small $\delta>0$; then, there exist a value for α (significantly higher than k) so that the above ratio satisfies $\frac{\mathrm{GREEDY}}{\mathrm{OPT}}\leq\frac{k+1}{2k}+\delta$. The theorem follows by picking k sufficiently large and δ sufficiently small. □

We can use the more sophisticated smooth greedy algorithm of Vondrák [24], which uses value queries to obtain the following.

Corollary 4.5. *There exists a $(1-1/e)$-approximation algorithm for the asymmetric binary matrix partition problem.*

One might hope that due to the particular form of functions R_i, better approximation guarantees might be possible using the $(1-1/e+\epsilon)$-approximation algorithm of Feige and Vondrák [12] which requires that demand queries of the form

> given agent i and a price q_j for every item $j\in[m]$, select the bundle S that maximizes the difference $R_i(S)-\sum_{j\in S}q_j$

can be answered in polynomial time. Unfortunately, in our setting, this is not the case in spite of the very specific form of the function R_i.

Lemma 4.6. *Answering demand queries associated with the asymmetric binary matrix partition problem are NP-hard.*

Proof. We use reduction from Partition to show that the following (very restricted) decision version DQ of a demand query is NP-hard.

> DQ: Given a $1 \times m$ binary matrix A, probabilities p_j and prices q_j for column $j \in [m]$, is there a set $S \subseteq [m]$ such that $R_i(S) - \sum_{j\in S} q_j \geq 5/18$?

We start from an instance of Partition consisting of a collection C of t items of integer size w_1, w_2, ..., w_t and the question of whether there exists a subset $Y \subseteq C$ of items such that

$$\sum_{j\in Y} w_j = \sum_{j\in C\setminus Y} w_j = \frac{1}{2}\sum_{j\in C} w_j.$$

Define $W = \sum_{j\in C} w_j$. Given this instance, we construct an instance of DQ with $m = t+1$ as follows. The binary matrix A consists of a single row that contains t 1-value entries with associated probabilities $\frac{w_1}{2W}$, $\frac{w_2}{2W}$, ..., $\frac{w_t}{2W}$ and a 0-value entry with associated probability $1/2$. Set the prices as $q_j = \frac{5w_j}{18W}$ for $j = 1, ..., t$ and $q_{t+1} = 0$.

By the definition of the function R_i, given a set $S \subseteq [t+1]$, we have

$$\begin{aligned} R_i(S) - \sum_{j\in S} q_j &= \frac{1}{2W}\sum_{j\in S\setminus\{t+1\}} w_j + \frac{\frac{1}{4W}\sum_{j\in[t]\setminus S} w_j}{\frac{1}{2} + \frac{1}{2W}\sum_{j\in[t]\setminus S} w_j} - \frac{5}{18W}\sum_{j\in S\setminus\{t+1\}} w_j \\ &= \frac{2}{9} - \frac{2}{9W}\sum_{j\in[t]\setminus S} w_j + \frac{\sum_{j\in[t]\setminus S} w_j}{2W + 2\sum_{j\in[t]\setminus S} w_j}. \end{aligned}$$

Now, consider the function $f(z) = \frac{2}{9} - \frac{2z}{9W} + \frac{z}{2W+2z}$; the equality above implies that

$$R_i(S) - \sum_{j\in S} q_j = f\left(\sum_{j\in[t]\setminus S} w_j\right).$$

By nullifying the derivative of function f, we obtain that it has a unique maximum at $z = W/2$. Since $f(W/2) = 5/18$, the instance of DQ is equivalent to asking whether there exists a set S such that $\sum_{j\in[t]\setminus S} w_j = W/2$, which is equivalent to asking whether there exists a set of items of total size $W/2$ in the instance of Partition. ☐

5 Discussion

In this work, we have focused on the binary version of the asymmetric matrix partition problem and presented improved approximation algorithms for uniform and non-uniform probability distributions. The approximation guarantees are superior to those in the previous work by Alon et al. [2]. Designing algorithms with even better approximation guarantees is a first obvious open problem.

Recall (see the example discussed in Section 1) that the motivation for the asymmetric matrix partition problem is from revenue maximization in take-it-or-leave-it sales. Admittedly, in the (uniform) binary case, the fact that the greedy partition schemes contain column-covering bundles makes it possible for a buyer to distinguish between cases in which she is actually offered an item she values as 1 (a singleton bundle with smooth value of 1) or 0 (a mixed bundle). This is clearly a drawback and asymmetric binary matrix partition (as studied here and in [2]) should not be used to model simple take-it-or-leave-it sales. One possible remedy could be to lower-bound the size of any bundle with non-zero value or require some symmetry among the bundles that contain any given zero column, so that no information about the item selected by nature is revealed to the buyer by the seller.

Still, we believe that asymmetric binary matrix partition is important as an algorithmically challenging problem and can provide insights to efficient solutions for revenue maximization. In this direction, the above issue does not seem to be as severe in the general asymmetric matrix partition. This is justified by the assumption that buyers do not know each other and information about the particular item that is selected to be sold is not as easy to be inferred. Again, the non-binary asymmetric matrix partition with additional constraints that guarantee no information revelation to the buyers deserves investigation.

References

[1] G. A. Akerlof. The market for 'lemons': Quality uncertainty and the market mechanism. *Quaterly Journal of Economics*, 84, pages 488–500, 1970.

[2] N. Alon, M. Feldman, I. Gamzu, and M. Tennenholtz. The asymmetric matrix partition problem. In *Proceedings of the 9th Conference on Web and Internet Economics (WINE)*, pages 1–14 2013.

[3] S. Athanassopoulos, I. Caragiannis, and C. Kaklamanis. Analysis of approximation algorithms for k-set cover using factor-revealing linear programs. *Theory of Computing Systems*, 45(3), pages 555–576, 2009.

[4] S. Athanassopoulos, I. Caragiannis, C. Kaklamanis, and M. Kyropoulou. An improved approximation bound for spanning star forest and color saving. In *Proceedings of the 34th International Symposium on Mathematical Foundations of Computer Science (MFCS)*, pages 90–101, 2009.

[5] I. Caragiannis. Wavelength management in WDM rings to maximize the number of connections. *SIAM Journal on Discrete Mathematics*, 23(2), pages 959–978, 2009.

[6] I. Caragiannis, C. Kaklamanis, and M. Kyropoulou. Tight approximation bounds for combinatorial frugal coverage algorithms. *Journal of Combinatorial Optimization*, 26(2), pages 292–309, 2013.

[7] V. Crawford and J. Sobel. Strategic information transmission. *Econometrica*, 50, pages 1431–1451, 1982.

[8] J. Cremer and R. P. McLean. Optimal selling strategies under uncertainty for a discriminating monopolist when demands are interdependent. *Econometrica*, 53, pages 345–361, 1985.

[9] J. Cremer and R. P. McLean. Full extraction of the surplus in bayesian and dominant strategy auctions. *Econometrica*, 56, pages 1247–1257, 1988.

[10] Y. Emek, M. Feldman, I. Gamzu, R. Paes Leme, and M. Tennenholtz. Signaling schemes for partition value maximization. In *Proceedings of the 13th ACM Conference on Electronic Commerce (EC)*, pages 514–531, 2012.

[11] U. Feige and S. Jozeph. Oblivious algorithms for the maximum directed cut problem. *Algorithmica*, 71(2), pages 404–428, 2015.

[12] U. Feige and J. Vondrák. The submodular welfare problem with demand queries. *Theory of Computing*, 6, pages 247–290, 2010.

[13] A. Ghosh, H. Nazerzadeh, and M. Sundararajan. Computing optimal bundles for sponsored search. In *Proceedings of the 3rd International Workshop on Web and Internet Economics (WINE)*, pages 576–583, 2007.

[14] G. H. Hardy, J. E. Littlewood, and G. Pólya. *Inequalities*, Cambridge University Press, 1934.

[15] K. Jain, M. Mahdian, E. Markakis, A. Saberi, and V. V. Vazirani. Greedy facility location algorithms analyzed using dual fitting with factor-revealing LP. *Journal of the ACM*, 50(6), pages 795–824, 2003.

[16] S. Khot, R. Lipton, E. Markakis, and A. Mehta. Inapproximability results for combinatorial auctions with submodular utility functions. *Algorithmica*, 52, pages 3–18, 2008.

[17] B. Lehmann, D. J. Lehmann, and N. Nisan. Combinatorial auctions with decreasing marginal utilities. *Games and Economic Behavior*, pages 270–296, 2006.

[18] J. Levin and P. Milgrom. Online advertising: Heterogeneity and conflation in market design. *American Economic Review*, 100, pages 603–607, 2010.

[19] M. Mahdian and Q. Yan. Online bipartite matching with random arrivals: an approach based on strongly factor-revealing LPs. In *Proceedings of the 43rd ACM Symposium on Theory of Computing (STOC)*, pages 597–606, 2011.

[20] P. Milgrom. Simplified mechanisms with an application to sponsored-search auctions. *Games and Economic Behavior*, 70, pages 62–70, 2010.

[21] P. R. Milgrom and R. J. Weber. A theory of auctions and competitive bidding. *Econometrica*, 50, pages 1089–1122, 1982.

[22] P. R. Milgrom and R. J. Weber. The value of information in a sealed-bid auction. *Journal of Mathematical Economics*, 10, pages 105–114, 1982.

[23] P. B. Miltersen and O. Sheffet. Send mixed signals - Earn more, work less. In *Proceedings of the 13th ACM Conference on Electronic Commerce (EC)*, pages 234–247, 2012.

[24] J. Vondrák. Optimal approximation for the submodular welfare problem in the value oracle model. In *Proceedings of the 40th ACM Symposium on Theory of Computing (STOC)*, pages 67-74, 2008.