DIMACS Technical Report 95-37
September 1995

Improved Approximation Guarantees for Packing and Covering Integer Programs[1]

by

Aravind Srinivasan[2,3]
Department of Information Systems and Computer Science
National University of Singapore
Lower Kent Ridge Road
Singapore 119260
Republic of Singapore

DIMACS is a partnership of Rutgers University, Princeton University, AT&T Research, Bellcore, and Bell Laboratories.
DIMACS is an NSF Science and Technology Center, funded under contract STC–91–19999; and also receives support from the New Jersey Commission on Science and Technology.

Abstract

Several important NP-hard combinatorial optimization problems can be posed as *packing/covering integer programs*; the *randomized rounding* technique of Raghavan & Thompson is a powerful tool to approximate them well. We present one elementary unifying property of all these integer programs (IPs), and use the FKG correlation inequality to derive an improved analysis of randomized rounding on them. This also yields a *pessimistic estimator*, thus presenting deterministic polynomial-time algorithms for them with approximation guarantees significantly better than those known.

Keywords. Approximation Algorithms, Combinatorial Optimization, Correlation Inequalities, Covering Integer Programs, De-randomization, Integer Programming, Linear Programming, Linear Relaxations, Packing Integer Programs, Positive Correlation, Randomized Rounding, Rounding Theorems.

[1]Preliminary versions of this work appeared in the *Proc. ACM Symposium on the Theory of Computing*, pages 268–276, 1995, and as DIMACS Technical Report 94-53.

[2]Post-doctoral Fellow.

[3]Part of this work was done at the National University of Singapore. Part of the work was done while at the School of Mathematics, Institute for Advanced Study, Princeton, NJ 08540, USA, supported by grant 93-6-6 of the Alfred P. Sloan Foundation to the Institute for Advanced Study. Part was done while visiting the Department of Computer Science, University of Warwick, Coventry CV4 7AL, UK, supported in part by the ESPRIT Basic Research Action Programme of the EC under contract No. 7141 (project ALCOM II). E-mail: **aravind@iscs.nus.sg**.

1 Introduction

Several important NP-hard combinatorial optimization problems such as basic problems on graphs and hypergraphs, can be posed as *packing/covering integer programs*; the *randomized rounding* technique of Raghavan & Thompson is a powerful tool to approximate them well [21]. We present an elementary property of all these IPs–*positive correlation*–and use the FKG inequality (Fortuin, Kasteleyn & Ginibre [10], Sarkar [22]) to derive an improved analysis of randomized rounding on them. Interestingly, this yields a *pessimistic estimator*, thus presenting deterministic polynomial algorithms for them with approximation guarantees significantly better than those known, in a unified way.

1.1 Previous work

Let Z_+ and $\Re_+$ denote the non-negative integers and the non-negative reals respectively. For a (column) vector v, let v^T denote its transpose, and v_i stand for its ith component. We first define the packing and covering integer programs.

Definition 1 *Given $A \in [0,1]^{n \times m}$, $b \in [1,\infty)^n$ and $c \in [0,1]^m$ with $max_j\ c_j = 1$, a packing (resp. covering) integer program PIP (resp. CIP) seeks to maximize (resp. minimize) $c^T \cdot x$ subject to $x \in Z_+^m$ and $Ax \leq b$ (resp. $Ax \geq b$). Furthermore if $A \in \{0,1\}^{n \times m}$, we assume that each entry of b is integral. We also define $B = min_i\ b_i$.*

Though there are usually no restrictions on the entries of A, b and c aside of non-negativity, it is easily seen that the above restrictions are without loss of generality (w.l.o.g.), because of the following. First, we may assume that $\forall i, j,\ A_{ij}$ is at most b_i. If this is not true for a PIP, then we may as well set $x_j := 0$; if this is not true for a CIP, we can just reset $A_{ij} := b_i$. Next, by scaling each row of A such that $\max_j\ A_{i,j} = 1$ for each row i and by scaling c so that $\max_j\ c_j = 1$, we get the above form for A, b and c. Finally, if $A \in \{0,1\}^{n \times m}$, then for a PIP, we can always reset $b_i := \lfloor b_i \rfloor$ for each i and for a CIP, reset $b_i := \lceil b_i \rceil$; hence the assumption on the integrality of each b_i, in this case.

Remark. The reader is requested to take note of the parameter B; it will occur frequently in the rest of the paper. Whenever we use the symbol B as a parameter for any given problem, it will be so since, in the "natural" PIP/CIP formulation of the problem, B will play the same role as it does in Definition 1.

As mentioned above, PIPs and CIPs model some basic problems in combinatorial optimization, but most of these problems are NP-hard; hence we are interested in efficient approximation algorithms for PIPs and CIPs, with a good performance guarantee. We now turn to an important technique for approximating integer linear programs–"relaxing" their integrality constraints, and considering the resulting linear program.

Definition 2 *The standard LP relaxation of PIPs/CIPs lets $x \in \Re_+^m$; given a PIP/CIP, x^* and y^* denote, resp., an optimal solution to, and the optimum value of, this relaxation. (For packing, we also allow constraints of the form $x_i \in \{0, 1, \ldots, d_i\}$, for any set of positive integers $\{d_i\}$; the LP relaxation sets $x_i \in [0, d_i]$ here.)*

Given a PIP or a CIP, we can solve its LP relaxation efficiently. However, how do we handle the possibility of possibly fractional entries in x^*? We need some mechanism to "round" fractional entries in x^* to integers, suitably. One possibility is to round every fractional value x_i^* to the closest integer, with some tie-breaking rule if x_i^* is half of an integer. However, it is known that such "thresholding" methods are of limited applicability.

A key technique to approximate a class of integer programming problems via a new rounding method–*randomized rounding*–was proposed in [21]. Given a positive real v, the idea is to look at its fractional part as a probability–round v to $\lfloor v \rfloor + 1$ with probability $v - \lfloor v \rfloor$, and round v to $\lfloor v \rfloor$ with probability $1 - v + \lfloor v \rfloor$. This has the nice property that the expected value of the result is v. How can we use this for packing and covering problems? Consider a PIP, for instance. Solve its LP relaxation and set $x_i' := x_i^*/\alpha$ for some parameter $\alpha > 1$ to be fixed later; this scaling down by α is done to boost the chance that the constraints in the PIP are all satisfied–recall that they are all $\leq$–constraints. Now define a random $z \in Z_+^m$, the outcome of randomized rounding, as follows. Independently for each i, set z_i to be $\lfloor x_i' \rfloor + 1$ with probability $x_i' - \lfloor x_i' \rfloor$, and $\lfloor x_i' \rfloor$ with probability $1 - (x_i' - \lfloor x_i' \rfloor)$.

We now need to show that all the constraints in the PIP are satisfied *and* that $c^T \cdot z$ is not "much below" y^*, with reasonable probability; we also need to choose α suitably. This is formalized in [21] as follows. As seen above, an important observation is that $E[z_i] = x_i'$. Hence,

$$E[(Az)_i] = (Ax')_i \leq b_i/\alpha$$

and

$$E[c^T \cdot z] = y^*/\alpha.$$

For some $\beta > 1$ to be fixed later, define events $E_1, E_2, \ldots, E_n$ by $E_i \equiv$ "$(Az)_i > b_i''$, and let $E_{n+1} \equiv$ "$c^T \cdot z < y^*/(\alpha\beta)''$. Now, z is an $(\alpha\beta)$–approximate solution to PIP if

$$\bigwedge_{i=1}^{n+1} \overline{E_i}$$

holds. How small a value for $(\alpha\beta)$ can we achieve? Bounding

$$Pr(\bigvee_{i=1}^{n+1} E_i) \leq \sum_{i=1}^{n+1} Pr(E_i), \tag{1}$$

we can pick $\alpha, \beta > 1$ such that $\sum_{i=1}^{n+1} Pr(E_i) < 1$ holds, using the Chernoff-Hoeffding (CH) bounds. This gives us an $(\alpha\beta)$–approximation z with nonzero probability, which is also made deterministic by Raghavan, using pessimistic estimators [19]. Similar ideas hold for CIPs–the fractions $\{x_i^*\}$ are scaled *up* by some $\alpha > 1$ here. Similar approximation bounds are derived through different methods by Plotkin, Shmoys & Tardos [18]. See Raghavan [20] for a survey of randomized rounding, and Crescenzi & Kann [7] for a comprehensive collection of results on NP-optimization problems.

Though randomized rounding is a unifying idea to derive good approximation algorithms, there are better approximation bounds for specific key problems such as set cover (Johnson [13], Lovász [14], Chvátal [6]), hypergraph matching (Aharoni, Erdős & Linial [1]) and file-sharing in distributed networks (Naor & Roth [17]), each derived through different means. One reason for this slack stems from bounding $Pr(\bigvee_{i=1}^{n+1} E_i)$ by $\sum_{i=1}^{n+1} Pr(E_i)$: to quote Raghavan [19],

> Throughout, we naively (?) sum the probabilities of all bad events–although these bad events are surely correlated. Can we prove a stronger result using algebraic properties (e.g., the rank) of the coefficient matrix? A tighter bound for the probabilistic existence proofs should lead to tighter approximation algorithms.

1.2 Proposed new method

We make progress in the above-suggested direction by exploiting an elementary property–*positive correlation*–of CIPs and PIPs. To motivate this idea, let us just take two constraints of a PIP, and let E_1 and E_2 be the corresponding bad events, as defined before. For instance, suppose E_1 is the event that $0.1z_1 + z_3 + 0.5z_4 + 0.9z_6 > 1.1$, and E_2 stands for the event that $0.4z_1 + 0.3z_2 + z_5 + 0.1z_6 > 1.2$, where the z_i are all *independent* 0-1 random variables. Now suppose we are given that $\overline{E_1}$ holds. Very roughly speaking, this seems to suggest that "many" among z_1, z_3, z_4 and z_6 were "small" (*i.e.*, zero), which seems to boost the chance that $\overline{E_2}$ holds, too. Formally, the claim is that $Pr(\overline{E_2}|\overline{E_1}) \geq Pr(\overline{E_2})$, *i.e.*, that $Pr(\overline{E_1} \bigwedge \overline{E_2}) \geq Pr(\overline{E_1}) \cdot Pr(\overline{E_2})$. This "intuitively clear" fact can then be easily generalized for us to guess that

$$\forall k \ \forall 1 \leq i_1 < i_2 < \ldots < i_k \leq n, \ Pr(\bigwedge_{j=1}^{k} \overline{E_{i_j}}) \geq \prod_{j=1}^{k} Pr(\overline{E_{i_j}}). \quad (2)$$

In other words, (2) claims that the constraints are *positively correlated*–given that all of any given subset of them are satisfied, the conditional probability that any other constraint is also satisfied, cannot go below its unconditional probability.

We prove (2), which seems plausible, using the FKG inequality. Thus,

$$Pr(\bigvee_{i=1}^{n+1} E_i) \leq Pr(\bigvee_{i=1}^{n} E_i) + Pr(E_{n+1}) \leq 1 - (\prod_{i=1}^{n}(1 - Pr(E_i))) + Pr(E_{n+1}), \quad (3)$$

which is always as good as, and most often much better than, (1). (For a detailed study of the FKG inequality, see, *e.g.*, Graham [11] and Chapter 6 of Alon, Spencer & Erdős [2].)

It is not hard to verify such a property for CIPs also. Why we have been so lucky as to have positive correlation among the constraints of PIPs and CIPs (a very desirable form of correlation)? The features of PIPs and CIPs which guarantee this are:

- All the entries of the matrix A are *non-negative*, and
- all the constraints "point" in the *same direction*.

Of course, it can also be shown that given that all of any given subset of the constraints are *violated*, the conditional probability that any other constraint is also violated, cannot go below its unconditional probability; but we will not have to deal with this situation! Also, such a nice correlation as given by (2) may not necessarily hold if the z_is are not independent.

More surprisingly, though this new approach usually only guarantees that z is a "good" approximation with very low (albeit positive) probability–in fact, it does not even seem to provide a randomized algorithm with any good success probability–the structure of PIPs and CIPs implies a sub-additivity property which yields a *pessimistic estimator* (a notion to be introduced in Section 2); we thus get deterministic polynomial-time algorithms achieving these improved approximation bounds. The problem in arriving at a good pessimistic estimator is that while the previous estimator $\sum_{i=1}^{n+1} Pr(E_i)$ (*i.e.*, the one used in [19] and in related papers) is upper-bounded by $E[Z]$ (for some random variable Z) on applying the CH bounds, such a fact does not seem to hold here. Nevertheless, the structure of CIPs/PIPs–in particular, the two simple properties itemized above–help in providing a good pessimistic estimator. This is a point that we would like to stress.

Thus we get, in a unified way, improved bounds on the *integrality gap*

$$\max\{(c^T \cdot z)/y^*, y^*/(c^T \cdot z)\}$$

and hence, improved approximation algorithms for all PIPs and CIPs. In particular, we improve on the above-mentioned results of [13, 14, 1, 17]; our bound is incomparable with that of [6].

1.3 Approximation bounds achieved

Our best improvements are for PIPs. For PIPs, the standard analysis of randomized rounding guarantees integral solutions of value $t_1 = \Omega(y^*/n^{1/B})$ and

$t_2 = \Omega(y^*/n^{1/(B+1)})$ respectively, if $A \in [0,1]^{n\times m}$ and $A \in \{0,1\}^{n\times m}$. Our method provides $\Omega(t_1^{B/(B-1)})$ and $\Omega(t_2^{(B+1)/B})$ bounds resp., thus improving well on the previous ones–*e.g.*, in the latter case if $y^* = \Theta(n)$ and $B = 1$, we get an integral solution of value $\Theta(n)$, as opposed to the previous $\Omega(\sqrt{n})$ bound. This method also gives Turán's classical theorem on independent sets in graphs [25] to within a constant factor.

An important packing problem where $A \in \{0,1\}^{n\times m}$ is *simple B–matching in hypergraphs* [14]: given a hypergraph with non-negative edge weights, finding a maximum weight collection of edges such that no vertex occurs in more than B of them. Usual hypergraph matching has $B = 1$, and is a well-known NP-hard problem. To our knowledge, the only known good bound for this problem, apart from the standard analysis of randomized rounding, was provided by the work of [1], which focused on the special case of *unweighted* edges. The methods of [1] can be used to show that if f is the minimum size of an edge in the hypergraph, then there exists an integral matching of value at least

$$\frac{(y^*)^2}{B^2 n - (f-1)(y^*)^2/\min(m,n)} \geq \frac{(y^*)^2}{B^2 n}.$$

While this matches our result to within a constant factor for $B = 1$, note that this bound *worsens* as B increases, while the standard analysis, as well as our present analysis, of randomized rounding in fact show that the integrality gap gets better (decreases) as B increases.

For covering, we prove an

$$1 + O(\max\{\ln(nB/y^*)/B, \sqrt{\ln(nB/y^*)/B}\}) \tag{4}$$

integrality gap, and derive the corresponding deterministic polynomial-time approximation algorithm. This improves on the

$$1 + O(\max\{(\ln n)/B, \sqrt{(\ln n)/B}\})$$

bound given by the standard analysis of randomized rounding. Also, Dobson [8] and Fisher & Wolsey [9] bound the performance of a natural greedy algorithm for CIPs in terms of the optimal *integral* solution. Our bound is incomparable with theirs, but for any given A, c, and the unit vector $b/||b||_2$ pointing in the direction of b, our bound is always better if B is more than a certain threshold $thresh(A,b,c)$. See Bertsimas & Vohra [4] for a detailed study of approximating CIPs; our work improves on all of their randomized rounding bounds except for their *weighted* CIPs (wherein it is not the case that $c_i = 1$ for all i) for which our bounds are incomparable with theirs.

An important subclass of the CIPs models the *unweighted set cover problem*: $\forall i,j$, $A_{i,j} \in \{0,1\}$, $b_i = 1$ and $c_j = 1$, here. The combinatorial interpretation is that we have a hypergraph $H = (V,E)$, and wish to pick a minimum cardinality collection of the edges so that every vertex is covered. (When viewed as an LP, this is the "dual" of the hypergraph matching problem.) The rows correspond to V and the columns, to E. Clearly, this problem requires that $x \in \{0,1\}^m$, which is not guaranteed by Definition 1; however, note that for this problem, any $x \in Z_+^m$ with $Ax \geq b$ trivially yields a $y \in \{0,1\}^m$ with $Ay \geq b$ and $c^T \cdot y \leq c^T \cdot x$ (set $y_i = \min(1, x_i)$).

For set cover, we tighten the constants in (4) to derive an

$$\ln(n/y^*) + O(\ln\ln(n/y^*)) + O(1)$$

approximation bound. The work of Lund & Yannakakis [15] and Bellare, Goldwasser, Lund & Russell [3] shows a constant $a > 0$ such that approximating this problem to within $a \ln n$ is likely to take super-polynomial time. However, this problem is important enough to study approximations parametrized by other parameters of A, b and c, that are always as good as and often much better than, $\Theta(\log n)$; for instance, the work of [13, 14, 6] shows a $\ln d + O(1)$ approximation bound, where d is the maximum column sum in A–note that $d \leq n$. Also since there is a trivial solution of size n for any set cover instance, n/y^* is a simple upper bound on the approximation ratio. Our bound is a further improvement– it is easily seen that $n/y^* \leq d$ always, and that there is a constant $\ell > 0$ such that for every non-decreasing function $f(n)$ with $1 \leq f(n) \leq \ell \ln n / \ln\ln n$, there exist families of (A, b, c) such that

$$\ln(n/y^*) \leq \min\{n/y^*, \ln d\}/f(n).$$

Thus our bound is never more than a multiplicative $(1 + o(1))$ or an additive $O(1)$ factor above the classical bound, and is usually much better; in the best case, our improvement is by $\Theta(\log n / \log\log n)$. (For instance, we can construct instances with $d = n^{\Theta(1)}$ and $y^* = n/\log^{\Theta(1)} n$, giving a $\Theta(\log n/\log\log n)$ improvement.)

Another noteworthy class of CIPs is related to the B–domination problem: given a (directed) graph G with n vertices, we want to place a minimum number of facilities on the nodes such that every node has at least B facilities in its out-neighborhood. This is also a key subproblem in sharing files in a distributed system [17]; under the assumption that G is undirected and letting Δ be its maximum degree, an

$$1 + O(\max\{\ln(\Delta)/B, \sqrt{\ln(\Delta)/B}\})$$

approximation bound is presented in [17], improving on the standard analysis of randomized rounding. Bound (4) improves further on this; in particular, even if G is directed with maximum in-degree Δ, (4) shows that the Naor-Roth bound holds. Furthermore, the comments regarding the $\Theta(\log n/\log\log n)$ improvement for set cover, hold even in the undirected case. All of this, in turn, provides better bounds for the file-sharing problem.

Thus, the two main contributions of this work are as follows. The first is the identification of a very desirable "correlation" property of all packing and covering integer programs, which enables one to prove, quite easily, improved bounds on the integrality gap for the linear relaxations of these problems. However, as shown in Section 4, this is often not constructive, since the probability of randomized rounding resulting in such good approximations can be (and usually is) negligibly small; Section 4 shows a simple family of instances where this "success probability" is as small as $\exp(-\Omega(n+m))$. The second idea, then, is to show that the structure of PIPs and CIPs in fact presents a suitable pessimistic estimator (see Section 2 for the definition), which, pleasingly, actually lets us come up with such approximations efficiently.

In Section 2, we present some basic notions such as large-deviation inequalities, the FKG inequality, and the notion of pessimistic estimators. Section 3 then handles PIPs. We devote Section 4 to the important problem of finding a maximum independent set problem on graphs by looking at it in the natural (and well-known) way as a PIP, and make some observations about this problem; these shine light on the strengths and weaknesses of our approach (and of related approaches). Section 5 handles CIPs; a good understanding of Section 3 is essential to read this section. Section 6 concludes.

2 Preliminaries

Let "r.v." abbreviate "random variable" and for any positive integer k, let $[k]$ denote the set $\{1, 2, \ldots, k\}$. If a universe $N = \{a_1, a_2, \ldots, a_\ell\}$ is understood, then for any $S \subseteq N$, $\chi(S)$ denotes its characteristic vector: $\chi(S) \in \{0,1\}^\ell$ with $\chi(S)_j = 1$ iff $a_j \in S$. For a sequence $s_1, s_2, \ldots$ and any integer $i \geq 1$, $s^{(i)}$ denotes the vector $(s_1, s_2, \ldots, s_i)$. In our usage, $s_1, s_2, \ldots$ could be a sequence of reals or of random variables. As usual, e denotes the base of the natural logarithm.

Remark. Though the following pages seem filled with formulae and calculations, many of them are routine. The real ideas of this work are contained in Lemmas 1, 5, and 6. The reader might even consider skipping the proofs of most of the rest of the lemmas, for the first reading.

We first recall the Chernoff-Hoeffding (CH) bounds, for the tail probabilities of sums of bounded independent r.v.s [5, 12]. Theorem 1 presents these tail bounds; see, *e.g.*, Motwani & Raghavan [16] for the proofs.

Theorem 1 *Let $X_1, X_2, \ldots, X_\ell$ be* independent *r.v.s, each taking values in $[0,1]$, with $R = \sum_{i=1}^{\ell} X_i$ and $E[R] = \mu$. Then for any $\delta \geq 0$,*

$$Pr(R \geq \mu(1+\delta)) < E[(1+\delta)^{R-\mu(1+\delta)}] \leq G(\mu,\delta) \doteq (e^\delta/(1+\delta)^{(1+\delta)})^\mu,$$

and if $0 \leq \delta < 1$, then

$$Pr(R \leq \mu(1-\delta)) < E[(1-\delta)^{R-\mu(1-\delta)}] \leq H(\mu,\delta) \doteq e^{-\mu\delta^2/2}. \qquad \square$$

It is easily seen that

Fact 1 *(a)* $G(\mu, \delta) \leq (e/(1+\delta))^{(1+\delta)\mu}$.

(b) $G(\mu, \delta) \leq e^{-\delta^2\mu/3}$ *if* $\delta \leq 1$.

(c) $G(\mu, \delta) \leq e^{-(1+\delta)\ln(1+\delta)\mu/4}$ *if* $\delta \geq 1$.

(d) If $0 < \mu_1 \leq \mu_2$, *then* $G(\mu_1, \delta) \geq G(\mu_2, \delta)$.

Call a family $\mathcal{F}$ of subsets of a set N *monotone increasing* (resp. *monotone decreasing*) if for all $S \subseteq T \subseteq N$, $S \in \mathcal{F}$ implies that $T \in \mathcal{F}$ (resp., $T \in \mathcal{F}$ implies that $S \in \mathcal{F}$). We next present Theorem 2, a special case of the powerful FKG inequality [10, 22]; for a proof, see, e.g., the proof of Theorem 3.2 in Chapter 6 of [2].

Theorem 2 *Given a finite set* $N = \{a_1, a_2, \ldots, a_\ell\}$ *and some* $p = (p_1, p_2, \ldots, p_\ell) \in [0,1]^\ell$, *suppose we pick a random* $Y \subseteq N$ *by placing each* a_i *in* Y independently, *with probability* p_i. *For any* $\mathcal{F} \subseteq 2^N$, *let* $Pr_p(\mathcal{F}) \doteq Pr(Y \in \mathcal{F})$. *Let* $F_1, F_2, \ldots, F_s \subseteq 2^N$ *be any sequence of monotone increasing families, and let* $G_1, G_2, \ldots, G_s \subseteq 2^N$ *be any sequence of monotone decreasing families. Then,*

$$Pr_p(\bigwedge_{i=1}^{s} F_i) \geq \prod_{i=1}^{s} Pr_p(F_i), \text{ and } Pr_p(\bigwedge_{i=1}^{s} G_i) \geq \prod_{i=1}^{s} Pr_p(G_i). \qquad \square$$

Finally, we recall the notion of *pessimistic estimators* [19]. For our purposes, we focus on the case of independent binary r.v.s. Let $X_1, X_2, \ldots, X_\ell \in \{0,1\}$ be independent r.v.s with $Pr(X_i = 1) = p_i$, for some $p \in [0,1]^\ell$. Suppose, for some implicitly defined $L \subseteq \{0,1\}^\ell$, that

$$Pr(X^{(\ell)} \in L) < 1.$$

How do we find some $v \in \{0,1\}^\ell - L$? Theorem 3 now presents the idea of pessimistic estimators applied to the method of conditional probabilities. See [19] for a detailed discussion and proof.

Notation 1 $\forall q \in [0,1]^\ell\ \forall i \in \{0\} \cup [\ell]\ \forall w \in \{0,1\}^i$, *let* $u(i, w, q) \doteq (w_1, w_2, \ldots w_i, q_{i+1}, q_{i+2}, \ldots, q_\ell)$, *and for any* $j \in \{0,1\}$, *define* $wj \in \{0,1\}^{i+1}$ *as* $(w_1, w_2, \ldots, w_i, j)$.

Returning to the X_is, p and L, we define

Definition 3 *A function* $U : [0,1]^\ell \to \Re^+$ *is a* pessimistic estimator *w.r.t.* $(X_1, \ldots, X_\ell)$ *and* L *if:*

(1) $U(p_1, p_2, \ldots, p_\ell) < 1$, *and*

(2) $\forall i \in \{0\} \cup [\ell]\ \forall w \in \{0,1\}^i$,

(a) $U(u(i,w,p)) \geq Pr(X^{(\ell)} \in L | X^{(i)} = w)$, *and*

(b) *if* $i \leq \ell - 1$, *then* $U(u(i,w,p))$ *is at least*

$$min\{U(u(i+1,w0,p)), U(u(i+1,w1,p))\}.$$

Theorem 3 *[19] Let an efficiently computable* U *be a pessimistic estimator w.r.t.* $(X_1, \ldots, X_\ell)$ *and* L*. Defining,* $\forall i \in \{0\} \cup [\ell - 1]$ $\forall w \in \{0,1\}^i$, $n(i,w) = j \in \{0,1\}$ *by*

$$U(u(i+1,wj,p)) = min\{U(u(i+1,w0,p)), U(u(i+1,w1,p))\}$$

by breaking ties arbitrarily, the following algorithm produces a $v \notin L$*:*
For $i := 0$ *to* $\ell - 1$ *do:* $\quad v_{i+1} := n(i, v^{(i)})$.

Proof. It is not hard to see by induction on i, that $\forall i \in \{0\} \cup [\ell]$,

$$Pr(X^{(\ell)} \in L | X^{(i)} = v^{(i)}) < 1.$$

Using this for $i = \ell$ in conjunction with property 2(a) of Definition 3, completes the proof. □

3 Approximating Packing Integer Programs

Let a PIP be given, conforming to Definition 1. We assume that $x \in Z_+^m$ is the constraint on x. (Clearly, even if we have constraints such as $x_i \in \{0,1,\ldots,d_i\}$, we will get identical bounds since scaling down by $\alpha > 1$ and then performing an randomized rounding cannot make $x_i \notin \{0,1,\ldots,d_i\}$.) Lemmas 1 and 6 are crucial, wherein the structure of PIPs is exploited. It is essential to read this section before reading Section 5–most proofs are omitted in Section 5 since they are very similar to the ones in this section.

We solve the LP relaxation, and let the scaling by α, events $E_1, E_2, \ldots, E_{n+1}$, and vectors z, x' etc. be as in Section 1.1; α and β will be determined later on. The main point of this section is to present a good candidate for a pessimistic estimator (see (5)), and to show that it indeed satisfies the conditions of Definition 3. We may then invoke Theorem 3 to show that not only do we get improved *existential* results on the integrality gap–that we can also constructivize the existence proof. The work of this section culminates in Theorem 4.

We first setup some notation, to formulate our "failure probability". For every $j \in [m]$, let $s_j = \lfloor x'_j \rfloor$, and $p_j = x'_j - s_j \in [0,1)$. Let A_i denote the ith row of A. Let $X_1, X_2, \ldots, X_m \in \{0,1\}$ be *independent* r.v.s with $Pr(X_j = 1) = p_j$ $\forall j \in [m]$, and let $X \doteq X^{(m)}$. It is clear that

$E_i \equiv$ "$A_i \cdot X > \mu_i(1+\delta_i)$" $\forall i \in [n]$, and that $E_{n+1} \equiv$ "$c^T \cdot X < \mu_{n+1}(1-\delta_{n+1})$", where $\mu_i = E[A_i \cdot X]$ and

$$\delta_i = (b_i - A_i \cdot s)/\mu_i - 1$$

for $i \in [n]$, $\mu_{n+1} = E[c^T \cdot X]$, and

$$\delta_{n+1} = 1 - (y^*/(\alpha\beta) - \ c^T \cdot s)/\mu_{n+1}.$$

It is readily verified that $\delta_i \geq 0 \ \forall i \in [n]$ and that $0 \leq \delta_{n+1} < 1$.

Our first objective is to prove (2) and hence (3), using Theorem 2; this will then suggest potential choices for a pessimistic estimator. In the notation of Theorem 2, $N = [m]$ and $Y = \{i \in N : X_i = 1\}$. For each $i \in [n]$, define $F_i \subseteq 2^N$ as

$$\{S \subseteq N : (A_i \cdot \chi(S)) \leq \mu_i(1 + \delta_i)\}.$$

A little reflection shows the crucial property that *each F_i is monotone decreasing.* Noting that $\overline{E_i} \equiv (Y \in F_i)$ for each i, we deduce (2) from Theorem 2. In fact, a similar proof shows that since the components of X are picked *independently*, we have

Lemma 1 *For any $j \in \{0\} \cup [m]$ and any $w \in \{0,1\}^j$,*

$$Pr(\bigvee_{i=1}^{n+1} E_i | X^{(j)} = w) \leq 1 - (\prod_{i=1}^{n} (1 - Pr(E_i | X^{(j)} = w))) + Pr(E_{n+1} | X^{(j)} = w). \ \square$$

Let

$$F_{n+1} = \{S \subseteq [m] : c^T \cdot \chi(S) \geq \mu_{n+1}(1 - \delta_{n+1})\}.$$

In the notation of Definition 3, the set to be avoided, L, is

$$\{x \in \{0,1\}^m : \exists i \in [n+1] \ \ \chi^{-1}(x) \notin F_i\}.$$

We are now ready to define a suitable pessimistic estimator; we first introduce some useful notation to avoid lengthy formulae.

Notation 2 *For all $i \in [n]$, $j \in \{0\} \cup [m]$ and $w \in \{0,1\}^j$, let*

$$h_i(j, w) \doteq E[(1 + \delta_i)^{A_i \cdot X - \mu_i(1+\delta_i)} | X^{(j)} = w],$$

$$f_i(j, w) \doteq E[(1 + \delta_i)^{A_i \cdot X - \mu_i(1+\delta_i)} | X^{(j+1)} = w0], \ \textit{and}$$

$$g_i(j, w) \doteq E[(1 + \delta_i)^{A_i \cdot X - \mu_i(1+\delta_i)} | X^{(j+1)} = w1].$$

When j and w are clear from the context, we might just refer to these as h_i, f_i and g_i.

From Theorem 1 and Lemma 1, a natural guess for a pessimistic estimator, $U(u(j,w,p))$, $\forall j \in \{0\} \cup [m]$ $\forall w \in \{0,1\}^j$, might be

$$1 - (\prod_{i=1}^{n}(1 - h_i(j,w))) + E[(1-\delta_{n+1})^{c^T \cdot X - \mu_{n+1}(1-\delta_{n+1})} | X^{(j)} = w].$$

However, this might complicate matters if $h_i(j,w) > 1$ and hence we first define

$$\begin{aligned} h'_i(j,w) &= \min\{h_i(j,w), 1\}, \\ f'_i(j,w) &= \min\{f_i(j,w), 1\}, \text{ and} \\ g'_i(j,w) &= \min\{g_i(j,w), 1\}. \end{aligned}$$

We now define $U(u(j,w,p))$, $\forall j \in \{0\} \cup [m]$ $\forall w \in \{0,1\}^j$, to be

$$1 - (\prod_{i=1}^{n}(1 - h'_i(j,w))) + E[(1-\delta_{n+1})^{c^T \cdot X - \mu_{n+1}(1-\delta_{n+1})} | X^{(j)} = w]. \quad (5)$$

To make progress toward proving that U is a pessimistic estimator w.r.t. X and L, we next upper-bound $Pr(E_i)$ for each i. Recall, by Theorem 1, that for each $i \in [n]$, $Pr(E_i) \le G(\mu_i, \delta_i)$; also, $Pr(E_{n+1}) \le H(y^*/\alpha, 1 - 1/\beta)$. Lemma 2 upper-bounds these quantities.

Lemma 2 *(a) For every* $i \in [n]$,

$$G(\mu_i, \delta_i) \le G(b_i/\alpha, \alpha - 1) \le G(B/\alpha, \alpha - 1).$$

(b) $H(\mu_{n+1}, \delta_{n+1}) \le H(y^*/\alpha, 1 - 1/\beta)$.

PROOF. (a) Note that $\mu_i + A_i \cdot s \le b_i/\alpha$, with $\mu_i, A_i \cdot s \ge 0$. Subject to these constraints and that $\alpha > 1$, we will show that $G(\mu_i, \delta_i)$ is maximized when $\mu_i = b_i/\alpha$ and $A_i \cdot s = 0$; this will prove (a). Now,

$$G(\mu_i, \delta_i) = \mu_i^{b_i - A_i \cdot s} e^{-\mu_i} (e/(b_i - A_i \cdot s))^{b_i - A_i \cdot s}. \quad (6)$$

If $A_i \cdot s$ is held fixed at some $\gamma \ge 0$, (6) is maximized at $\mu_i = \Delta \doteq b_i/\alpha - \gamma$, under the constraint that $\mu_i \in [0, \Delta]$. Thus,

$$G(\mu_i, \delta_i) \le e^{b_i - b_i/\alpha}((b_i/\alpha - A_i \cdot s)/(b_i - A_i \cdot s))^{b_i - A_i \cdot s},$$

which is readily shown to be maximized when $A_i \cdot s = 0$. A similar proof holds for (b). □

Now that we have good tail bounds, we set $\alpha, \beta > 1$ such that $(\alpha\beta)$ is "small" and such that for the PIP,

$$U(p_1, p_2, \ldots, p_m) < 1$$

(property (1) of Definition 3). Note that the bound of Lemma 3 makes sense only if $B > 1$. Lemma 4 handles the common case where $A_{i,j} \in \{0,1\}\ \forall i,j$, to get improved bounds which, in particular, work even if $B = 1$. We have not attempted to optimize the constants.

Lemma 3 *There exist constants $c_1 \geq 3$ and $c_2 \geq 1$ for PIP, such that if $\alpha = c_1(c_2 n/y^*)^{1/(B-1)}$ and $\beta = 2$, then $U(p_1, p_2, \ldots, p_m) < 1$.*

PROOF. By Lemma 2, it suffices to show that $H(y^*/\alpha, 1/2) < (1 - G(B/\alpha, \alpha - 1))^n$. Furthermore, Fact 1(a) shows that

$$e^{-y^*/(8\alpha)} < (1 - e^{-B(\ln\alpha - 1)})^n$$

suffices. Now since $B \geq 1$ and $\ln\alpha - 1 \geq \ln 3 - 1 > 0$, there exists a fixed $d > 0$ such that

$$1 - e^{-B(\ln\alpha - 1)} \geq e^{-de^{-B(\ln\alpha-1)}}$$

and hence, it suffices if $y^*/(8\alpha) > nde^{-B(\ln\alpha - 1)}$. Solving for α gives the claimed bound. □

Lemma 4 *There exists a constant $c_1 \geq 3$ for PIP instances with $A_{i,j} \in \{0,1\}$ $\forall i,j$, such that if $\alpha = c_1(n/y^*)^{1/B}$ and $\beta = 2$, then $U(p_1, p_2, \ldots, p_m) < 1$.*

PROOF. Note, since $A_{i,j} \in \{0,1\}$, that for any $i \in [n]$, $((Az)_i > b_i) \rightarrow (Az)_i \geq b_i + 1$. Hence, B essentially gets replaced by $B+1$ in Lemma 3, leading to the strengthened bounds. □

As remarked in the introduction, it can be seen that the bounds (on the approximation ratio $(\alpha\beta)$) of Lemmas 3 and 4 significantly strengthen the corresponding bounds achievable by the standard analysis of randomized rounding.

At this point, we have exhibited suitable α and β such that our function U satisfies properties (1) and 2(a) of Definition 3. We now turn to proving property 2(b), which is more interesting. Before showing Lemma 6 which proves this, we first establish a simple lemma which facilitates the proof of Lemma 6.

Lemma 5 *For all $i \in [n]$, $j \in \{0\} \cup [m]$ and $w \in \{0,1\}^j$,*

(i) $0 \leq f'_i(j,w) \leq g'_i(j,w) \leq 1$, *and*

(ii) $h'_i(j,w) \geq (1 - p_{j+1})f'_i(j,w) + p_{j+1}g'_i(j,w)$.

PROOF. We drop the parameters j and w for the rest of the proof. Part (i) is easily seen. For part (ii), we first note that

$$0 \leq f_i \leq g_i \quad \text{and} \quad h_i = (1 - p_{j+1})f_i + p_{j+1}g_i, \tag{7}$$

by the definition of these quantities. Now if $h_i < 1$ and $g_i \leq 1$, then $f_i < 1$ by (7) and hence part (ii) above follows from (7), with equality. Instead if $h_i < 1$

and $g_i > 1$, note again that $f_i < 1$ and furthermore, that $g_i > g'_i = 1$; thus, part (ii) follows from (7). Finally if $h_i \geq 1$, note that $h'_i = 1$, $g'_i = 1$, and that $f'_i \leq 1$, implying (ii) again. □

Remark. In most previous constructions of pessimistic estimators for various analyses, equality actually holds in part (ii) of Lemma 5 (as opposed to our "$\geq$"). This then makes it quite easy to prove that the function on hand is a valid pessimistic estimator. Our task is made more challenging because of this (significant) change in our case.

Lemma 6 *For any $j \in \{0\} \cup [m-1]$ and for all $w \in \{0,1\}^j$,*

$$U(u(j,w,p)) \geq (1-p_{j+1})U(u(j+1,w0,p)) + p_{j+1}U(u(j+1,w1,p)).$$

Thus in particular,

$$U(u(j,w,p)) \geq min\{U(u(j+1,w0,p)), U(u(j+1,w1,p))\}.$$

Proof. Let $r \doteq p_{j+1}$, for convenience. Note that

$$\begin{aligned} E[(1-\delta_{n+1})^{c^T \cdot X - \mu_{n+1}(1-\delta_{n+1})} | X^{(j)} = w] \;\; = \;\; & (1-r)E[(1-\delta_{n+1})^{c^T \cdot X - \mu_{n+1}(1-\delta_{n+1})} | X^{(j+1)} = w0] + \\ & rE[(1-\delta_{n+1})^{c^T \cdot X - \mu_{n+1}(1-\delta_{n+1})} | X^{(j+1)} = w1]. \end{aligned}$$

Omitting the parameters j and w in f_i, g_i etc., it is thus sufficient to show that

$$\prod_{i=1}^{n}(1-h'_i) \leq (1-r)\prod_{i=1}^{n}(1-f'_i) + r\prod_{i=1}^{n}(1-g'_i).$$

Thus from Lemma 5(ii) and since $h'_i \leq 1$, it suffices to show that

$$\prod_{i=1}^{n}(1-(1-r)f'_i - rg'_i) \leq (1-r)\prod_{i=1}^{n}(1-f'_i) + r\prod_{i=1}^{n}(1-g'_i), \tag{8}$$

which we now prove by induction on n.

Equality holds in (8) for the base case $n=1$. We now prove (8) by assuming its analogue for $n-1$, *i.e.*, show that

$$((1-r)\prod_{i=1}^{n-1}(1-f'_i) + r\prod_{i=1}^{n-1}(1-g'_i))(1-(1-r)f'_n - rg'_n) \leq (1-r)\prod_{i=1}^{n}(1-f'_i) + r\prod_{i=1}^{n}(1-g'_i).$$

Simplifying, we need to show that

$$r(1-r)(g'_n - f'_n)\left(\prod_{i=1}^{n-1}(1-f'_i) - \prod_{i=1}^{n-1}(1-g'_i)\right) \geq 0, \tag{9}$$

which holds in view of Lemma 5(i). □

By now, we have fulfilled all the requirements of Definition 3 and thus present

Theorem 4 *There exist constants $c_3, c_4 > 0$ such that given any PIP conforming to the notation of Definition 1, we can produce, in deterministic polynomial time, a feasible solution to it, of value at least*

$$c_3(c_4 y^*/n^{1/B})^{B/(B-1)}.$$

If $A \in \{0,1\}^{m \times n}$, the guarantee on the solution value is at least

$$c_3(y^*/n^{1/(B+1)})^{(B+1)/B}.$$

PROOF. Lemmas 3 and 4 show property (1) of Definition 3. Properties 2(a) and 2(b) of Definition 3 are shown by Lemmas 1 and 6 respectively. Theorem 3 now completes the proof. □

4 The Maximum Independent Set Problem on Graphs

We consider the classical NP-hard problem of finding a maximum independent set (MIS) in a given undirected graph $G = (V, E)$, and pose it naturally as a packing problem. Though we do not get improved approximation algorithms for this problem, a few observations on this important problem are relevant, as we shall see shortly.

Turán's classical theorem [25] shows that G always has an independent set of size at least $|V|^2/(2|E| + |V|)$; such a set can also be found in polynomial time. The standard packing formulation described below, combined with our approach, shows the existence of an independent set of size $\Omega(|V|^2/|E|)$. The constant factor hidden in the $\Omega(\cdot)$ is weaker than that of Turán's theorem however– our reason for presenting this result is just to show that our approach proves a few other known results too, in a unified way. We remark that we do not use the standard notation of graphs having n vertices and m edges, as it will go against our notation for PIPs and CIPs–the packing formulation has $|E|$ constraints and $|V|$ variables.

Define an indicator variable $x_i \in \{0,1\}$ for each vertex i, for the presence of vertex i in the independent set (IS). Subject to the constraint that $x_i + x_j \leq 1$ for every edge (i,j), we want to maximize $\sum_i x_i$. For specific problems like this, we can get better bounds than does the analysis for Theorem 4, which uses the general CH bounds. The fractional solution $x_i^* = 1/2$ for each i, is optimal to within a factor of 2. Suppose we scale x^* down by some $\alpha > 1$ and do the randomized rounding as before. Then for any given edge (i,j), $Pr(z_i + z_j > 1) \leq 1/(4\alpha^2)$, a bound much better than the CH bound. Analysis as above then shows that $\alpha = \Theta(|E|/|V|)$ and $\beta = \Theta(1)$ suffice, thus producing an IS of size $\Omega(y^*/(\alpha\beta)) = \Omega(|V|^2/|E|)$.

One reason for our considering the MIS problem is to show that the failure probability given by (3) can be extremely close to (though strictly smaller than)

1. This would then underscore the importance of the fact that a pessimistic estimator can be constructed for PIPs and CIPs. Suppose the graph $G = (V, E)$ is a line on the N vertices $1, 2, \ldots, N$, and that each vertex independently picks a random bit for itself with the bit being one with probability q, for some $q \in [0, 1]$. Let p_N be the probability that no two adjacent vertices choose the bit "1". Setting $q = 1/(2\alpha) = \Theta(|V|/|E|) = \Theta(1)$ above, it is then clear that the probability that randomized rounding (with the above values for α and β) picks a valid IS in G, equals p_N. We now proceed to show that p_N is exponentially small in N, validating our point.

Computing p_N by induction on N is standard. Let a_N (resp., b_N) denote the probability that not only do no pair of adjacent vertices both choose "1", but also that vertex N chooses the bit "1" (resp., 0). Note that $p_N = a_N + b_N$. The recurrences

$$\begin{aligned} a_{N+1} &= q b_N \\ b_{N+1} &= (1-q)(a_N + b_N) \end{aligned}$$

are immediate. Letting $\nu \doteq \sqrt{(1-q)(1+3q)}$, it can then be seen that

$$a_N = \frac{q}{\nu}\left(\left(\frac{1-q+\nu}{2}\right)^N - \left(\frac{1-q-\nu}{2}\right)^N\right).$$

Using the facts $b_N = a_{N+1}/q$ and $p_N = a_N + b_N$, we then see that $p_N = \exp(-\Omega(N))$, *i.e.*, extremely small. Thus, the success probability of randomized rounding with our chosen values for α and β can be (and usually is) extremely small, motivating the need for a good pessimistic estimator.

The MIS problem also illustrates the well-known fact that linear relaxations are not tight in general. As seen above, this problem always has a fractional solution lying between $|V|/2$ and $|V|$. However, the graph G can have its independence number to be any integer in $[|V|]$ and hence, the integrality gap of this LP formulation can be quite bad. Furthermore, recent breakthrough work has shown that the MIS cannot be approximated fast to within any factor better than $|V|^\epsilon$ for some fixed $\epsilon > 0$, unless some unexpected containment result holds in complexity theory. This shows that we cannot expect very good approximation algorithms for all PIPs.

5 Approximating Covering Integer Programs

Given a CIP conforming to Definition 1, we show how to get a good approximation algorithm for it. Since most ideas here are very similar to those of Section 3, we borrow a lot of notation from there, skim over most details and just present the essential differences.

The idea here is to solve the LP relaxation, and for an $\alpha > 1$ to be fixed later, to set $x'_j = \alpha x^*_j$, for each $j \in [m]$. We then construct a random integral solution

z by setting, *independently* for each $j \in [m]$, $z_j = \lfloor x'_j \rfloor + 1$ with probability $x'_j - \lfloor x'_j \rfloor$, and $z_j = \lfloor x'_j \rfloor$ with probability $1 - (x'_j - \lfloor x'_j \rfloor)$. Let A_i, s_i, and $X_1, X_2, \ldots, X_m$ be as in Section 3. The bad events now are

$$E_i \equiv \text{“}A_i \cdot X < \mu_i(1-\delta_i)\text{''} \;\forall i \in [n], \text{ and}$$

$$E_{n+1} \equiv \text{“}c^T \cdot X > \mu_{n+1}(1+\delta_{n+1})\text{''},$$

where $\mu_i = E[A_i \cdot X]$ and

$$\delta_i = 1 - (b_i - A_i \cdot s)/\mu_i$$

for $i \in [n]$, $\mu_{n+1} = E[c^T \cdot X]$, and

$$\delta_{n+1} = (y^* \alpha\beta - \; c^T \cdot s)/\mu_{n+1} - 1.$$

Analogously to PIPs, $0 \le \delta_i < 1 \;\forall i \in [n]$ and $\delta_{n+1} \ge 0$.

For any $i \in [m]$, let

$$F_i = \{S \subseteq [m] : A_i \cdot \chi(S) \ge \mu_i(1-\delta_i)\}.$$

Each of these families is *monotone increasing* now, and thus Theorem 2 again guarantees Lemma 1, for the present definition of $E_1, E_2, \ldots, E_{n+1}$ also.

Suppose we define, given some $j \in \{0\} \cup [m-1]$ and some $w \in \{0,1\}^j$, $h'_i(j,w)$, $f'_i(j,w)$ and $g'_i(j,w)$ for every $i \in [n]$ analogously as in Notation 2:

$$h'_i(j,w) \doteq \min\{1, E[(1-\delta_i)^{A_i \cdot X - \mu_i(1-\delta_i)} | X^{(j)} = w]\},$$

$$f'_i(j,w) \doteq \min\{1, E[(1-\delta_i)^{A_i \cdot X - \mu_i(1-\delta_i)} | X^{(j+1)} = w0]\}, \text{ and}$$

$$g'_i(j,w) \doteq \min\{1, E[(1-\delta_i)^{A_i \cdot X - \mu_i(1-\delta_i)} | X^{(j+1)} = w1]\}.$$

As can be expected, the pessimistic estimator $U(u(j,w,p))$, $\forall j \in \{0\} \cup [m] \;\forall w \in \{0,1\}^j$, is now

$$1 - (\prod_{i=1}^{n}(1 - h'_i(j,w))) + E[(1+\delta_{n+1})^{c^T \cdot X - \mu_{n+1}(1+\delta_{n+1})} | X^{(j)} = w]. \tag{10}$$

Now for the analogue of the important Lemma 6. It is easily checked that Lemma 5(ii) holds again and that instead of part (i) of Lemma 5, we have

$$0 \le g'_i \le f'_i \le 1. \tag{11}$$

Thus, (11) guarantees (9) even now! This shows that Lemma 6 holds for the current definition of U also.

Thus to establish that U is a pessimistic estimator, we only have to exhibit, as do Lemmas 3 and 4, $\alpha, \beta > 1$ which ensure that $U(p_1, \ldots, p_m) < 1$. We first present a lemma similar to Lemma 2, whose proof is simple and omitted.

Lemma 7 *For all $i \in [n]$, $Pr(E_i) \leq H(B\alpha, 1 - 1/\alpha)$. Also, $Pr(E_{n+1}) \leq G(y^*\alpha, \beta - 1)$.*

We now present the main theorem on covering problems. Since set cover is an important problem, we present the precise approximation bound for this problem as a distinct part of the theorem.

Theorem 5 *Given a CIP conforming to the notation of Definition 1, we can produce, in deterministic polynomial time, a feasible solution to it with value at most*

$$y^*(1 + O(max\{\ln(nB/y^*)/B, \sqrt{\ln(nB/y^*)/B}\})).$$

For the unweighted set cover problem, we can improve this to

$$y^*(\ln(n/y^*) + O(\ln\ln(n/y^*)) + O(1)).$$

Proof. For general CIPs, there are two cases: $\ln(nB/y^*)/B$ is at least one or at most one. In the former case, we set $\alpha = \Theta(\ln(nB/y^*)/B)$ and $\beta = \Theta(1)$. For the latter case, we set both α and β to be of the form

$$1 + O(\sqrt{(\ln(n/y^*) + O(1))/B}).$$

The proofs follow from standard CH bound analysis using Theorem 1 and Fact 1 with Lemma 7, and the details are omitted.

For the important unweighted set cover problem (see Section 1.3 for the definition), we observe that for any $i \in [n]$, E_i holds iff $A_i \cdot s = A_i \cdot X = 0$; this makes the calculations easier. If A_i has j non-zeroes (ones) in it, say in columns $\ell_1, \ell_2, \ldots, \ell_j$, then it is not hard to see that $Pr(A_i \cdot X = 0)$ is maximized when

$$x'_{\ell_k} = \frac{\alpha}{j}, \text{ for all } k \in [j].$$

Thus,

$$Pr(E_i) \leq (1 - \alpha/j)^j < e^{-\alpha}$$

and hence, by Lemma 7, it suffices to pick $\alpha, \beta \geq 1$ such that

$$G(y^*\alpha, \beta - 1) \leq (1 - e^{-\alpha})^n. \tag{12}$$

Now since $\alpha \geq 1$, $d = e\ln(e/(e-1))$ satisfies

$$1 - e^{-\alpha} \geq e^{-de^{-\alpha}}.$$

Also if we agree to make $\beta \leq 2$, we then have

$$G(y^*\alpha, \beta - 1) \leq e^{-y^*\alpha(\beta-1)^2/3},$$

by Fact 1(b). So from (12), it suffices to choose $\alpha \geq 1$ and $1 \leq \beta \leq 2$ such that

$$nde^{-\alpha} \leq y^* \alpha(\beta - 1)^2/3. \tag{13}$$

It can now be verified that by choosing

$$\alpha = \ln(n/y^*) + a_1 \ln\ln(n/y^*) + O(1) \text{ and}$$

$$\beta = 1 + (\ln(n/y^*))^{-a_2}$$

for some suitable positive constants a_1 and a_2, we will satisfy (13). Hence, the approximation guarantee $\alpha\beta$ can be made as small as

$$\ln(n/y^*) + O(\ln\ln(n/y^*)) + O(1).$$

□

It is also worth looking at some concrete improvements brought about by Theorem 5, over existing algorithms. In the case of unweighted set cover, suppose $d \leq n$ is the maximum column sum–the maximum cardinality of any edge in the given hypergraph. Then, by just summing up all the constraints, we can see that

$$y^* d \geq n. \tag{14}$$

Thus, our approximation bound for the set cover problem–see the second statement of Theorem 5–is never more by a multiplicative $(1 + o(1))$ or an additive $O(1)$ factor above the classical bound of

$$\min\{n/y^*, \ln d + O(1)\}.$$

On the other hand, $n/y^* \ll d$ is quite likely, and it is easy to construct set cover instances with

$$\min\{n/y^*, \ln d\} = \Theta(\log n/\log\log n)\ln(n/y^*).$$

For instance, we can arrange for just a few edges to have the maximum edge size of $n^{\Theta(1)}$, while keeping y^* as high as

$$n/\log^{\Theta(1)} n.$$

Thus in the best case, we get a $\Theta(\log n/\log\log n)$ factor improvement in the approximation ratio. An important case of the unweighted set cover problem is the *dominating set* problem: given a (directed) graph G, the problem is to pick a minimum number of vertices such that for every one vertex v, at least one vertex in $v \cup Out(v)$ is picked, where $Out(v)$ denotes the out-neighborhood of v.

We next consider a more general domination-type problem on graphs, modeling a class of location problems. Given a (directed) graph G with n nodes and some integral parameter $B \geq 1$, we have to place the smallest possible number of facilities on the nodes of G, so that every node has at least B facilities in its out-neighborhood–multiple facilities at the same node are allowed.

For the case where G is *undirected* with maximum degree Δ, an approximation bound of

$$1 + O(\max\{\ln(\Delta)/B, \sqrt{\ln(\Delta)/B}\})$$

is presented in [17], improving on the

$$1 + O(\max\{\ln(n)/B, \sqrt{\ln(n)/B}\})$$

bound given by the standard analysis of randomized rounding. For us, Theorem 5 gives a bound of

$$1 + O(\max\{\ln(nB/y^*)/B, \sqrt{\ln(nB/y^*)/B}\}).$$

Even if G is directed, this new bound is as good or better than

$$1 + O(\max\{\ln(\Delta_{in})/B, \sqrt{\ln(\Delta_{in})/B}\}),$$

where Δ_{in} denotes the maximum in-degree of G; this is easily seen from the fact that

$$y^* \geq nB/\Delta_{in},$$

which follows from the same reasoning as for (14). We thus get a generalization of the Naor-Roth result. In the case of undirected graphs, it is not hard to show families of graphs for which the present bound is better than that of Naor & Roth's by a factor of upto $\Theta(\log n/\log\log n)$.

In addition to its independent interest, the above problem is a crucial subproblem in the following file-sharing problem in distributed networks [17]. Given an undirected graph G with maximum degree Δ and a file F of B bits, F must be stored in some way at the nodes of G, such that every node can recover F by examining the contents of its neighbor's memories; the aim is to minimize the total amount of memory used. (Note that solving the above domination problem is not sufficient for this task.) An approximation bound of

$$1 + O(\max\{\ln(\Delta)/B, \sqrt{\ln(\Delta)/B}\}) \tag{15}$$

is presented in [17] for this problem. Letting y^* be the optimum of the above domination problem on G, we derive an approximation bound of

$$1 + O(\max\{\ln(\Delta)/B, \sqrt{\ln(nB/y^*)/B}\}),$$

which is always as good as (15), and better if $B \gg \ln(\Delta)$.

6 Concluding Remarks

We have presented a simple but very useful property of all packing and covering integer programs–*positive correlation.* This naturally suggests a better way of analyzing the performance of randomized rounding on PIPs and CIPs. However, the provable probability of success–of satisfying all the constraints *and* delivering a very good approximation–can be extremely low; so, in itself, this approach may just prove an existential result. Fortunately, the structure of PIPs and CIPs in fact suggests a pessimistic estimator, thus converting this existence proof into a (deterministic) polynomial-time algorithm. In our view, this is very interesting, and gives evidence of the utility of de-randomization techniques. A common objection to de-randomization is that often, it converts a fast randomized algorithm that has a good probability of success, to a somewhat slower deterministic algorithm. However, note that the opposite is true here! The randomized algorithm suggested by the existence proof can have an extremely low probability of success; second, solving the LP relaxation heavily dominates the running time, and the time for running the de-randomization is comparatively negligible. (This observation about running the LP relaxation, also suggests that in practice, it would be better to quickly get an approximately optimal solution to the LP relaxation, since we are anyway dealing with approximate solutions.)

Another conclusion is that studying correlations helps; this is a well-known fact in number theory and statistical physics, for instance. In the case of PIPs and CIPs, we have benefited from the fact that the constraints "help each other", by being positively correlated. The precise reasons for such a correlation are spelled out in Section 1.2. It is a challenging open question to use the structure of correlations in more complicated scenarios; one such problem is the *set discrepancy problem* [23, 2]. Given a system of n subsets $S_1, S_2, \ldots, S_n$ of a ground set A with n elements, the problem is to come up with a function $\psi : A \rightarrow \{-1, 1\}$, such that the *discrepancy*

$$\mathrm{disc}(\psi) \doteq \max_{i \in [n]} |\psi(S_i)|$$

is "small", where

$$\psi(S_i) = \sum_{j \in S_i} \psi(j).$$

While randomized rounding and the method of conditional probabilities can be used to produce a ψ with discrepancy $O(\sqrt{n \log n})$ [23, 2], a classical *non-constructive* result of Spencer shows the existence of a ψ with $\mathrm{disc}(\psi) = O(\sqrt{n})$ [24]. This is best possible, and it is an important open problem to make this constructive. If we write down the natural integer programming formulation for this problem, we can see that each constraint is positively correlated with some subsets of the constraints, and negatively correlated with others. (There is the associated observation that in several IPs with both $\leq$ and $\geq$ constraints, the $\leq$

constraints are often positively correlated amongst each other; similarly for the $\geq$ constraints. This idea could potentially bring improvements in some cases.) It would be very interesting if such more complicated forms of correlation can be used to get a constructive result here.

Yet another potential room for improvement lies in lower-bounding, in the context of (2), the ratio

$$Pr(\bigwedge_{i=1}^{n} \overline{E_i}) / \prod_{i=1}^{n} Pr(\overline{E_i}),$$

at least for some particular classes of PIPs/CIPs. We know this ratio to be at least one, by (2); a better lower bound (at least for particular problems) will lead to better bounds on the integrality gap. Roughly speaking, such better lower bounds seem plausible especially for PIPs/CIPs wherein "several" columns have "several" nonzero entries, *i.e.*, in situations where there is heavy (positive) correlation among the constraints of the IP. This could however be a difficult problem.

How far can such ideas be pushed? In the general setting of all PIPs and CIPs, not much progress seems to be possible along these lines, as shown in Section 4. It would however be very interesting to improve our bounds for particular important problems such as for the edge-disjoint paths problem on graphs. Furthermore, it would be very interesting to study the correlations involved in other relaxation approaches such as semi-definite programming relaxations.

Finally, as we had seen before, our bounds are incomparable with known results for some *weighted* CIPs, *e.g.*, those considered in [6, 4]. It would be interesting if our method could be extended to include these results also.

Acknowledgements

We thank Moni Naor, Babu Narayanan and David Shmoys for their valuable comments. Thanks in particular to Prabhakar Raghavan for his insightful suggestions and pointers to the literature. Thanks are also due to the STOC 1995 program committee and anonymous referee(s) for their helpful comments on the written style.

References

[1] R. Aharoni, P. Erdős, and N. Linial. Optima of dual integer linear programs. *Combinatorica*, 8:13–20, 1988.

[2] N. Alon, J. H. Spencer, and P. Erdős. *The Probabilistic Method.* Wiley–Interscience Series, John Wiley & Sons, Inc., New York, 1992.

[3] M. Bellare, S. Goldwasser, C. Lund, and A. Russell. Efficient probabilistically checkable proofs and applications to approximation. In *Proc. ACM Symposium on Theory of Computing*, pages 294–304, 1993.

[4] D. Bertsimas and R. Vohra. Linear programming relaxations, approximation algorithms and randomization; a unified view of covering problems. Technical Report OR 285–94, Massachusetts Institute of Technology, 1994.

[5] H. Chernoff. A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations. *Annals of Mathematical Statistics*, 23:493–509, 1952.

[6] V. Chvátal. A greedy heuristic for the set covering problem. *Mathematics of Operations Research*, 4:233–235, 1979.

[7] P. Crescenzi and V. Kann. A compendium of NP optimization problems. Technical Report SI/RR-95/02, Department of Computer Science, University of Rome "La Sapienza", 1995.

[8] G. Dobson. Worst-case analysis of greedy heuristics for integer programming with nonnegative data. *Mathematics of Operations Research*, 7:515–531, 1982.

[9] M. L. Fisher and L. A. Wolsey. On the greedy heuristic for continuous covering and packing problems. *SIAM J. on Algebraic and Discrete Methods*, 3:584–591, 1982.

[10] C. M. Fortuin, J. Ginibre, and P. N. Kasteleyn. Correlational inequalities for partially ordered sets. *Communications of Mathematical Physics*, 22:89–103, 1971.

[11] R. L. Graham. Application of the FKG Inequality and its Relatives. In A. Bachem, M. Grötschel and B. Korte Ed., *Mathematical Programming: The State of the Art*, Springer–Verlag, 1983.

[12] W. Hoeffding. Probability inequalities for sums of bounded random variables. *American Statistical Association Journal*, 58:13–30, 1963.

[13] D. S. Johnson. Approximation algorithms for combinatorial problems. *Journal of Computer and System Sciences*, 9:256–278, 1974.

[14] L. Lovász. On the ratio of optimal integral and fractional covers. *Discrete Mathematics*, 13:383–390, 1975.

[15] C. Lund and M. Yannakakis. On the hardness of approximating minimization problems. *Journal of the ACM*, 41:960–981, 1994.

[16] R. Motwani and P. Raghavan. *Randomized Algorithms.* Cambridge University Press, 1995.

[17] M. Naor and R. M. Roth. Optimal file sharing in distributed networks. *SIAM Journal on Computing*, 24:158–183, 1995.

[18] S. A. Plotkin, D. B. Shmoys, and É. Tardos. Fast approximation algorithms for fractional packing and covering problems. In *Proc. IEEE Symposium on Foundations of Computer Science*, pages 495–504, 1991. To appear in *Mathematics of Operations Research.*

[19] P. Raghavan. Probabilistic construction of deterministic algorithms: approximating packing integer programs. *Journal of Computer and System Sciences*, 37:130–143, 1988.

[20] P. Raghavan. Randomized approximation algorithms in combinatorial optimization. In *Proc. FST & TCS Conference*, pages 300–317, 1994. Lecture Notes in Computer Science 880, Springer-Verlag, Berlin.

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

[22] T. K. Sarkar. Some lower bounds of reliability. Technical Report 124, Department of Operations Research and Statistics, Stanford University, 1969.

[23] J. H. Spencer. *Ten Lectures on the Probabilistic Method.* SIAM, Philadelphia, 1987.

[24] J. H. Spencer. Six standard deviations suffice. *Transactions of the American Mathematical Society*, 289:679–706, 1985.

[25] P. Turán. On an extremal problem in graph theory. *Matematicko Fizicki Lapok*, 48:436–452, 1941. (See, for instance, pages 81 and 82 in N. Alon, J. H. Spencer and P. Erdős, *The Probabilistic Method*, John Wiley & Sons, Inc., 1992.)