The Asymmetric Matrix Partition Problem

Noga Alon[1], Michal Feldman[2], Iftah Gamzu[3], and Moshe Tennenholtz[4]

[1] Tel Aviv University and Microsoft Research. Email: nogaa@tau.ac.il.
[2] Hebrew University of Jerusalem and Harvard University. Email: mfeldman@seas.harvard.edu.
[3] Yahoo! Research. Email: iftah.gamzu@yahoo.com.
[4] Microsoft Research and Technion-Israel Institute of Technology. Email: moshet@microsoft.com.

Abstract. An instance of the asymmetric matrix partition problem consists of a matrix $A \in \mathbb{R}_+^{n \times m}$ and a probability distribution p over its columns. The goal is to find a partition scheme that maximizes the resulting partition value. A partition scheme $\mathcal{S} = \{\mathcal{S}_1, \ldots, \mathcal{S}_n\}$ consists of a partition $\mathcal{S}_i$ of $[m]$ for each row i of the matrix. The partition $\mathcal{S}_i$ can be interpreted as a smoothing operator on row i, which replaces the value of each entry in that row with the expected value in the partition subset that contains it. Given a scheme $\mathcal{S}$ that induces a smoothed matrix A', the partition value is the expected maximum column entry of A'.

We establish that this problem is already APX-hard for the seemingly simple setting in which A is binary and p is uniform. We then demonstrate that a constant factor approximation can be achieved in most cases of interest. Later on, we discuss the symmetric version of the problem, in which one must employ an identical partition for all rows, and prove that it is essentially trivial. Our matrix partition problem draws its interest from several applications like broad matching in sponsored search advertising and information revelation in market settings. We conclude by discussing the latter application in depth.

1 Introduction

An instance of the *asymmetric matrix partition* problem consists of a matrix $A \in \mathbb{R}_+^{n \times m}$ of non-negative values and a probability distribution p over its columns, namely, $p \in [0,1]^m$ such that $\sum_{j=1}^{m} p_j = 1$. The objective is to find a partition scheme $\mathcal{S}$ that maximizes the resulting partition value $v_{\mathcal{S}}$. A *partition scheme* $\mathcal{S} = \{\mathcal{S}_1, \ldots, \mathcal{S}_n\}$ consists of a partition $\mathcal{S}_i$ of $[m] = \{1, \ldots, m\}$ for each row i of the matrix, namely, $\mathcal{S}_i$ is a collection of pairwise disjoint subsets $S_{i1}, \ldots, S_{ik_i} \subseteq [m]$ such that $S_{i1} \dot\cup \cdots \dot\cup S_{ik_i} = [m]$. Note that the partitions within a scheme may be different, and hence, it is referred to as an *asymmetric* scheme. The partition $\mathcal{S}_i$ can be interpreted as a smoothing operator on row i, which replaces the value of each entry in that row with the expected value in the partition subset that contains it. Formally, the smoothed value for each $j \in S_{ik}$ is $A'_{ij} = \sum_{\ell \in S_{ik}} p_\ell A_{i\ell} / \sum_{\ell \in S_{ik}} p_\ell$. Given a partition scheme $\mathcal{S}$ that induces a smoothed matrix A', the resulting *partition value* is the expected maximum column entry, that is, $v_{\mathcal{S}} = \sum_{j \in [m]} p_j \cdot \max_{i \in [n]} A'_{ij}$. The *contribution* of a column j to the partition value is $p_j \cdot \max_{i \in [n]} A'_{ij}$, and similarly, $\text{argmax}_{i \in [n]} A'_{ij}$ is referred to as the entry of column j that contributes to the partition value.

For the purpose of illustrating the above setting, let us focus on the simple scenario in which the input instance consists of an $n \times n$ matrix such that all the entries in the first column have a value of 1 and all remaining entries have a value of 0. Furthermore, the probability distribution over the columns of this matrix is uniform. One partition scheme that naturally comes to mind is the identity scheme, which results in a smoothed matrix that is identical to the original matrix. This identity scheme sets all the partitions to consist of singletons, namely, each $\mathcal{S}_i = \{\{1\}, \ldots, \{n\}\}$. One can easily validate that the resulting partition value in this case is $1/n$. Another extreme partition scheme is the one in which all partitions consist of one subset, that is, each $\mathcal{S}_i = \{[n]\}$. This scheme gives rise to a smoothed matrix in which all the entries of each row have the same value. In our case, all the entries of the resulting matrix are $1/n$, and accordingly, it is easy to validate that the partition value is again $1/n$. Finally, one can demonstrate that there is a partition scheme that exhibits a significant improvement over the above-mentioned schemes. This scheme consists of the partitions $\mathcal{S}_i = \{\{1, i\}, [n] \setminus \{1, i\}\}$, namely, it joins together the 1-value of each row $i \neq 1$ with the 0-value of column i in that row, resulting in a smoothed value of $1/2$ for both entries. One can verify that the resulting partition value in this case is roughly $1/2$. The above scenario is presented in the figure below.

$$\begin{pmatrix} 1 & 0 & \cdots & 0 \\ 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 0 & \cdots & 0 \end{pmatrix} \qquad \begin{pmatrix} \boxed{1} & 0 & \cdots & 0 \\ \boxed{1} & \boxed{0} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ \boxed{1} & 0 & \cdots & \boxed{0} \end{pmatrix} \qquad \begin{pmatrix} 1 & 0 & \cdots & 0 \\ \frac{1}{2} & \frac{1}{2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ \frac{1}{2} & 0 & \cdots & \frac{1}{2} \end{pmatrix}$$

Fig. 1. Given the input matrix on the left with a uniform distribution over its columns, one can utilize the partition scheme illustrated on the middle, and obtain the smoothed matrix on the right. Note that the boxes in each row of the middle matrix represent entries that are joined together in the same subset; the remaining entries of each row are clustered together in a different subset.

Application I: personalized broad matching in sponsored search advertising. The asymmetric matrix partition problem draws its interest from several applications. One such application relates to sponsored search advertising, namely, advertising on a web search result page, where the ads are driven by the originating query. In the basic model, there are advertisers, each of which has keywords relevant to her ad. Each advertiser also associates some valuation with each of her keywords, indicating the gain she derives when a user clicks on her ad. This valuation underlies a bid that the advertiser reports to the search engine, expressing the maximum amount that she is willing to pay for a click. When a user queries the search engine for some keyword, the engine runs an auction among all the advertisers interested in that keyword. The advertiser that wins this auction is allocated the ad slot, and she is required to pay some amount if the user clicks on her ad. This amount is determined by her bid and the payment rule of the engine.

Advertisers can realistically only identify a small set of keywords due to the effort involved, and therefore, search engines recently introduced broad matching. This feature enables an advertiser to automatically target a broader range of queries that the search engine deems relevant to match her ad, and not only the keywords specified by her. Such relevant queries can be modifications of the specified keywords (like synonyms, singular and plural forms, misspellings, reordering, etc.), or can even be a completely different set of keywords, which are conceptually related to the specified keywords. This feature clearly has potential to help advertisers reach wider audience, while spending less time on building their keyword lists. On the other hand, a search engine can utilize the flexibility in expanding the set of keywords specified by an advertiser to optimize its revenue. Understanding the power of flexibility in broad matching seems an interesting research goal.

We consider a stylized non-strategic version of broad matching. In the underlying setting, there is a single ad slot, and a set of advertisers, each of which interested in one keyword from a set of possible keywords $\{k_1, \ldots, k_m\}$, where keyword k_j is queried by users with probability p_j. The search engine keeps a relevance distance measure between keywords, $\alpha(i, j)$, that has the following semantics: if an advertiser has valuation v for her specified keyword k_i, then her valuation for each keyword k_j is $v \cdot \alpha(i, j)$. The goal is to develop a personalized broad matching scheme that maximizes the expected revenue of the search engine. Specifically, we are interested in a scheme that assigns each advertiser a partition of keywords to disjoint subsets, such that all keywords in each subset are automatically bid with the expected valuation in that subset whenever a user queries a keyword from that subset. We assume that the search engine knows the valuation that each advertiser has for her specified keyword, i.e., a non-strategic setting in which there is no need to incentivize the advertisers, and a winning advertiser pays her expected valuation. Consequently, given a query, the search engine selects the advertiser that has the highest bid. One can validate that our asymmetric matrix partition problem captures the problem of designing a personalized broad matching that maximizes the expected revenue.

Application II: signaling in take-it-or-leave-it sales. Another application of the asymmetric matrix partition problem relates to a question of information revelation in market settings. In many sale scenarios, a seller has much more accurate information about an item for sale than the buyers. As an example, consider a used-car dealer or an Internet liquidation site, both of which receive or purchase items for sale. The seller in these scenarios may have quite adequate information about the particular item for sale (e.g., by checking it in detail), while the potential buyers may only have probabilistic information about the item, relying, for example, on some publicly-available statistical information. It seems of the essence to study how a seller can utilize her informational superiority to optimize her revenue.

The above-mentioned scenario can be modeled by considering a take-it-or-leave-it sale of a probabilistic item among multiple buyers. More precisely, a single item is chosen randomly from a set of m possible items according to some known probability distribution p, and the seller approaches a buyer with a monetary offer of delivering the item for a specified payment. There are n buyers, each of which has her own valuation for every item in the set. While the buyers only know the probability distribution over the items, the seller knows the actual realization of the probabilistic item. In an attempt to increase her revenue, the seller may partially reveal some information about the item to the buyers. The question that concerns us is how much information should the seller reveal to every buyer in order to maximize her expected revenue.

The information revelation is materialized by means of a buyer-specific signaling scheme. For each buyer, the seller partitions the set of items into pairwise disjoint subsets, and reports this partition to the buyer. After the signaling scheme has been declared, an item j is randomly chosen by nature, and the seller reveals to each buyer i, the subset that contains j according to i's partition. Upon being signaled, a buyer can update her belief regarding the probability distribution p conditioned on the choice of some item in her signaled subset. A key assumption in our model is that the buyers are unaware of the environment, namely, each buyer knows her own valuation and partition, but is unaware of the existence of the other buyers and their associated valuations and partitions. Hence, the conditional probability of every item j that is contained in a signaled subset is the ratio between p_j and the overall probability of items in that subset, and 0 in case j is not in the signaled subset. It is clear that the maximal take-it-or-leave-it offer that a buyer will accept is her expected valuation for the item under the new probability distribution induced by the received signal. Consequently, upon the realization of an item, the seller will choose to make such offer to a buyer that has the highest expected valuation. One can validate that our asymmetric matrix partition problem captures the task of designing a buyer-specific signaling scheme that maximizes the expected revenue.

Our contribution. We begin by studying the asymmetric matrix partition problem when the input matrix is binary, namely, $A \in \{0,1\}^{n \times m}$. We prove that this seemingly simple setting is already APX-hard when the probability distribution p is uniform. Specifically, we show a gap-preserving reduction that proves that the problem is NP-hard to approximate to within a factor of 1.0001. We also establish that the binary setting admits a constant factor approximation; thus, settling the complexity of this setting to within constant factors. In particular, we demonstrate that there is a 1.775-approximation algorithm when p is uniform, and there is a 13-approximation algorithm when p is arbitrary. We further study several interesting special scenarios. For example, we prove that when the number of rows n is fixed then the uniform distribution case can be solved to optimality in polynomial-time, whereas the general distribution case remains NP-hard even when $n = 4$. This result separates the uniform distribution setting from the general distribution setting. The specifics of these results are presented in Section 2. We then consider the problem in its utmost generality, that is, when the input matrix $A \in \mathbb{R}_+^{n \times m}$ consists of arbitrary non-negative values. We present a 2-approximation algorithm for the case that p is uniform, and a logarithmic approximation when p is arbitrary under some practical assumptions. These results appear in Section 3. Later on, in Section 4, we discuss the symmetric version of our problem in which one must employ an identical partition for all rows. We demonstrate that this problem is essentially trivial, and establish a tight bound on the advantage that asymmetric schemes have over symmetric ones. Finally, we formally model the application of signaling in take-it-or-leave-it sales with its connection to our problem, and discuss some of our modeling decisions. These application details are provided in Section 5. Due to space constraints, most proofs are omitted from this extended abstract and may be found in the full version of the paper.

2 The Binary Matrix Case

In this section, we study the problem when the input matrix is binary, namely, $A \in \{0,1\}^{n \times m}$. We prove that this setting is APX-hard even when the probability distribution p is uniform. We also establish that this setting admits a constant factor approximation; thus, settling the complexity of this setting to within constant factors. We further study several interesting special scenarios.

We begin by introducing a notation and terminology that will be used in the remainder of this section. Let $C^+ = \{j \in [m] : \exists i \text{ such that } A_{ij} = 1\}$ be the set of columns that consist of at least one 1-value entry, and $C^0 = [m] \setminus C^+$ be the set of remaining all-zero columns. Moreover, let $r = \sum_{j \in C^+} p_j$ be the total probability of the columns in C^+. Similarly, we denote the set of columns that have a 1-value entry in row i by $C_i^+ = \{j \in [m] : A_{ij} = 1\}$, and use $r_i = \sum_{j \in C_i^+} p_j$ to denote their total probability. We say that a partition scheme $\mathcal{S}$ *covers* C^+ if it covers each of the columns in C^+. A column $j \in C^+$ is said to be *covered* by $\mathcal{S}$ if there is some row i such that $A_{ij} = 1$ and the partition scheme consists of a singleton subset of column j in row i, namely, $\{j\} \in \mathcal{S}_i$. Note that a partition scheme that covers C^+ can be easily computed in polynomial-time. Finally, we say that a subset is *mixed* if it consists of both 1-value and 0-value entries.

We now turn to identify several interesting structural properties of partition schemes for the binary case. These properties will be utilized later when establishing our primary technical results.

Lemma 1. *Let S, T be two disjoint subsets of columns, and let S^+ (resp., T^+) and S^0 (resp., T^0) be the respective 1-value entries and 0-value entries of S (resp., T) in some row i. Suppose that all the entries of S^0 and T^0 contribute to the partition value when the partition of row i consists of S and T. Then, the overall contribution of those entries when the partition consists of a unified subset $S \cup T$ is at least as large.*

Note that a useful corollary of the above lemma is that given some fixed covering of C^+, the optimal way to complete the partition scheme is to join together all the remaining 1-value entries of each row in a single subset with some additional 0-value entries. We can also utilize the above lemma and prove the following.

Lemma 2. *Given an instance of the asymmetric matrix partition problem in which A is binary and p is uniform, there is an optimal solution that covers C^+.*

2.1 A uniform distribution

We study the binary matrix setting when the distribution over the columns is uniform, namely, each $p_j = 1/m$. We establish that this seemingly simple setting is already APX-hard, that is, it is NP-hard to approximate to within some constant. On the algorithmic side, we identify a simple algorithmic procedure that guarantees 2-approximation, and then develop an algorithm that achieves a better approximation ratio. We also prove that the case that the number of rows n is constant can be solved to optimality in polynomial-time. We emphasize that for simplicity of presentation, we neglect the uniform probability term $1/m$ from the partition value contribution terms in the rest of this subsection.

Theorem 1. *Given an instance of the asymmetric matrix partition problem in which A is binary and p is uniform, it is NP-hard to attain an approximation ratio better than* 1.0001.

Approximation algorithms We begin by presenting a simple 2-approximation algorithm for the problem under consideration. Later on, we develop a different algorithm that attains an improved approximation ratio. Our 2-approximation algorithm begins by covering C^+. This ensures that the contribution of the columns of C^+ to the resulting partition value is exactly r. Subsequently, the algorithm goes over the rows, one after the other, and for each row that has ℓ remaining 1-value entries (after the covering), it creates a subset in that row that consists of these entries and ℓ entries of distinct all-zero columns. In case there are no more all-zero columns left to match to some row then this step ends. Finally, all remaining entries of each row are clustered together.

For the purpose of analyzing this algorithm, notice that a straight-forward upper bound on the partition value of the optimal scheme is $\mathrm{OPT} \leq \min\{1, \sum_{i=1}^n r_i\}$. Now, consider the following two complementary cases: (case 1) if we matched every all-zero column to some row, then the contribution of each column j is 1 if $j \in C^+$, or at least $1/2$ if $j \in C^0$. Hence, the partition value is at least $r + (1-r)/2 \geq 1/2 \geq \mathrm{OPT}/2$; (case 2) if we did not match all all-zero columns, then the partition value is at least $r + (\sum_{i=1}^n r_i - r)/2 \geq \sum_{i=1}^n r_i/2 \geq \mathrm{OPT}/2$. This implies that the partition scheme achieves 2-approximation.

A greedy completion procedure. Before we turn to improve the above algorithm, we study the following greedy procedure that given a fixed covering of C^+ completes the partition scheme by matching all-zero columns to partition subsets. The *greedy procedure* begins by associating a subset S_i to each row i. This subset is initialized with all the columns corresponding to 1-value entries in row i that were not used in the covering of C^+. Then, it proceeds by going over the all-zero columns, one after the other, and adding a column to the subset S_i that maximizes the marginal contribution from the all-zero columns. Specifically, the *marginal contribution* of some all-zero column that is added to a subset that already consists of x and y columns corresponding to 0-value entries and 1-value entries, respectively, is

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

Note that $\Delta(x,y) \geq 0$ for any non-negative x, y, and that Δ is monotonically non-increasing in x for any fixed y, that is, $\Delta(x,y) \geq \Delta(x+1,y)$. The following lemma establishes that once C^+ is covered in some way, the greedy procedure yields the optimal contribution from the all-zero columns. Notice that this result implies, in conjunction with Lemma 2, that the computational hardness of the underlying setting of the problem resides in finding the right way to cover C^+.

Lemma 3. *Given some fixed covering of C^+, the greedy procedure yields the optimal contribution from the all-zero columns.*

The greedy procedure can be leveraged to construct a 1.775-approximation algorithm for our problem. We emphasize that our main effort is to improve upon the previous 2-approximation algorithm, and we have not tried to optimize the constants in our analysis. Let $r^* = 0.127$ and $\sigma^* = 2(1 - r^*)/3 = 0.582$. The algorithm computes a partition scheme according to the following cases, which depend on the values of r and $\sum_{i=1}^{n} r_i$ in a given instance:

Case I: when $r \geq r^*$. The algorithm first covers C^+ in some arbitrary way. Then, the algorithm goes over the rows, one after the other, and for each row that has ℓ remaining 1-value entries after the covering, it creates a subset in that row that consists of these entries and ℓ entries of distinct all-zero columns. Note that in case there are no more all-zero columns left to match to some row then this step ends. Finally, all the remaining entries of each row are clustered together.

Case II: when $r < r^*$ and $\sum_{i=1}^{n} r_i \leq \sigma^*$. The algorithm forms a subset on top of every 1-value entry. Specifically, given a 1-value entry (i, j), the algorithm forms a subset in row i that consists of column j and some additional distinct all-zero columns. Half of the 1-value entries are clustered together with two distinct two all-zero columns, and the other half of the 1-value entries are clustered together with a single all-zero column. Then, all the remaining entries of each row are clustered together.

Case III: when $r < r^*$ and $\sum_{i=1}^{n} r_i > \sigma^*$. The algorithm executes the previously-mentioned greedy procedure over the given instance (without covering C^+ first). After this procedure ends, all the remaining entries of each row are clustered together.

Theorem 2. *Given an instance of the asymmetric matrix partition problem in which A is binary and p is uniform, our algorithm computes a partition scheme whose resulting partition value is a 1.775-approximation for the optimal one.*

An optimal solution for a fixed number of rows We prove that an optimal partition scheme can be computed in polynomial-time when the number of rows n is fixed. We emphasize that this result separates the uniform distribution setting from the general distribution setting since we establish that the latter setting is NP-hard in Theorem 4.

Theorem 3. *Given an instance of the asymmetric matrix partition problem in which A is binary and p is uniform, an optimal partition scheme can be constructed in polynomial-time when n is fixed.*

2.2 A general distribution

We next study the binary matrix setting when the distribution over the columns is arbitrary. We first demonstrate that this setting is NP-hard even when the number of rows is fixed. This result separates this setting from the uniform distribution setting, which admits a polynomial-time optimal solution when the number of rows is fixed. Later on, we present a constant factor approximation algorithm for this setting.

Theorem 4. *Given an instance of the asymmetric matrix partition problem in which A is binary and p is general, and a positive number α, it is NP-hard to determine if there is a partition scheme whose resulting partition value is at least α, even when $n = 4$.*

An approximation algorithm We develop a polynomial-time constant approximation algorithm for the problem under consideration. Specifically, we present three algorithms whose performance depends on different parameters of the input instance and the optimal solution. We then demonstrate that given any input instance, one of these algorithms is guaranteed to compute a 13-approximation partition scheme. Hence, by executing all three algorithms and selecting the scheme that attains the maximal resulting partition value, we achieve a 13-approximation solution.

Recall that C^+ denotes the set of columns that consist of at least one 1-value entry, C^0 is the set of remaining all-zero columns, and C_i^+ marks the set of columns having 1 in row i. Furthermore, recall that r is the total probability of columns in C^+, and r_i is the total probability of columns in C_i^+. Note that the optimal partition value can be trivially bounded by $\mathrm{OPT} \leq r + \mathrm{OPT}_0$, where OPT_0 is the overall contribution of the all-zero columns of C^0 to the optimal partition value.

Algorithm 1. The first algorithm attains to the case in which the input instance has a relatively large r, namely, $r \geq \mathrm{OPT}_0/12$. In this case, a constant approximation can be obtained by simply covering C^+. That is, for every column $j \in C^+$, the algorithm arbitrarily selects some row i such that $A_{ij} = 1$ and forms a singleton subset of entry j in row i. Then, all the remaining entries of each row are clustered together. One can easily validate that the resulting scheme has a partition value which is a 13-approximation to the optimal one. This follows since the resulting partition value is at least r, while $\mathrm{OPT} \leq r + \mathrm{OPT}_0 \leq 13r$.

In the remainder of the subsection, we focus on the case that r is relatively small, namely, $r < \mathrm{OPT}_0/12$. Since r is small, we concentrate on designing partition schemes that yield high contribution from the all-zero columns. We say that an all-zero column j is *large* for a row i in case $p_j \geq r_i$; otherwise, j is said to be *small* for i. We consider two complementary cases and develop constant factor approximation algorithms for both. The first case is when a large fraction of the contribution of all-zero columns to the optimal partition value comes from such columns that are large for the rows that realize their contribution.

Algorithm 2. The algorithm begins by constructing an undirected bipartite graph $G = (V_R, V_L, E)$ with a weight function $w : E \to \mathbb{R}_+$ on its edges. Specifically, V_R is a set of n vertices that correspond to the rows, V_L is a set of $|C^0|$ vertices that correspond to the all-zero columns, and $E = \{(i,j) \in V_R \times V_L : r_i \leq p_j\}$ is the edge set. Moreover, the weight function sets $w(i,j) = r_i$, for every $(i,j) \in E$. With these definitions in mind, the algorithm finds a maximal weighted matching M with respect to the constructed bipartite graph. Then, for each edge $(i,j) \in M$, the algorithm forms the subset $C_i^+ \cup \{j\}$ in row i. Subsequently, all the remaining entries of each row are clustered together.

Lemma 4. *Algorithm 2 computes a partition scheme that yields at least $1/2$ of the optimal contribution of all-zero columns that are large for the rows that realize their contribution.*

Lemma 4 implies that in case that at least $1/6$ of the optimal contribution of all-zero columns comes from such columns that are large for the rows that realize their contribution then we obtain a 13-approximation solution. Formally, one can utilize Lemma 4 to claim that the partition value of the computed scheme is at least $\mathrm{OPT}_0/12$. On the other hand, $\mathrm{OPT} \leq r + \mathrm{OPT}_0 \leq 13/12 \cdot \mathrm{OPT}_0$, where the last inequality follows from the assumption that $r < \mathrm{OPT}_0/12$. We now turn to consider the remaining case in which at least $5/6$ of the optimal contribution of all-zero columns comes from such columns that are small for the rows that realize their contribution.

Algorithm 3. Similarly to the previous algorithm, this algorithm forms two subsets for each row; one mixed subset and an additional subset that consists of the remaining row entries. The mixed subset of row i consists of the columns in C_i^+ and some additional all-zeros columns. To decide which all-zeros columns are added to each mixed subset, the algorithm goes over the all-zero columns in an arbitrary order, and adds the column j to the mixed subset of row i if (1) j is small for i and (2) the total probability of the all-zero columns already added to this subset is no more than r_i. We emphasize that each column is added to at most one mixed subset, and a column is neglected only if the algorithm cannot add it to any of the mixed subsets.

Lemma 5. *Algorithm 3 computes a partition scheme that yields at least* $1/10$ *of the optimal contribution of all-zero columns that are small for the rows that realize their contribution.*

Lemma 5 implies that in case that at least $5/6$ of the optimal contribution of all-zero columns comes from such columns that are small for the rows that realize their contribution then we also attain a 13-approximation solution. More precisely, one can utilize Lemma 5 to claim that the partition value of the computed scheme is at least $\mathrm{OPT}_0/12$, while $\mathrm{OPT} \leq r + \mathrm{OPT}_0 \leq 13/12 \cdot \mathrm{OPT}_0$. Reviewing the algorithms and the case analysis, we can conclude with the following theorem.

Theorem 5. *Given an instance of the asymmetric matrix partition problem in which A is binary and p is general, there is an algorithm that computes a partition scheme whose resulting partition value is a* 13*-approximation for the optimal one.*

3 The General Matrix Case

In this section, we study the problem in its utmost generality, i.e., when the input matrix $A \in \mathbb{R}_+^{n \times m}$ consists of arbitrary non-negative values. We develop a constant factor approximation algorithm for the case that the probability distribution p over the columns is uniform, and a logarithmic approximation for the general case under some practical assumptions.

3.1 A uniform distribution

We present an algorithm that computes a partition scheme whose resulting partition value is a 2-approximation for the optimal one. Let M be the set of m largest entries in the matrix A. Our algorithm first forms a singleton subset of every entry $(i, j) \in M$ that is maximal for the corresponding column. In case there are several maximal entries for some column then one of them is selected arbitrarily. We say that column j was *covered* if there was an entry (i, j) that was clustered as a singleton. Subsequently, for every entry $(i, j) \in M$ that was not clustered in the first step, the algorithm forms a subset in row i consisting of column j and a distinct column that was not covered. Then, all the remaining entries of each row are clustered together.

Theorem 6. *Given an instance of the asymmetric matrix partition problem in which A is general and p is uniform, our algorithm computes a partition scheme whose resulting partition value is a 2-approximation for the optimal one.*

3.2 A general distribution

We present an algorithm that achieves a logarithmic approximation under some practical assumptions. Specifically, the algorithm achieves $O(\log m)$-approximation if the column probabilities are at most polynomially small, namely, when each $p_j \geq 1/m^c$ for some constant c.

Let $A_{\max}$ be the value of the largest entry of an input matrix A. Our algorithm begins by manipulating the matrix A to construct the matrix B as follows: All the entries whose value is smaller than $A_{\max}/m^{c+2}$ are replaced by 0, and all the values of the remaining entries are rounded down to the closest power of 2. For example, if $2^{-k} \leq A_{ij} < 2^{1-k}$ then $B_{ij} = 2^{-k}$. Notice that after this manipulation, the matrix B is populated with at most $K = O(\log m^c) = O(\log m)$ types of positive values $\{v_1, \ldots, v_K\}$ in addition to a 0-value. As a result, we can express the matrix B as a sum of K matrices where the kth matrix consists of the values $\{0, v_k\}$. That is, the kth matrix has a value of v_k in each entry that B has a value of v_k, and 0 in all remaining entries. Each of the K matrices, together with the probability distribution p, can be considered to be an instance of our problem with a binary matrix and a general distribution. Hence, we can apply the algorithm from Theorem 5 on each of these K instances to obtain K partition schemes. Finally, the algorithm selects the partition scheme that obtains the maximal partition value from the original instance.

Theorem 7. *Given an instance of the asymmetric matrix partition problem in which A and p are general such that each $p_j \geq 1/m^c$ for some constant c, our algorithm computes a partition scheme whose resulting partition value is a $O(\log m)$-approximation for the optimal one.*

4 Symmetric Partition Schemes

In this section, we discuss the symmetric version of our matrix partition problem, and most notably, compare between the performance guarantees of symmetric and asymmetric partition schemes. The *symmetric matrix partition* problem is identical to the asymmetric matrix partition problem with the exception that the underlying partition scheme must be symmetric. A *symmetric* partition scheme consists of a single partition $\mathcal{S}$ of $[m]$ that is used as the smoothing operator of all the rows. A variant of the symmetric matrix partition problem has been studied in a series of works [7, 3, 4, 14]; A more detailed discussion is given in Section 5.

An easy argument shows that the symmetric matrix partition problem is essentially trivial as the partition scheme that consists only of singletons always achieves the optimal partition value. To establish this argument, suppose by way of contradiction that the partition scheme of singletons does not attain an optimal outcome. Consider the optimal partition scheme $\mathcal{S}$. This scheme must consist of a subset $S \in \mathcal{S}$ whose cardinality is greater than 1. Notice that the contribution of all the columns in S to the resulting partition value is realized in the same row i. The overall contribution of those columns is exactly $\sum_{j \in S} p_j A_{ij}$. Now, observe that if one replaces the instance of S in the optimal partition scheme with the collection of singleton subsets of the columns in S, the overall contribution of the columns in S may only improve to $\sum_{j \in S} p_j \cdot \max_{i \in [n]} A_{ij}$, and the contribution of any other column in $[m] \setminus S$ does not change. Applying this argument repeatedly as long as $\mathcal{S}$ has subsets whose cardinality is greater than 1 results in an optimal partition scheme that consists only of singletons; a contradiction.

In light of this state of affairs, we next focus on quantifying the advantage that asymmetric partition schemes have over symmetric schemes. Given an instance of our matrix partition problem, let $\mathrm{OPT}_{\mathrm{sym}}$ and $\mathrm{OPT}_{\mathrm{asym}}$ denote the optimal partition values that can be achieved by symmetric and asymmetric partition schemes, respectively. Clearly, $\mathrm{OPT}_{\mathrm{asym}}/\mathrm{OPT}_{\mathrm{sym}} \geq 1$. However, we are also interested to establish a tight upper bound on this ratio.

Lemma 6. *Given a matrix partition instance in which A and p are general, the ratio $\mathrm{OPT}_{\mathrm{asym}}/\mathrm{OPT}_{\mathrm{sym}} \leq m$. Furthermore, there are instances for which the ratio can be arbitrarily close to m.*

5 An Application: Signaling in Take-It-Or-Leave-It Sales

In this section, we formally model the application of signaling in take-it-or-leave-it sales, and explain its connection to our asymmetric matrix partition problem. Later on, we discuss the previous literature on signaling and some of our modeling decisions.

5.1 The model

A *probabilistic single-item sale* is formally depicted by a valuations matrix $A \in \mathbb{R}_{+}^{n \times m}$, and a probability distribution p over its columns. More precisely, there are n *agents* and m distinct indivisible *items*. Each entry A_{ij} of the matrix captures the *valuation* of the row-agent i for the column-item j. We assume that each agent knows her valuation vector but is *unaware* of the rest of the valuation matrix. A single item j is chosen by nature according to the distribution p and then offered for sale. This one-time sale is conducted via a personalized *take-it-or-leave-it* rule: The *seller* gives a take-it-or-leave-it offer to some agent i. If the selected agent is interested, the chosen item is sold to her for the suggested price.

A signaling scheme. Although the agents know the distribution p, they do not know its actual realization, which is only observed by the seller. In an attempt to increase her expected *revenue*, the seller may partially reveal the realization to the agents. This is performed via the following *asymmetric signaling scheme*: For every agent i, the seller partitions the items into a collection of pairwise disjoint subsets $S_{i1}\dot\cup\cdots\dot\cup S_{ik_i} = [m]$, and reports this partition to agent i; we denote the partition of agent i by $\mathcal{S}_i$. Crucially, the seller can use different partitions for different agents, i.e., a buyer-specific signaling scheme. We emphasize that the seller decides on a signaling scheme prior to nature's random choice of an item. When item j is randomly chosen, every agent i is *signaled* with the subset S_{ik} that contains j. The agent can then update her belief to the probability distribution p conditioned on the choice of some item in S_{ik}. In other words, each agent i knows that none of the items in $[m] \setminus S_{ik}$ was chosen, and can calculate the conditional probability $\mathbb{P}(j : S_{ik}) = p_j/\mathbb{P}(S_{ik})$, for every $j \in S_{ik}$.

The optimization problem. Consider some probabilistic single-item sale $\langle A, p\rangle$ and a signaling scheme $\mathcal{S} = (\mathcal{S}_1, \ldots, \mathcal{S}_n)$. Clearly, the maximal take-it-or-leave-it offer that agent i will accept under the signal S_{ik} is given by $\mathbb{E}_p[A_{ij} : S_{ik}]$. Therefore, given an asymmetric signaling scheme $\mathcal{S}$, when item j is randomly chosen, the seller will choose to make a take-it-or-leave-it offer to agent i that maximizes $\mathbb{E}_p[A_{ij} : S_{ik}]$, where S_{ik} is the subset that contains j for agent i. In what follows, we denote by $S^i(j)$ the subset S_{ik} of agent i that contains j. Hence, the expected *revenue* of the seller is given by

$$\sum_{j\in[m]} p_j \cdot \max_{i\in[n]} \left\{ \sum_{\ell\in S^i(j)} \mathbb{P}(\ell : S^i(j)) \cdot A_{i\ell} \right\} = \sum_{j\in[m]} p_j \cdot \max_{i\in[n]} \left\{ \frac{\sum_{\ell\in S^i(j)} p_\ell A_{i\ell}}{\sum_{\ell\in S^i(j)} p_\ell} \right\} . \quad (1)$$

This raises the following combinatorial optimization problem: given a probabilistic single-item sale $\langle A, p\rangle$, construct the asymmetric signaling scheme $\mathcal{S}$ that maximizes the expected revenue.

We believe that the mapping of the above problem to the asymmetric matrix partition problem is straightforward. Yet, we wish to emphasize that the expression that is maximized in Equation 1 is essentially the smoothed value A'_{ij}, which was defined when we formalized the problem.

5.2 Related work

The literature on signaling in economics is very broad. Our approach can be viewed as related to the study of strategic information transmission, originated in the seminal work of Crawford and Sobel [2]. More specifically, our approach deals with the idea that a seller knows some information about the valuations of the buyers, via information about the item, and may use strategic information transmission to exploit this knowledge [11]. As in that work, our model depart from the classic literature of Milgrom and Weber [12, 13], who showed the superiority of full revelation of information. Information revelation in online markets has been recently studied also in [9], where it is shown that in an environment with multiple publishers, a publisher may prefer not to share user information with the advertiser, due to information leakage, where the advertiser may target the same user through a cheaper publisher. Our approach has some of the flavor of the work on the value of information in conflicts [10]. One special distinction of the current work is its focus on algorithmic issues.

This work is also closely related to the study of revenue maximization via signaling in second-price auctions [7, 3, 4, 14]. The are few fundamental differences between the model considered by those papers and ours. First, rather than a take-it-or-leave-it sale, the sale is conducted by means of a *second-price* auction; i.e., each agent places her bid and the chosen item is sold to the bidder that placed the highest bid for the price of the second highest bid. Second, rather than a *buyer-specific asymmetric* signaling scheme, the signaling is performed via a *symmetric* partition, where the auctioneer partitions the items into pairwise disjoint clusters and reports this partition to all the bidders.

A point of interest in our approach is the assumption of unawareness. Classical economic and game-theoretic approaches assume that buyers are aware of other buyers and the take-it-or-leave-it offers they may be given. As a result, the (deduced) probabilistic information that each buyer holds about the item may be affected by her awareness to the cases when she is given an offer versus the cases other buyers are given an offer. While this approach is natural, we believe it is interesting to consider the complementary attitude of competition-unaware buyers, who disregard the existence of other buyers. This approach clearly gives much power to the seller. Indeed, the fact that decision-makers may be unaware of aspects of a strategic situation, and in particular, of actions and even existence of other players is a puzzle game theorists were concerned with. Most efforts so far have been concentrated on trying to find general models that incorporate such reasoning. For an example of the modeling challenges encountered when considering unaware agents, one may consult the work of Halpren and Rego [8] on extensive games with possibly unaware players, or the work by Feinberg [5] on games with unawareness. Our approach is complementary, as it emphasizes the combinatorial and algorithmic issues that arise in such settings.

5.3 Awareness vs. unawareness

A natural question one may ask is what would be the ramifications when considering a setting in which the agents are aware of one another, and more generally, when the whole setting is common-knowledge. Interestingly, in what follows, we observe that in the latter case, the seller maximizes her revenue by fully revealing all information, essentially revealing the realization of the probabilistic item. This result is in the spirit of the famous 'Linkage Principle' of Milgrom and Weber [12, 13].

In a *competition-aware* model, each agent is aware of the valuations of other agents and the signaling scheme that the seller runs. As a result, any agent can calculate, for each item, which agent will be given the take-it-or-leave-it offer and in which price. Suppose some agent i is signaled a subset S. How would she evaluate her expected value? Clearly, agent i should compute the expectation only over the items $j \in S$ such that she would be given the take-it-or-leave-it offer. For that reason, when analyzing an asymmetric partition scheme in the competition-aware model, it can be assumed without loss of generality that if some subset is a winning subset for some item then it is a winning subset for all its items. One can also verify that this implies that there is only one winning subset for each agent. Using these observations, we next show that in the competition-aware model, an optimal signaling scheme obtains the same expected revenue as a signaling scheme that is symmetric and partitions the items into singleton subsets. Conceptually, this implies that the best interest of the seller is to fully reveal which item arrived when the buyers are competition-aware.

Lemma 7. *In the competition-aware model, the optimal asymmetric signaling scheme obtains the same expected revenue as a signaling scheme that is symmetric and partitions the items into singleton subsets.*

References

1. M. Chlebík and J. Chlebíková. Approximation Hardness for Small Occurrence Instances of NP-Hard Problems. *Italian Conference on Algorithms and Complexity* 2003, 152–164.
2. V. Crawford and J. Sobel. Strategic Information Transmission. Econometrica, 50(6), 1431-1451, 1982.
3. Y. Emek, M. Feldman, I. Gamzu, R. Paes Leme and M. Tennenholtz. Signaling schemes for revenue maximization. *ACM Conference on Electronic Commerce* 2012, 514–531.
4. Y. Emek, M. Feldman, I. Gamzu and M. Tennenholtz. Signaling Schemes for Revenue Maximization. *AdAuctions Workshop*, 2011.
5. Y. Feinberg. Games with Unawareness. Manuscript, 2009.
6. M. R. Garey and D. S. Johnson. Computers and Intractability: A Guide to the Theory of NP-Completeness. W. H. Freeman, 1979.
7. A. Ghosh, H. Nazerzadeh and M. Sundararajan. Computing optimal bundles for sponsored search. *International Conference on Internet and Network Economics*, 2007, 576–583.
8. J. Y. Halpern and L. C. Rego. Generalized Solution Concepts in Games with Possibly Unaware Players. International Journal of Game Theory, 41(1), 131–155, 2012.
9. M. Mahdian, A. Ghosh, R. P. McAfee and S. Vassilvitskii. To match or not to match: economics of cookie matching in online advertising. *ACM Conference on Electronic Commerce*, 2012, 741–753.
10. M. Kamien, Y. Tauman and S. Zamir. Information Transmission. Game Theory and Applications, 273-281. Academic Press.
11. T. R. Kaplan and S. Zamir. The Strategic Use of Seller Information. Private-Value Auctions, Hebrew University Center For Rationality Working Paper No. 221, 2000.
12. P. R. Milgrom and R. J. Weber. A Theory of Auctions and Competitive Bidding. Econometrica, 50(5), 1089-1122, 1982.
13. P. R. Milgrom and R. J. Weber. The Value of Information in a Sealed-Bid Auction. Journal of Mathematical Economics, 10(1), 105-114, 1982.
14. P. B. Miltersen and O. Sheffet. Send mixed signals: earn more, work less. *ACM Conference on Electronic Commerce* 2012, 234–247.