Near-Optimal Asymmetric Binary Matrix Partitions

Fidaa Abed[1], Ioannis Caragiannis[2], and Alexandros A. Voudouris[2(✉)]

[1] Technische Universität Berlin, Berlin, Germany
[2] Computer Technology Institute and Press "Diophantus" and Department of Computer Engineering and Informatics, University of Patras, 26504 Rion, Greece
voudouris@ceid.upatras.gr

Abstract. We study the asymmetric binary matrix partition problem that was recently introduced by Alon et al. (WINE 2013) to model the impact of asymmetric information on the revenue of the seller in take-it-or-leave-it sales. 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 binary matrix partition problem* (ABMP) proposed by Alon et al. [2]. Consider a matrix $A \in \{0, 1\}^{n \times m}$ 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 $1 \leq k \leq k_i$) 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

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.

G.F. Italiano et al. (Eds.): MFCS 2015, Part II, LNCS 9235, pp. 1–13, 2015.
DOI: 10.1007/978-3-662-48054-0_1

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

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

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,

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

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

Alon et al. [2] were the first to consider the asymmetric matrix partition problem. They prove that the problem is APX-hard and provide a 0.563- and a 1/13-approximation for uniform and non-uniform probability distributions, respectively. They also consider input matrices with non-negative non-binary entries and present a 1/2- and an $\Omega(1/\log m)$-approximation algorithm for uniform and non-uniform distributions, respectively. This interesting combinatorial optimization problem has apparent relations to revenue maximization in *take-it-or-leave-it sales*. For example, consider a setting with m items and n potential buyers. Each buyer has a value for each item. Nature selects at random (according to some probability distribution) an item for sale and, then, the seller approaches the highest valuation buyer and offers the item to her at a price equal to her valuation. 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. After nature's random choice, the seller can reveal to each buyer the bundle that contains the realization, thus enabling her to update her valuation beliefs. The relation of this problem to asymmetric matrix partition should be apparent. Interestingly, 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 [17], 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 [15] and [16] 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. [12] 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 [19] consider fractional bundling schemes for signaling.

In this work, we focus on the simplest binary case of asymmetric matrix partition which has been proved to be APX-hard. 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 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 in a series of papers on variants of set cover (e.g. [3–6]) by the second author; 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. [14] and then apply Vondrák's smooth greedy algorithm [20] to achieve a $(1 - 1/e)$-approximation. Vondrák's algorithm is optimal in the value query model as Khot et al. [13] have proved. In a more powerful model where it is assumed that demand queries can be answered efficiently, Feige and Vondrák [11] 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 Sect. 2. Then, we present our 9/10-approximation algorithm for the uniform case in Sect. 3 and our $(1 - 1/e)$-approximation algorithm for the non-uniform case in Sect. 4. Due to lack of space, most proofs are omitted.

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 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* and corresponds to a set of columns that intersect with row i at both 1- and 0-value entries.

Let us examine the following partition scheme B and its corresponding smooth matrix A^B defined by Eq. (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	1/2	1/2	1/2	1/2
	1/2	1/2	1	0	1	0
	0	2/3	2/3	0	2/3	0
$\max_i A_{ij}^B$	1/2	2/3	1	1/2	1	1/2

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

By Eq. (2), the partition value is $25/36$ and it can be further improved. Observe that the leftmost zero-column is included in two mixed bundles (in rows 1 and 2), and the mixed bundle in row 3 contains a one-column that is covered by a column-covering bundle in row 2 and intersects with row 3 at a 0-value entry. Let us modify these 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	2/3	2/3	2/3	1/2	1/2
	1/2	1/2	1	0	1	0
	0	1	1	0	0	0
$\max_i A_{ij}^{B'}$	1/2	1	1	2/3	1	1/2

The partition value is $7/9 > 25/36$. By merging the mixed bundles $\{2, 3, 4\}$ and $\{5, 6\}$ in row 1, we obtain a partition value $47/60 > 7/9$. Observe that the contribution of column 4 to the partition value decreases but, overall, the partition value increases due to the increase in the contribution of column 6. Actually, such merges never decrease the partition value (see Lemma 1).

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	3/5	3/5	3/5	3/5	3/5
	1/2	1/2	1	0	1	0
	0	1	1	0	0	0
$\max_i A^{B''}_{ij}$	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	3/5	3/5	3/5	3/5	3/5
	2/3	2/3	2/3	0	1	0
	0	1	1	0	0	0
$\max_i A^{B'''}_{ij}$	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 such observations into the next statement.

Lemma 1 (Alon et al. [2]). *Given a uniform instance of ABMP 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 $j \in A^0$, 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 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.

The proof of Lemma 1 in [2] extensively uses the fact that the instance is uniform. Unfortunately, as we will see later in Sect. 4, the lemma (in particular, property P1) does not hold for non-uniform instances. Luckily, it turns out that non-uniform instances also exhibit some structure 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 $v_i : 2^{[m]} \rightarrow \mathbb{R}^+$ 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 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 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 $x \notin T$.*

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. 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 given by the quantity

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

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. Alon et al. [2] proved that, in the uniform case, this greedy completion procedure yields the maximum contribution from the zero-columns to the partition value among all partition schemes that include a given full cover. We extensively use this property as well as the fact that $\Delta(x, y)$ is non-decreasing with respect to y.

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. In the rest of this section, we will show that this simple algorithm obtains an approximation ratio of 9/10; we will also show that our analysis is tight. Even though our algorithm is purely combinatorial, our analysis exploits linear programming duality.

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

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, the additional partition value obtained during the greedy phase of the algorithm is lower-bounded by the partition value we would have by creating bundles containing exactly one one-column and either $\lceil 1/\rho \rceil$ or $\lfloor 1/\rho \rfloor$ zero-columns.

Lemma 3. *If $\rho < 1$, then the partition value obtained by the algorithm is at least* 0.97 *times the optimal one.*

Proof. 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. For each appearance of a one-column in a mixed bundle in the partition scheme computed by the algorithm (observe that the total number of such appearances is exactly $\rho m(1-r)$), we include this one-column in a mixed bundle together with either $\lceil 1/\rho \rceil$ or $\lfloor 1/\rho \rfloor$ distinct zero-columns. By Lemma 2, this process yields an optimal partition value if $1/\rho$ is an integer. Otherwise, denote by $x = m(1-r)(1-\rho\lfloor 1/\rho \rfloor)$ the number of mixed bundles containing $\lceil 1/\rho \rceil$ zero-columns. Then, the number of mixed bundles containing $\lfloor 1/\rho \rfloor$ zero-columns will be $\rho m(1-r) - x = m(1-r)(\rho\lceil 1/\rho \rceil - 1)$. Observe that the smooth value of a zero-column is $\frac{1}{1+\lceil 1/\rho \rceil}$ in the first case and $\frac{1}{1+\lfloor 1/\rho \rfloor}$ in the second case. Hence, we can bound the partition value obtained by the algorithm as follows:

$$\mathrm{ALG} \geq r + (1-r)(1-\rho\lfloor 1/\rho \rfloor)\frac{\lceil 1/\rho \rceil}{1+\lceil 1/\rho \rceil} + (1-r)(\rho\lceil 1/\rho \rceil - 1)\frac{\lfloor 1/\rho \rfloor}{1+\lfloor 1/\rho \rfloor}.$$

Now, assuming that $\rho \in (\frac{1}{k+1}, \frac{1}{k})$ for some integer $k \geq 1$, we have that $\lfloor 1/\rho \rfloor = k$ and $\lceil 1/\rho \rceil = k+1$ and, hence,

$$\mathrm{ALG} \geq r + (1-r)\frac{1+\rho k(k+1)}{(k+1)(k+2)}.$$

Using Lemma 2, we have

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

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

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

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

For the case $\rho \geq 1$, we use completely different arguments. We will reason about the solution produced by the algorithm by considering a particular decomposition of the set of mixed bundles computed in the greedy phase. Then, using again the observation of Alon et al. [2, Lemma 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.

The decomposition is defined as follows. It takes as input a bundle with y zero-columns and x one-columns and decomposes it into y bundles containing exactly one zero-column and either $\lfloor x/y \rfloor$ or $\lceil x/y \rceil$ one-columns. Note that if x/y is not an integer, there will be $x - y\lfloor x/y \rfloor$ bundles with $\lceil x/y \rceil$ one-columns. The solution obtained after the decomposition of the solution returned by the algorithm has a very special structure as our next lemma suggests.

Lemma 4. *There exists an integer $s \geq 1$ such that each bundle in the decomposition has at least s and at most $3s$ one-columns.*

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, 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 $\mathrm{LP}(r, \rho, s)$:

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

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) to the question of how inefficient the algorithm can be. In particular, given an instance with parameters r and ρ, the quantity $\min_{\text{int } s \geq 1} \mathrm{LP}(r, \rho, s)$ yields a lower bound on the contribution of the zero-columns to the partition value and $r + \min_{\text{int } s \geq 1} \mathrm{LP}(r, \rho, s)$ 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 5. *For every $r \in [0, 1]$ and $\rho \geq 1$, $r + \min_{\text{int } s \geq 1} \mathrm{LP}(r, \rho, s) \geq \frac{9}{10}\mathrm{OPT}$.*

The next statement summarizes the discussion above.

Theorem 1. *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 2. *There exists an instance of the uniform ABMP for which the greedy algorithm computes a partition scheme with value (at most)* $9/10$ *of the optimal one.*

Proof. Consider the ABMP instance 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 $5/6$. The 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 1 does not hold any more as the following statement shows.

Lemma 6. *For every* $\epsilon > 0$*, there exists an instance of ABMP 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 ABMP instance 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 $\beta > 2$. There are four partition schemes containing a full cover (depending on the rows that contain the column-covering bundle of the first two columns) and, in each of them, the zero-column is bundled together with a 1-value entry. 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
$p_j \cdot \max_i A^B_{ij}$	$\frac{1}{\beta+3}$	$\frac{1}{\beta+3}$	$\frac{1}{\beta+3}$	$\frac{\beta}{(\beta+1)(\beta+3)}$

Now, 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, A_{32} and A_{33} are bundled together in row 3 and 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 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. The theorem follows by selecting β sufficiently large for any $\epsilon > 0$. □

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

Lemma 7. *Consider an ABMP instance 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 and the following one:*

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.

What Lemma 7 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 or $j \in A^0_i$ and j belongs to the mixed bundle of row i. The 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. Hence, the problem of computing the optimal partition scheme is equivalent to deciding the row from which each column contributes to the partition value.

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 the 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 $R_i(S)$ 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 8. *For every row i, the function R_i is non-decreasing and submodular.*

Lehmann et al. [14] studied the submodular welfare maximization problem and provided a simple algorithm that yields a 1/2-approximation of the optimal welfare. Their algorithm considers the items one by one and assigns item j to the 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). Whenever considering a zero-column, the algorithm includes it to a mixed bundle so that the increase in the partition value is maximized. Using the terminology of [2], 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 for this algorithm.

Theorem 3. *The greedy algorithm for ABMP has a tight approximation ratio of* $1/2$.

We can use the more sophisticated smooth greedy algorithm of Vondrák [20], which uses value queries to obtain the following.

Corollary 1. *There exists a* $(1 - 1/e)$*-approximation algorithm for ABMP.*

One might hope that better approximation guarantees might be possible using the $(1 - 1/e + \epsilon)$-approximation algorithm of Feige and Vondrák [11] which requires that demand queries of the form "given 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 9. *Answering demand queries associated with ABMP are NP-hard.*

References

1. Akerlof, G.A.: The market for lemons: quality uncertainty and the market mechanism. Quaterly J. Econ. **84**, 488–500 (1970)
2. Alon, N., Feldman, M., Gamzu, I., Tennenholtz, M.: The asymmetric matrix partition problem. In: Chen, Y., Immorlica, N. (eds.) WINE 2013. LNCS, vol. 8289, pp. 1–14. Springer, Heidelberg (2013)
3. Athanassopoulos, S., Caragiannis, I., Kaklamanis, C.: Analysis of approximation algorithms for k-set cover using factor-revealing linear programs. Theor. Comput. Sys. **45**(3), 555–576 (2009)
4. Athanassopoulos, S., Caragiannis, I., Kaklamanis, C., Kyropoulou, M.: An improved approximation bound for spanning star forest and color saving. In: Královič, R., Niwiński, D. (eds.) MFCS 2009. LNCS, vol. 5734, pp. 90–101. Springer, Heidelberg (2009)
5. Caragiannis, I.: Wavelength management in WDM rings to maximize the number of connections. SIAM J. Discrete Math. **23**(2), 959–978 (2009)
6. Caragiannis, I., Kaklamanis, C., Kyropoulou, M.: Tight approximation bounds for combinatorial frugal coverage algorithms. J. Comb. Optim. **26**(2), 292–309 (2013)
7. Crawford, V., Sobel, J.: Strategic information transmission. Econometrica **50**, 1431–1451 (1982)
8. Cremer, J., McLean, R.P.: Optimal selling strategies under uncertainty for a discriminating monopolist when demands are interdependent. Econometrica **53**, 345–361 (1985)
9. Cremer, J., McLean, R.P.: Full extraction of the surplus in bayesian and dominant strategy auctions. Econometrica **56**, 1247–1257 (1988)
10. Emek, Y., Feldman, M., Gamzu, I., Paes Leme, R., Tennenholtz, M.: Signaling schemes for partition value maximization. In: 13th ACM Conference on Electronic Commerce (EC), pp. 514–531 (2012)
11. Feige, U., Vondrák, J.: The submodular welfare problem with demand queries. Theor. Comput. **6**, 247–290 (2010)
12. Ghosh, A., Nazerzadeh, H., Sundararajan, M.: Computing optimal bundles for sponsored search. In: Deng, X., Graham, F.C. (eds.) WINE 2007. LNCS, vol. 4858, pp. 576–583. Springer, Heidelberg (2007)
13. Khot, S., Lipton, R., Markakis, E., Mehta, A.: Inapproximability results for combinatorial auctions with submodular utility functions. Algorithmica **52**, 3–18 (2008)
14. Lehmann, B., Lehmann, D.J., Nisan, N.: Combinatorial auctions with decreasing marginal utilities. Games Econ. Behav. **55**, 270–296 (2006)
15. Levin, J., Milgrom, P.: Online advertising: Heterogeneity and conflation in market design. Am. Econ. Rev. **100**, 603–607 (2010)
16. Milgrom, P.: Simplified mechanisms with an application to sponsored-search auctions. Games Econ. Behav. **70**, 62–70 (2010)
17. Milgrom, P.R., Weber, R.J.: A theory of auctions and competitive bidding. Econometrica **50**, 1089–1122 (1982)
18. Milgrom, P.R., Weber, R.J.: The value of information in a sealed-bid auction. J. Math. Econ. **10**, 105–114 (1982)
19. Miltersen, P. B., Sheffet, O.: Send mixed signals - Earn more, work less. In: 13th ACM Conference on Electronic Commerce (EC), pp. 234–247 (2012)
20. Vondrák, J.: Optimal approximation for the submodular welfare problem in the value oracle model. In: 40th ACM Symposium on Theory of Computing (STOC), pp. 67–74 (2008)