THEORY OF COMPUTING, Volume 8 (2012), pp. 533–565
www.theoryofcomputing.org

Solving Packing Integer Programs via Randomized Rounding with Alterations*

Nikhil Bansal† Nitish Korula‡ Viswanath Nagarajan
Aravind Srinivasan§

Received: July 31, 2011; published: October 28, 2012.

Abstract: We give new approximation algorithms for packing integer programs (PIPs) by employing the method of randomized rounding combined with *alterations*. Our first result is a simpler approximation algorithm for general PIPs which matches the best known bounds, and which admits an efficient parallel implementation. We also extend these results to a *multi-criteria* version of PIPs.

Our second result is for the class of packing integer programs (PIPs) that are *column sparse*, i. e., where there is a specified upper bound k on the number of constraints that each variable appears in. We give an $(\mathrm{e}k + o(k))$-approximation algorithm for k-column sparse PIPs, improving over previously known $O(k^2)$-approximation ratios. We also generalize our result to the case of maximizing non-negative monotone *submodular* functions over k-column sparse packing constraints, and obtain an $\left(\frac{\mathrm{e}^2 k}{\mathrm{e}-1} + o(k)\right)$-approximation algorithm. In obtaining this result, we prove a new property of submodular functions that generalizes the fractional subadditivity property, which might be of independent interest.

ACM Classification: F.2.2

AMS Classification: 68W25, 90C05, 90C10

Key words and phrases: approximation algorithms, packing integer programs, submodular functions

*This is a full version that contains results from two previous conference papers, [30] and [4].

†Part of this work was while the author was at IBM T.J. Watson Research Center.

‡This work was done while the author was at the Department of Computer Science of the University of Illinois, and was supported by a University of Illinois dissertation completion fellowship.

§Part of this work done while at Bell Laboratories, Lucent Technologies, Murray Hill, NJ 07974. Supported in part by NSF ITR Award CNS-0426683, NSF Award CNS-0626636 and NSF Award CNS 1010789.

 DOI: 10.4086/toc.2012.v008a024

1 Introduction

Packing integer programs (PIPs) are those integer programs of the form:

$$\max\left\{w^T x \,\middle|\, Ax \le b,\ x \in \{0,1\}^n\right\}, \quad \text{where } w \in \mathbb{R}_+^n,\, b \in \mathbb{R}_+^m \text{ and } A \in \mathbb{R}_+^{m\times n}.$$

Above, n is the number of variables/columns, m is the number of rows/constraints, A is the matrix of *sizes*, b is the *capacity* vector, and w is the *weight* vector. PIPs are a large and varied class of problems occupying a central place in combinatorial optimization. They include problems which can be solved optimally (e. g., matching) or approximated well (e. g., the knapsack problem). At the other end of the spectrum, they include problems that are effectively inapproximable, such as the classic independent set problem, which is NP-hard to approximate within a factor of $n^{1-\varepsilon}$ [33], while an n-approximation is trivial.

Many algorithms for combinatorial problems that are modeled as PIPs rely heavily on the structure of the problem; though these problem-specific techniques often lead to good results, they may not be more broadly applicable. Thus, there is also a need for general techniques that can be applied to a large class of problems. One natural candidate is *randomized rounding*, which involves first solving a linear programming relaxation to obtain a fractional solution x^*, and then converting/rounding x^* to a true integral solution by setting variable j to one with probability x_j^* divided by a suitable scaling factor. However, the difficulty in applying this technique to all packing problems is that for some instances, unless the scaling factor is extremely large (resulting in a solution with low expected weight), the probability of obtaining a feasible solution is extremely small.

In this paper, we take the approach of *randomized rounding with alteration*. The method of alteration is a key branch of the probabilistic method: do a random construction (allowing the result to not satisfy all our requirements), alter the constructed random structure appropriately, and argue that this achieves our goal (typically in expectation or with high probability). Applied to solving PIPs, we obtain a fractional solution x^* and independently select each variable j with probability proportional to x_j^*. This may leave some of the packing constraints unsatisfied; we then argue that a few variables can be deleted to leave us with a feasible solution of comparable weight. Our alteration/deletion procedure ensures that we always obtain a feasible integral solution.

We illustrate the usefulness of this method by first using it to obtain a simple algorithm matching the best approximation guarantee known for general PIPs [29]. We then consider the special case of *k-column sparse PIPs* (denoted k-CS-PIP), which are PIPs where the number of non-zero entries in each column of matrix A is at most k. This is a fairly general class and models several basic problems such as k-set packing [16] and independent set in graphs with degree at most k.

In addition to being simple and general, our rounding algorithm has several advantages. First, it is inherently parallelizable: in the initial rounding step, we independently select each variable j according to its fractional value, and in the subsequent alteration step, each constraint independently marks certain variables for deletion. Thus, we obtain RNC-approximation algorithms. Second, our algorithm to convert a fractional solution to an integer solution does not depend on the objective function. This allows us to (i) obtain good solutions to *multi-criteria* (multiple objective) packing problems and (ii) extend some of our results to the *submodular* setting; we elaborate on both of these subsequently.

Notation Before stating our results in full, we fix some useful notation. For any integer $t \geq 1$, we use $[t] := \{1,2,\ldots,t\}$. We index items (i. e., columns) by $j \in [n]$ and constraints (i. e., rows) by $i \in [m]$. Given a matrix $A \in [0,1]^{m\times n}$, vectors $b \in [1,\infty)^m$ and $w \in \mathbb{R}_+^n$, we seek to solve the following packing integer program:

$$\max\left\{\sum_{j=1}^{n} w_j x_j \;\middle|\; \sum_{j=1}^{n} a_{ij}\cdot x_j \leq b_i, \forall i \in [m];\; x_j \in \{0,1\}, \forall j \in [n]\right\}.$$

It is well-known that PIPs do not require any assumptions on A,b,w other than non-negativity: the assumptions $b \in [1,\infty)^m$ and $A \in [0,1]^{m\times n}$ are without loss of generality, see, e. g., [29]. We also assume by scaling that $\max_{j\in[n]} w_j = 1$; thus the optimal value of a PIP is always at least one. We refer to A as a *size* matrix, with a_{ij} denoting the size of item j in constraint i. A useful parameter in studying PIPs is $B = \min_{i\in[m]} b_i$, which measures the relative "slack" between item sizes and the capacity constraints. The natural LP relaxation of a PIP simply replaces the integrality constraint $x \in \{0,1\}^n$ by $x \in [0,1]^n$.

The special case when the sizes $A \in \{0,1\}^{m\times n}$ is called $\{0,1\}$-PIP. Often one can obtain stronger results for $\{0,1\}$-PIPs. In this case we may assume, without loss of generality, that vector b is also integral.

Note that a PIP may also have general upper-bound constraints of the form "$x_j \leq d_j$" for some integer d_j, instead of unit upper-bounds $x_j \in \{0,1\}$. Such constraints are only "easier" to deal with; intuitively, rounding a variable with fractional value 26.3 to either 26 or 27 is a less delicate operation than rounding 0.3 to either 0 or 1. Formally (as shown next), given an LP-based ρ-approximation for a PIP with unit upper-bounds, one can obtain a $(\rho+1)$-approximation for the same PIP even with general upper-bounds. Consider any PIP $\mathcal{P}$ with general upper-bounds. The algorithm first solves the natural LP relaxation of $\mathcal{P}$ to obtain fractional solution $y \in \mathbb{R}_+^n$. Let $z \in \mathbb{Z}_+^n$ and $x \in [0,1]^n$ be defined as: $z_j = \lfloor y_j \rfloor$ and $x_j = y_j - \lfloor y_j \rfloor$ for all $j \in [n]$; note that $w^T y = w^T z + w^T x$. Clearly z is a feasible integral solution to $\mathcal{P}$. Moreover x is a feasible fractional solution to $\mathcal{P}$ even with *unit upper-bounds*. Hence using the LP-based ρ-approximation, we obtain a feasible integral solution $\bar{x} \in \{0,1\}^n$ with $w^T \bar{x} \geq (1/\rho)\cdot w^T x$. It can be seen by simple calculation that the better of z and $\bar{x}$ is a $(\rho+1)$-approximate solution relative to the natural LP relaxation for $\mathcal{P}$. Based on this observation, throughout the paper we assume that the only integrality constraints are of the form $x_j \in \{0,1\}$.

We say that item j *participates* in constraint i if $a_{ij} > 0$. For each $j \in [n]$, let $N(j) := \{i \in [m] \mid a_{ij} > 0\}$ be the set of constraints that j participates in. In a k-column sparse PIP, we have $|N(j)| \leq k$ for each $j \in [n]$. Finally, for each constraint i, we let $P(i)$ denote the set of items participating in this constraint. Note that $|P(i)|$ can be arbitrarily large even for column-sparse PIPs.

1.1 Our results and techniques

1.1.1 RNC-approximation algorithms for packing integer programs

We present the first RNC-approximation algorithms for PIPs that match the currently best known sequential approximation guarantees [29] to within a constant factor.

Theorem 1.1. *There is an RNC-algorithm for PIPs achieving an $O\left(1+(K_1 m/y^*)^{1/(B-1)}\right)$-approximation ratio, where y^* is the optimal value of its LP relaxation and $K_1 \geq 1$ is a constant. For $\{0,1\}$-PIPs, the approximation ratio improves to $O\left(1+(K_1 m/y^*)^{1/B}\right)$.*

Remark Our analysis seems to require the constant $K_1 > 1$, in order to handle the case where B may be more than 1 by a tiny amount. However, for the case of $\{0,1\}$-PIPs, we note that K_1 can be eliminated from the statement of Theorem 1.1 by absorbing into the "$O(1)$" term.

Before describing our approach, we outline the algorithm in [29], which is also based on randomized rounding. Concretely, suppose we have a PIP, with each x_j required to lie in $\{0,1\}$. Let x^* denote a near-optimal solution to the LP relaxation, with objective value $y^* = \sum_j w_j x_j^*$. For a certain $\lambda \geq 1$ (which is the factor in Theorem 1.1), round each x_j independently: to 1 with probability x_j^*/λ and to 0 with probability $1 - x_j^*/\lambda$. It is shown in [29] that a suitable choice of λ ensures, with positive probability, that: (i) all constraints are obeyed, and (ii) the objective function is at least as large as, say, one-third its expected value y^*/λ. However, the catch is that this "positive probability" can be exponentially small in the input size of the problem: an explicit example of this is shown in [29]. Thus, if we just conduct a randomized rounding, we will produce a solution guaranteed in [29] with positive probability, but we are not assured of a randomized *polynomial-time* algorithm for constructing such a solution. It is then shown in [29] that the method of conditional probabilities can be adapted to the situation and that we can efficiently round the variables one-by-one, achieving the existential bound.[1] However, this approach seems inherently sequential, and hence it appears difficult to develop an RNC version of it.

It is known that the LP relaxation of PIPs can be solved in NC, losing only a $(1 + (\log(m+n))^{-C})$ factor in the objective value for any desired constant $C > 0$ [22]. So the key step is to efficiently round this fractional solution to an integer solution. As mentioned previously, we tackle this problem through randomized rounding followed by alteration. The randomized rounding step uses parameters similar to those of [29]; the probability of all constraints being satisfied is positive, but can be tiny. To deal with the unsatisfied constraints, we apply a parallel *greedy* technique that modifies some variables to enforce the constraints, and then argue that the objective function does not change much in the process; our specific greedy approach is critical to making this argument work. Altogether, this yields our RNC approximation algorithm for PIPs.

Multi-criteria optimization There has been considerable research on *multi-criteria* (approximation) algorithms in scheduling, network design, routing and other areas: see, e. g., [28, 10, 19]. The motivation is to study how to balance *multiple* objective functions under a given system of constraints, modeling the fact that different participating individuals/organizations may have differing objectives. One abstract framework for this is in the setting of PIPs: given the constraints of a PIP, a *set* of non-negative vectors $\{w_1, w_2, \ldots, w_\ell\}$, and a feasible solution x^* to the LP relaxation of the PIP's constraints, do there exist "good" integral solutions x (and can we compute them efficiently) that achieve a "balance" among the different weight functions w_i? For instance, given a PIP's constraints, we could ask for (approximately) the smallest $\alpha \geq 1$ such that there exists an integral feasible x with $w_i \cdot x \geq (w_i \cdot x^*)/\alpha$ for all i. We show (Theorem 2.3) how our alteration method helps expand the set of situations here for which good RNC approximation algorithms exist; it is crucial here that our alteration step depends only on the constraint matrix A and not on the objective function. In Section 2.3 we also show why earlier approaches such as that of [29] cannot achieve the bounds we get here.

[1]This is one of the very few situations known to us where a probabilistic argument does not yield a good randomized algorithm, but spurs the development of an efficient deterministic algorithm.

1.1.2 k-column-sparse packing integer programs

Recall that a k-column-sparse PIP is one in which each item j participates in at most k constraints (i. e., $\forall j, |N(j)| \leq k$). We present improved approximation bounds in this case.

Theorem 1.2. *There is a randomized* $(\mathrm{e}k + o(k))$*-approximation algorithm for k-*CS*-*PIP*.*

This improves over the previously best known $O(k^2)$ approximation ratio, due independently to Chekuri, Ene and Korula [9] and Pritchard and Chakrabarty [25].

Our algorithm is based on solving a *strengthened* version of the natural LP relaxation of k-CS-PIP, and then performing randomized rounding followed by suitable alterations. We use essentially the same greedy alteration step as for general PIPs. A similar approach can be used with the natural LP relaxation for k-CS-PIP, obtained by simply dropping the integrality constraints on the variables; this gives a slightly weaker $8k$-approximation bound. To obtain the $\mathrm{e}k + o(k)$ bound, we construct a stronger LP relaxation by adding additional valid inequalities for k-CS-PIP. The analysis of our rounding procedure is based on exploiting these additional constraints and using the positive correlation between various probabilistic events via the FKG inequality.

Our result is almost the best possible that one can hope for using the LP based approach. We show that the integrality gap of the strengthened LP is at least $2k-1$, so our analysis is tight up to a small constant factor $\mathrm{e}/2 \approx 1.36$ for large values of k. Even without restricting to LP based approaches, an $O(k)$ approximation is nearly best possible since it is NP-hard to obtain an $o(k/\log k)$-approximation for the special case of k-set packing [15].

Our second main result for column-sparse packing is for the more general problem of maximizing a monotone submodular function over packing constraints that are k-column sparse. This problem is a common generalization of maximizing a submodular function over (i) k-dimensional knapsack [20], and (ii) intersection of k partition matroids [13].

Theorem 1.3. *There is a randomized algorithm for maximizing any non-negative monotone submodular function (given by a value oracle) over k-column sparse packing constraints, achieving approximation ratio*

$$\left(\frac{\mathrm{e}^2}{\mathrm{e}-1}\right) k + o(k)\,.$$

Our algorithm uses the continuous greedy algorithm of Vondrák [31] in conjunction with our randomized rounding plus alteration based approach. However, it turns out that the analysis of the approximation guarantee is much more intricate: in particular, we need a generalization of a result of Feige [11] that shows that submodular functions are also *fractionally subadditive*. See Section 4 for a statement of the new result, Theorem 4.3, and related context. This generalization is based on an interesting connection between submodular functions and the FKG inequality. We believe that this result might be of further use in the study of submodular optimization.

Finally, we also obtain improved results for k-CS-PIP when capacities are large relative to the sizes, i. e., as a function of the parameter $B := \min_i b_i$.

Theorem 1.4. *There is an approximation algorithm for k-*CS*-*PIP *achieving approximation ratio*

$$4\mathrm{e} \cdot \left(\mathrm{e}\,\lfloor B \rfloor\, k + c\right)^{1/\lfloor B \rfloor}$$

and an approximation algorithm for maximizing any non-negative monotone submodular function over k-column sparse packing constraints achieving approximation ratio

$$\frac{4\mathrm{e}^2}{\mathrm{e}-1}\cdot\left(\mathrm{e}\lfloor B\rfloor k+c\right)^{1/\lfloor B\rfloor};$$

here $c>0$ is a fixed constant.

Notice that we always have $B \geq 1$, so the approximation factor here is $O\left(k^{1/\lfloor B\rfloor}\right)$. We also show that this result is tight up to constant factors, relative to its LP relaxation.

1.2 Related previous work

As already mentioned, the best known approximation bounds for general PIPs are due to [29]; this obtained the ratios stated in Theorem 1.1 using a *sequential* algorithm. Simple randomized rounding, as in [26], yields an approximation ratio of $O(m^{1/B})$ for general PIPs and $O(m^{1/(B+1)})$ for $\{0,1\}$-PIPs. The crux of [29] is that much better bounds can be obtained via an analysis of correlations.[2] It is shown in [18] that the improved bounds for $\{0,1\}$-PIPs can be extended to the special case of *column restricted PIPs*, where in each column, all the non-zero entries are identical.

The *column sparse* case of PIPs (k-CS-PIP) was explicitly considered only recently, by Pritchard [24]. In a somewhat surprising result, [24] gave an algorithm for k-CS-PIP where the approximation ratio only depends on k; this is very useful when k is small. This result is surprising because in contrast, no such guarantee is possible for k-row sparse PIPs. In particular, the independent set problem on general graphs is a 2-row sparse PIP, but is NP-hard to approximate within $n^{1-\varepsilon}$ (for any $\varepsilon>0$) [33]. Pritchard's algorithm [24] had an approximation ratio of $2^k\cdot k^2$. This was based on iteratively solving an LP relaxation, and then applying a randomized selection procedure. Subsequently, in independent work, Chekuri et al. [9] and Pritchard and Chakrabarty [25] showed that this final step could be derandomized, yielding an improved bound of $O(k^2)$. All these previous results crucially use the structural properties of basic feasible solutions of the LP relaxation. However, as stated above, our $O(k)$-approximation algorithm is based on randomized rounding with alterations and does not use properties of basic solutions. This is crucial for the submodular maximization version of the problem, as a solution to the fractional relaxation there does not have these properties.

Related issues have been considered in discrepancy theory, where the goal is to round a fractional solution to a k-column sparse linear program so that the capacity violation for any constraint is minimized. A celebrated result of Beck and Fiala [5] shows that the capacity violation is at most $O(k)$. A major open question in discrepancy theory is whether the above bound can be improved to $O(\sqrt{k})$, or even $O(k^{1-\varepsilon})$ for a fixed $\varepsilon>0$. While the result of [24] uses techniques similar to that of [5], a crucial difference in our problem is that no constraint can be violated at all. In fact, at the end of Section 3, we show another crucial qualitative difference between discrepancy and k-CS-PIP.

Various special cases of k-CS-PIP have been extensively studied. One important special case is the *k-set packing* problem, where, given a collection of sets of cardinality at most k, the goal is to find the maximum weight sub-collection of mutually disjoint sets. This is equivalent to k-CS-PIP where the

[2] Recall the parameter $B=\min_i b_i$. If, say, $B=1$ and $A\in\{0,1\}^{m\times n}$ for a given PIP, [26] shows how to construct an integral solution of value $v=\Omega(y^*/\sqrt{m})$; [29] constructs a solution of value $\Omega(v^2)$ in this case.

constraint matrix A has only 0-1 entries. Note that for $k=2$ this is *maximum weight matching* which can be solved in polynomial time, and for $k=3$ the problem becomes APX-hard [15]. After a long line of work [16, 2, 8, 6], the best approximation ratio known for this problem is $((k+1)/2)+\varepsilon$ obtained using local search techniques [6]. An improved bound of $(k/2)+\varepsilon$ is also known [16] for the unweighted case, i. e., when the weight vector $w=\mathbf{1}$. It is also known that the natural LP relaxation for this problem has integrality gap at least $k-1+1/k$, and in particular this holds for the projective plane instance of order $k-1$. Hazan, Safra and Schwartz [15] showed that k-set packing is NP-hard to approximate within $\Omega(k/\log k)$. Another special case of k-CS-PIP is the independent set problem in graphs with maximum degree at most k. This is equivalent to k-CS-PIP where the constraint matrix A is 0-1, and each row is 2-sparse. This problem has an $O(k\log\log k/\log k)$-approximation [14], and is $\Omega(k/\log^2 k)$-hard to approximate [3], assuming the Unique Games Conjecture [17].

There is a large body of work on constrained maximization of submodular functions; we only cite the relevant papers here. Călinescu, Chekuri, Pál and Vondrák [7] introduced a continuous relaxation (called the *multi-linear extension* or extension-by-expectation) of submodular functions. They gave an elegant $e/(e-1)$-approximation algorithm for maximizing this continuous relaxation over any "downward monotone" polytope $\mathcal{P}$, as long as there is a polynomial-time algorithm for maximizing linear functions over $\mathcal{P}$; see also [31]. We use this continuous relaxation in our algorithm for submodular maximization over k-sparse packing constraints. As noted earlier, k-sparse packing constraints generalize both k-partition matroids and k-dimensional knapsacks. Nemhauser, Wolsey and Fisher [13] gave a $(k+1)$-approximation for submodular maximization over the intersection of k partition matroids. When k is constant, Lee, Mirrokni, Nagarajan and Sviridenko [21] improved this to $k+\varepsilon$; very recently, Ward [32] improved this further to $(k+3)/2$ (which also holds under more general constraints introduced by [12]). Kulik, Shachnai and Tamir [20] gave an $(e/(e-1)+\varepsilon)$-approximation for submodular maximization over k-dimensional knapsacks when k is constant; if k is part of the input, the best known approximation bound is $O(k)$.

1.3 Organization

In Section 2, we study general PIPs and prove Theorem 1.1. This also introduces the framework of randomized rounding with alteration. Additionally, in Section 2.3, we describe how our approach extends to packing problems with multiple objectives.

In Section 3, we move on to column sparse PIPs. We begin with the natural LP relaxation for k-CS-PIP and describe a simple algorithm with approximation ratio $8k$. We then present a stronger relaxation, and use it to obtain the better result in Theorem 1.2. We also present an integrality gap of $2k-1$ for this strengthened LP, implying that our result is almost tight.

In Section 4, we consider k-column sparse packing problems over a submodular objective and prove Theorem 1.3. In Section 5, we deal with k-CS-PIP when the sizes of items may be small relative to the capacities of constraints and prove Theorem 1.4.

We remark that we do not include some of the results from the conference paper [30] here, including the extensive details of space-bounded derandomization and the Group Steiner Tree problem: this is so as to not distract from the main ideas presented here.

2 Improved RNC algorithm for general PIPs

Preliminaries We first recall the Chernoff-Hoeffding bounds [23, 1]. Let $Z_1, Z_2, \ldots, Z_\ell$ be *independent* random variables, each taking values in $[0,1]$. Let $Z = \sum_i Z_i$ and $\mathbb{E}[Z] = \mu$. Then,

$$\Pr[Z \geq t \cdot \mu] \;\leq\; G(\mu, t) \;:=\; (\mathrm{e}/t)^{t \cdot \mu}\,, \quad \forall t \geq 1\,. \tag{2.1}$$

Throughout this paper, all logarithms are taken to the base 2.

As described before, the randomized rounding involves selecting item j with probability x_j^*/λ for a suitable scaling parameter $\lambda \geq 1$; the choice of λ is determined by the bounds above. Recall that $y^* = \sum_j w_j x_j^*$ denotes the optimal LP objective value for a PIP; also recall the parameter $B = \min_i b_i$. As in [29], the scaling parameter λ is chosen to be:

$$K_0 \cdot \left(1 + (K_1 m / y^*)^{1/B}\right) \qquad \text{if } A \in \{0,1\}^{m \times n}\text{, and} \tag{2.2}$$

$$K_0 \cdot \left(1 + (K_1 m / y^*)^{1/(B-1)}\right) \qquad \text{otherwise,} \tag{2.3}$$

where $K_0 > 1$ and $K_1 \geq 1$ are constants that will be fixed later. We present a randomized algorithm that finds an integral solution of value at least $\Omega(y^*/\lambda)$. Note that the approximation guarantee $O(\lambda)$ is better for the case $A \in \{0,1\}^{m \times n}$ than for the general case where $A \in [0,1]^{m \times n}$. As mentioned earlier, this approximation bound was first obtained by [29]. However, our analysis is much simpler, and our algorithm admits an efficient parallel implementation, i. e., an RNC algorithm.

Section 2.1 presents our algorithm that involves randomized rounding followed by alteration; and in Section 2.2 we give the analysis.

2.1 The basic algorithm

Given a PIP, for any desired constant $C > 0$, there is an NC algorithm to find a feasible solution x^* to its LP relaxation with objective function value at least $1 - (\log(m+n))^{-C}$ times the optimal LP objective value [22].

Randomized rounding Given the vector x^*, we choose a suitable λ according to (2.2)-(2.3). Then we do the following randomized rounding: independently for each j, round x_j to 1 with probability x_j^*/λ and to 0 with probability $1 - x_j^*/\lambda$. Let $X_j \in \{0,1\}$ denote the outcome of the rounding for each item $j \in [n]$. This step is basically the same as in [26, 29]. However, some constraints of the PIP may be violated, and we alter $\vec{X} = \langle X_1, \ldots, X_n \rangle$ to correct these violations.

Alteration For each violated constraint $i \in [m]$, we run the following process $\mathcal{F}_i$ to enforce the constraint. Let L_i be the permutation $\langle \sigma(i,1), \sigma(i,2), \ldots, \sigma(i,n) \rangle$ of $[n]$ such that:

$$a_{i,\sigma(i,1)} \;\geq\; a_{i,\sigma(i,2)} \;\geq\; \cdots \;\geq\; a_{i,\sigma(i,n)}$$

where ties are broken arbitrarily. So L_i orders all items in non-increasing order of their size in the ith constraint. Now, process $\mathcal{F}_i$ traverses items in the order L_i and for each item $j = \sigma(i,k)$, sets $X_j \leftarrow 0$ if

$$\sum_{\ell=k}^{n} a_{i,\sigma(i,\ell)} \cdot X_{\sigma(i,\ell)} \;>\; b_i\,.$$

In other words, $\mathcal{F}_i$ rounds down items in the order L_i until constraint i is satisfied.

It is easy to see that our alteration can be implemented in NC by a parallel prefix computation. Also, this is done in parallel for all constraints i that were violated; these parallel threads $\mathcal{F}_1, \ldots, \mathcal{F}_m$ do not interact with each other. After this alteration, all constraints will be satisfied. The alteration is greedy in the sense that each $\mathcal{F}_i$, guided by its permutation L_i, tries to alter as few variables as possible.

For each variable $j \in [n]$ let R_j denote the outcome of randomized rounding, i. e., the value *before* alteration; and let X_j denote its value *after* the alteration. So $\vec{X}$ is the final integral solution.

Example Suppose the ith constraint is $0.4x_2 + 0.5x_3 + 0.3x_5 + 0.5x_7 + 0.35x_8 \leq 1$, and that randomized rounding set $X_3 := 0$ and each of X_2, X_5, X_7 and X_8 to 1. Then the ith constraint will round down precisely X_7 and X_2 to zero. (Now, some other constraint may round down X_8, in which case we could potentially revisit the ith constraint and try to set, say, X_2 back to 1. We do not analyze such possible optimizations.)

2.2 Analysis of the algorithm

Recall the notation from above. Suppose we run the "randomized rounding and alteration" of Section 2.1, with λ as in (2.2) and (2.3). Let U_i be the random variable denoting the number of variables altered in row $i \in [m]$, by process $\mathcal{F}_i$. (Suppose $\mathcal{F}_i$ stops at some index j in the list L_i. Note that $\mathcal{F}_{i'}$, for some $i' \neq i$, may turn some variable $X_{j'}$ from 1 to 0, with j' appearing *after* j in L_i. Such variables j' are *not* counted in calculating U_i.)

Notation (i) For $k = 1, 2, \ldots$, let $L_{i,k}$ be the sub-list of L_i such that for all $j \in L_{i,k}$, $a_{i,j} \in (2^{-k}, 2^{-k+1}]$. Note that L_i is the concatenation of $L_{i,1}, L_{i,2}, \ldots$. (ii) Define $L_{i,\geq k}$ to be the set obtained by collecting together the elements of $L_{i,t}$, for all $t \geq k$. (iii) Let Z_i denote the largest k for which some element of $L_{i,k}$ was reset from 1 to 0 by $\mathcal{F}_i$; set $Z_i = 0$ if the ith constraint was not violated. (iv) A $\{0,1\}$-PIP is one where the sizes $A \in \{0,1\}^{m \times n}$.

Good upper-tail bounds for the U_i will be crucial to our analysis: Lemma 2.1 provides such bounds. Parts (a) and (b) of Lemma 2.1 will respectively help handle the cases of "large" Z_i and "small" Z_i. In the statement of the lemma, the function G is from the Chernoff bound inequality (2.1).

Lemma 2.1. *Fix a PIP and a fractional solution x^*; let $i \in [m]$, and $k, y \geq 1$ be integers. Then:*

a. $\Pr[Z_i \geq k] \;\leq\; G(b_i 2^{k-1}/\lambda, \lambda)$.

b. If $\lambda \geq 3$, then $\Pr[Z_i = k \wedge U_i \geq y] \;\leq\; O\big((\mathrm{e}/\lambda)^{b_i + 0.5(\lceil y/k \rceil - 1)}\big)$.

c. For $\{0,1\}$-PIPs, $\Pr[U_i \geq y] \;\leq\; G\big(b_i/\lambda, \lambda(1 + y/b_i)\big)$.

Proof. Part (a): $Z_i \geq k$ implies that $\sum_{j \in L_{i,\geq k}} a_{i,j} R_j > b_i$, i. e., that $\sum_{j \in L_{i,\geq k}} 2^{k-1} a_{i,j} R_j > 2^{k-1} b_i$. Since $\mathbb{E}[R_j] = x_j^*/\lambda$, $\mathbb{E}[\sum_j a_{i,j} R_j] \leq b_i/\lambda$; so

$$\mathbb{E}\left[\sum_{j \in L_{i,\geq k}} 2^{k-1} a_{i,j} R_j\right] \;\leq\; b_i 2^{k-1}/\lambda\,.$$

Also, $2^{k-1} a_{i,j} \in [0,1]$ for all $j \in L_{i,\geq k}$. Bound (2.1) now completes the proof.

Part (b): Let $U_{i,\ell}$ denote the number of elements of $L_{i,\ell}$ that are altered by $\mathcal{F}_i$. Now, $U_i = \sum_\ell U_{i,\ell}$; also, if $Z_i = k$, then $U_{i,\ell} = 0$ for $\ell > k$. So, the event $\{Z_i = k \wedge U_i \geq y\}$ implies the existence of $\ell \in [k]$ with $U_{i,\ell} \geq \lceil y/k \rceil$. This in turn implies that

$$\sum_{j \in L_{i,\geq \ell}} a_{i,j} R_j \; > \; b_i + (\lceil y/k \rceil - 1) 2^{-\ell},$$

since resetting any element of $L_{i,\ell}$ from 1 to 0 decreases $\sum_j a_{i,j} R_j$ by more than $2^{-\ell}$. Let $\theta := \lceil y/k \rceil - 1$ for notational convenience; note that $\theta \geq 0$ since $y \geq 1$. So,

$$\Pr[Z_i = k \wedge U_i \geq y] \; \leq \; \sum_{\ell=1}^{k} \Pr\Bigg[\sum_{j \in L_{i,\geq \ell}} 2^{\ell-1} a_{i,j} R_j \; > \; b_i 2^{\ell-1} + \theta/2\Bigg]. \tag{2.4}$$

We have that

$$\mathbb{E}\Bigg[\sum_{j \in L_{i,\geq \ell}} 2^{\ell-1} a_{i,j} R_j\Bigg] \; \leq \; b_i 2^{\ell-1}/\lambda\,,$$

and that $2^{\ell-1} a_{i,j} \in [0,1]$ for all $j \in L_{i,\geq \ell}$. Using (2.1) to bound (2.4), we get the bound

$$\begin{aligned}
\sum_{\ell=1}^{k} G\big(b_i 2^{\ell-1}/\lambda, \lambda(1+\theta 2^{-\ell}/b_i)\big) \; &\leq \; \sum_{\ell=1}^{k} \left(\frac{\mathrm{e}}{\lambda(1+\theta 2^{-\ell}/b_i)}\right)^{b_i 2^{\ell-1}+0.5\theta} \\
&\leq \; \sum_{\ell=1}^{k} (\mathrm{e}/\lambda)^{b_i 2^{\ell-1}+0.5\theta} \; = \; O\Big((\mathrm{e}/\lambda)^{b_i+0.5\theta}\Big).
\end{aligned}$$

The second inequality uses $b_i \geq 1$ and $\theta \geq 0$; the equality is by $\lambda \geq 3$ and a geometric summation.
Part (c): Here, $U_i \geq y$ if and only if $\sum_j a_{i,j} R_j \geq b_i + y$; we now employ (2.1).

This concludes the proof of Lemma 2.1. □

Remark Observe how the greedy nature of $\mathcal{F}_i$ helps much in establishing Lemma 2.1.

Lemma 2.2. *There are constants $K_2, K_3, K_4 > 0$ such that the following hold. Fix any PIP and any $i \in [m]$; suppose $\lambda \geq 3$. Define $p = (\mathrm{e}/\lambda)^B$ for general PIPs, and $p = (\mathrm{e}/\lambda)^{B+1}$ for $\{0,1\}$-PIPs. Then:*

a. For any integer $y \geq 4$, $\Pr[U_i \geq y] \; \leq \; K_2\, p \cdot \mathrm{e}^{-K_3 y/\log y}$.

b. $\mathbb{E}[U_i] \leq K_4 p$.

Proof. Part (a): In case of $\{0,1\}$-PIPs, a stronger version of part (a) easily follows from part (c) of Lemma 2.1:

$$\begin{aligned}
\Pr[U_i \geq y] \; &\leq \; G\big(b_i/\lambda, \lambda(1+y/b_i)\big) \; = \; \left[\frac{\mathrm{e}}{\lambda(1+y/b_i)}\right]^{b_i+y} \; \leq \; (\mathrm{e}/\lambda)^{b_i+1} \cdot (\mathrm{e}/\lambda)^{y-1} \\
&\leq \; p \cdot (\mathrm{e}/\lambda)^{y/2} \; \leq \; p \cdot \mathrm{e}^{-c_1 \cdot y},
\end{aligned}$$

where $c_1 > 0$ is a constant. In the above we use $\lambda \geq 3 > \mathrm{e}$; also, the second-last inequality used $p = (\mathrm{e}/\lambda)^{B+1} \geq (\mathrm{e}/\lambda)^{b_i+1}$ and $y \geq 2$.

Now consider general PIPs, where $p = (\mathrm{e}/\lambda)^B$. Choose $r = \lfloor \log y \rfloor + 1$. We have:

$$\Pr[U_i \geq y] \;=\; \sum_{k \geq 1} \Pr[Z_i = k \wedge U_i \geq y] \;\leq\; \sum_{k=1}^{r} \Pr[Z_i = k \wedge U_i \geq y] + \Pr[Z_i \geq r+1]\,. \tag{2.5}$$

Using part (a) of Lemma 2.1 we have,

$$\Pr[Z_i \geq r+1] \;\leq\; G(b_i \cdot 2^r/\lambda, \lambda) \;\leq\; (\mathrm{e}/\lambda)^{b_i \cdot y} \;\leq\; (\mathrm{e}/\lambda)^{b_i} \cdot (\mathrm{e}/\lambda)^{y/2} \;\leq\; p \cdot \mathrm{e}^{-c_2 \cdot y}\,, \tag{2.6}$$

where $c_2 > 0$ is a constant. The second inequality uses $r \geq \log y$; the third inequality uses $b_i \geq 1$, $y \geq 2$ and $\mathrm{e}/\lambda < 1$; the last inequality uses $p = (\mathrm{e}/\lambda)^B$ and $B \leq b_i$.

Next we upper bound the summation $\sum_{k=1}^{r} \Pr[Z_i = k \wedge U_i \geq y]$. Notice that since $y \geq 4$, $r = \lfloor \log y \rfloor + 1$ and $1 \leq k \leq r$, we have $\lceil y/k \rceil - 1 \geq y/r - 1 \geq y/(4 \log y)$. Applying part (b) of Lemma 2.1,

$$\begin{aligned}\sum_{k=1}^{r} \Pr[Z_i = k \wedge U_i \geq y] \;&\leq\; O(1) \sum_{k=1}^{r} (\mathrm{e}/\lambda)^{b_i + 0.5(\lceil y/k \rceil - 1)} \;\leq\; O(1) \cdot r \cdot (\mathrm{e}/\lambda)^{b_i + \frac{y}{8 \log y}} \\ &\leq\; O(1) \cdot (\mathrm{e}/\lambda)^{b_i} \cdot \mathrm{e}^{c_3 (\log\log y - \frac{y}{8 \log y})} \;\leq\; O(1) \cdot (\mathrm{e}/\lambda)^{b_i} \cdot \mathrm{e}^{-c_4 \cdot \frac{y}{\log y}}\end{aligned} \tag{2.7}$$

where $c_3, c_4 > 0$ are constants. Since $p = (\mathrm{e}/\lambda)^B \geq (\mathrm{e}/\lambda)^{b_i}$, combining (2.5), (2.6) and (2.7) implies part (a).

Part (b): Note that

$$\mathbb{E}[U_i] \;=\; \sum_{y \geq 1} \Pr[U_i \geq y] \;\leq\; 3 \cdot \Pr[U_i \geq 1] + \sum_{y \geq 4} \Pr[U_i \geq y]\,.$$

Applying part (a) to the second term,

$$\sum_{y \geq 4} \Pr[U_i \geq y] \;\leq\; K_2 \cdot p \sum_{y \geq 4} \mathrm{e}^{-K_3 \cdot \frac{y}{\log y}} \;\leq\; K_4' \cdot p\,,$$

where $K_4' > 0$ is a constant.

Next we show that $\Pr[U_i \geq 1] \leq p$, which suffices to prove part (b). For a $\{0,1\}$-PIP, by part (c) of Lemma 2.1,

$$\Pr[U_i \geq 1] \;\leq\; G(b_i/\lambda, \lambda + \lambda/b_i) \;\leq\; (\mathrm{e}/\lambda)^{b_i+1} \;\leq\; (\mathrm{e}/\lambda)^{B+1} = p\,.$$

For a general PIP, using the Chernoff bound (2.1),

$$\Pr[U_i \geq 1] \;=\; \Pr\Big[\sum_j a_{i,j} \cdot R_j > b_i\Big] \;\leq\; G(b_i/\lambda, \lambda) \;=\; (\mathrm{e}/\lambda)^{b_i} \;\leq\; (\mathrm{e}/\lambda)^B \;=\; p\,. \qquad \square$$

We are now ready to analyze our RNC alteration algorithm. The expected objective value after the randomized rounding is y^*/λ. Since $w_j \leq 1$ for all j, the expected reduction in the objective value caused by the alteration is at most $\sum_{i \in [m]} \mathbb{E}[U_i] \leq K_4 m p$, by part (b) of Lemma 2.2. (This is an overcount since the same altered variable j may get counted by several U_i.) For a general PIP, by (2.3) we have:

$$K_4 m \cdot p \;=\; K_4 m \cdot (\mathrm{e}/\lambda)^B = \frac{y^*}{\lambda} \cdot \left(\mathrm{e} K_4 \cdot \frac{m}{y^*} \cdot (\mathrm{e}/\lambda)^{B-1} \right) \;\leq\; \frac{y^*}{\lambda} \cdot \left(\mathrm{e} K_4 \, (\mathrm{e}/K_0)^{B-1} \cdot \frac{m}{y^*} \frac{y^*}{K_1 m} \right) \leq \frac{y^*}{2\lambda}\,,$$

by setting $K_0 \geq \mathrm{e}$ and $K_1 \geq 2\mathrm{e}K_4$. Similarly, for a $\{0,1\}$-PIP, using $p = (\mathrm{e}/\lambda)^{B+1}$ and the choice of λ from (2.2), we obtain $K_4\, m \cdot p \leq y^*/(2\lambda)$.

Thus, the expected final objective value is at least $y^*/\lambda - K_4 mp \geq y^*/(2\lambda)$. This proves Theorem 1.1.

2.3 Multi-criteria PIPs

We now work with multi-criteria PIPs, generalizing the results from the previous subsection. The basic setting is as follows: suppose, as in a PIP, we are given a system of m linear constraints $Ax \leq b$, subject to each $x_j \in \{0,1\}$. Furthermore, instead of just one objective function, suppose we are given a collection of non-negative vectors $\{\vec{w}_1, \vec{w}_2, \ldots, \vec{w}_\ell\}$. The question is: given a feasible solution $\vec{x}^*$ to the LP relaxation of the constraints, do good *integral* feasible solutions $\vec{x}$ exist that can be computed/approximated efficiently and that achieve a "good balance" among the different $\vec{w}_i$? For instance, we focus in this section on the case where all the $\vec{w}_i$ have equal importance; so, we could ask for approximating the smallest $\alpha \geq 1$ such that there exists an integral feasible $\vec{x}$ with $\vec{w}_i \cdot \vec{x} \geq (\vec{w}_i \cdot \vec{x}^*)/\alpha$ for all weight vectors w_i.

We now show how our algorithm and analysis from above can be extended. For simplicity, we consider here the case where all the values $\vec{w}_i \cdot \vec{x}^*$ are of the same "order of magnitude": say, in the range $[y^*/2, y^*]$. Our result also holds in the general case, where we use $y^* = \min_{i \in [\ell]} \vec{w}_i \cdot \vec{x}^*$. Basically, Theorem 2.3 says that we can get essentially the same approximation guarantee of $O(\lambda)$ as for a single objective, even if we have up to $\exp(C_1 y^*/\lambda)$ objective functions $\vec{w}_i$. (See (2.2), (2.3) for the value of λ.)

Theorem 2.3. *There are constants $C_1 > 0$, $K_0 > 1$ and $K_1 \geq 1$ such that the following holds. Suppose we are given:*

a. *the constraints $Ax \leq b$ of a PIP;*

b. *a feasible solution $\vec{x}^*$ to the LP relaxation of these constraints, and*

c. *a collection of non-negative vectors $\{\vec{w}_1, \vec{w}_2, \ldots, \vec{w}_\ell\}$ such that $\max_{j \in [n]} w_i(j) \leq 1$ for all $i \in [\ell]$, and there exists y^* with $\vec{w}_i \cdot \vec{x}^* \in [y^*/2, y^*]$ for all i.*

Let λ be as in (2.2) and (2.3), and suppose ℓ, the number of given $\vec{w}_i$, is at most $\exp(C_1 y^*/\lambda)$. Then, there exists an integral feasible solution $\vec{z}$ to the constraints in (a), such that $\vec{w}_i \cdot \vec{z} \geq \vec{w}_i \cdot \vec{x}^*/(2\lambda)$ for all i; furthermore, we can compute such a $\vec{z}$ in RNC.

Proof. We use the same notation as previously; as described earlier, the basic algorithm is to conduct a randomized rounding with $\Pr[x_j = 1] = x_j^*/\lambda$ for each j, and then to conduct our alteration. Recall that $R_j \in \{0,1\}$ (for each $j \in [n]$) denotes the outcome of randomized rounding, before the alteration step. For each $i \in [\ell]$, we have $\mathbb{E}[\vec{w}_i \cdot \vec{R}] = \vec{w}_i \cdot \vec{x}^*/\lambda \geq y^*/(2\lambda)$, by assumption (c) of the theorem. Using a lower-tail Chernoff bound ([23, Theorem 4.2]) shows that for a certain absolute constant $C' > 0$,

$$\Pr[\vec{w}_i \cdot \vec{R} \leq \vec{w}_i \cdot \vec{x}^*/(1.5\lambda)] \;\leq\; \exp(-C' y^*/\lambda)\,.$$

So, the probability of existence of some $i \in [\ell]$ for which "$\vec{w}_i \cdot \vec{R} \leq \vec{w}_i \cdot \vec{x}^*/(1.5\lambda)$" holds, is at most

$$\ell \cdot \exp(-C' y^*/\lambda) \;\leq\; \exp(-(C' - C_1) y^*/\lambda)\,. \tag{2.8}$$

We may assume that $y^*/\lambda \geq (\ln 2)/C_1$, since otherwise $\ell \leq \exp(C_1 y^*/\lambda)$ implies that $\ell = 1$, leaving us with the "one objective function" case, which we have handled before. Therefore we have from (2.8) that

$$\Pr\left[\exists i \in [\ell] : \vec{w}_i \cdot \vec{R} \leq \vec{w}_i \cdot \vec{x}^*/(1.5\lambda)\right] \;\leq\; \exp\left(-(C' - C_1)\cdot(\ln 2)/C_1\right). \tag{2.9}$$

Lemma 2.2(b) shows that $\mathbb{E}[\sum_{i\in[m]} U_i] \leq K_4 mp$ (also recall definition of p). By Markov's inequality,

$$\Pr\left[\sum_{i\in[m]} U_i \geq K_4 C_2 mp\right] \;\leq\; \frac{1}{C_2}$$

for any $C_2 \geq 1$. Thus if $C_1 < C'$ and C_2 is a large enough constant such that the sum of $1/C_2$ and (2.9) is less than, and bounded away, from 1, then with positive constant probability, we have: (i) for all $i \in [\ell]$, $\vec{w}_i \cdot \vec{R} > \vec{w}_i \cdot \vec{x}^*/(1.5\lambda)$ and (ii) the total number of altered variables is at most $K_4 C_2 mp$. Let $\vec{X}$ denote the values of variables after alteration. Then with constant probability, for all $i \in [\ell]$, the value $\vec{w}_i \cdot \vec{X} \geq \vec{w}_i \cdot \vec{x}^*/(1.5\lambda) - K_4 C_2 mp$, since $\max_{j\in[n]} w_i(j) \leq 1$. This can be ensured to be at least $\vec{w}_i \cdot \vec{x}^*/(2\lambda)$ by taking K_0 and K_1 large enough, using the definition of p from Lemma 2.2, and using the fact that $\vec{w}_i \cdot \vec{x}^* \geq y^*/2$ for all i. Thus we have proved Theorem 2.3. □

The primary reason why the above works is that the tail bound on $\sum_{i\in[m]} U_i$ works in place of approaches such as that of [29] which try to handle the very low (in some cases exponentially small) probability of satisfying *all* the m constraints $Ax \leq b$. We note that one can use the FKG inequality (see Theorem 3.4)—instead of a union bound—in (2.8) to get somewhat better bounds on ℓ than the bound $\exp(C_1 y^*/\lambda)$.

3 Approximation algorithm for k-CS-PIP

Problem definition and notation In k-CS-PIP, we are given a PIP in which each item j participates in at most k constraints; the goal is to find a maximum-weight subset of items that satisfies all the constraints. For each $j \in [n]$, recall that $N(j) := \{i \in [m] \mid a_{ij} > 0\}$ is the set of constraints that j participates in; since the PIP is k-column sparse PIP, we have $|N(j)| \leq k$ for each $j \in [n]$. For notational convenience, throughout Sections 3 and 4, we assume (without loss of generality, by scaling) that the *right-hand side of each constraint in a PIP is one*. As before, we continue to have each $a_{ij} \leq 1$. In Sections 3 and 4, we measure performance of our algorithms in terms of the column sparsity k.

Before presenting our algorithm, we describe a (seemingly correct) alteration step that does not quite work. Understanding why this easier algorithm fails gives useful insight into the design of the correct alteration step. Then we present a simple analysis of an $8k$-approximation algorithm (Section 3.1), and obtain the improved result (Theorem 1.2) in Section 3.2.

A strawman algorithm Consider the following algorithm. Let x^* be some feasible solution to the natural LP relaxation of k-CS-PIP (i. e., dropping integrality). For each element $j \in [n]$, select it independently at random with probability $x_j^*/(2k)$. Let $\mathcal{S}$ be the chosen set of items. For any constraint $i \in [m]$, if it is violated, then discard all items in $\mathcal{S} \cap P(i)$ (i. e., items $j \in \mathcal{S}$ for which $a_{ij} > 0$).

Since the probabilities are scaled down by $2k$, by Markov's inequality any one constraint i is violated with probability at most $1/(2k)$. Hence, any single constraint will discard its items with probability at most $1/2k$. By the k-sparse property, each element can be discarded by at most k constraints, and hence by union bound over those k constraints, it is discarded with probability at most $k \cdot (1/2k) = 1/2$. Since an element is chosen in $\mathcal{S}$ with probability $x_j^*/2k$, this appears to imply that it lies in the overall solution with probability at least $x_j^*/(4k)$, implying that the proposed algorithm is a $4k$ approximation.

However, the above argument is not correct. Consider the following example. Suppose there is a single constraint (and so $k = 1$),

$$x_1 + \frac{1}{M}x_2 + \frac{1}{M}x_3 + \frac{1}{M}x_4 + \cdots + \frac{1}{M}x_M \leq 1$$

where $M \gg 1$ is a large integer. Clearly, setting $x_j^* = 1/2$ for $i = 1, \ldots, M$ is a feasible solution. Now consider the execution of the strawman algorithm. Note that whenever item 1 is chosen in $\mathcal{S}$, it is very likely that some item other than 1 will also be chosen (since $M \gg 1$ and we pick each item independently with probability $x_j^*/2k = 1/4$); in this case, item 1 would be discarded. Thus the final solution will *almost always not contain* item 1, violating the claim that it lies in the final solution with probability at least $x_1^*/4k = 1/8$.

The key point is that we must consider the probability of an item being discarded by some constraint, *conditional* on it being chosen in the set $\mathcal{S}$. Thus, the alteration step described above is not good because for item 1 in the above example, the probability it is discarded is close to one, not at most half. This is not a problem if either all item sizes are small (i. e., say $a_{ij} \leq 1/2$), or all item sizes are large (say $a_{ij} \approx 1$). The algorithm we analyze shows that the difficult case is indeed when some constraints contain both large and small items, as in the example above.

3.1 A simple algorithm for k-CS-PIP

In this subsection, we use the natural LP relaxation for k-CS-PIP (i. e., dropping the integrality condition) and obtain an $8k$-approximation algorithm. An item $j \in [n]$ is called *big* for constraint $i \in [m]$ if $a_{ij} > 1/2$; and j is *small* for constraint i if $0 < a_{ij} \leq 1/2$. The algorithm first solves the LP relaxation to obtain an optimal fractional solution x^*. Then we round to an integral solution as follows. With foresight, set $\alpha = 4$.

1. Sample each item $j \in [n]$ independently with probability $x_j^*/(\alpha k)$. Let $\mathcal{S}$ denote the set of chosen items. We call an item in $\mathcal{S}$ an $\mathcal{S}$-item.

2. For each item j, mark j (for deletion) if, for any constraint $i \in N(j)$, either:

   - $\mathcal{S}$ contains some *other* item $j' \in [n] \setminus \{j\}$ which is big for constraint i or
   - the sum of sizes of $\mathcal{S}$-items that are small for i exceeds 1 (i. e., the capacity).

3. Delete all marked items, and return $\mathcal{S}'$, the set of remaining items.

Analysis We will show that this algorithm gives an $8k$-approximation.

Lemma 3.1. *Solution $\mathcal{S}'$ is feasible with probability one.*

Proof. Consider any fixed constraint $i \in [m]$.

1. Suppose there is some $j' \in \mathcal{S}'$ that is big for i. Then the algorithm guarantees that j' will be the only item in $\mathcal{S}'$ (either small or big) that participates in constraint i: consider any other $\mathcal{S}$-item j participating in i; j must have been deleted from $\mathcal{S}$ because $\mathcal{S}$ contains another item (namely j') that is big for constraint i. Thus, j' is the only item in $\mathcal{S}'$ participating in constraint i, and so the constraint is trivially satisfied, as all sizes ≤ 1.

2. The other case is when all items in $\mathcal{S}'$ are small for i. Let $j \in \mathcal{S}'$ be some item that is small for i (if there are none such, then constraint i is trivially satisfied). Since j was not deleted from $\mathcal{S}$, it must be that the total size of $\mathcal{S}$-items that are small for i did not exceed 1. Now, $\mathcal{S}' \subseteq \mathcal{S}$, and so this condition is also true for items in $\mathcal{S}'$.

Thus every constraint is satisfied by solution $\mathcal{S}'$ and we obtain the lemma. □

We now prove the main lemma.

Lemma 3.2. *For any item $j \in [n]$, the probability* $\Pr[j \in \mathcal{S}' \mid j \in \mathcal{S}] \geq 1 - 2/\alpha$. *Equivalently, the probability that item j is deleted from $\mathcal{S}$ conditioned on it being chosen in $\mathcal{S}$ is at most $2/\alpha$.*

Proof. For any item j and constraint $i \in N(j)$, let $\mathcal{B}_{ij}$ denote the event that j is marked for deletion from $\mathcal{S}$ because there is some other $\mathcal{S}$-item that is big for constraint i. Let $\mathcal{G}_i$ denote the event that the total size of $\mathcal{S}$-items that are small for constraint i exceeds 1. For any item $j \in [n]$ and constraint $i \in N(j)$, we will show that:

$$\Pr[\mathcal{B}_{ij} \mid j \in \mathcal{S}] + \Pr[\mathcal{G}_i \mid j \in \mathcal{S}] \leq \frac{2}{\alpha k}. \tag{3.1}$$

We prove (3.1) using the following intuition: the total extent to which the LP selects items that are *big* for any constraint cannot be more than 2 (each big item has size at least $1/2$); therefore, $\mathcal{B}_{ij}$ is unlikely to occur since we scaled down probabilities by factor αk. Ignoring for a moment the conditioning on $j \in \mathcal{S}$, event $\mathcal{G}_i$ is also unlikely, by Markov's inequality. But items are selected for $\mathcal{S}$ independently, so if j is big for constraint i, then its presence in $\mathcal{S}$ does not affect the event $\mathcal{G}_i$ at all. If j is small for constraint i, then *even if* $j \in \mathcal{S}$, the total size of $\mathcal{S}$-items is unlikely to exceed 1.

We now prove (3.1) formally, using some care to save a factor of 2. Let $B(i)$ denote the set of items that are big for constraint i, and $Y_i := \sum_{\ell \in B(i)} x_\ell^*$. Recall that constraints are scaled so that $b_i = 1$; so the LP constraint $\sum_j a_{ij} \cdot x_j^* \leq 1$ implies that $Y_i \leq 2$ (since each $\ell \in B(i)$ has size $a_{i\ell} > 1/2$). Now by a union bound,

$$\Pr[\mathcal{B}_{ij} \mid j \in \mathcal{S}] \leq \frac{1}{\alpha k} \sum_{\ell \in B(i) \setminus \{j\}} x_\ell^* \leq \frac{Y_i}{\alpha k}. \tag{3.2}$$

Now, let $S(i)$ denote the set of items that are small for constraint i. Notice that if j is big for i, then event $\mathcal{G}_i$ occurs only if the total i-size of $\mathcal{S}$-items in $S(i) = S(i) \setminus j$ exceeds 1. On the other hand, if j is small for i, then event $\mathcal{G}_i$ occurs only if the total i-size of $\mathcal{S}$-items in $S(i)$ exceeds 1; since $a_{ij} \le 1/2$, this means the total i-size of $\mathcal{S}$-items in $S(i) \setminus j$ must exceed $1/2$. Thus, whether j is big or small,

$$\Pr[\mathcal{G}_i \mid j \in \mathcal{S}] \le \Pr\left[i\text{-size of items } \mathcal{S} \cap (S(i) \setminus j) > 1/2 \;\middle|\; j \in \mathcal{S}\right].$$

We now bound the right-hand side in the above inequality. Using the LP constraint i,

$$\sum_{\ell \in S(i) \setminus j} a_{i\ell} \cdot x_\ell^* \le 1 - \sum_{\ell \in B(i)} a_{i\ell} \cdot x_\ell^* \le 1 - \frac{Y_i}{2}. \tag{3.3}$$

Since each item j' is chosen into $\mathcal{S}$ with probability $x_{j'}^*/(\alpha k)$, inequality (3.3) implies that the expected i-size of $\mathcal{S} \cap (S(i) \setminus j)$ is at most

$$\frac{1}{\alpha k}\left(1 - \frac{Y_i}{2}\right).$$

By Markov's inequality, the probability that the i-size of these items exceeds $1/2$ is at most

$$\frac{2}{\alpha k}\left(1 - \frac{Y_i}{2}\right).$$

Since items are chosen independently and $j \notin S(i) \setminus j$,

$$\Pr\left[i\text{-size of items } \mathcal{S} \cap (S(i) \setminus j) > 1/2 \;\middle|\; j \in \mathcal{S}\right] \le \frac{2}{\alpha k}\left(1 - \frac{Y_i}{2}\right).$$

Thus,

$$\Pr[\mathcal{G}_i \mid j \in \mathcal{S}] \le \frac{2}{\alpha k}\left(1 - \frac{Y_i}{2}\right) = \frac{2}{\alpha k} - \frac{Y_i}{\alpha k}.$$

Combined with inequality (3.2) we obtain (3.1):

$$\Pr[\mathcal{B}_{ij} \mid j \in \mathcal{S}] + \Pr[\mathcal{G}_i \mid j \in \mathcal{S}] \le \frac{Y_i}{\alpha k} + \Pr[\mathcal{G}_i \mid j \in \mathcal{S}] \le \frac{Y_i}{\alpha k} + \frac{2}{\alpha k} - \frac{Y_i}{\alpha k} = \frac{2}{\alpha k}.$$

To see that (3.1) implies the lemma, for any item j, simply take the union bound over all $i \in N(j)$. Thus, the probability that j is deleted from $\mathcal{S}$ conditional on it being chosen in $\mathcal{S}$ is at most $2/\alpha$. Equivalently, $\Pr[j \in \mathcal{S}' \mid j \in \mathcal{S}] \ge 1 - 2/\alpha$. □

We are now ready to present the $8k$-approximation:

Theorem 3.3. *There is a randomized $8k$-approximation algorithm for k-CS-PIP.*

Proof. First observe that our algorithm always outputs a feasible solution (Lemma 3.1). To bound the objective value, recall that $\Pr[j \in \mathcal{S}] = x_j^*/(\alpha k)$ for all $j \in [n]$. Hence Lemma 3.2 implies that

$$\Pr[j \in \mathcal{S}'] = \Pr[j \in \mathcal{S}] \cdot \Pr[j \in \mathcal{S}' \mid j \in \mathcal{S}] \ge \frac{x_j^*}{\alpha k} \cdot \left(1 - \frac{2}{\alpha}\right)$$

for all $j \in [n]$. Finally using linearity of expectation and $\alpha = 4$, we obtain the theorem. □

Remark: The analysis above only uses Markov's inequality conditioned on a single item being chosen in set $\mathcal{S}$. Thus a pairwise independent distribution suffices to choose the set $\mathcal{S}$, and hence the algorithm can be easily derandomized. Furthermore, it is easy to see that this algorithm can be implemented in RNC.

3.2 A stronger LP, and improved approximation

We now present our strengthened LP and the $(\mathrm{e}k+o(k))$-approximation algorithm for k-CS-PIP. We will make use of the FKG inequality in the following form; see [1] for a proof.

Theorem 3.4. *Given a set $[N]=\{1,2,\ldots,N\}$ and $p_1,\ldots,p_N\in[0,1]$, let $\mathcal{I}\subseteq[N]$ be obtained by choosing each $j\in[N]$ into $\mathcal{I}$ independently with probability p_j. Let $g_1,g_2,\ldots,g_\ell : 2^{[N]}\to\mathbb{R}_+$ be decreasing set functions. Then:*

$$\mathbb{E}\left[\Pi_{i=1}^{\ell}\, g_i(\mathcal{I})\right] \;\geq\; \Pi_{i=1}^{\ell}\,\mathbb{E}[g_i(\mathcal{I})]\,.$$

Recall that a set function g is *decreasing* if for any subsets $A\subseteq B$, we have $g(A)\geq g(B)$. A useful special case of this theorem is when the set functions are $0-1$ valued: in this case they correspond to indicator random variables and the expectations above are just probabilities.

Stronger LP relaxation An item j is called *big* for constraint i if $a_{ij}>1/2$. For each constraint $i\in[m]$, let $B(i)=\{j\in[n]\mid a_{ij}>1/2\}$ denote the set of big items. Since no two items that are big for some constraint can be chosen in an integral solution, the inequality $\sum_{j\in B(i)} x_j\leq 1$ is valid for each $i\in[m]$. The strengthened LP relaxation that we consider is as follows.

$$\max \quad \sum_{j=1}^{n} w_j x_j \tag{3.4}$$

$$\text{such that} \quad \sum_{j=1}^{n} a_{ij}\cdot x_j \leq 1 \qquad \forall i\in[m]\,, \tag{3.5}$$

$$\sum_{j\in B(i)} x_j \leq 1 \qquad \forall i\in[m]\,, \tag{3.6}$$

$$0\leq x_j\leq 1 \qquad \forall j\in[n]\,. \tag{3.7}$$

Algorithm The algorithm obtains an optimal solution x^* to the LP relaxation (3.4)-(3.7), and rounds it to an integral solution $\mathcal{S}'$ as follows (parameter α will be set to 1 later).

1. Pick each item $j\in[n]$ independently with probability $x_j^*/(\alpha k)$. Let $\mathcal{S}$ denote the set of chosen items.

2. For any item j and constraint $i\in N(j)$, let E_{ij} denote the event that the items $\{j'\in\mathcal{S}\mid a_{ij'}\geq a_{ij}\}$ have total size (in constraint i) exceeding one. Mark j for deletion if E_{ij} occurs for any $i\in N(j)$.

3. Return set $\mathcal{S}'\subseteq\mathcal{S}$ consisting of all items $j\in\mathcal{S}$ not marked for deletion.

Note the rule for deleting an item from $\mathcal{S}$. In particular, whether item j is deleted from constraint i only depends on items that are at least as large as j in constraint i. Observe that this alteration step considers items in reverse order to the algorithm for general PIPs in Section 2.

Analysis It is clear that $\mathcal{S}'$ is feasible with probability one. The main lemma is the following, where we show that each item appears in $\mathcal{S}'$ with good probability.

Lemma 3.5. *For every item $j \in [n]$ and constraint $i \in N(j)$, we have*

$$\Pr[E_{ij} \mid j \in \mathcal{S}] \;\le\; \frac{1}{\alpha k}\left(1+\left(\frac{2}{\alpha k}\right)^{1/3}\right).$$

Proof. Let $\ell := (4\alpha k)^{1/3}$. We classify items in relation to constraints as:

- Item $j \in [n]$ is *big* for constraint $i \in [m]$ if $a_{ij} > \frac{1}{2}$.
- Item $j \in [n]$ is *medium* for constraint $i \in [m]$ if $\frac{1}{\ell} \le a_{ij} \le \frac{1}{2}$.
- Item $j \in [n]$ is *tiny* for constraint $i \in [m]$ if $a_{ij} < \frac{1}{\ell}$.

For any constraint $i \in [m]$, let $B(i), M(i), T(i)$ respectively denote the set of big, medium, tiny items for i. In the next three claims, we bound $\Pr[E_{ij} \mid i \in \mathcal{S}]$ when item j is big, medium, and tiny respectively.

Claim 3.6. *For any $j \in [n]$ and $i \in [m]$ such that item j is big for constraint i,*

$$\Pr[E_{ij} \mid j \in \mathcal{S}] \;\le\; \frac{1}{\alpha k}.$$

Proof. The event E_{ij} occurs if some item that is at least as large as j for constraint i is chosen in $\mathcal{S}$. Since j is big in constraint i, E_{ij} occurs only if some big item other than i is chosen for S. Now by the union bound, the probability that some item from $B(i)\setminus\{j\}$ is chosen into $\mathcal{S}$ is:

$$\Pr\big[(B(i)\setminus\{j\})\cap\mathcal{S}\neq\emptyset \;\big|\; j\in\mathcal{S}\big] \;\le\; \sum_{j'\in B(i)\setminus\{j\}} \frac{x^*_{j'}}{\alpha k} \;\le\; \frac{1}{\alpha k}\sum_{j'\in B(i)} x^*_{j'} \;\le\; \frac{1}{\alpha k},$$

where the last inequality follows from the new LP constraint (3.6) on big items for j. □

Claim 3.7. *For any $j \in [n]$, $i \in [m]$ such that item j is medium for constraint i,*

$$\Pr[E_{ij} \mid j \in \mathcal{S}] \;\le\; \frac{1}{\alpha k}\left(1+\frac{\ell^2}{2\alpha k}\right).$$

Proof. Here, if event E_{ij} occurs then it must be that either some big item is chosen or (otherwise) at least two medium items other than j are chosen, i. e., E_{ij} implies that either

$$\mathcal{S}\cap B(i)\neq\emptyset \quad\text{or}\quad \big|\mathcal{S}\cap(M(i)\setminus\{j\})\big|\ge 2.$$

This is because j together with any *one* other medium item is not enough to reach the capacity of constraint i. (Since j is medium, we do not consider tiny items for constraint i in determining whether j should be deleted.)

Just as in Claim 3.6, we have that the probability some big item for i is chosen is at most $1/(\alpha k)$, i. e.,

$$\Pr\left[\mathcal{S} \cap B(i) \neq \emptyset \mid j \in \mathcal{S}\right] \leq \frac{1}{\alpha k}.$$

Now consider the probability that $|\mathcal{S} \cap (M(i) \setminus \{j\})| \geq 2$, conditioned on $j \in \mathcal{S}$. We will show that this probability is much smaller than $1/\alpha k$. Since each item $h \in M(i) \setminus \{j\}$ is chosen independently with probability $x_h^*/(\alpha k)$ (even given $j \in \mathcal{S}$):

$$\Pr\left[\left|\mathcal{S} \cap (M(i) \setminus \{j\})\right| \geq 2 \;\middle|\; j \in \mathcal{S}\right] \leq \frac{1}{2} \cdot \left(\sum_{h \in M(i)} \frac{x_h^*}{\alpha k}\right)^2 \leq \frac{\ell^2}{2\alpha^2 k^2}$$

where the last inequality follows from the fact that

$$1 \geq \sum_{h \in M(i)} a_{ih} \cdot x_h^* \geq \frac{1}{\ell} \sum_{h \in M(i)} x_h^*$$

(recall each item in $M(i)$ has size at least $1/\ell$). Combining these two cases, we have the desired upper bound on $\Pr[E_{ij} \mid j \in \mathcal{S}]$. □

Claim 3.8. *For any $j \in [n]$, $i \in [m]$ such that item j is tiny for constraint i,*

$$\Pr[E_{ij} \mid j \in \mathcal{S}] \leq \frac{1}{\alpha k}\left(1 + \frac{2}{\ell}\right).$$

Proof. Since j is tiny, if event E_{ij} occurs then the total size (in constraint i) of items $\mathcal{S} \setminus \{j\}$ is greater than $1 - 1/\ell$. So,

$$\Pr[E_{ij} \mid j \in \mathcal{S}] \leq \Pr\left[\sum_{h \in \mathcal{S} \setminus \{j\}} a_{ih} > 1 - \frac{1}{\ell}\right] \leq \frac{1}{\alpha k} \cdot \frac{\ell}{\ell - 1} \leq \frac{1}{\alpha k}\left(1 + \frac{2}{\ell}\right)$$

where the first inequality follows from the above observation and the fact that $\mathcal{S} \setminus \{j\}$ is independent of the event $j \in \mathcal{S}$, the second is Markov's inequality, and the last uses $\ell \geq 2$. □

Thus, for any item j and constraint $i \in N(j)$,

$$\Pr[E_{ij} \mid j \in \mathcal{S}] \leq \frac{1}{\alpha k} \max\left\{\left(1 + \frac{2}{\ell}\right), \left(1 + \frac{\ell^2}{2\alpha k}\right)\right\}.$$

From the choice of $\ell = (4\alpha k)^{1/3}$, which makes the probability in Claims 3.7 and 3.8 equal, we obtain the lemma. □

We now prove the main result of this section

Lemma 3.9. *For each $j \in [n]$,*

$$\Pr[j \in \mathcal{S}' \mid j \in \mathcal{S}] \geq \left(1 - \frac{1}{\alpha k}\left(1 + \left(\frac{2}{\alpha k}\right)^{1/3}\right)\right)^k.$$

Proof. For any item j and constraint $i \in N(j)$, the conditional event $(\neg E_{ij} \mid j \in \mathcal{S})$ is a *decreasing* function over the choice of items in set $[n] \setminus \{j\}$. Thus, by the FKG inequality (Theorem 3.4), for any fixed item $j \in [n]$, the probability that no event $(E_{ij} \mid j \in \mathcal{S})$ occurs is:

$$\Pr\left[\bigwedge_{i \in N(j)} \neg E_{ij} \;\middle|\; j \in \mathcal{S}\right] \geq \prod_{i \in N(j)} \Pr[\neg E_{ij} \mid j \in \mathcal{S}].$$

From Lemma 3.5,

$$\Pr[\neg E_{ij} \mid j \in \mathcal{S}] \geq 1 - \frac{1}{\alpha k}\left(1 + \left(\frac{2}{\alpha k}\right)^{1/3}\right).$$

As each item is in at most k constraints, we obtain the theorem. □

Now, by setting $\alpha = 1$,[3] we have

$$\Pr[j \in \mathcal{S}] = 1/k \quad \text{and} \quad \Pr[j \in \mathcal{S}' \mid j \in \mathcal{S}] \geq \frac{1}{\mathrm{e} + o(1)},$$

which immediately implies Theorem 1.2. □

Remark This algorithm can be derandomized using conditional expectation and pessimistic estimators, since we can exactly compute estimates of the relevant probabilities. Also, the algorithm can be easily implemented in RNC, as in Section 2.

Integrality gap of LP (3.4)-(3.7) Recall that the LP relaxation for the k-set packing problem has an integrality gap of $k - 1 + 1/k$, as shown by the instance given by the projective plane of order $k - 1$, where $k - 1$ is a prime power. If we have the same size-matrix and set each capacity to $2 - \varepsilon$, this directly implies an integrality gap arbitrarily close to $2(k - 1 + 1/k)$ for the (weak) LP relaxation for k-CS-PIP. This is because the LP can set each $x_j = (2 - \varepsilon)/k$ hence obtaining a profit of $(2 - \varepsilon)(k - 1 + 1/k)$, while the integral solution can only choose one item. However, for our stronger LP relaxation (3.4)-(3.7) used in this section, this example does not work and the projective plane instance only implies a gap of $k - 1 + 1/k$ (note that here each item is big in every constraint that it appears in).

However, using a different instance of k-CS-PIP, we show that even the stronger LP relaxation has an integrality gap at least $2k - 1$. Consider the instance on $n = m = 2k - 1$ items and constraints defined as follows. We view the indices $[n] = \{0, 1, \cdots, n - 1\}$ as integers modulo n. The weights $w_j = 1$ for all $j \in [n]$. The sizes are:

$$\forall i, j \in [n], \qquad a_{ij} := \begin{cases} 1 & \text{if } i = j, \\ \varepsilon & \text{if } j \in \{i + 1 \bmod n, \ldots, i + k - 1 \bmod n\}, \\ 0 & \text{otherwise,} \end{cases}$$

where $\varepsilon > 0$ is arbitrarily small, in particular $\varepsilon \ll 1/(nk)$.

[3] Note that this is optimal only asymptotically. In the case of $k = 2$, for instance, it is better to choose $\alpha \approx 2.8$ which results in an approximation factor ≈ 11.6; a better bound is known [27] for this special case.

Observe that setting $x_j = 1 - k\varepsilon$ for all $j \in [n]$ is a feasible fractional solution to the strengthened LP (3.4)-(3.7); each constraint has only one big item and so the new constraint (3.6) is satisfied. Thus the optimal LP value is at least $(1-k\varepsilon)\cdot n \approx n = 2k-1$.

On the other hand, we claim that the optimal integral solution can only choose one item and hence has value 1. For the sake of contradiction, suppose that it chooses two items $j,h \in [n]$. Then there is some constraint i (either $i = j$ or $i = h$) that implies either $x_j + \varepsilon \cdot x_h \le 1$ or $x_h + \varepsilon \cdot x_j \le 1$; in either case constraint i would be violated.

Thus the integrality gap of the LP we consider is at least $2k-1$, for every $k \ge 1$.

Bad example for possible generalization A natural extension of the k-CS-PIP result is to consider PIPs where the ℓ_1-norm of each column is upper-bounded by k (when capacities are all-ones). We observe that unlike k-CS-PIP, the LP relaxation for this generalization has an $\Omega(n)$ integrality gap. The example has $m = n$; sizes $a_{ii} = 1$ for all $j \in [n]$, and $a_{ij} = 1/n$ for all $j \ne i$; and all weights one. The ℓ_1-norm of each column is at most 2. Clearly, the optimal integral solution has value one. On the other hand, picking each item to the extent of $1/2$ is a feasible LP solution of value $n/2$.

This integrality gap is in sharp contrast to the results on *discrepancy of sparse matrices*, where the classic Beck-Fiala bound of $O(k)$ applies also to matrices with entries in $[-1,1]$, just as well as $\{-1,0,1\}$ entries; here k denotes an upper-bound on the ℓ_1-norm of the columns.

4 Submodular objective functions

We now consider the more general case where the objective we seek to maximize is an arbitrary non-negative *monotone submodular function*. A function $f : 2^{[n]} \to \mathbb{R}_+$ on groundset $[n] = \{1,\ldots,n\}$ is called submodular if for every $A,B \subseteq [n]$, $f(A\cup B) + f(A\cap B) \le f(A) + f(B)$. Function f is monotone if for each $A \subseteq B \subseteq [n]$, $f(A) \le f(B)$. The problem we consider is:

$$\max\left\{ f(T) \;\middle|\; \sum_{j\in T} a_{ij} \le b_i, \forall i \in [m]; T \subseteq [n] \right\}. \tag{4.1}$$

As is standard when dealing with submodular functions, we only assume *value-oracle* access to the function (i. e., the algorithm can query any subset $T \subseteq [n]$, and it obtains the function value $f(T)$ in constant time). Again, we let k denote the column-sparseness of the underlying constraint matrix. In this section we prove Theorem 1.3, an $O(k)$-approximation algorithm for Problem (4.1). The algorithm is similar to that for k-CS-PIP (where the objective was additive), and involves the following two steps.

1. Solve (approximately) a suitable continuous relaxation of (4.1). This step follows directly from the algorithm of [31, 7].

2. Perform the randomized rounding with alteration described in Section 3. Although the algorithm is the same as for additive functions, the analysis requires considerably more work. In the process, we also establish a new property of submodular functions that generalizes *fractional subadditivity* [11].

The non-negativity and monotonicity properties of the submodular function are used in both these steps.

Solving the continuous relaxation The *extension-by-expectation* (also called the *multi-linear* extension) of a submodular function f is a continuous function $F : [0,1]^n \to \mathbb{R}_+$ defined as follows:

$$F(x) := \sum_{T \subseteq [n]} \left(\Pi_{j \in T}\, x_j \cdot \Pi_{j \notin T}\, (1 - x_j) \right) \cdot f(T) \,.$$

Note that $F(x) = f(x)$ for $x \in \{0,1\}^n$ and hence F is an extension of f. Even though F is a non-linear function, using the continuous greedy algorithm from [7], we can obtain a $(1 - 1/\mathrm{e})$-approximation algorithm to the following *fractional relaxation* of (4.1).

$$\max \left\{ F(x) \;\middle|\; \sum_{j=1}^{n} a_{ij} \cdot x_j \le b_i, \forall i \in [m];\; 0 \le x_j \le 1, \forall j \in [n] \right\} . \tag{4.2}$$

In order to apply the algorithm from [7], one needs to solve in polynomial time the problem of maximizing a *linear* objective over the constraints

$$\left\{ \sum_{j=1}^{n} a_{ij} \cdot x_j \le b_i,\ \forall i \in [m];\ 0 \le x_j \le 1,\ \forall j \in [n] \right\} .$$

This is indeed possible since it is a linear program on n variables and m constraints.

The rounding algorithm The rounding algorithm is identical to that for k-CS-PIP. Let x denote any feasible solution to Problem (4.2). We apply the rounding algorithm for the additive case (from the previous section), to first obtain a (possibly infeasible) solution $\mathcal{S} \subseteq [n]$ and then a *feasible integral solution* $\mathcal{S}' \subseteq [n]$. In the rest of this section, we prove the performance guarantee of this algorithm.

Fractional Subadditivity The following is a useful lemma (see Feige [11]) showing that submodular functions are also fractionally subadditive.

Lemma 4.1 ([11]). *Let $\mathcal{U}$ be a set of elements and $\{\mathcal{A}_t \subseteq \mathcal{U}\}$ be a collection of subsets with non-negative weights $\{\lambda_t\}$ such that $\sum_{t \mid j \in \mathcal{A}_t} \lambda_t = 1$ for all elements $j \in \mathcal{U}$. Then, for any submodular function $f : 2^{\mathcal{U}} \to \mathbb{R}_+$, we have $f(\mathcal{U}) \le \sum_t \lambda_t f(\mathcal{A}_t)$.*

The above result can be used to show that (the infeasible solution) $\mathcal{S}$ has good profit in expectation.

Lemma 4.2. *For any $x \in [0,1]^n$ and $0 \le p \le 1$, let set $\mathcal{S}$ be constructed by selecting each item $j \in [n]$ independently with probability $p \cdot x_j$. Then, $\mathbb{E}[f(\mathcal{S})] \ge pF(x)$. In particular, this implies that our rounding algorithm that forms set $\mathcal{S}$ by independently selecting each element $j \in [n]$ with probability $x_j/(\alpha k)$ satisfies $\mathbb{E}[f(\mathcal{S})] \ge F(x)/(\alpha k)$.*

Proof. Consider the following equivalent procedure for constructing $\mathcal{S}$: first, construct $\mathcal{S}_0$ by selecting each item j with probability x_j. Then construct $\mathcal{S}$ by retaining each element in $\mathcal{S}_0$ independently with probability p.

By definition $\mathbb{E}[f(\mathcal{S}_0)] = F(x)$. For any fixed set $T \subseteq [n]$, consider the outcomes for set $\mathcal{S}$ *conditioned* on $\mathcal{S}_0 = T$; the set $\mathcal{S} \subseteq \mathcal{S}_0$ is a random subset such that $\Pr[j \in \mathcal{S} \mid \mathcal{S}_0 = T] = p$ for all $i \in T$. Thus by Lemma 4.1, we have $\mathbb{E}[f(\mathcal{S}) \mid \mathcal{S}_0 = T] \geq p \cdot f(T)$. Hence:

$$\mathbb{E}[f(\mathcal{S})] \;=\; \sum_{T\subseteq[n]} \Pr[\mathcal{S}_0 = T]\cdot\mathbb{E}[f(\mathcal{S}) \mid \mathcal{S}_0 = T] \;\geq\; \sum_{T\subseteq[n]} \Pr[\mathcal{S}_0 = T]\cdot p\, f(T) \;=\; p\,\mathbb{E}[f(\mathcal{S}_0)] \;=\; p\cdot F(x)\,.$$

Thus we obtain the lemma. □

However, we cannot complete the analysis using the approach in Lemma 3.9. The problem is that even though $\mathcal{S}$ (which is chosen by random sampling) has good expected profit, $\mathbb{E}[f(\mathcal{S})] = \Omega(1/k)F(x)$ (from Lemma 4.2), it may happen that the alteration step used to obtain $\mathcal{S}'$ from $\mathcal{S}$ throws away essentially all the profit. This was not an issue for linear objective functions since our alteration procedure guarantees that $\Pr[j \in \mathcal{S}' | j \in \mathcal{S}] = \Omega(1)$ for each $j \in [n]$, and if f is linear, this implies $\mathbb{E}[f(\mathcal{S}')] = \Omega(1)\,\mathbb{E}[f(\mathcal{S})]$. However, this property is not enough for general submodular functions. Consider the following:

Example Let set $\mathcal{S} \subseteq [n]$ be drawn from the following distribution:

- With probability $1/2n$, $\mathcal{S} = [n]$.
- For each $j \in [n]$, $\mathcal{S} = \{j\}$ with probability $1/2n$.
- With probability $1/2 - 1/2n$, $\mathcal{S} = \emptyset$.

Now define $\mathcal{S}' = \mathcal{S}$ if $\mathcal{S} = [n]$, and $\mathcal{S}' = \emptyset$ otherwise. Note that for each $j \in [n]$, we have

$$\Pr[j \in \mathcal{S}' \mid j \in \mathcal{S}] \;=\; 1/2 \;=\; \Omega(1)\,.$$

However, consider the profit with respect to the "coverage" submodular function f, where $f(T) = 1$ if $T \neq \emptyset$ and is 0 otherwise. We have $\mathbb{E}[f(\mathcal{S})] = 1/2 + 1/2n$, but $\mathbb{E}[f(\mathcal{S}')]$ is only $1/2n \ll \mathbb{E}[f(\mathcal{S})]$.

Remark Note that if $\mathcal{S}'$ itself was chosen randomly from $\mathcal{S}$ such that $\Pr[j \in \mathcal{S}' \mid \mathcal{S} = T] = \Omega(1)$ for *every* $T \subseteq [n]$ *and* $j \in T$, then we would be done by Lemma 4.1. Unfortunately, this is too much to hope for. In our rounding procedure, for any particular choice of $\mathcal{S}$, set $\mathcal{S}'$ is a fixed subset of $\mathcal{S}$; and there could be (bad) sets $\mathcal{S}$, where after the alteration step we end up with sets $\mathcal{S}'$ such that $|\mathcal{S}'| \ll |\mathcal{S}|$.

However, it turns out that we can use the following two additional properties beyond just marginal probabilities to argue that $\mathcal{S}'$ has reasonable profit. First, the sets $\mathcal{S}$ constructed by our algorithm are drawn from a product distribution on the items; in contrast, the example above does not have this property. Second, our alteration procedure has the following "monotonicity" property: suppose $T_1 \subseteq T_2 \subseteq [n]$, and $j \in \mathcal{S}'$ when $\mathcal{S} = T_2$. Then we are guaranteed that $j \in \mathcal{S}'$ when $\mathcal{S} = T_1$. That is, if $\mathcal{S}$ contains additional items, it is more likely that j will be discarded by the alteration. The above example does not satisfy this property either. That these properties suffice is proved in Corollary 4.4. Roughly speaking, the intuition is that, since f is submodular, the marginal contribution of item j to $\mathcal{S}$ is largest when $\mathcal{S}$ is "small," and this is also the case when j is most likely to be retained for $\mathcal{S}'$. That is, for every $j \in [n]$, both $\Pr[j \in \mathcal{S}' \mid j \in \mathcal{S}]$ and the marginal contribution of j to $f(\mathcal{S})$ are *decreasing* functions of $\mathcal{S}$. To show Corollary 4.4 we need the following generalization of Feige's subadditivity lemma.

Theorem 4.3. *Let $[n]$ denote a groundset and $x \in [0,1]^n$. $\mathcal{B} \subseteq [n]$ is a random subset chosen according to the product distribution with probabilities x. $\mathcal{A}$ is a random subset of $\mathcal{B}$ chosen as follows. For each $B \subseteq [n]$ there is a distribution $\mathcal{D}_B$ on subsets of B: conditional on $\mathcal{B} = B$, subset $\mathcal{A}$ is drawn according to $\mathcal{D}_B$. Suppose that the following conditions hold for some $\beta > 0$:*

Marginal Property:

$$\forall j \in [n], \quad \Pr[j \in \mathcal{A} \mid j \in \mathcal{B}] \;\geq\; \beta\,. \tag{4.3}$$

Monotonicity Property:

$$\forall B \subseteq B' \subseteq [n],\, \forall j \in B, \quad \Pr[j \in \mathcal{A} \mid \mathcal{B} = B] \;\geq\; \Pr\left[j \in \mathcal{A} \mid \mathcal{B} = B'\right]\,. \tag{4.4}$$

Then, for any non-negative monotone submodular function f, we have $\mathbb{E}[f(\mathcal{A})] \geq \beta \cdot \mathbb{E}[f(\mathcal{B})]$.

Proof. The proof is by induction on n, the size of the groundset. The base case of $n = 1$ is straightforward. So suppose $n \geq 2$. For any subsets $A \subseteq B \subseteq [n]$ such that $n \in A$, by submodularity we have that $f(A) \geq f(B) - f(B \setminus \{n\}) + f(A \setminus \{n\})$. So,

$$\begin{aligned}
\mathbb{E}[f(\mathcal{A})] &= \mathbb{E}\left[f(\mathcal{A}) \cdot \mathbf{1}_{n\in\mathcal{A}}\right] + \mathbb{E}\left[f(\mathcal{A}) \cdot \mathbf{1}_{n\notin\mathcal{A}}\right] \\
&\geq \mathbb{E}\left[(f(\mathcal{B}) - f(\mathcal{B}\setminus n)) \cdot \mathbf{1}_{n\in\mathcal{A}}\right] + \mathbb{E}\left[f(\mathcal{A}\setminus n) \cdot \mathbf{1}_{n\in\mathcal{A}}\right] + \mathbb{E}\left[f(\mathcal{A}) \cdot \mathbf{1}_{n\notin\mathcal{A}}\right] \\
&= \mathbb{E}\left[(f(\mathcal{B}) - f(\mathcal{B}\setminus n)) \cdot \mathbf{1}_{n\in\mathcal{A}}\right] + \mathbb{E}\left[f(\mathcal{A}\setminus n)\right]\,.
\end{aligned} \tag{4.5}$$

We will show that the first expectation in (4.5) is at least $\beta \cdot \mathbb{E}[f(\mathcal{B}) - f(\mathcal{B}\setminus n)]$ and the second expectation is at least $\beta \cdot \mathbb{E}[f(\mathcal{B}\setminus n)]$, which suffices to prove the theorem.

Bounding $\mathbb{E}\left[(f(\mathcal{B}) - f(\mathcal{B}\setminus n)) \cdot \mathbf{1}_{n\in\mathcal{A}}\right]$ For any set $Y \subseteq [n-1]$, define the following two functions:

$$g(Y) := f(Y \cup \{n\}) - f(Y), \quad \text{and} \quad h(Y) := \Pr[n \in \mathcal{A} \,|\, \mathcal{B} = Y \cup \{n\}]\,.$$

Clearly, both g and h are non-negative. Note that g is a decreasing set function due to submodularity of f. Moreover, function h is also decreasing by the monotonicity property with $j = n$. Now,

$$\begin{aligned}
\mathbb{E}\left[(f(\mathcal{B}) - f(\mathcal{B}\setminus n)) \cdot \mathbf{1}_{n\in\mathcal{A}}\right] &= \sum_{Y\subseteq[n-1]} (f(Y\cup\{n\}) - f(Y)) \cdot \Pr[n \in \mathcal{A} \mid \mathcal{B} = Y \cup \{n\}] \cdot \Pr[\mathcal{B} = Y \cup \{n\}] \\
&= \sum_{Y\subseteq[n-1]} g(Y) \cdot h(Y) \cdot \Pr[\mathcal{B}\setminus n = Y] \cdot x_n \\
&= x_n \cdot \mathbb{E}[g(\mathcal{B}\setminus n) \cdot h(\mathcal{B}\setminus n)]\,.
\end{aligned} \tag{4.6}$$

The first equality uses the fact that $f(\mathcal{B}) - f(\mathcal{B}\setminus n) = 0$ when $n \notin \mathcal{B}$. The second equality uses the definition of g and h, and that $\Pr[n \in \mathcal{B}] = x_n$ independent of $\mathcal{B}\setminus n$.

Let $\mathcal{C} \subseteq [n-1]$ denote a random set with product distribution of marginal probabilities $\{x_j\}_{j=1}^{n-1}$. Notice that,

$$\begin{aligned}
\mathbb{E}[h(\mathcal{C})] &= \sum_{Y\subseteq[n-1]} \Pr[\mathcal{C} = Y] \cdot \Pr[n \in \mathcal{A} \,|\, \mathcal{B} = Y \cup \{n\}] \\
&= \sum_{Y\subseteq[n-1]} \frac{\Pr[\mathcal{B} = Y \cup \{n\}]}{x_n} \cdot \Pr[n \in \mathcal{A} \,|\, \mathcal{B} = Y \cup \{n\}] \;=\; \frac{\Pr[n \in \mathcal{A}]}{x_n} \;\geq\; \beta\,.
\end{aligned} \tag{4.7}$$

The second equality is by the simple coupling $\mathcal{C} = \mathcal{B} \setminus n$, and the inequality uses the marginal property with $j = n$. Moreover,

$$\mathbb{E}[f(\mathcal{B}) - f(\mathcal{B} \setminus n)] \;=\; \sum_{Y \subseteq [n-1]} \Pr[\mathcal{B} = Y \cup \{n\}] \cdot (f(Y \cup \{n\}) - f(Y)) \;=\; x_n \cdot \mathbb{E}[g(\mathcal{C})]\,. \tag{4.8}$$

Applying the FKG inequality (Theorem 3.4) on the *decreasing set functions* g and h, it follows that

$$\mathbb{E}[g(\mathcal{C}) \cdot h(\mathcal{C})] \;\geq\; \mathbb{E}[g(\mathcal{C})] \cdot \mathbb{E}[h(\mathcal{C})] \;\geq\; \beta \cdot \mathbb{E}[g(\mathcal{C})] \;=\; \frac{\beta}{x_n} \cdot \mathbb{E}[f(\mathcal{B}) - f(\mathcal{B} \setminus n)]\,,$$

where the second inequality is by (4.7) and the equality by (4.8). Since $\mathbb{E}[g(\mathcal{C}) \cdot h(\mathcal{C})]$ is the same as $\mathbb{E}[g(\mathcal{B} \setminus n) \cdot h(\mathcal{B} \setminus n)]$, using (4.6) with the above, we obtain:

$$\mathbb{E}\left[(f(\mathcal{B}) - f(\mathcal{B} \setminus n)) \cdot \mathbf{1}_{n \in \mathcal{A}}\right] \;\geq\; \beta \cdot \mathbb{E}[f(\mathcal{B}) - f(\mathcal{B} \setminus n)]\,. \tag{4.9}$$

Bounding $\mathbb{E}\left[f(\mathcal{A} \setminus n)\right]$ Here we use the inductive hypothesis on groundset $[n-1]$. Let $\mathcal{B}'$ denote a random subset of $[n-1]$ having the product distribution with probabilities $\{x_j\}_{j=1}^{n-1}$. Define $\mathcal{A}' \subseteq \mathcal{B}'$ as follows. For each $C \subseteq [n-1]$, the distribution $\mathcal{D}'_C$ of $\mathcal{A}'$ conditional on $\mathcal{B}' = C$ is: with probability $1 - x_n$ it returns random subset $A' \leftarrow \mathcal{D}_C$, and with probability x_n it returns $A'' \setminus \{n\}$ where $A'' \leftarrow \mathcal{D}_{C \cup \{n\}}$. We now verify the two required properties.

From the definition of $\mathcal{A}'$, for any $C \subseteq [n-1]$ and $j \in C$, we have

$$\Pr[j \in \mathcal{A}' \,|\, \mathcal{B}' = C] \;=\; x_n \cdot \Pr[j \in \mathcal{A} \,|\, \mathcal{B} = C \cup \{n\}] + (1 - x_n) \cdot \Pr[j \in \mathcal{A} \,|\, \mathcal{B} = C]\,.$$

The monotonicity property now follows directly from that of the original instance.

To see the marginal property, for any $j \in [n-1]$,

$$\begin{aligned}
\Pr[j \in \mathcal{A}'] \;&=\; \sum_{Y \subseteq [n-1]} \Pr[(j \in \mathcal{A}') \wedge (\mathcal{B}' = Y)] \\
&=\; \sum_{Y \subseteq [n-1]} \Pr[\mathcal{B}' = Y] \cdot \big(x_n \cdot \Pr[j \in \mathcal{A} \,|\, \mathcal{B} = Y \cup \{n\}] + (1 - x_n) \cdot \Pr[j \in \mathcal{A} \,|\, \mathcal{B} = Y]\big) \\
&=\; \sum_{Y \subseteq [n-1]} \big(\Pr[(j \in \mathcal{A}) \wedge (\mathcal{B} = Y \cup \{n\})] + \Pr[(j \in \mathcal{A}) \wedge (\mathcal{B} = Y)]\big) \\
&=\; \Pr[j \in \mathcal{A}] \;\geq\; \beta \cdot x_j\,.
\end{aligned}$$

The second equality uses the definition of distribution $\mathcal{D}'_Y$, the third equality is by coupling $\mathcal{B}' = \mathcal{B} \setminus n$, and the inequality is by the marginal property of the original instance on $[n]$. Since $\Pr[j \in \mathcal{B}'] = x_j$, this proves the marginal property of the new instance.

Thus, using Theorem 1.3 inductively, $\mathbb{E}[f(\mathcal{A}')] \geq \beta \cdot \mathbb{E}[f(\mathcal{B}')]$. Using the coupling $\mathcal{B}' = \mathcal{B} \setminus n$ and $\mathcal{A}' = \mathcal{A} \setminus n$, we have:

$$\mathbb{E}[f(\mathcal{A} \setminus n)] \;\geq\; \beta \cdot \mathbb{E}[f(\mathcal{B} \setminus n)]\,. \tag{4.10}$$

Combining (4.5), (4.9) and (4.10) completes the proof. □

Remark It is easy to see that Theorem 4.3 generalizes Lemma 4.1: let $x_j = 1$ for each $j \in [n]$. The distribution $\mathcal{D}_{[n]} \equiv \langle \mathcal{A}_t, \lambda_t / \sum_\ell \lambda_\ell \rangle_t$ is associated with $B = [n]$. For all other $B' \subsetneq [n]$, the distribution $\mathcal{D}_{B'} \equiv \langle B', 1 \rangle$. The monotonicity condition is trivially satisfied. By the assumption in Lemma 4.1, the Marginal property holds with $\beta = 1/\sum_\ell \lambda_\ell$. Thus Theorem 4.3 applies and yields the conclusion in Lemma 4.1.

Corollary 4.4. *Let $\mathcal{S}$ be a random set drawn from a product distribution on $[n]$. Let $\mathcal{S}'$ be another random set where for each choice of $\mathcal{S}$, set $\mathcal{S}'$ is some subset of $\mathcal{S}$. Suppose that for each $j \in [n]$,*

- $\Pr_{\mathcal{S}}[j \in \mathcal{S}' \mid j \in \mathcal{S}] \geq \beta$, *and*
- *for all $T_1 \subseteq T_2$ with $T_1 \ni j$, if $j \in \mathcal{S}'$ when $\mathcal{S} = T_2$ then $j \in \mathcal{S}'$ when $\mathcal{S} = T_1$.*

Then $\mathbb{E}[f(\mathcal{S}')] \geq \beta \cdot \mathbb{E}[f(\mathcal{S})]$.

Proof. This is immediate from Theorem 4.3; we use $\mathcal{B} = \mathcal{S}$ and simply associate the single set distribution (i. e., $\mathcal{A} = \mathcal{S}'$) for each $\mathcal{S}$. The two conditions stated above on $\mathcal{S}'$ imply the Marginal and Monotonicity properties respectively; and Theorem 4.3 yields $\mathbb{E}[f(\mathcal{S}')] \geq \beta \, \mathbb{E}[f(\mathcal{S})]$. □

We are now ready to prove the performance guarantee of our algorithm. Observe that our rounding algorithm satisfies the hypothesis of Corollary 4.4 with $\beta = 1/(\mathrm{e} + o(1))$, when parameter $\alpha = 1$. Moreover, by Lemma 4.2, it follows that $\mathbb{E}[f(\mathcal{S})] \geq F(x)/(\alpha k)$. Thus,

$$\mathbb{E}[f(\mathcal{S}')] \;\geq\; \frac{1}{\mathrm{e} + o(1)} \, \mathbb{E}[f(\mathcal{S})] \;\geq\; \frac{1}{\mathrm{e}k + o(k)} \cdot F(x) \, .$$

Combined with the fact that x is an $\mathrm{e}/(\mathrm{e}-1)$-approximate solution to the continuous relaxation (4.2), we have proved Theorem 1.3. □

5 k-CS-PIP algorithm for general B

In this section, we obtain substantially better approximation guarantees for k-CS-PIP when the capacities are large relative to the sizes: we will prove Theorem 1.4.

Here, it is convenient to scale the sizes so that for every constraint $i \in [m]$, $\max_{j \in P(i)} a_{ij} = 1$. Recall from the introduction, that the parameter we use to measure the relative "slack" between sizes and capacities is $B = \min_{i \in [m]} b_i \geq 1$. We consider the k-CS-PIP problem as a function of both k and B, and obtain an improved approximation ratio of $O\big((\lfloor B \rfloor k)^{1/\lfloor B \rfloor}\big) = O(k^{1/\lfloor B \rfloor})$; we also give a matching integrality gap (for every k and $B \geq 1$) for the natural LP relaxation. Previously, Pritchard [24] studied k-CS-PIP when $B > k$ and obtained a ratio of $(1 + k/B)/(1 - k/B)$; in contrast, we obtain asymptotically improved approximation ratios even when $B = 2$.

Set $\alpha := 4\mathrm{e} \cdot (\lfloor B \rfloor k)^{1/\lfloor B \rfloor}$. The algorithm first solves the natural LP relaxation for k-CS-PIP to obtain fractional solution x. Then it proceeds as follows.

1. Sample each item $j \in [n]$ independently with probability x_j/α. Let $\mathcal{S}$ denote the set of chosen items.

2. Define *t-sizes* as follows: for every item j and constraint $i \in N(j)$, round up a_{ij} to

$$t_{ij} \in \{2^{-a} \mid a \in \mathbb{Z}_{\geq 0}\},$$

   the next larger power of 2.

3. For any item j and constraint $i \in N(j)$, let E_{ij} denote the event that the items $\{j' \in \mathcal{S} \mid t_{ij'} \geq t_{ij}\}$ have total t-size (in constraint i) exceeding b_i. Mark j for deletion if E_{ij} occurs for any $i \in N(j)$.

4. Return set $\mathcal{S}' \subseteq \mathcal{S}$ consisting of all items $j \in \mathcal{S}$ not marked for deletion.

Note the differences from the algorithm in Section 3: the scaling factor for randomized rounding is smaller, and the alteration step is more intricate (it uses slightly modified sizes). It is clear that $\mathcal{S}'$ is a feasible solution with probability one, since the original sizes are at most the t-sizes.

The approximation guarantee is proved using the following lemma.

Lemma 5.1. *For each $j \in [n]$,*

$$\Pr[j \in \mathcal{S}' \mid j \in \mathcal{S}] \;\geq\; \left(1 - \frac{1}{k\lfloor B \rfloor}\right)^k.$$

Proof. Fix any $j \in [n]$ and $i \in N(j)$. Recall that E_{ij} is the event that items $\{j' \in \mathcal{S} \mid t_{ij'} \geq t_{ij}\}$ have total t-size (in constraint i) greater than b_i.

We first bound $\Pr[E_{ij} \mid j \in \mathcal{S}]$. Let $t_{ij} = 2^{-\ell}$, where $\ell \in \mathbb{N}$. Observe that all the t-sizes that are at least $2^{-\ell}$ are actually integral multiples of $2^{-\ell}$ (since they are all powers of two). Let

$$\mathcal{I}_{ij} \;=\; \{j' \in [n] \mid t_{ij'} \geq t_{ij}\} \setminus \{j\},$$

and $Y_{ij} := \sum_{j' \in \mathcal{I}_{ij}} t_{ij'} \cdot \mathbf{1}_{j' \in \mathcal{S}}$ where $\mathbf{1}_{j' \in \mathcal{S}}$ are indicator random variables. The previous observation implies that Y_{ij} is always an integral multiple of $2^{-\ell}$. Note that

$$\begin{aligned}\Pr[E_{ij} \mid j \in \mathcal{S}] \;&=\; \Pr\left[Y_{ij} > b_i - 2^{-\ell} \;\middle|\; j \in \mathcal{S}\right]\\ &\leq\; \Pr\left[Y_{ij} > \lfloor b_i \rfloor - 2^{-\ell} \;\middle|\; j \in \mathcal{S}\right] \;=\; \Pr\left[Y_{ij} \geq \lfloor b_i \rfloor \mid j \in \mathcal{S}\right],\end{aligned}$$

where the last equality uses the fact that Y_{ij} is always a multiple of $2^{-\ell}$. Since each item is included into $\mathcal{S}$ independently, we also have $\Pr[Y_{ij} \geq \lfloor b_i \rfloor \mid j \in \mathcal{S}] = \Pr[Y_{ij} \geq \lfloor b_i \rfloor]$. Now Y_{ij} is the sum of independent $[0,1]$ random variables with mean:

$$\mathbb{E}[Y_{ij}] \;=\; \sum_{j' \in \mathcal{I}_{ij}} t_{ij'} \cdot \Pr[j' \in \mathcal{S}] \;\leq\; \sum_{j'=1}^{n} t_{ij'} \cdot \frac{x_{j'}}{\alpha} \;\leq\; \frac{2}{\alpha} \sum_{j'=1}^{n} a_{ij'} \cdot x_{j'} \;\leq\; \frac{2}{\alpha} b_i.$$

Choose δ such that $(\delta+1) \cdot \mathbb{E}[Y_{ij}] = \lfloor b_i \rfloor$, i. e., (using $b_i \geq 1$),

$$\delta + 1 \;=\; \frac{\lfloor b_i \rfloor}{\mathbb{E}[Y_{ij}]} \;\geq\; \frac{\alpha \lfloor b_i \rfloor}{2 \cdot b_i} \;\geq\; \frac{\alpha}{4}.$$

Now using the Chernoff bound (2.1), we have:

$$\Pr[Y_{ij} \geq \lfloor b_i \rfloor] \;=\; \Pr\left[Y_{ij} \geq (1+\delta)\cdot \mathbb{E}[Y_{ij}]\right] \;\leq\; \left(\frac{\mathrm{e}}{\delta+1}\right)^{\lfloor b_i \rfloor} \;\leq\; \left(\frac{4\,\mathrm{e}}{\alpha}\right)^{\lfloor b_i \rfloor} \;\leq\; \left(\frac{4\,\mathrm{e}}{\alpha}\right)^{\lfloor B \rfloor}.$$

The last inequality uses the fact that $b_i \geq B$. Finally, by the choice of $\alpha = 4\mathrm{e}\cdot(\lfloor B \rfloor k)^{1/\lfloor B \rfloor}$,

$$\Pr[E_{ij} \mid j \in \mathcal{S}] \;\leq\; \Pr[Y_{ij} \geq \lfloor b_i \rfloor] \;\leq\; \frac{1}{k\lfloor B \rfloor}. \tag{5.1}$$

As in the proof of Lemma 3.9, for any fixed item $j \in [n]$, the conditional events $\{E_{ij} \mid j \in \mathcal{S}\}_{i \in N(j)}$ are positively correlated. Thus using (5.1) and the FKG inequality (Theorem 3.4),

$$\Pr[j \in \mathcal{S}' \mid j \in \mathcal{S}] \;=\; \Pr\left[\bigwedge_{i \in N(j)} \neg E_{ij} \;\middle|\; j \in \mathcal{S}\right] \;\geq\; \prod_{i \in N(j)} \Pr[\neg E_{ij} \mid j \in \mathcal{S}] \;\geq\; \left(1 - \frac{1}{k\lfloor B \rfloor}\right)^{k}.$$

This completes the proof of the lemma. □

Notice that for any $M \geq 2$, we have

$$\begin{aligned}(1-1/M)^{-M} \;&=\; \exp(-M\cdot \ln(1-1/M)) \;=\; \exp\big(M\cdot(1/M + 1/(2M^2) + 1/(3M^3) + \cdots)\big)\\ &=\; \mathrm{e} + \mathrm{e}/(2M) + O(1/M^2) \;\leq\; \mathrm{e} + c/M\end{aligned}$$

for some constant $c > 0$. By Lemma 5.1 and using $M = k\lfloor B \rfloor$ in this inequality,

$$\Pr[j \in \mathcal{S}' \mid j \in \mathcal{S}] \;\geq\; \left(1 - \frac{1}{k\lfloor B \rfloor}\right)^{k} \;\geq\; \left(\mathrm{e} + \frac{c}{k\lfloor B \rfloor}\right)^{-1/\lfloor B \rfloor}.$$

Since $\Pr[j \in \mathcal{S}] = x_j/\alpha$, plugging in the value of α, we obtain

$$\Pr[j \in \mathcal{S}'] \;\geq\; \frac{x_j}{4\mathrm{e}\cdot\big(\mathrm{e}\lfloor B \rfloor k + c\big)^{1/\lfloor B \rfloor}}.$$

This proves the first part of Theorem 1.4 (for linear objectives).

This algorithm can also be used for maximizing *submodular* functions over such packing constraints (parameterized by k and B). Again, we would first (approximately) solve the continuous relaxation using [7], and perform the above randomized rounding and alteration. Corollary 4.4 can be used with Lemma 5.1 to obtain a $\big(4\mathrm{e}^2/(\mathrm{e}-1)\big)\cdot\big(\mathrm{e}\lfloor B \rfloor k + c\big)^{1/\lfloor B \rfloor}$-approximation algorithm.

This completes the proof of Theorem 1.4. □

Integrality gap for general B We show that the natural LP relaxation for k-CS-PIP has an $\Omega(k^{1/\lfloor B \rfloor})$ integrality gap for every $B \geq 1$, matching the above approximation ratio up to constant factors. For any $B \geq 1$, let $t := \lfloor B \rfloor$. We construct an instance of k-CS-PIP with n columns and $m = \binom{n}{t+1}$ constraints. For all $j \in [n]$, weight $w_j = 1$. For every $(t+1)$-subset $C \subseteq [n]$, there is a constraint $i(C)$ involving the variables in C: set $a_{i(C),j} = 1$ for all $j \in C$, and $a_{i(C),j} = 0$ for $j \notin C$. For each constraint $i \in [m]$, the capacity $b_i = B$. Note that the column sparsity

$$k = \binom{n-1}{t} \le \left(\frac{ne}{t}\right)^t .$$

Setting $x_j = 1/2$ for all $j \in [n]$ is a feasible fractional solution. Indeed, each constraint is occupied to extent

$$\frac{t+1}{2} \le \frac{B+1}{2} \le B$$

(since $B \ge 1$). Thus the optimal LP value is at least $n/2$. On the other hand, the optimal integral solution has value at most t. Suppose for contradiction that the solution contains some $t+1$ items, indexed by $C \subseteq [n]$. Then consider the constraint $i(C)$, which is occupied to extent $t+1 = \lfloor B \rfloor + 1 > B$, this contradicts the feasibility of the solution. Thus the integral optimum is t, and the integrality gap for this instance is at least

$$\frac{n}{2t} \ge \frac{1}{2e} k^{1/\lfloor B \rfloor} .$$

Acknowledgments Aravind Srinivasan thanks Moses Charikar, Chandra Chekuri, Magnús Halldórsson, Goran Konjevod, Seffi Naor, Jaikumar Radhakrishnan, R. Ravi and K.V. Subrahmanyam for helpful discussions. Nitish Korula thanks Chandra Chekuri and Alina Ene for detailed discussions on k-CS-PIP. We thank Deeparnab Chakrabarty and David Pritchard for discussions and sharing a copy of [25]. We also thank Jan Vondrák and Chandra Chekuri for pointing out an error in the earlier proof of Theorem 1.3, which prompted us to prove Theorem 4.3. Our thanks also to the IPCO and ToC referees for their helpful suggestions, which improved the presentation of the paper.

References

[1] NOGA ALON AND JOEL SPENCER: *The Probabilistic Method.* Wiley-Interscience, 2008. 540, 549

[2] ESTHER M. ARKIN AND REFAEL HASSIN: On local search for weighted k-set packing. *Math. Oper. Res.*, 23(3):640–648, 1998. Preliminary version in ESA'97. [doi:10.1287/moor.23.3.640] 539

[3] PER AUSTRIN, SUBHASH KHOT, AND MULI SAFRA: Inapproximability of vertex cover and independent set in bounded degree graphs. *Theory of Computing*, 7(3):27–43, 2011. Preliminary version in CCC'09. [doi:10.4086/toc.2011.v007a003] 539

[4] NIKHIL BANSAL, NITISH KORULA, VISWANATH NAGARAJAN, AND ARAVIND SRINIVASAN: On k-column sparse packing programs. In *Proc. 14th Ann. Conf. on Integer Programming and Combinatorial Optimization (IPCO '10)*, pp. 369–382, 2010. [doi:10.1007/978-3-642-13036-6_28] 533

[5] JÓZSEF BECK AND TIBOR FIALA: "Integer-making" theorems. *Discrete Appl. Math.*, 3(1):1–8, 1981. [doi:10.1016/0166-218X(81)90022-6] 538

[6] PIOTR BERMAN: A $d/2$ approximation for maximum weight independent set in d-claw free graphs. *Nord. J. Comput.*, 7(3):178–184, 2000. Preliminary version in SWAT'00. 539

[7] GRUIA CALINESCU, CHANDRA CHEKURI, MARTIN PÁL, AND JAN VONDRÁK: Maximizing a monotone submodular function subject to a matroid constraint. *SIAM J. Comput.*, 40(6):1740–1766, 2011. Combines results from IPCO'07 and STOC'08. [doi:10.1137/080733991] 539, 553, 554, 560

[8] BARUN CHANDRA AND MAGNÚS M. HALLDÓRSSON: Greedy local improvement and weighted set packing approximation. *J. Algorithms*, 39(2):223–240, 2001. Preliminary version in SODA'99. [doi:10.1006/jagm.2000.1155] 539

[9] CHANDRA CHEKURI, ALINA ENE, AND NITISH KORULA: Personal communication, 2009. 537, 538

[10] FUNDA ERGÜN, RAKESH K. SINHA, AND LISA ZHANG: QoS routing with performance-dependent costs. In *19th Ann. Joint Conf. IEEE Computer and Communications Societies (INFOCOM'00)*, pp. 137–146, 2000. [doi:10.1109/INFCOM.2000.832182] 536

[11] URIEL FEIGE: On maximizing welfare when utility functions are subadditive. *SIAM J. Comput.*, 39(1):122–142, 2009. Preliminary version in STOC'06. [doi:10.1137/070680977] 537, 553, 554

[12] MORAN FELDMAN, JOSEPH (SEFFI) NAOR, ROY SCHWARTZ, AND JUSTIN WARD: Improved approximations for k-exchange systems - (extended abstract). In *Proc. 19th Ann. European Symp. on Algorithms (ESA'11)*, pp. 784–798, 2011. [doi:10.1007/978-3-642-23719-5_66] 539

[13] M. L. FISHER, G. L. NEMHAUSER, AND L. A. WOLSEY: An analysis of approximations for maximizing submodular set functions II. *Math. Prog. Study*, 8:73–87, 1978. [doi:10.1007/BFb0121195] 537, 539

[14] ERAN HALPERIN: Improved approximation algorithms for the vertex cover problem in graphs and hypergraphs. *SIAM J. Comput.*, 31(5):1608–1623, 2002. Preliminary version in SODA'00. [doi:10.1137/S0097539700381097] 539

[15] ELAD HAZAN, SHMUEL SAFRA, AND ODED SCHWARTZ: On the complexity of approximating k-set packing. *Comput. Complexity*, 15(1):20–39, 2006. Preliminary version in APPROX'03. [doi:10.1007/s00037-006-0205-6] 537, 539

[16] COR A. J. HURKENS AND ALEXANDER SCHRIJVER: On the size of systems of sets every t of which have an SDR, with an application to the worst-case ratio of heuristics for packing problems. *SIAM J. Discrete Math.*, 2(1):68–72, 1989. [doi:10.1137/0402008] 534, 539

[17] SUBHASH KHOT: On the power of unique 2-prover 1-round games. In *Proc. 34th STOC*, pp. 767–775. ACM Press, 2002. [doi:10.1145/509907.510017] 539

[18] STAVROS G. KOLLIOPOULOS AND CLIFFORD STEIN: Approximating disjoint-path problems using packing integer programs. *Math. Program.*, 99(1):63–87, 2004. Preliminary version in IPCO'98. [doi:10.1007/s10107-002-0370-6] 538

[19] JOCHEN KÖNEMANN AND R. RAVI: A matter of degree: Improved approximation algorithms for degree-bounded minimum spanning trees. *SIAM J. Comput.*, 31(6):1783–1793, 2002. Preliminary version in STOC'00. [doi:10.1137/S009753970036917X] 536

[20] ARIEL KULIK, HADAS SHACHNAI, AND TAMI TAMIR: Maximizing submodular set functions subject to multiple linear constraints. In *Proc. 20th Ann. ACM-SIAM Symp. on Discrete Algorithms (SODA'09)*, pp. 545–554. ACM Press, 2009. [ACM:1496770.1496830] 537, 539

[21] JON LEE, VAHAB S. MIRROKNI, VISWANATH NAGARAJAN, AND MAXIM SVIRIDENKO: Maximizing nonmonotone submodular functions under matroid or knapsack constraints. *SIAM J. Discrete Math.*, 23(4):2053–2078, 2010. Preliminary version in STOC'09. [doi:10.1137/090750020] 539

[22] MICHAEL LUBY AND NOAM NISAN: A parallel approximation algorithm for positive linear programming. In *Proc. 25th STOC*, pp. 448–457. ACM Press, 1993. [doi:10.1145/167088.167211] 536, 540

[23] RAJEEV MOTWANI AND PRABHAKAR RAGHAVAN: *Randomized Algorithms*. Cambridge University Press, 1995. 540, 544

[24] DAVID PRITCHARD: Approximability of sparse integer programs. In *Proc. 17th Ann. European Symp. on Algorithms (ESA'09)*, pp. 83–94. Springer, 2009. [doi:10.1007/978-3-642-04128-0_8] 538, 558

[25] DAVID PRITCHARD AND DEEPARNAB CHAKRABARTY: Approximability of sparse integer programs. *Algorithmica*, 61(1):75–93, 2011. [doi:10.1007/s00453-010-9431-z] 537, 538, 561

[26] PRABHAKAR RAGHAVAN: Probabilistic construction of deterministic algorithms: Approximating packing integer programs. *J. Comput. System Sci.*, 37(2):130–143, 1988. Preliminary version in FOCS'86. [doi:10.1016/0022-0000(88)90003-7] 538, 540

[27] F. BRUCE SHEPHERD AND ADRIAN VETTA: The demand-matching problem. *Math. Oper. Res.*, 32(3):563–578, 2007. Preliminary version in IPCO'02. [doi:10.1287/moor.1070.0254] 552

[28] DAVID B. SHMOYS AND ÉVA TARDOS: An approximation algorithm for the generalized assignment problem. *Math. Program.*, 62(1-3):461–474, 1993. [doi:10.1007/BF01585178] 536

[29] ARAVIND SRINIVASAN: Improved approximation guarantees for packing and covering integer programs. *SIAM J. Comput.*, 29(2):648–670, 1999. Preliminary version in STOC'95. [doi:10.1137/S0097539796314240] 534, 535, 536, 538, 540, 545

[30] ARAVIND SRINIVASAN: New approaches to covering and packing problems. In *Proc. 12th Ann. ACM-SIAM Symp. on Discrete Algorithms (SODA'01)*, pp. 567–576. ACM Press, 2001. [ACM:365411.365535] 533, 539

[31] JAN VONDRÁK: Optimal approximation for the submodular welfare problem in the value oracle model. In *Proc. 40th STOC*, pp. 67–74. ACM Press, 2008. [doi:10.1145/1374376.1374389] 537, 539, 553

[32] JUSTIN WARD: A $(k+3)/2$-approximation algorithm for monotone submodular k-set packing and general k-exchange systems. In *Proc. 29th Symp. Theoretical Aspects of Comp. Sci. (STACS'12)*, pp. 42–53. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik, 2012. [doi:10.4230/LIPIcs.STACS.2012.42] 539

[33] DAVID ZUCKERMAN: Linear degree extractors and the inapproximability of max clique and chromatic number. *Theory of Computing*, 3(1):103–128, 2007. Preliminary version in STOC'06. [doi:10.4086/toc.2007.v003a006] 534, 538

AUTHORS

Nikhil Bansal
Dept. of Mathematics and Computer Science
Eindhoven University of Technology, the Netherlands
n.bansal@tue.nl
http://www.win.tue.nl/~nikhil

Nitish Korula
Research Scientist
Google Inc.
nitish@google.com
http://research.google.com/pubs/NitishKorula.html

Viswanath Nagarajan
Research Staff Member
IBM T. J. Watson Research Center
viswanath@us.ibm.com
http://researcher.ibm.com/person/us-viswanath

Aravind Srinivasan
Dept. of Computer Science and UMIACS
University of Maryland
College Park, MD
srin@cs.umd.edu
http://www.cs.umd.edu/~srin

ABOUT THE AUTHORS

NIKHIL BANSAL received his Ph. D. from CMU in 2003; advised by Avrim Blum where he benefited greatly from his wisdom and caring. His research has focused on approximation and online algorithms for scheduling and other optimization problems. He grew up all over India, living in more than a dozen cities. After living in the U. S. for several years, he recently moved to the Netherlands, where he is still getting used to the idea that he can bike to anywhere without being run over by speeding maniacs.

NITISH KORULA received his Ph. D. in 2012 at the University of Illinois, where he was advised by Chandra Chekuri, and his B. E. in 2004 from the Birla Institute of Technology & Science, Pilani, where he benefited from the guidance of Sundar Balasubramaniam. After graduating, he joined Google as a Research Scientist in New York, a striking change from Champaign-Urbana, Pilani, and Vellore, where he grew up. His research interests are in approximation algorithms, particularly for network design and other graph problems, online algorithms, and combinatorics.

VISWANATH NAGARAJAN received his M. S. (2006) and Ph. D. (2009) degrees from Carnegie Mellon University under the supervision of R. Ravi. He thanks R. Ravi and Anupam Gupta for their guidance in his early research career. After graduating from CMU, he became a Research Staff Member in the algorithms group at the IBM T. J. Watson Research Center. His research interests are in approximation algorithms and combinatorial optimization. He grew up in Chennai, Kochi, and Mumbai, where he also received his Bachelors degree from the Indian Institute of Technology Bombay.

ARAVIND SRINIVASAN received his Ph. D. from Cornell University in 1993 under the guidance of David B. Shmoys. In addition to David Shmoys, he thanks Joel Spencer for mentoring him during his early career. He is a Professor at the University of Maryland, College Park. His research interests include algorithms, probabilistic methods, networks, and computational approaches to sustainability. He grew up in the city of Chennai (formerly Madras) in India, and his native language is Tamil. He is motivated by humor, and is inspired by Nature.