ℓ_1-sparsity Approximation Bounds for Packing Integer Programs *

Chandra Chekuri Kent Quanrud Manuel R. Torres

February 26, 2019

Abstract

We consider approximation algorithms for packing integer programs (PIPs) of the form $\max\{\langle c, x\rangle : Ax \leq b, x \in \{0,1\}^n\}$ where c, A, and b are nonnegative. We let $W = \min_{i,j} b_i/A_{i,j}$ denote the width of A which is at least 1. Previous work by Bansal et al. [1] obtained an $\Omega(\frac{1}{\Delta_0^{1/\lfloor W \rfloor}})$-approximation ratio where Δ_0 is the maximum number of nonzeroes in any column of A (in other words the ℓ_0-column sparsity of A). They raised the question of obtaining approximation ratios based on the ℓ_1-column sparsity of A (denoted by Δ_1) which can be much smaller than Δ_0. Motivated by recent work on covering integer programs (CIPs) [4, 6] we show that simple algorithms based on randomized rounding followed by alteration, similar to those of Bansal et al. [1] (but with a twist), yield approximation ratios for PIPs based on Δ_1. First, following an integrality gap example from [1], we observe that the case of $W = 1$ is as hard as maximum independent set even when $\Delta_1 \leq 2$. In sharp contrast to this negative result, as soon as width is strictly larger than one, we obtain positive results via the natural LP relaxation. For PIPs with width $W = 1 + \epsilon$ where $\epsilon \in (0, 1]$, we obtain an $\Omega(\epsilon^2/\Delta_1)$-approximation. In the large width regime, when $W \geq 2$, we obtain an $\Omega((\frac{1}{1+\Delta_1/W})^{1/(W-1)})$-approximation. We also obtain a $(1-\epsilon)$-approximation when $W = \Omega(\frac{\log(\Delta_1/\epsilon)}{\epsilon^2})$.

1 Introduction

Packing integer programs (abbr. PIPs) are an expressive class of integer programs of the form:

$$\text{maximize } \langle c, x\rangle \text{ over } x \in \{0,1\}^n \text{ s.t. } Ax \leq b,$$

where $A \in \mathbb{R}_{\geq 0}^{m\times n}$, $b \in \mathbb{R}_{\geq 0}^m$ and $c \in \mathbb{R}_{\geq 0}^n$ all have nonnegative entries[1]. Many important problems in discrete and combinatorial optimization can be cast as special cases of PIPs. These include the maximum independent set in graphs and hypergraphs, set packing, matchings and b-matchings, knapsack (when $m = 1$), and the multi-dimensional knapsack. The maximum independent set problem (MIS), a special case of PIPs, is NP-hard and unless $P = NP$ there is no $n^{1-\epsilon}$-approximation where n is the number of nodes in the graph [9, 17]. For this reason it is meaningful to consider special cases and other parameters that control the difficulty of PIPs. Motivated by the fact that MIS admits a simple $\frac{1}{\Delta(G)}$-approximation where $\Delta(G)$ is the maximum degree of G, previous work considered approximating PIPs based on the maximum number of nonzeroes in any column of A (denoted by Δ_0); note that when MIS is written as a PIP, Δ_0 coincides with $\Delta(G)$. As another example, when maximum weight matching is written as a PIP, $\Delta_0 = 2$. Bansal et al. [1] obtained a simple and clever algorithm that achieved an $\Omega(1/\Delta_0)$-approximation for PIPs via the natural LP relaxation; this improved previous work of Pritchard [12, 13] who was the first to obtain an approximation for PIPs only as a function of Δ_0. Moreover, the rounding algorithm in [1] can be viewed as a

*C. Chekuri and K. Quanrud supported in part by NSF grant CCF-1526799. M. Torres supported in part by fellowships from NSF and the Sloan Foundation. University of Illinois, Urbana-Champaign, IL 61801. {chekuri, quanrud2, manuelt2}@illinois.edu

[1] We can allow the variables to have general integer upper bounds instead of restricting them to be boolean. As observed in [1], one can reduce this more general case to the $\{0,1\}$ case without too much loss in the approximation.

contention resolution scheme which allows one to get similar approximation ratios even when the objective is submodular [1, 5]. It is well-understood that PIPs become easier when the entries in A are small compared to the packing constraints b. To make this quantitative we consider the well-studied notion called the *width* defined as $W := \min_{i,j:A_{i,j}>0} b_i/A_{i,j}$. Bansal et al. obtain an $\Omega((\frac{1}{\Delta_0})^{1/\lfloor W \rfloor})$-approximation which improves as W becomes larger. Although they do not state it explicitly, their approach also yields a $(1-\epsilon)$-approximation when $W = \Omega(\frac{1}{\epsilon^2}\log(\Delta_0/\epsilon))$.

Δ_0 is a natural measure for combinatorial applications such as MIS and matchings where the underlying matrix A has entries from $\{0,1\}$. However, in some applications of PIPs such as knapsack and its multi-dimensional generalization which are more common in resource-allocation problems, the entries of A are arbitrary rational numbers (which can be assumed to be from the interval $[0,1]$ after scaling). In such applications it is natural to consider another measure of column-sparsity which is based on the ℓ_1 norm. Specifically we consider Δ_1, the maximum column sum of A. Unlike Δ_0, Δ_1 is not scale invariant so one needs to be careful in understanding the parameter and its relationship to the width W. For this purpose we normalize the constraints $Ax \leq b$ as follows. Let $W = \min_{i,j:A_{i,j}>0} b_i/A_{i,j}$ denote the width as before (we can assume without loss of generality that $W \geq 1$ since we are interested in integer solutions). We can then scale each row A_i of A separately such that, after scaling, the i'th constraint reads as $A_i x \leq W$. After scaling all rows in this fashion, entries of A are in the interval $[0,1]$, and the maximum entry of A is equal to 1. Note that this scaling process does not alter the original width. We let Δ_1 denote the maximum column sum of A after this normalization and observe that $1 \leq \Delta_1 \leq \Delta_0$. In many settings of interest $\Delta_1 \ll \Delta_0$. We also observe that Δ_1 is a more robust measure than Δ_0; small perturbations of the entries of A can dramatically change Δ_0 while Δ_1 changes minimally.

Bansal et al. raised the question of obtaining an approximation ratio for PIPs as a function of only Δ_1. They observed that this is not feasible via the natural LP relaxation by describing a simple example where the integrality gap of the LP is $\Omega(n)$ while Δ_1 is a constant. In fact their example essentially shows the existence of a simple approximation preserving reduction from MIS to PIPs such that the resulting instances have $\Delta_1 \leq 2$; thus no approximation ratio that depends only on Δ_1 is feasible for PIPs unless $P = NP$. These negative results seem to suggest that pursuing bounds based on Δ_1 is futile, at least in the worst case. However, the starting point of this paper is the observation that both the integrality gap example and the hardness result are based on instances where the width W of the instance is arbitrarily close to 1. We demonstrate that these examples are rather brittle and obtain several positive results when we consider $W \geq (1+\epsilon)$ for any fixed $\epsilon > 0$.

1.1 Our results

Our first result is on the hardness of approximation for PIPs that we already referred to. The hardness result suggests that one should consider instances with $W > 1$. Recall that after normalization we have $\Delta_1 \geq 1$ and $W \geq 1$ and the maximum entry of A is 1. We consider three regimes of W and obtain the following results, all via the natural LP relaxation, which also establish corresponding upper bounds on the integrality gap.

(i) $1 < W \leq 2$. For $W = 1+\epsilon$ where $\epsilon \in (0,1]$ we obtain an $\Omega(\frac{\epsilon^2}{\Delta_1})$-approximation.

(ii) $W \geq 2$. We obtain an $\Omega((\frac{1}{1+\frac{\Delta_1}{W}})^{1/(W-1)})$-approximation which can be simplified to $\Omega((\frac{1}{1+\Delta_1})^{1/(W-1)})$ since $W \geq 1$.

(iii) A $(1-\epsilon)$-approximation when $W = \Omega(\frac{1}{\epsilon^2}\log(\Delta_1/\epsilon))$.

Our results establish approximation bounds based on Δ_1 that are essentially the same as those based on Δ_0 as long as the width is not too close to 1. We describe randomized algorithms which can be derandomized via standard techniques. The algorithms can be viewed as contention resolution schemes, and via known techniques [1, 5], the results yield corresponding approximations for submodular objectives; we omit these extensions in this version.

All our algorithms are based on a simple randomized rounding plus alteration framework that has been successful for both packing and covering problems. Our scheme is similar to that of Bansal et al. at a high level but we make a simple but important change in the algorithm and its analysis. This is inspired by recent work on covering integer programs [4] where ℓ_1-sparsity based approximation bounds from [6] were simplified.

1.2 Other related work

We note that PIPs are equivalent to the multi-dmensional knapsack problem. When $m = 1$ we have the classical knapsack problem which admits a very efficient FPTAS (see [2]). There is a PTAS for any fixed m [7] but unless $P = NP$ an FPTAS does not exist for $m = 2$.

Approximation algorithms for PIPs in their general form were considered initially by Raghavan and Thompson [14] and refined substantially by Srinivasan [15]. Srinivasan obtained approximation ratios of the form $\Omega(1/n^W)$ when A had entries from $\{0,1\}$, and a ratio of the form $\Omega(1/n^{1/\lfloor W \rfloor})$ when A had entries from $[0,1]$. Pritchard [12] was the first to obtain a bound for PIPs based solely on the column sparsity parameter Δ_0. He used iterated rounding and his initial bound was improved in [13] to $\Omega(1/\Delta_0^2)$. The current state of the art is due to Bansal et al. [1]. Previously we ignored constant factors when describing the ratio. In fact [1] obtains a ratio of $(1 - o(1)\frac{e-1}{e^2 \Delta_0})$ by strengthening the basic LP relaxation.

In terms of hardness of approximation, PIPs generalize MIS and hence one cannot obtain a ratio better than $n^{1-\epsilon}$ unless $P = NP$ [9, 17]. Building on MIS, [3] shows that PIPs are hard to approximate within a $n^{\Omega(1/W)}$ factor for any constant width W. Hardness of MIS in bounded degree graphs [16] and hardness for k-set-packing [10] imply that PIPs are hard to approximate to within $\Omega(1/\Delta_0^{1-\epsilon})$ and to within $\Omega((\log \Delta_0)/\Delta_0)$ when Δ_0 is a sufficiently large constant. These hardness results are based on $\{0,1\}$ matrices for which Δ_0 and Δ_1 coincide.

There is a large literature on deterministic and randomized rounding algorithms for packing and covering integer programs and connections to several topics and applications including discrepancy theory. ℓ_1-sparsity guarantees for covering integer programs were first obtained by Chen, Harris and Srinivasan [6] partly inspired by [8].

2 Hardness of approximating PIPs as a function of Δ_1

Bansal et al. [1] showed that the integrality gap of the natural LP relaxation for PIPs is $\Omega(n)$ even when Δ_1 is a constant. One can use essentially the same construction to show the following theorem.

Theorem 1. *There is an approximation preserving reduction from MIS to instances of PIPs with $\Delta_1 \leq 2$.*

Proof. Let $G = (V, E)$ be an undirected graph without self-loops and let $n = |V|$. Let $A \in [0,1]^{n \times n}$ be indexed by V. For all $v \in V$, let $A_{v,v} = 1$. For all $uv \in E$, let $A_{u,v} = A_{v,u} = 1/n$. For all the remaining entries in A that have not yet been defined, set these entries to 0. Consider the following PIP:

$$\text{maximize } \langle x, \mathbf{1} \rangle \text{ over } x \in \{0,1\}^n \text{ s.t. } Ax \leq 1. \tag{1}$$

Let S be the set of all feasible integral solutions of (1) and $\mathcal{I}$ be the set of independent sets of G. Define $g : S \to \mathcal{I}$ where $g(x) = \{v : x_v = 1\}$. To show g is surjective, consider a set $I \in \mathcal{I}$. Let y be the characteristic vector of I. That is, y_v is 1 if $v \in I$ and 0 otherwise. Consider the row in A corresponding to an arbitrary vertex u where $y_u = 1$. For all $v \in V$ such that v is a neighbor to u, $y_v = 0$ as I is an independent set. Thus, as the nonzero entries in A of the row corresponding to u are, by construction, the neighbors of u, it follows that the constraint corresponding to u is satisfied in (1). As u is an arbitrary vertex, it follows that y is a feasible integral solution to (1) and as $I = \{v : y_v = 1\}$, $g(y) = I$.

Define $h : S \to \mathbb{N}_0$ such that $h(x) = |g(x)|$. It is clear that $\max_{x \in S} h(x)$ is equal to the optimal value of (1). Let I_{max} be a maximum independent set of G. As g is surjective, there exists $z \in S$ such that $g(z) = I_{max}$. Thus, $\max_{x \in S} h(x) \geq |I_{max}|$. As $\max_{x \in S} h(x)$ is equal to the optimum value of (1), it follows that a β-approximation for PIPs implies a β-approximation for maximum independent set.

Round-and-Alter Framework: input A, b, and α
let x be the optimum fractional solution of the natural LP relaxation
for $j \in [n]$, set x'_j to be 1 independently with probability αx_j and 0 otherwise
$x'' \leftarrow x'$
for $i \in [m]$ **do**
find $S \subseteq [n]$ such that setting $x'_j = 0$ for all $j \in S$ would satisfy $\langle e_i, Ax' \rangle \le b_i$
for all $j \in S$, set $x''_j = 0$
end for
return x''

Figure 1: Randomized rounding with alteration framework.

Furthermore, we note that for this PIP, $\Delta_1 \le 2$, thus concluding the proof. □

Unless $P = NP$, MIS does not admit a $n^{1-\epsilon}$-approximation for any fixed $\epsilon > 0$ [9, 17]. Hence the preceding theorem implies that unless $P = NP$ one cannot obtain an approximation ratio for PIPs solely as a function of Δ_1.

3 Round and alter framework

The algorithms in this paper have the same high-level structure. The algorithms first scale down the fractional solution x by some factor α, and then randomly round each coordinate independently. The rounded solution x' may not be feasible for the constraints. The algorithm alters x' to a feasible x'' by considering each constraint separately in an arbitrary order; if x' is not feasible for constraint i some subset S of variables are chosen to be set to 0. Each constraint corresponds to a knapsack problem and the framework (which is adapted from [1]) views the problem as the intersection of several knapsack constraints. A formal template is given in Figure 1. To make the framework into a formal algorithm, one must define α and how to choose S in the for loop. These parts will depend on the regime of interest.

For an algorithm that follows the round-and-alter framework, the expected output of the algorithm is $\mathbb{E}\left[\langle c, x'' \rangle\right] = \sum_{j=1}^{n} c_j \cdot \Pr[x''_j = 1]$. Independent of how α is defined or how S is chosen, $\Pr[x''_j = 1] = \Pr[x''_j = 1 | x'_j = 1] \cdot \Pr[x'_j = 1]$ since $x''_j \le x'_j$. Then we have

$$\mathbb{E}[\langle c, x'' \rangle] = \alpha \sum_{j=1}^{n} c_j x_j \cdot \Pr[x''_j = 1 | x'_j = 1].$$

Let E_{ij} be the event that x''_j is set to 0 when ensuring constraint i is satisfied in the for loop. As x''_j is only set to 0 if at least one constraint sets x''_j to 0, we have

$$\Pr[x''_j = 0 | x'_j = 1] = \Pr\left[\bigcup_{i \in [m]} E_{ij} | x'_j = 1\right] \le \sum_{i=1}^{m} \Pr[E_{ij} | x'_j = 1].$$

Combining these two observations, we have the following lemma, which applies to all of our subsequent algorithms.

Lemma 2. *Let $\mathcal{A}$ be a randomized rounding algorithm that follows the round-and-alter framework given in Figure 1. Let x' be the rounded solution obtained with scaling factor α. Let E_{ij} be the event that x''_j is set to 0 by constraint i. If for all $j \in [n]$ we have $\sum_{i=1}^{m} \Pr[E_{ij} | x'_j = 1] \le \gamma$, then $\mathcal{A}$ is an $\alpha(1 - \gamma)$-approximation for PIPs.*

We will refer to the quantity $\Pr[E_{ij} | x'_j = 1]$ as the *rejection probability* of item j in constraint i. We will also say that constraint i *rejects* item j if x''_j is set to 0 in constraint i.

```
round-and-alter-by-sorting(A, b, α₁):
  let x be the optimum fractional solution of the natural LP relaxation
  for j ∈ [n], set x'_j to be 1 independently with probability α₁x_j and 0 otherwise
  x'' ← x'
  for i ∈ [m] do
    sort and renumber such that A_{i,1} ≤ ··· ≤ A_{i,n}
    s ← max{ℓ ∈ [n] : Σ_{j=1}^{ℓ} A_{i,j} x'_j ≤ b_i}
    for each j ∈ [n] such that j > s, set x''_j = 0
  end for
  return x''
```

Figure 2: Round-and-alter in the large width regime. Each constraint sorts the coordinates in increasing size and greedily picks a feasible set and discards the rest.

4 The large width regime: $W \geq 2$

In this section, we consider PIPs with width $W \geq 2$. Recall that we assume $A \in [0,1]^{m \times n}$ and $b_i = W$ for all $i \in [m]$. Therefore we have $A_{i,j} \leq W/2$ for all i, j and from a knapsack point of view all items are "small". We apply the round-and-alter framework in a simple fashion where in each constraint i the coordinates are sorted by the coefficents in that row and the algorithm chooses the largest prefix of coordinates that fit in the capacity W and the rest are discarded. We emphasize that this sorting step is crucial for the analysis and differs from the scheme in [1]. Figure 2 describes the formal algorithm.

The key property for the analysis: The analysis relies on obtaining a bound on the rejection probability of coordinate j by constraint i. Let X_j be the indicator variable for j being chosen in the first step. We show that $\Pr[E_{ij} \mid X_j = 1] \leq cA_{ij}$ for some c that depends on the scaling factor α. Thus coordinates with smaller coefficients are less likely to be rejected. The total rejection probability of j, $\sum_{i=1}^{m} \Pr[E_{ij} \mid X_j = 1]$, is proportional to the column sum of coordinate j which is at most Δ_1.

The analysis relies on the Chernoff bound, and depending on the parameters, one needs to adjust the analysis. In order to highlight the main ideas we provide a detailed proof for the simplest case and include the proofs of the other cases in the appendix.

4.1 An $\Omega(1/\Delta_1)$-approximation algorithm

We show that round-and-alter-by-sorting yields an $\Omega(1/\Delta_1)$-approximation if we set the scaling factor $\alpha_1 = \frac{1}{c_1 \Delta_1}$ where $c_1 = 4e^{1+1/e}$.

The rejection probability is captured by the following main lemma.

Lemma 3. *Let $\alpha_1 = \frac{1}{c_1 \Delta_1}$ for $c_1 = 4e^{1+1/e}$. Let $i \in [m]$ and $j \in [n]$. Then we have $\Pr[E_{ij} | X_j = 1] \leq \frac{A_{i,j}}{2\Delta_1}$ in the algorithm* round-and-alter-by-sorting(A, b, α_1).

Proof. At iteration i of round-and-alter-by-sorting, after the set $\{A_{i,1}, \ldots, A_{i,n}\}$ is sorted, the indices are renumbered so that $A_{i,1} \leq \cdots \leq A_{i,n}$. Note that j may now be a different index j', but for simplicity of notation we will refer to j' as j. Let $\xi_\ell = 1$ if $x'_\ell = 1$ and 0 otherwise. Let $Y_{ij} = \sum_{\ell=1}^{j-1} A_{i,\ell} \xi_\ell$.

If E_{ij} occurs, then $Y_{ij} > W - A_{i,j}$, since x''_j would not have been set to zero by constraint i otherwise. That is,

$$\Pr[E_{ij} | X_j = 1] \leq \Pr[Y_{ij} > W - A_{i,j} | X_j = 1].$$

The event $Y_{ij} > W - A_{i,j}$ does not depend on x'_j. Therefore,

$$\Pr[Y_{ij} > W - A_{i,j} | X_j = 1] \leq \Pr[Y_{ij} \geq W - A_{i,j}].$$

To upper bound $\mathbb{E}[Y_{ij}]$, we have

$$\mathbb{E}[Y_{ij}] = \sum_{\ell=1}^{j-1} A_{i,\ell} \cdot \Pr[X_\ell = 1] \leq \alpha_1 \sum_{\ell=1}^{n} A_{i,\ell} x_\ell \leq \alpha_1 W.$$

As $A_{i,j} \leq 1$, $W \geq 2$, and $\alpha_1 < 1/2$, we have $\frac{(1-\alpha_1)W}{A_{i,j}} > 1$. Using the fact that $A_{i,j}$ is at least as large as all entries $A_{i,j'}$ for $j' < j$, we satisfy the conditions to apply the Chernoff bound in Theorem 13. This implies

$$\Pr[Y_{ij} > W - A_{i,j}] \leq \left(\frac{\alpha_1 e^{1-\alpha_1} W}{W - A_{i,j}}\right)^{(W-A_{i,j})/A_{i,j}}.$$

Note that $\frac{W}{W-A_{i,j}} \leq 2$ as $W \geq 2$. Because $e^{1-\alpha_1} \leq e$ and by the choice of α_1, we have

$$\left(\frac{\alpha_1 e^{1-\alpha_1} W}{W - A_{i,j}}\right)^{(W-A_{i,j})/A_{i,j}} \leq (2e\alpha_1)^{(W-A_{i,j})/A_{i,j}} = \left(\frac{1}{2e^{1/e}\Delta_1}\right)^{(W-A_{i,j})/A_{i,j}}.$$

Then we prove the final inequality in two parts. First, we see that $W \geq 2$ and $A_{i,j} \leq 1$ imply that $\frac{W-A_{i,j}}{A_{i,j}} \geq 1$. This implies

$$\left(\frac{1}{2\Delta_1}\right)^{(W-1)/A_{i,j}} \leq \frac{1}{2\Delta_1}.$$

Second, we see that

$$(1/e^{1/e})^{(W-A_{i,j})/A_{i,j}} \leq (1/e^{1/e})^{1/A_{i,j}} \leq A_{i,j}$$

for $A_{i,j} \leq 1$, where the first inequality holds because $W - A_{i,j} \geq 1$ and the second inequality holds by Lemma 14. This concludes the proof. □

Theorem 4. *When setting $\alpha_1 = \frac{1}{c_1\Delta_1}$ where $c_1 = 4e^{1+1/e}$,* round-and-alter-by-sorting(A, b, α_1) *is a randomized $(\alpha_1/2)$-approximation algorithm for PIPs with width $W \geq 2$.*

Proof. Fix $j \in [n]$. By Lemma 3 and the definition of Δ_1, we have

$$\sum_{i=1}^{m} \Pr[E_{ij} | X_j = 1] \leq \sum_{i=1}^{m} \frac{A_{i,j}}{2\Delta_1} \leq \frac{1}{2}.$$

By Lemma 2, which shows that upper bounding the sum of the rejection probabilities by γ for every item leads to an $\alpha_1(1-\gamma)$-approximation, we get the desired result. □

4.2 An $\Omega(\frac{1}{(1+\Delta_1/W)^{1/(W-1)}})$-approximation

We improve the bound from the previous section by setting $\alpha_1 = \frac{1}{c_2(1+\Delta_1/W)^{1/(W-1)}}$ where $c_2 = 4e^{1+2/e}$. Note that the scaling factor becomes larger as W increases. The analysis of the following lemma is similar to that of Lemma 3 and is therefore left for the appendix.

Lemma 5. *Let $\alpha_1 = \frac{1}{c_2(1+\Delta_1/W)^{1/(W-1)}}$ for $c_2 = 4e^{1+2/e}$. Let $i \in [m]$ and $j \in [n]$. Then in the algorithm* round-and-alter-by-sorting(A, b, α_1), *we have $\Pr[E_{ij} | X_j = 1] \leq \frac{A_{i,j}}{2\Delta_1}$.*

If we replace Lemma 3 with Lemma 5 in the proof of Theorem 4, we obtain the following stronger guarantee.

Theorem 6. *When setting $\alpha_1 = \frac{1}{c_2(1+\Delta_1/W)^{1/(W-1)}}$ where $c_2 = 4e^{1+2/e}$, for PIPs with width $W \geq 2$,* round-and-alter-by-sorting(A, b, α_1) *is a randomized $(\alpha_1/2)$-approximation.*

4.3 A $(1 - O(\epsilon))$-approximation when $W \geq \Omega(\frac{1}{\epsilon^2} \ln(\frac{\Delta_1}{\epsilon}))$

In this section, we give a randomized $(1-O(\epsilon))$-approximation for the case when $W \geq \Omega(\frac{1}{\epsilon^2} \ln(\frac{\Delta_1}{\epsilon}))$. We use the algorithm round-and-alter-by-sorting in Figure 2 with the scaling factor $\alpha_1 = 1 - \epsilon$. The analysis follows the same structure as the analyses for the lemmas bounding the rejection probabilities from the previous sections. The proof can be found in the appendix.

Lemma 7. *Let $0 < \epsilon < \frac{1}{e}$, $\alpha_1 = 1 - \epsilon$, and $W = \frac{2}{\epsilon^2} \ln(\frac{\Delta_1}{\epsilon}) + 1$. Let $i \in [m]$ and $j \in [n]$. Then in* round-and-alter-by-sorting(A, b, α_1)*, we have* $\Pr[E_{ij} | X_j = 1] \leq e \cdot \frac{\epsilon A_{i,j}}{\Delta_1}$.

Lemma 7 implies that we can upper bound the sum of the rejection probabilities for any item j by $e\epsilon$, leading to the following theorem.

Theorem 8. *Let $0 < \epsilon < \frac{1}{e}$ and $W = \frac{2}{\epsilon^2} \ln(\frac{\Delta_1}{\epsilon}) + 1$. When setting $\alpha_1 = 1 - \epsilon$ and $c = e + 1$,* round-and-alter-by-sorting(A, b, α_1) *is a randomized $(1 - c\epsilon)$-approximation algorithm.*

Proof. Fix $j \in [n]$. By Lemma 7 and the definition of Δ_1,

$$\sum_{i=1}^{m} \Pr[E_{ij} | X_j = 1] \leq \sum_{i=1}^{m} \frac{e\epsilon A_{i,j}}{\Delta_1} \leq e\epsilon.$$

By Lemma 2, which shows that an upper bound on the rejection probabilities of γ leads to an $\alpha_1(1 - \gamma)$-approximation, we have an $\alpha_1(1-e\epsilon)$-approximation. Then note that $\alpha_1(1-e\epsilon) = (1-\epsilon)(1-e\epsilon) \geq 1-(e+1)\epsilon$. This concludes the proof. $\square$

5 The small width regime: $W = (1 + \epsilon)$

We now consider the regime when the width is small. Let $W = 1+\epsilon$ for some $\epsilon \in (0, 1]$. We cannot apply the simple sorting based scheme that we used for the large width regime. We borrow the idea from [1] in splitting the coordinates into big and small in each constraint; now the definition is more refined and depends on ϵ. Moreover, the small coordinates and the big coordinates have their own reserved capacity in the constraint. This is crucial for the analysis. We provide more formal details below.

We set α_2 to be $\frac{\epsilon^2}{c_3 \Delta_1}$ where $c_3 = 8e^{1+2/e}$. The alteration step differentiates between "small" and "big" coordinates as follows. For each $i \in [m]$, let $S_i = \{j : A_{i,j} \leq \epsilon/2\}$ and $B_i = \{j : A_{i,j} > \epsilon/2\}$. We say that an index j is *small* for constraint i if $j \in S_i$. Otherwise we say it is *big* for constraint i when $j \in B_i$. For each constraint, the algorithm is allowed to pack a total of $1 + \epsilon$ into that constraint. The algorithm separately packs small indices and big indices. In an ϵ amount of space, small indices that were chosen in the rounding step are sorted in increasing order of size and greedily packed until the constraint is no longer satisfied. The big indices are packed by arbitrarily choosing one and packing it into the remaining space of 1. The rest of the indices are removed to ensure feasibility. Figure 3 gives pseudocode for the randomized algorithm round-alter-small-width which yields an $\Omega(\epsilon^2/\Delta_1)$-approximation.

It remains to bound the rejection probabilities. Recall that for $j \in [n]$, we define X_j to be the indicator random variable $\mathbb{1}(x'_j = 1)$ and E_{ij} is the event that j was rejected by constraint i.

We first consider the case when index j is big for constraint i. Note that it is possible that there may not exist any big indices for a given constraint. The same holds true for small indices.

Lemma 9. *Let $\epsilon \in (0, 1]$ and $\alpha_2 = \frac{\epsilon^2}{c_3 \Delta_1}$ where $c_3 = 8e^{1+2/e}$. Let $i \in [m]$ and $j \in B_i$. Then in* round-alter-small-width$(A, b, \epsilon, \alpha_2)$*, we have* $\Pr[E_{ij} | X_j = 1] \leq \frac{A_{i,j}}{2\Delta_1}$.

Proof. Let $\mathcal{E}$ be the event that there exists $j' \in B_i$ such that $j' \neq j$ and $X_{j'} = 1$. Observe that if E_{ij} occurs and $X_j = 1$, then it must be the case that at least one other element of B_i was chosen in the rounding step. Thus,

$$\Pr[E_{ij} | X_j = 1] \leq \Pr[\mathcal{E}] \leq \sum_{\substack{\ell \in B_i \\ \ell \neq j}} \Pr[X_\ell = 1] \leq \alpha_2 \sum_{\ell \in B_i} x_\ell,$$

```
round-alter-small-width(A, b, ε, α₂):
  let x be the optimum fractional solution of the natural LP relaxation
  for j ∈ [n], set x'_j to be 1 independently with probability α₂x_j and 0 otherwise
  x'' ← x'
  for i ∈ [m] do
    if |S_i| = 0 then
      s ← 0
    else
      sort and renumber such that A_{i,1} ≤ ··· ≤ A_{i,n}
      s ← max{ℓ ∈ S_i : Σ_{j=1}^{ℓ} A_{i,j} x'_j ≤ ε}
    end if
    if |B_i| = 0, then t = 0, otherwise let t be an arbitrary element of B_i
    for each j ∈ [n] such that j > s and j ≠ t, set x''_j = 0
  end for
  return x''
```

Figure 3: By setting the scaling factor $\alpha_2 = \frac{\epsilon^2}{c\Delta_1}$ for a sufficiently large constant c, round-alter-small-width is a randomized $\Omega(\epsilon^2/\Delta_1)$-approximation for PIPs with width $W = 1+\epsilon$ for some $\epsilon \in (0,1]$ (see Theorem 11).

where the second inequality follows by the union bound. Observe that for all $\ell \in B_i$, we have $A_{i,\ell} > \epsilon/2$. By the LP constraints, we have $1+\epsilon \geq \sum_{\ell\in B_i} A_{i,\ell}x_\ell > \frac{\epsilon}{2}\cdot\sum_{\ell\in B_i} x_\ell$. Thus, $\sum_{\ell\in B_i} x_\ell \leq \frac{1+\epsilon}{\epsilon/2} = 2/\epsilon + 2$.

Using this upper bound for $\sum_{\ell\in B_i} x_\ell$, we have

$$\alpha_2 \sum_{\ell\in B_i} x_\ell \leq \frac{\epsilon^2}{c_3\Delta_1}\left(\frac{2}{\epsilon}+2\right) \leq \frac{4\epsilon}{c_3\Delta_1} \leq \frac{A_{i,j}}{2\Delta_1},$$

where the second inequality utilizes the fact that $\epsilon \leq 1$ and the third inequality holds because $c_3 \geq 16$ and $A_{i,j} > \epsilon/2$. □

Next we consider the case when index j is small for constraint i. The analysis here is similar to that in the preceding section with width at least 2. The proof is left for the appendix.

Lemma 10. *Let $\epsilon \in (0,1]$ and $\alpha_2 = \frac{\epsilon^2}{c_3\Delta_1}$ where $c_3 = 8e^{1+2/e}$. Let $i \in [m]$ and $j \in S_i$. Then in* round-alter-small-width$(A, b, \epsilon, \alpha_2)$*, we have* $\Pr[E_{ij}|X_j = 1] \leq \frac{A_{i,j}}{2\Delta_1}$.

As Lemma 10 shows that the rejection probability is small, we can prove the following approximation guarantee much like in Theorems 4 and 6.

Theorem 11. *Let $\epsilon \in (0,1]$. When setting $\alpha_2 = \frac{\epsilon^2}{c_3\Delta_1}$ for $c_3 = 8e^{1+2/e}$, for PIPs with width $W = 1+\epsilon$,* round-alter-small-width$(A, b, \epsilon, \alpha_2)$ *is a randomized $(\alpha_2/2)$-approximation algorithm .*

Proof. Fix $j \in [n]$. Then by Lemmas 9 and 10 and the definition of Δ_1, we have

$$\sum_{i=1}^{m} \Pr[E_{ij}|X_j = 1] \leq \sum_{i=1}^{m} \frac{A_{i,j}}{2\Delta_1} \leq \frac{1}{2}.$$

Recall that Lemma 2 gives an $\alpha_2(1-\gamma)$-approximation where γ is an upper bound on the sum of the rejection probabilities for any item. This concludes the proof. □

References

[1] Bansal, N., Korula, N., Nagarajan, V., and Srinivasan, A. Solving packing integer programs via randomized rounding with alterations. *Theory of Computing 8*, 24 (2012), 533–565.

[2] Chan, T. M. Approximation schemes for 0-1 knapsack. In *1st Symposium on Simplicity in Algorithms* (2018).

[3] Chekuri, C., and Khanna, S. On multidimensional packing problems. *SIAM journal on computing 33*, 4 (2004), 837–851.

[4] Chekuri, C., and Quanrud, K. On approximating (sparse) covering integer programs. In *Proceedings of the Thirtieth Annual ACM-SIAM Symposium on Discrete Algorithms* (2019), SIAM, pp. 1596–1615.

[5] Chekuri, C., Vondrák, J., and Zenklusen, R. Submodular function maximization via the multilinear relaxation and contention resolution schemes. *SIAM Journal on Computing 43*, 6 (2014), 1831–1879.

[6] Chen, A., Harris, D. G., and Srinivasan, A. Partial resampling to approximate covering integer programs. In *Proceedings of the twenty-seventh annual ACM-SIAM symposium on Discrete algorithms* (2016), Society for Industrial and Applied Mathematics, pp. 1984–2003.

[7] Frieze, A., and Clarke, M. Approximation algorithms for the m-dimensional 0-1 knapsack problem: worst-case and probabilistic analyses. *European Journal of Operational Research 15*, 1 (1984), 100–109.

[8] Harvey, N. J. A note on the discrepancy of matrices with bounded row and column sums. *Discrete Mathematics 338*, 4 (2015), 517–521.

[9] Håstad, J. Clique is hard to approximate within $n^{1-\epsilon}$. *Acta Mathematica 182*, 1 (1999), 105–142.

[10] Hazan, E., Safra, S., and Schwartz, O. On the complexity of approximating k-set packing. *Computational Complexity 15*, 1 (2006), 20–39.

[11] Mitzenmacher, M., and Upfal, E. *Probability and computing: Randomized algorithms and probabilistic analysis*. Cambridge university press, 2005.

[12] Pritchard, D. Approximability of sparse integer programs. In *European Symposium on Algorithms* (2009), Springer, pp. 83–94.

[13] Pritchard, D., and Chakrabarty, D. Approximability of sparse integer programs. *Algorithmica 61*, 1 (Sep 2011), 75–93.

[14] Raghavan, P., and Thompson, C. D. Randomized rounding: a technique for provably good algorithms and algorithmic proofs. *Combinatorica 7*, 4 (1987), 365–374.

[15] Srinivasan, A. Improved approximation guarantees for packing and covering integer programs. *SIAM Journal on Computing 29*, 2 (1999), 648–670.

[16] Trevisan, L. Non-approximability results for optimization problems on bounded degree instances. In *Proceedings of the thirty-third annual ACM symposium on Theory of computing* (2001), ACM, pp. 453–461.

[17] Zuckerman, D. Linear degree extractors and the inapproximability of max clique and chromatic number. In *Proceedings of the thirty-eighth annual ACM symposium on Theory of computing* (2006), ACM, pp. 681–690.

Appendix

A Chernoff Bounds and Useful Inequalities

The following standard Chernoff bound is used to obtain a more convenient Chernoff bound in Theorem 13. The proof of Theorem 13 follows directly from choosing δ such that $(1+\delta)\mu = W - \beta$ and applying Theorem 12. We include the proof for convenience.

Theorem 12 ([11])**.** *Let $X_1, \ldots, X_n$ be independent random variables where X_i is defined on $\{0, \beta_i\}$, where $0 < \beta_i \le \beta \le 1$ for some β. Let $X = \sum_i X_i$ and denote $\mathbb{E}[X]$ as μ. Then for any $\delta > 0$,*

$$\Pr[X \ge (1+\delta)\mu] \le \left(\frac{e^{\delta}}{(1+\delta)^{1+\delta}}\right)^{\mu/\beta}$$

Theorem 13. *Let $X_1, \ldots, X_n \in [0, \beta]$ be independent random variables for some $0 < \beta \le 1$. Suppose $\mu = \mathbb{E}[\sum_i X_i] \le \alpha W$ for some $0 < \alpha < 1$ and $W \ge 1$ where $(1-\alpha)W > \beta$. Then*

$$\Pr\left[\sum_i X_i > W - \beta\right] \le \left(\frac{\alpha e^{1-\alpha} W}{W - \beta}\right)^{(W-\beta)/\beta}.$$

Proof. Since the right-hand side is increasing in α, it suffices to assume $\mu = \alpha W$. Choose δ such that $(1+\delta)\mu = W - \beta$. Then $\delta = (W - \beta - \mu)/\mu$. Because $\mu = \alpha W$ and since $(1-\alpha)W > \beta$, we have $\delta = ((1-\alpha)W - \beta)/\mu > 0$. We apply the standard Chernoff bound in Theorem 13 to obtain

$$\Pr\left[\sum_i X_i > W - \beta\right] = \Pr\left[\sum_i X_i > (1+\delta)\mu\right] \le \left(\frac{e^{\delta}}{(1+\delta)^{1+\delta}}\right)^{\mu/\beta}.$$

Because $1 + \delta = (W - \beta)/\mu$ and $\delta = (W - \beta - \mu)/\mu$,

$$\left(\frac{e^{\delta}}{(1+\delta)^{1+\delta}}\right)^{\mu/\beta} = \left(\frac{e^{W-\beta-\mu}}{((W-\beta)/\mu)^{W-\beta}}\right)^{1/\beta}.$$

Exponentiating the denominator,

$$\left(\frac{e^{W-\beta-\mu}}{((W-\beta)/\mu)^{W-\beta}}\right)^{1/\beta} = \exp\left(\frac{1}{\beta}\left(W - \beta - \mu + (W-\beta)\ln\left(\frac{\mu}{W-\beta}\right)\right)\right)$$

As $\mu = \alpha W$,

$$\exp\left(\frac{1}{\beta}\left(W - \beta - \mu + (W-\beta)\ln\left(\frac{\mu}{W-\beta}\right)\right)\right) = \exp\left(\frac{1}{\beta}\left((1-\alpha)W - \beta + (W-\beta)\ln\left(\frac{\alpha W}{W-\beta}\right)\right)\right)$$

We can rewrite the exponent to show that

$$\exp\left(\frac{1}{\beta}\left((1-\alpha)W - \beta - (W-\beta)\ln\left(\frac{W-\beta}{\alpha W}\right)\right)\right) \le \left(\frac{\alpha e^{1-\alpha} W}{W-\beta}\right)^{(W-\beta)/\beta}.$$

□

The following three lemmas are used in the proofs bounding the rejection probabilities for different regimes of width. The inequalities are easily verified via calculus. The proofs are included for the sake of completeness.

Lemma 14. *Let $x \in (0,1]$. Then $(1/e^{1/e})^{1/x} \leq x$.*

Proof. Taking logs of both sides of the stated inequality and rearranging, it suffices to show that $\ln(1/e^{1/e}) \leq x \ln x$ for $x > 0$. $x \ln x$ is convex and its minimum is $-1/e$ at $x = 1/e$. Since $\ln(1/e^{1/e}) = -1/e$, the inequality holds. □

Lemma 15. *Let $y \geq 2$ and $x \in (0,1]$. Then $x/y \geq (1/e^{2/e})^{y/2x}$.*

Proof. We start with a simple rewriting of the statement. After taking logs and rearranging, it is sufficient to show

$$(x/y)\ln(x/y) \geq (1/2)\ln(1/e^{2/e}) = -1/e.$$

Replacing x/y with z, we see that it suffices the prove $z \ln z \geq -1/e$ for $0 < z \leq 1/2$. We note that $x \ln x$ is convex and its minimum is $-1/e$ at $x = 1/e$. Thus, $z \ln z \geq -1/e$. This concludes the proof. □

Lemma 16. *Let $0 < \epsilon \leq 1$ and $x \in (0,1]$. Then $\epsilon x/2 \geq (\epsilon/e^{2/e})^{1/x}$.*

Proof. To start, let $d = e^{2/e}/2$ and observe that $d > 1$. We first do a change of variables, replacing $\epsilon/2$ with ϵ and x with x/ϵ. If we take a log of both sides, then our reformulated goal is to show that

$$x \ln x \geq \epsilon \ln(\epsilon/d)$$

for $0 < \epsilon \leq 1/2$ and $x \in (0, \epsilon]$. Letting $f(y) = y \ln y$ and $g(y) = y \ln(y/d)$, we want to show that $f(x) \geq g(\epsilon)$. We will proceed by cases.

First, suppose $0 < \epsilon \leq d/e$. It is easy to show that f is decreasing on $(0, 1/e]$ and increasing on $[1/e, \infty)$ and that g is decreasing on $(0, d/e]$ and increasing on $[d/e, \infty)$. As f is decreasing on $(0, 1/e]$, for $0 < \epsilon \leq 1/e$, we have $f(x) \geq f(\epsilon)$ as $x \leq \epsilon$. As $d > 1$, it follows that $f(\epsilon) \geq g(\epsilon)$. Therefore, $f(x) \geq g(\epsilon)$ for $0 < \epsilon \leq 1/e$. Furthermore, as g is decreasing on $[1/e, d/e]$ and f is increasing on $[1/e, d/e]$, we have $f(x) \geq g(\epsilon)$ for $0 < \epsilon \leq d/e$.

For the second case, suppose $d/e < \epsilon \leq 1/2$. Note that the minimum of f on the interval $(0, 1/2]$ is $f(1/e) = -1/e$. Thus, it would suffice to show that $g(\epsilon) \leq -1/e$. As we noted previously that g is increasing on $[d/e, 1/2]$, it would suffice to show that $g(1/2) \leq -1/e$. By definition of g, we see $g(1/2) = -1/e$. Therefore, $f(x) \geq g(\epsilon)$. This concludes the proof. □

B Skipped Proofs

B.1 Proof of Lemma 5

Proof. The proof proceeds similarly to the proof of Lemma 3. Since $\alpha_1 < 1/2$, everything up to and including the application of the Chernoff bound there applies. This gives that for each $i \in [m]$ and $j \in [n]$,

$$\Pr[E_{ij} | X_j = 1] \leq (2e\alpha_1)^{(W - A_{i,j})/A_{i,j}} .$$

By choice of α_1, we have

$$(2e\alpha_1)^{(W - A_{i,j})/A_{i,j}} = \left(\frac{1}{2e^{2/e}(1 + \Delta_1/W)^{1/(W-1)}}\right)^{(W - A_{i,j})/A_{i,j}}$$

We prove the final inequality in two parts. First, note that $\frac{W - A_{i,j}}{A_{i,j}} \geq W - 1$ since $A_{i,j} \leq 1$. Thus,

$$\left(\frac{1}{2(1 + \Delta_1/W)^{1/(W-1)}}\right)^{(W - A_{i,j})/A_{i,j}} \leq \frac{1}{2^{W-1}(1 + \Delta_1/W)} \leq \frac{W}{2\Delta_1}.$$

Second, we see that

$$\left(\frac{1}{e^{2/e}}\right)^{(W - A_{i,j})/A_{i,j}} \leq \left(\frac{1}{e^{2/e}}\right)^{W/2A_{i,j}} \leq \frac{A_{i,j}}{W}$$

for $A_{i,j} \leq 1$, where the first inequality holds because $W \geq 2$ and the second inequality holds by Lemma 15. □

B.2 Proof of Lemma 7

Proof. Renumber indices so that $A_{i,1} \leq \cdots \leq A_{i,n}$ and if the index of j changes to j', we still refer to j' as j. Let $Y_{ij} = \sum_{\ell=1}^{j-1} A_{i,\ell}\xi_\ell$ where $\xi_\ell = 1$ if $x'_\ell = 1$ and 0 otherwise. We first note that

$$\Pr[E_{ij} | X_j = 1] \leq \Pr[Y_{ij} > W - A_{i,j}].$$

By the choice of α_1 and the fact that $A_{i,j} \leq 1$ and $W = \frac{2}{\epsilon^2}\ln(\frac{\Delta_1}{\epsilon}) + 1$, we have $((1-\alpha_1)W)/A_{i,j} \geq \epsilon W = \frac{2}{\epsilon}\ln(\frac{\Delta_1}{\epsilon}) + \epsilon$. A direct argument via calculus shows $\frac{2}{\epsilon}\ln(\frac{\Delta_1}{\epsilon}) + \epsilon > 1$ for $\epsilon \in (0, \frac{1}{e})$. Thus, $(1-\alpha_1)W > A_{i,j}$.

By the LP constraints, $\mathbb{E}[Y_{ij}] \leq \alpha_1 W$. Then as $A_{i,j'} \leq A_{i,j}$ for all $j' < j$, we can apply the Chernoff bound in Theorem 13 to obtain

$$\Pr[Y_{ij} \geq W - A_{i,j}] \leq \left(\frac{\alpha_1 e^{1-\alpha_1} W}{W - A_{i,j}}\right)^{(W-A_{i,j})/A_{i,j}}.$$

As $A_{i,j} \leq 1$,

$$\left(\frac{W}{W - A_{i,j}}\right)^{(W-A_{i,j})/A_{i,j}} \leq \left(\frac{W}{W-1}\right)^{W-1} \leq e,$$

where the last inequality follows from the fact that $(1 - 1/z)^{z-1} \geq 1/e$ for all $z \geq 1$. Then

$$\left(\frac{\alpha_1 e^{1-\alpha_1} W}{W - A_{i,j}}\right)^{(W-A_{i,j})/A_{i,j}} \leq e \cdot \left(\alpha_1 e^{1-\alpha_1}\right)^{(W-A_{i,j})/A_{i,j}}.$$

By the choice of α_1,

$$e \cdot \left(\alpha_1 e^{1-\alpha_1}\right)^{(W-A_{i,j})/A_{i,j}} = e \cdot ((1-\epsilon)e^{\epsilon})^{(W-A_{i,j})/A_{i,j}}.$$

For $0 < \epsilon < \frac{1}{e}$, we have $1 - \epsilon \leq \exp(-\epsilon - \frac{\epsilon^2}{2})$. As $W = \frac{2}{\epsilon^2}\ln(\frac{\Delta_1}{\epsilon}) + 1$ and $A_{i,j} \leq 1$,

$$e \cdot ((1-\epsilon)e^{\epsilon})^{(W-A_{i,j})/A_{i,j}} \leq e \cdot \left(e^{-\epsilon^2/2}\right)^{\frac{2}{\epsilon^2}\ln(\frac{\Delta_1}{\epsilon})} \leq e \cdot \exp\left(-\frac{\ln(\frac{\Delta_1}{\epsilon})}{A_{i,j}}\right).$$

Observe that $\frac{1}{A_{i,j}} - \ln(\frac{e}{A_{i,j}}) \geq 0$. For $A_{i,j} \in [0,1]$, a direct argument shows $\frac{\ln(t)}{A_{i,j}} - \ln(\frac{t}{A_{i,j}})$ is increasing in t for $t \geq e$. As $\Delta_1/\epsilon > e$, we have $\frac{\ln(\frac{\Delta_1}{\epsilon})}{A_{i,j}} \geq \ln(\frac{\Delta_1}{\epsilon A_{i,j}})$. Therefore,

$$e \exp\left(-\frac{\ln(\frac{\Delta_1}{\epsilon})}{A_{i,j}}\right) \leq e \exp\left(-\ln\left(\frac{\Delta_1}{\epsilon A_{i,j}}\right)\right) = \frac{e\epsilon A_{i,j}}{\Delta_1}.$$

This concludes the proof. ∎

B.3 Proof of Lemma 10

Proof. Renumber the indices so that $A_{i,1} \leq \cdots \leq A_{i,n}$. Note that the index j might have changed to j' but for simplicity we refer to j' as j. Let $\xi_\ell = 1$ if $x'_\ell = 1$ and 0 otherwise. Let $Y_{ij} = \sum_{\ell=1}^{j-1} A_{i,\ell}\xi_\ell$. We have

$$\Pr[E_{ij} | X_j = 1] \leq \Pr[Y_{ij} \geq \epsilon - A_{i,j}].$$

Let $A'_{i,\ell} = \frac{2}{\epsilon} \cdot A_{i,\ell}$ for $\ell \in [j]$. As $A_{i,\ell} \leq \epsilon/2$ for all $\ell \in [j]$, we have $A'_{i,\ell} \in [0,1]$. Let $Y'_{ij} = \sum_{\ell=1}^{j-1} A'_{i,\ell}\xi_\ell$. Then

$$\Pr[Y_{ij} \geq \epsilon - A_{i,j}] = \Pr[Y'_{ij} \geq 2 - A'_{i,j}].$$

To upper bound $\mathbb{E}[Y'_{ij}]$, we have

$$\mathbb{E}[Y'_{ij}] = \sum_{\ell=1}^{j-1} A'_{i,\ell} \cdot \Pr[X_\ell = 1] \leq \frac{2\alpha_2}{\epsilon} \sum_{\ell=1}^{n} A_{i,\ell} x_\ell \leq \frac{2\alpha_2(1+\epsilon)}{\epsilon} = \frac{2\epsilon(1+\epsilon)}{c_3\Delta_1}.$$

Let $\alpha'_2 = \frac{2\epsilon}{c_3\Delta_1}$ and $W = 2$. Then $\mathbb{E}[Y'_{ij}] \leq \alpha'_2 W$. As $\alpha'_2 < 1/2$ and $A'_{i,j} \leq 1$, we see that $((1-\alpha)W)/A'_{i,j} > 1$. Therefore, as $A'_{i,\ell} \leq A'_{i,j}$ for all $\ell < j$, we can apply the Chernoff bound in Theorem 13 to obtain

$$\Pr[Y'_{ij} \geq 2 - A'_{i,j}] \leq \left(\frac{\alpha'_2 e^{1-\alpha'_2} W}{W - A'_{i,j}}\right)^{(W-A'_{i,j})/A'_{i,j}}.$$

Observe that $e^{1-\alpha'_2} \leq e$ and $\frac{W}{W-A'_{i,j}} \leq 2$ since $W = 2$ and $A'_{i,j} \leq 1$. By our choice of α'_2,

$$\left(\frac{\alpha'_2 e^{1-\alpha'_2} W}{W - A'_{i,j}}\right)^{(W-A'_{i,j})/A'_{i,j}} \leq (2e\alpha'_2)^{(W-A'_{i,j})/A'_{i,j}} = \left(\frac{\epsilon}{2e^{2/e}\Delta_1}\right)^{(W-A'_{i,j})/A'_{i,j}}$$

We prove the final inequality in two parts. First, we note that $\frac{W-A'_{i,j}}{A'_{i,j}} \geq 1$ since $W = 2$ and $A'_{i,j} \leq 1$. Then

$$\left(\frac{1}{2\Delta_1}\right)^{(W-A'_{i,j})/A'_{i,j}} \leq \frac{1}{2\Delta_1}.$$

Second, we observe $\frac{W-A'_{i,j}}{A'_{i,j}} \geq 1/A'_{i,j}$ since $W = 2$ and $A'_{i,j} \leq 1$. Then we can apply Lemma 16 to obtain

$$(\epsilon/e^{2/e})^{(W-A'_{i,j})/A'_{i,j}} \leq (\epsilon/e^{2/e})^{1/A'_{i,j}} \leq \frac{\epsilon A'_{i,j}}{2}.$$

We have shown $\Pr[E_{ij} | X_j = 1] \leq \frac{\epsilon A'_{i,j}}{4\Delta_1}$. Since $A'_{i,j} = A_{i,j} \cdot \frac{2}{\epsilon}$, the result follows. $\square$