IMPROVED APPROXIMATION GUARANTEES FOR PACKING AND COVERING INTEGER PROGRAMS*

ARAVIND SRINIVASAN†

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 linear programs, 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.

Key words. approximation algorithms, combinatorial optimization, correlation inequalities, covering integer programs, derandomization, integer programming, linear programming, linear relaxations, packing integer programs, positive correlation, randomized rounding, rounding theorems

AMS subject classifications. 90C27, 68R05, 60C05, 90B80, 90C05, 90C10

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 [30]. We present an elementary property (*positive correlation*) of all these integer linear programs (ILPs); we then use the FKG inequality (Fortuin, Kasteleyn & Ginibre [14], Sarkar [31]) 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 the ith component of v. We first define the packing and covering integer programs.

DEFINITION 1.1. *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 $B = \min_i\ b_i$; we assume without loss of generality that $B \geq 1$.*

Though there are usually no restrictions on the entries of A, b and c beyond 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

*A preliminary version of this work appeared in the *Proc. ACM Symposium on the Theory of Computing*, pages 268–276, 1995.

†School of Computing, National University of Singapore, Singapore 119260, Republic of Singapore. Part of this work was done at the National University of Singapore, supported in part by National University of Singapore Academic Research Fund Grant RP970607. 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 of this work was done while at DIMACS (NSF Center for Discrete Mathematics and Theoretical Computer Science), supported by NSF grant NSF-STC91-19999 to DIMACS and by support to DIMACS from the New Jersey Commission on Science and Technology. 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). aravind@comp.nus.edu.sg

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 (in particular, we get $B \geq 1$ since $A_{i,j} \leq b_i$ for all i, j). 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, B will play the same role in the "natural" PIP/CIP formulation of the problem, as it does in Definition 1.1.

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 1.2. *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 $x_i \in \{0, 1, \ldots, d_i\}$, where d_i is some positive integer; 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 [30]. 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 [30] as follows. In what follows, let $\Pr(\mathcal{E})$ denote the probability of an event $\mathcal{E}$, and $\mathbf{E}[X]$ denote the expected value of random variable X. As seen above, an important observation is that $\mathbf{E}[z_i] = x'_i$. Hence,

$$\mathbf{E}[(Az)_i] = (Ax')_i \leq b_i/\alpha, \ 1 \leq i \leq n; \quad \mathbf{E}[c^T \cdot z] = y^*/\alpha.$$

For some $\beta > 1$ to be fixed later, define $(n+1)$ "bad" events

$$E_i \equiv ((Az)_i > b_i), \ 1 \leq i \leq n; \quad E_{n+1} \equiv (c^T \cdot z < y^*/(\alpha\beta)). \tag{1.1}$$

Now, z is an $(\alpha\beta)$–approximate solution to the PIP if

$$\bigwedge_{i=1}^{n+1} \overline{E_i} \neq \phi \tag{1.2}$$

holds. How small a value for $(\alpha\beta)$ can we achieve while still guaranteeing (1.2)? Bounding

$$\Pr(\bigvee_{i=1}^{n+1} E_i) \leq \sum_{i=1}^{n+1} \Pr(E_i), \tag{1.3}$$

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 [28]. 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 [27]. See Raghavan [29] for a survey of randomized rounding, and Crescenzi & Kann [9] for a comprehensive collection of approximation 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 [21], Lovász [23], Chvátal [8]), hypergraph matching (Aharoni, Erdős & Linial [1]) and file-sharing in distributed networks (Naor & Roth [26]), each derived through different means. (For instance, for the set cover problem which will be defined soon, let n denote the number of vertices of a given hypergraph H, and $d \leq n$ be the maximum number of vertices in any edge of H. The analysis of randomized rounding via (1.3) for appropriately defined events E_i yields an approximation of $O(\ln n)$. However, an $O(\log d)$ approximation is provided in [21], [23], [8].) 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 [28],

> 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 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 to:

$$\forall k \; \forall 1 \leq i_1 < i_2 < \ldots < i_k \leq n, \; \Pr(\bigwedge_{j=1}^{k} \overline{E_{i_j}}) \geq \prod_{j=1}^{k} \Pr(\overline{E_{i_j}}). \tag{1.4}$$

In other words, (1.4) 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 (1.4), which seems plausible, using the FKG inequality. Thus,

$$(1.5)\quad \Pr(\bigvee_{i=1}^{n+1} E_i) \leq \Pr(\bigvee_{i=1}^{n} E_i) + \Pr(E_{n+1}) \leq 1 - (\prod_{i=1}^{n}(1 - \Pr(E_i))) + \Pr(E_{n+1}),$$

which is always as good as, and most often much better than, (1.3). (For a detailed study of the FKG inequality, see, e.g., Graham [15] and Chapter 6 of Alon, Spencer & Erdős [3].)

It is not hard to verify such a property for CIPs also. Why have we been so lucky as to have positive correlation among the constraints of PIPs and CIPs? 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, 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, (1.4) may not hold if the z_is are not independent.

This approach usually only guarantees that z is a "good" approximation with very low (albeit positive) probability; it does not even seem to provide a randomized algorithm with any good success probability. However, the structure of PIPs and CIPs implies a sub-additivity property that yields a *pessimistic estimator* (a notion to be recalled in §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 [28] and in related papers) is upper-bounded by $\mathbf{E}[Z]$ (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 [21, 23, 1, 26]; our bound is incomparable with that of [8].

Positive correlation has been used in the context of network reliability by Esary & Proschan [11]; see, e.g., Chapter 4.1 in [32]. It has also been used for the set cover problem, an important covering problem to be defined below, by Bertsimas & Vohra [5].

1.3. Approximation bounds achieved. Our best improvements are for PIPs. For PIPs, the standard analysis of randomized rounding, i.e., the usage of (1.3), 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 $A \in \{0,1\}^{n\times m}$. (For completeness, we prove this in the appendix.) Our method provides $\Omega(\min\{y^*, (K_0 t_1)^{B/(B-1)}\})$ and $\Omega(\min\{y^*, t_2^{(B+1)/B}\})$ bounds resp., where $0 < K_0 < 1$ is an absolute constant, thus improving on the previous solutions. For instance, if $A \in \{0,1\}^{n\times m}$, $y^* = \Theta(n)$ 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 [39] to within a constant factor.

An important packing problem where $A \in \{0,1\}^{n \times m}$ is *simple B–matching in hypergraphs* [23]: 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

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

This matches our result to within a constant factor for $B = 1$. However, this bound *worsens* as B increases, while the standard analysis, as well as our present analysis, of randomized rounding show that the integrality gap decreases as B increases. (An interesting point is that approximation algorithms for packing/covering often have an approximation ratio that is independent of y^*. The work of [1], and our work here, are exceptions.)

For covering, we prove an

$$1 + \mathrm{O}(\max\{\ln(nB/y^*)/B, \sqrt{\ln(2\lceil nB/y^* \rceil)/B}\}) \tag{1.6}$$

integrality gap, and derive the corresponding deterministic polynomial-time approximation algorithm. (To parse the "$\ln(2\lceil t \rceil)$" term for $t > 0$, first take "t large" (say, $t \geq 2$) to be the typical case, wherein the term equals $\ln t + \Theta(1)$. If t is "small", say $t < 2$, then the term is $\Theta(1)$.) This improves on the

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

bound given by the standard analysis of randomized rounding. Also, Dobson [10] and Fisher & Wolsey [13] bound the performance of a natural greedy algorithm for CIPs; their results are as follows. For a given CIP, let $\gamma_1 = \min_i \max_j A_{i,j}/c_j$, and $\gamma_2 = \max_j (\sum_{i=1}^n A_{i,j}/c_j)$. Then, the work of [13] shows that the greedy algorithm produces a solution of value at most $y^*(1 + \ln(\gamma_2/\gamma_1))$. Alternatively, if each row of the linear system $Ax \geq b$ is scaled so that the minimum nonzero entry in the row is at least 1, then if OPT denotes the value of the optimal *integral* solution to the CIP, it is shown in [10] that the greedy algorithm produces a solution of value at most $OPT(1 + \ln(\max_j \sum_{i=1}^n A_{i,j}))$. However, more general covering problems that have constraints of the form $x_j \leq u_j$, are also considered in [10]. Our bound is incomparable with these previous bounds, but for any given (A, b, c), our bound is always better for CIP instances parametrized by $(A, \lambda b, c)$ if the scalar λ is more than a certain threshold $\mathrm{thresh}(A, b, c)$. See [5] 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 1.1; however, 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 (1.6) to derive a $\ln(n/y^*) + \ln\ln(n/y^*) + O(1)$ approximation bound. Recent work of Feige, improving earlier results of [24, 4], shows that for any fixed $\epsilon > 0$, approximating this problem to within $(1-\epsilon)\ln n$ is likely to take super-polynomial time [12]. However, this problem is important enough to study approximations parametrized by other parameters of A, b and c, that are at least as good as and often much better than, $\Theta(\log n)$; for instance, the work of [21, 23, 8] 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 $(1 + o(1))$ factor above the classical bound, and is usually much better; in the best case, our improvement is by $\Theta(\log n / \log\log n)$. (We can construct, e.g., 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 [26]; 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 [26], improving on the standard analysis of randomized rounding. Bound (1.6) improves further on this; in particular, even if G is directed with maximum in-degree Δ, (1.6) 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. First, we identify 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 § 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 contribution is in showing that the structure of PIPs and CIPs presents a suitable pessimistic estimator (see § 2 for the definition), which allows one to obtain such approximations efficiently.

In § 2, we present some basic notions such as large-deviation inequalities, the FKG inequality, and the idea of pessimistic estimators. Section 3 covers PIPs. We

devote § 4 to the important problem of finding a maximum independent set problem on graphs by looking at it in the usual way as a PIP. Its analysis shows the strengths and weaknesses of both our approach and related approaches. Section 5 deals with CIPs; a good understanding of § 3 is essential for 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. For any i, any vector $w \in \{0,1\}^i$, and for $j \in \{0,1\}$, we define $wj \in \{0,1\}^{i+1}$ as $(w_1, w_2, \ldots, w_i, j)$. As usual, e denotes the base of the natural logarithm. We let $\exp(x)$ denote e^x.

Remark. The sequel is filled with formulae and calculations, many of them routine. The real ideas of this work are contained in Lemmas 3.1, 3.5, and 3.6. The reader might 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 [7, 19]. Theorem 2.1 presents these tail bounds; see, e.g., Motwani & Raghavan [25] for the proofs.

THEOREM 2.1. *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 $\mathbf{E}[R] = \mu$. For any $\delta \geq 0$,*

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

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

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

The following fact is easily seen:

FACT 1.

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

(b) $G(\mu, \delta) \leq \exp(-\delta^2\mu/3)$ if $\delta \leq 1$.

(c) $G(\mu, \delta) \leq \exp(-(1+\delta)\ln(1+\delta)\mu/4)$ 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 2.2, a special case of the powerful FKG inequality [14, 31]; for a proof, see, e.g., Chapter 6 of [3].

THEOREM 2.2. *Given a 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,*

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

Finally, we recall the notion of *pessimistic estimators* [28]. 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 $\Pr(X^{(\ell)} \in L) < 1$. Our goal is to find some $v \in \{0,1\}^\ell - L$ efficiently. Theorem 2.4 presents the idea of pessimistic estimators applied to the method of conditional probabilities, as a means to achieve this goal. See [28] for a detailed discussion and proof. We define

DEFINITION 2.3. *A function $U : [0,1]^\ell \rightarrow \Re^+$ 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) $U(w_1, \ldots, w_i, p_{i+1}, \ldots, p_\ell) \geq \Pr(X^{(\ell)} \in L | X^{(i)} = w)$, and
(b) if $i \leq \ell - 1$, then $U(w_1, \ldots, w_i, p_{i+1}, \ldots, p_\ell)$ is at least as large as

$$\min\{U(w_1, \ldots, w_i, 0, p_{i+2}, \ldots, p_\ell), U(w_1, \ldots, w_i, 1, p_{i+2}, \ldots, p_\ell)\}.$$

THEOREM 2.4. **([28])** *Let U be a pessimistic estimator w.r.t. $(X_1, \ldots, X_\ell)$ and L. The following algorithm produces a vector $v \notin L$:*
For $i := 0$ **to** $\ell - 1$ **do**
if $U(v_1, \ldots, v_i, 0, p_{i+2}, \ldots, p_\ell) \leq U(v_1, \ldots, v_i, 1, p_{i+2}, \ldots, p_\ell)$ **then** $v_{i+1} := 0$, **else** $v_{i+1} := 1$.

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 2.3, completes the proof. □

The efficiency of the algorithm depends on how efficiently we can compute U. Also note that for the proof of Theorem 2.4, it suffices if property 2(a) of Definition 2.3 holds just for $i = \ell$.

3. Approximating Packing Integer Programs. Let a PIP be given, conforming to Definition 1.1. 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 a randomized rounding cannot make $x_i \notin \{0, 1, \ldots, d_i\}$.) Lemmas 3.1 and 3.6 are crucial, wherein the structure of PIPs is exploited. It is essential to read this section before reading § 5–some proofs are omitted in § 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 § 1.1; α and β will be determined later. The main point of this section is to show that good integral solutions exist, and to present a candidate for a pessimistic estimator (see (3.3)). We show that this estimator satisfies the conditions of Definition 2.3, and invoke Theorem 2.4 to show that we can constructivize the existence proof for the improved integrality gap. The work of this section culminates in Theorem 3.7.

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

$$E_i \equiv \text{“}A_i \cdot X > \mu_i(1 + \delta_i)\text{''} \ \forall i \in [n], \text{ and that } E_{n+1} \equiv \text{“}c^T \cdot X < \mu_{n+1}(1 - \delta_{n+1})\text{''},$$

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

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

for $i \in [n]$, $\mu_{n+1} = \mathbf{E}[c^T \cdot X]$, 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 $\delta_{n+1} \geq 0$. Also, since $c^T \cdot X \geq 0$ with probability 1, we may assume that β is not so large that $\delta_{n+1} \geq 1$: if $\delta_{n+1} \geq 1$, then $\Pr(E_{n+1}) = \Pr(c^T \cdot X < \mu_{n+1}(1-\delta_{n+1})) = 0$, and the bad event E_{n+1} will never happen. Thus, we may assume that $0 \leq \delta_{n+1} < 1$.

Our first objective is to prove (1.4) and hence (1.5), using Theorem 2.2; this will also suggest a choice for a pessimistic estimator. In the notation of Theorem 2.2, $N = [m]$ and $Y = \{i \in N : X_i = 1\}$. For each $i \in [n]$, define $F_i \subseteq 2^N$ as

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

It is easy to see that *each F_i is monotone decreasing*. Since $\overline{E_i} \equiv (Y \in F_i)$ for each i, we deduce (1.4) from Theorem 2.2. In fact, a similar proof shows that since the components of X are picked *independently*, we have

LEMMA 3.1. *For any $j \in \{0\} \cup [m]$ and any $w \in \{0,1\}^j$,*

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

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 2.3, the set to be avoided, L, is

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

We next upper-bound $\Pr(E_i)$ for each i. Recall, by Theorem 2.1, that for each $i \in [n]$, $\Pr(E_i) \leq G(\mu_i, \delta_i)$; also, $\Pr(E_{n+1}) \leq H(\mu_{n+1}, \delta_{n+1})$. Lemma 3.2 upper-bounds these quantities.

LEMMA 3.2. *(a) For every $i \in [n]$, $G(\mu_i, \delta_i) \leq G(b_i/\alpha, \alpha - 1) \leq G(B/\alpha, \alpha - 1)$. (b) $H(\mu_{n+1}, \delta_{n+1}) \leq H(y^*/\alpha, 1 - 1/\beta)$.*

Proof. (a) Note that $\mu_i + A_i \cdot s \leq b_i/\alpha$, with $\mu_i, A_i \cdot s \geq 0$. 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,

$$G(\mu_i, \delta_i) = \mu_i^{b_i - A_i \cdot s} \exp(-\mu_i)(e/(b_i - A_i \cdot s))^{b_i - A_i \cdot s}. \tag{3.1}$$

If $A_i \cdot s$ is held fixed at some $\gamma \geq 0$, (3.1) is maximized at $\mu_i = \Delta \doteq b_i/\alpha - \gamma$, under the constraint that $\mu_i \in [0, \Delta]$. Thus,

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

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, $1 - (\prod_{i=1}^{n}(1 - \Pr(E_i))) + \Pr(E_{n+1}) < 1$ holds. This, combined with setting $j = 0$ in Lemma 3.1, will show that there exists a feasible integral solution with objective function value at least $y^*/(\alpha\beta)$. Observe that the bound of Lemma 3.3

makes sense only if $B > 1$. Lemma 3.4 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 3.3. *There exist constants $K_1 \geq 3$ and $K_2 \geq 1$ such that for any PIP, if α is taken as $K_1 \cdot \max\{1, (K_2 n/y^*)^{1/(B-1)}\}$ and $\beta = 2$, then $1 - (\prod_{i=1}^{n}(1 - \Pr(E_i))) + \Pr(E_{n+1}) < 1$.*

Proof. Suppose we agree to guarantee

$$\alpha \geq 3, \tag{3.2}$$

and $\beta = 2$. By Lemma 3.2, it suffices to show that $H(y^*/\alpha, 1/2) < (1 - G(B/\alpha, \alpha - 1))^n$. Furthermore, Fact 1(a) shows that

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

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

$$1 - \exp(-B(\ln\alpha - 1)) \geq \exp(-d\exp(-B(\ln\alpha - 1)))$$

and hence, it suffices if $y^*/(8\alpha) > nd\exp(-B(\ln\alpha - 1))$. Simplifying, we see that

$$\alpha \geq K_1(K_2 n/y^*)^{1/(B-1)}$$

suffices, for certain positive absolute constants $K_1 \geq 3$ and $K_2 \geq 1$. Thus, to satisfy the requirement (3.2), it will suffice to take $\alpha = K_1 \cdot \max\{1, (K_2 n/y^*)^{1/(B-1)}\}$. □

LEMMA 3.4. *There exists a constant $K_1 \geq 3$ for PIP instances with $A_{i,j} \in \{0,1\}$ $\forall i,j$, such that if $\alpha = K_1 \cdot \max\{1, (n/y^*)^{1/B}\}$ and $\beta = 2$, then $1 - (\prod_{i=1}^{n}(1 - \Pr(E_i))) + \Pr(E_{n+1}) < 1$.*

Proof. Since $A_{i,j} \in \{0,1\}$, we may assume w.l.o.g. that each b_i is an integer, as mentioned in Definition 1.1. Again, since $A_{i,j} \in \{0,1\}$, we have, 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 3.3, leading to the strengthened bounds. (We have eliminated the constant K_2 here, since $K_2^{1/B} \leq K_2$; hence, this term can be absorbed in the constant K_1.) □

As remarked in the introduction, it can be seen that the bounds (on the integrality gap ($\alpha\beta$)) of Lemmas 3.3 and 3.4 significantly strengthen the corresponding bounds achievable by the standard analysis of randomized rounding. However, as will be seen in §4, the probability of outputting such an integral solution through randomized rounding, can be as small as $\exp(-\Omega(n+m))$; hence, showing a good bound on the integrality gap alone, is not satisfactory from the algorithmic viewpoint. We now show that there is indeed a suitable pessimistic estimator; we first introduce some notation to avoid lengthy formulae.

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

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

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

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

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 2.1 and Lemma 3.1, a natural guess for a pessimistic estimator, $U(w_1, \ldots, w_j, p_{j+1}, \ldots, p_m)$, $\forall j \in \{0\} \cup [m]$ $\forall w \in \{0,1\}^j$, might be

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

But this might complicate matters if $h_i(j,w) > 1$; so we first define $h'_i(j,w) = \min\{h_i(j,w), 1\}$, $f'_i(j,w) = \min\{f_i(j,w), 1\}$, and $g'_i(j,w) = \min\{g_i(j,w), 1\}$. We now let $U(w_1, \ldots, w_j, p_{j+1}, \ldots, p_m)$, $\forall j \in \{0\} \cup [m]$ $\forall w \in \{0,1\}^j$, to be

$$1 - (\prod_{i=1}^{n}(1 - h'_i(j,w))) + \mathbf{E}[(1-\delta_{n+1})^{c^T \cdot X - \mu_{n+1}(1-\delta_{n+1})} | X^{(j)} = w]. \tag{3.3}$$

At this point, we have exhibited suitable α and β such that our function U satisfies properties (1) and 2(a) of Definition 2.3. We now turn to proving property 2(b), the proof of which is more interesting. Before showing Lemma 3.6 which proves this, we first establish a simple lemma which facilitates the proof of Lemma 3.6.

LEMMA 3.5. *For all $i \in [n]$, $j \in \{0\} \cup [m]$ and $w \in \{0,1\}^j$,*

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

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

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

$$0 \le f_i \le g_i \quad \text{and} \quad h_i = (1 - p_{j+1}) f_i + p_{j+1} g_i, \tag{3.4}$$

by the definition of these quantities. If $h_i < 1$ and $g_i \le 1$, then $f_i < 1$ by (3.4) and hence part (ii) above follows from (3.4), with equality. 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 (3.4). Finally if $h_i \ge 1$, note that $h'_i = 1$, $g'_i = 1$, and that $f'_i \le 1$, implying (ii) again. □

Remark. In many constructions of pessimistic estimators for various analyses, equality holds in part (ii) of Lemma 3.5 (as opposed to our "$\ge$"). This makes it easy to prove that the function on hand is a valid pessimistic estimator. Our task is made more challenging because of this change in our case.

LEMMA 3.6. *$\forall j \in \{0\} \cup [m-1]$ $\forall w \in \{0,1\}^j$, $U(w_1, \ldots, w_j, p_{j+1}, p_{j+2}, \ldots, p_m) \ge (1 - p_{j+1}) U(w_1, \ldots, w_j, 0, p_{j+2}, \ldots, p_m) + p_{j+1} U(w_1, \ldots, w_j, 1, p_{j+2}, \ldots, p_m)$. Thus, in particular, $U(w_1, \ldots, w_j, p_{j+1}, p_{j+2}, \ldots, p_m)$ is at least as high as*

$$\min\{U(w_1, \ldots, w_j, 0, p_{j+2}, \ldots, p_m), U(w_1, \ldots, w_j, 1, p_{j+2}, \ldots, p_m)\}.$$

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

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

equals the sum of two terms:

$$(1 - r)\mathbf{E}[(1-\delta_{n+1})^{c^T \cdot X - \mu_{n+1}(1-\delta_{n+1})} | X^{(j+1)} = w0]$$

and

$$r\mathbf{E}[(1-\delta_{n+1})^{c^T\cdot X-\mu_{n+1}(1-\delta_{n+1})}|X^{(j+1)}=w1].$$

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 3.5(ii) and since $h_i'\leq 1$, it suffices to show that

$$\prod_{i=1}^{n}(1-(1-r)f_i'-rg_i')\leq(1-r)\prod_{i=1}^{n}(1-f_i')+r\prod_{i=1}^{n}(1-g_i'), \tag{3.5}$$

which we now prove by induction on n.

Equality holds in (3.5) for the base case $n=1$. We now prove (3.5) 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

$$r(1-r)(g_n'-f_n')\left(\prod_{i=1}^{n-1}(1-f_i')-\prod_{i=1}^{n-1}(1-g_i')\right)\geq 0, \tag{3.6}$$

which holds in view of Lemma 3.5(i). □

Since all the requirements of Definition 2.3 are satisfied, we have

THEOREM 3.7. *There exist constants $K_3, K_4 > 0$ such that given any PIP conforming to the notation of Definition 1.1, we can produce, in deterministic polynomial time, a feasible solution to it, of value at least*

$$K_3\cdot\min\{y^*,(K_4y^*/n^{1/B})^{B/(B-1)}\}.$$

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

$$K_3\cdot\min\{y^*,(y^*/n^{1/(B+1)})^{(B+1)/B}\}.$$

Proof. Lemmas 3.3 and 3.4 show property (1) of Definition 2.3. Properties 2(a) and 2(b) of Definition 2.3 are shown by Lemmas 3.1 and 3.6 respectively. Theorem 2.4 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 formulate it 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 [39] 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 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 3.7, 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 considering the MIS problem is to show that the failure probability given by (1.5) can be extremely close to (though strictly smaller than) 1. This would 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 always tight. 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 work of Håstad, building on earlier research, has shown that the MIS cannot be approximated in polynomial time to within any factor better than $|V|^{1-\epsilon}$ for any fixed $\epsilon > 0$, unless some unexpected containment results holds in computational complexity theory [17]. This shows that we cannot expect even reasonably good approximation algorithms for *all* PIPs.

5. Approximating Covering Integer Programs. Given a CIP conforming to Definition 1.1, we show how to get a good approximation algorithm for it. Since the ideas here are very similar to those of § 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 § 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 $\mu_i = \mathbf{E}[A_i \cdot X]$ and

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

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

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

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

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

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

Each of these families is *monotone increasing* now, and thus Theorem 2.2 again guarantees Lemma 3.1, 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 1:

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

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

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

As can be expected, the pessimistic estimator $U(w_1, \ldots, w_j, p_{j+1}, \ldots, p_m)$, $\forall j \in \{0\} \cup [m]$ $\forall w \in \{0,1\}^j$, is now

$$1 - (\prod_{i=1}^{n} (1 - h'_i(j,w))) + \mathbf{E}[(1+\delta_{n+1})^{c^T \cdot X - \mu_{n+1}(1+\delta_{n+1})} | X^{(j)} = w]. \tag{5.1}$$

Now for the analogue of the important Lemma 3.6. It is easily checked that Lemma 3.5(ii) holds again and that instead of part (i) of Lemma 3.5, we have

$$0 \leq g'_i \leq f'_i \leq 1. \tag{5.2}$$

Thus, (5.2) guarantees (3.6) even now! This shows that Lemma 3.6 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 3.3 and 3.4, $\alpha, \beta > 1$ which ensure that $U(p_1, \ldots, p_m) < 1$. We first present a lemma similar to Lemma 3.2, whose simple proof is omitted.

LEMMA 5.1. *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 unweighted set cover is an important problem, we present the precise approximation bound for this problem as a distinct part of the theorem. For this problem, the term "$\ln \ln(n/y^*)$" appears in our approximation bound with the understanding that we are working in the "nontrivial" range where $y^* \leq n/e$, say. Since there is always a trivial integral solution of value at most n for this problem, the case $y^* > n/e$ admits an e-approximation.

THEOREM 5.2. *(a) Given a CIP conforming to the notation of Definition 1.1, we can produce, in deterministic polynomial time, a feasible solution to it with value at most*

$$y^*(1 + \mathrm{O}(\max\{\ln(nB/y^*)/B, \sqrt{\ln(2\lceil nB/y^* \rceil)/B}\})).$$

(b) For unweighted set cover, we can improve this to $y^*(\ln(n/y^*) + \ln\ln(n/y^*) + \mathrm{O}(1))$.

Proof. We start with the simple claim that for $t \in (0, 1/e)$, $1 - t > \exp(-1.25t)$. To see this, let $u = e(1 - \ln(e-1)) = 1.246\cdots$. Consider the function $t \mapsto 1 - t - \exp(-ut)$; we now show that for $t \in (0, 1/e)$, this function is positive. The function vanishes at $t = 0$, and its derivative is $u\exp(-ut) - 1$. So as t increases from 0, the function increases initially and then decreases. Thus since the function vanishes at $t = 1/e$, it is positive for $t \in (0, 1/e)$.

Recall from the above discussion that $(1 - \max_{i \in [n]} \Pr(E_i))^n > G(y^*\alpha, \beta - 1)$ suffices. In all cases here, we will ensure that

$$\max_{i \in [n]} \Pr(E_i) < 1/e \text{ and } 1 < \beta \leq 2. \tag{5.3}$$

Thus, by the above claim and from Fact 1(b), it will suffice to show that

$$1.25n \max_{i \in [n]} \Pr(E_i) \leq y^*\alpha(\beta - 1)^2/3. \tag{5.4}$$

We now choose α and β appropriately for each of our cases.

(a) For general CIPs, we will use the bound $\max_{i \in [n]} \Pr(E_i) \leq H(B\alpha, 1 - 1/\alpha)$ of Lemma 5.1. So, assuming $H(B\alpha, 1 - 1/\alpha) < 1/e$ and $1 < \beta \leq 2$ hold, (5.4) shows that it will be enough to show

$$1.25n\exp(-B\alpha(1 - 1/\alpha)^2/2) \leq y^*\alpha(\beta - 1)^2/3. \tag{5.5}$$

For general CIPs, we will consider three cases: $\ln(nB/y^*) > B$, $3 \leq \ln(nB/y^*) \leq B$, and $\ln(nB/y^*) < 3$ (the first and third cases are not necessarily mutually exclusive).

Suppose $\ln(nB/y^*) > B$. Since $B \geq 1$, this shows that $nB/y^* \geq e$. We take $\alpha = 4\ln(nB/y^*)/B$ and $\beta = 2$. Since $\alpha \geq 4$, $(1 - 1/\alpha) \geq 3/4$; so it is easy to check that $H(B\alpha, 1 - 1/\alpha) < 1/e$. Thus in order to prove (5.5), it is enough to show

$$1.25n(y^*/(nB))^{9/8} \leq 4y^*\ln(nB/y^*)/(3B), \quad \text{i.e.,} \quad 1.25(y^*/(nB))^{1/8} \leq 4\ln(nB/y^*)/3,$$

which is true since $nB/y^* \geq e$. Thus, the approximation bound $\alpha\beta$ in this case is $\mathrm{O}(\ln(nB/y^*)/B)$.

Next, suppose $3 \leq \ln(nB/y^*) \leq B$. So $nB/y^* \geq e^3$. We choose $\alpha = 1 + 3\sqrt{\ln(nB/y^*)/B}$ and $\beta = 1 + \sqrt{\ln(nB/y^*)/B} \leq 2$. We have

$$\begin{aligned}\exp(-B\alpha(1-1/\alpha)^2/2) &= \exp(-B(\alpha-1)^2/(2\alpha)) \\ &= \exp(-9\ln(nB/y^*)/(2\alpha)) \\ &\leq \exp(-9\ln(nB/y^*)/8),\end{aligned}$$

since $\alpha \leq 4$. So we see that $\exp(-B\alpha(1-1/\alpha)^2/2) < 1/e$. Hence by (5.5), it is enough to show

$$1.25n(y^*/(nB))^{9/8} \leq y^*\alpha\ln(nB/y^*)/(3B),$$

which follows from the facts that $\alpha \geq 1$ and $nB/y^* \geq e^3$. So the approximation bound $\alpha\beta$ is $1 + \mathrm{O}(\sqrt{\ln(nB/y^*)/B})$ here.

Now consider the third case $\ln(nB/y^*) < 3$. Choose $\alpha = 1 + 10/\sqrt{B}$ and $\beta = 1 + 1/\sqrt{B} \leq 2$. Now,

$$\exp(-B\alpha(1-1/\alpha)^2/2) = \exp(-50/\alpha) \leq \exp(-50/11),$$

since $\alpha \leq 11$. So $\exp(-B\alpha(1-1/\alpha)^2/2) < 1/e$. Hence, again by (5.5), it suffices to show that

$$1.25n\exp(-50/11) \leq y^*\alpha/(3B),$$

which follows from the facts that: (i) $\alpha \geq 1$ and (ii) $nB/y^* \leq \exp(3) \leq \exp(50/11)/3.75$. So $\alpha\beta = 1 + \mathrm{O}(1/\sqrt{B})$ in this case.

(b) Thus we have proved the theorem for general CIPs. For the important unweighted set cover problem (see § 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 $x'_{\ell_k} = \alpha/j$, for all $k \in [j]$. Thus, $\Pr(E_i) \leq (1-\alpha/j)^j < \exp(-\alpha)$ and hence, by (5.4), it suffices to pick $\alpha > 1$ and $1 < \beta \leq 2$ such that

$$1.25n\exp(-\alpha) \leq y^*\alpha(\beta-1)^2/3. \tag{5.6}$$

It can now be verified that by choosing

$$\alpha = \ln(n/y^*) + \ln\ln(n/y^*) + \mathrm{O}(1) \text{ and } \beta = 1 + (\ln(n/y^*))^{-1},$$

we will satisfy (5.6). Hence, the approximation guarantee $\alpha\beta$ can be made as small as $\ln(n/y^*) + \ln\ln(n/y^*) + \mathrm{O}(1)$. □

It is also worth looking at some concrete improvements brought about by Theorem 5.2, 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

$$y^*d \geq n. \tag{5.7}$$

Thus, our approximation bound for the set cover problem–see the second statement of Theorem 5.2–is never more by a multiplicative $(1 + o(1))$ factor above the classical

bound of $\min\{n/y^*, \ln d + \mathrm{O}(1)\}$. 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 $v \cup \mathrm{Out}(v)$ is picked, where $\mathrm{Out}(v)$ 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 u has at least B facilities in $u \cup \mathrm{Out}(u)$; multiple facilities at the same node are allowed. Let us call this the B-domination problem on G.

For the case where G is *undirected* with maximum degree Δ, an approximation bound of $1 + \mathrm{O}(\max\{\ln(\Delta)/B, \sqrt{\ln(\Delta)/B}\})$ is presented in [26], improving on the $1 + \mathrm{O}(\max\{\ln(n)/B, \sqrt{\ln(n)/B}\})$ bound given by the standard analysis of randomized rounding. Our Theorem 5.2 gives a bound of

$$1 + \mathrm{O}(\max\{\ln(nB/y^*)/B, \sqrt{\ln(2\lceil nB/y^*\rceil)/B}\}).$$

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

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

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 (5.7). 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 up to $\Theta(\log n/\log\log n)$.

In addition to its independent interest, the above B-domination problem is a crucial sub-problem in the following file-sharing problem in distributed networks [26]. Given an undirected graph $G = (V, E)$ with n vertices and maximum degree Δ and a file F of k 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 and 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.) Letting y^* be the optimum of the (obvious) linear relaxation of the k-domination problem on G, a memory allocation that uses a total of

$$1 + \mathrm{O}(\max\{\ln(\Delta)/k, \sqrt{\ln(\Delta)/k}\}) \tag{5.8}$$

bits is presented in [26]. (Note that y^* is a lower bound on the total amount of memory needed.) Briefly, the approach of [26] works as follows. Let $B = k + 3\lceil \log_2 \Delta \rceil + 1$. Given a feasible integral solution (vector) $x = [x_v]_{v \in V}$ to the B-domination problem on G, the work of [26] shows how to come up with a feasible memory allocation for

F, using a total of $t = \sum_{v \in V} x_v$ bits. Recall that our bound for CIPs shows a way of achieving

$$t = y^*(B/k)(1 + \mathrm{O}(\max\{\ln(nB/y^*)/B, \sqrt{\ln(2\lceil nB/y^*\rceil)/B}\})),$$

which equals

$$y^*(1 + \mathrm{O}(\max\{\ln(\Delta)/k, \sqrt{\ln(2\lceil nk/y^*\rceil)/k}\})).$$

This is always as good as (5.8), and better if $k \gg \ln(\Delta)$.

5.1. Applications to a facility location problem. We now consider another type of canonical facility location problem, *uncapacitated facility location*. Given a directed graph $G = (V, E)$ with non-negative vertex-costs $\{d(j)\}$ and non-negative edge-lengths $\{c(i, j)\}$, let $\mathrm{Out}(v)$ denote the out-neighborhood of vertex v. Suppose it costs $d(j)$ units to locate a facility at vertex j. The problem is to place facilities on the nodes of G such that for every vertex v, at least one vertex in $\mathrm{Out}(v)$ houses a facility; the aim is to minimize the the sum of the total weight of the sites of the facilities and the total distance (in terms of edge-lengths) traveled by all the vertices to their closest facilities. By scaling, we assume w.l.o.g. that all the $d(j)$ and $c(i, j)$ lie in $[0, 1]$. $\mathrm{O}(\log n)$ approximation bounds are known for this problem (Hochbaum [18], [5]). There is a known LP relaxation for the problem to be described below. Letting $n = |V|$ and y^* be the optimal value of this relaxation, we now show how to obtain an integral solution of value $\mathrm{O}(y^*(1 + \log(\lceil n/y^*\rceil)))$.

We start with a known ILP formulation of the problem (Kolen & Tamir [22]). Let $\Delta(i) \doteq |\mathrm{Out}(i)|$. For each $i \in V$, sort the set $\{c(i, j) : j \in \mathrm{Out}(i)\}$ in non-decreasing order to obtain a sequence $c'(i, 1), c'(i, 2), \ldots, c'(i, \Delta(i))$. Following [22], consider the following variables. For all $i \in V$ and all $j \leq \Delta(i) - 1$, let

- $x_i \in \{0, 1\}$ be the indicator variable for a facility being placed at i, and
- $z_{i,j} \in \{0, 1\}$ be the indicator variable for no facility being at a distance of at most $c'(i, j)$ from i.

Then, it is not hard to see that the following ILP formulation of [22] is valid:

$$\text{minimize } \sum_{j \in V} d_j x_j + \sum_{i \in V} (c'(i, 1) + \sum_{j=1}^{\Delta(i)-1} (c'(i, j + 1) - c'(i, j)) z_{i,j}) \text{ subject to}$$

$$\forall i \in V \ \forall k \leq \Delta(i) - 1, \quad z_{i,k} + \sum_{j:\ (i,j) \in E,\ c(i,j) \leq c'(i,k)} x_j \geq 1; \tag{5.9}$$

$$\forall i \in V, \quad \sum_{j:\ (i,j) \in E} x_j \geq 1; \tag{5.10}$$

$$\forall i \in V \ \forall k \leq \Delta(i) - 1, \quad x_i \in \{0, 1\}, \quad z_{i,k} \in \{0, 1\}. \tag{5.11}$$

Relaxing each x_i and $z_{i,j}$ to be a real in $[0, 1]$, we get an LP relaxation. Let y^* denote the optimum value of this relaxation, and $\{x^*, z^*\}$ be an optimal solution to the relaxation. We now show how to round this to a feasible solution for the above ILP. Let $m = |E|$. Note that a direct application of our covering results will lead to an integral solution of value $\mathrm{O}(y^*(1 + \log(\lceil m/y^*\rceil)))$, since there are $\mathrm{O}(m)$ constraints above. We now improve this to $\mathrm{O}(y^*(1 + \log(\lceil n/y^*\rceil)))$.

For each (i, j), set $z'_{i,j} := 0$ if $z^*_{i,j} < 1/2$, and $z'_{i,j} := 1$ if $z^*_{i,j} \geq 1/2$. For each i, set $x'_i = \min\{2x^*_i, 1\}$. It is not hard to check that this is a feasible fractional solution

to the LP with objective function value at most $2y^*$; crucially, we have rounded all the $z_{i,j}$ to 0 and 1. Let us see how to round the vector x' to a $\{0,1\}$-vector. Fix $i \in V$. It is easily seen that $z^*_{i,1} \geq z^*_{i,2} \geq \cdots$. Let $k(i)$ be the maximum k for which $z^*_{i,k} \geq 1/2$ (if no such k exists, $k(i) = 0$); for all $k \leq k(i)$, (5.9) will anyway be satisfied irrespective of our rounding of x', since $z'_{i,k} = 1$. So it is not hard to check that in order to satisfy (5.9) and (5.10), it suffices to have

$$\sum_{j:\ (i,j)\in E,\ c(i,j)\leq c'(i,k(i)+1)} x_j \geq 1.$$

Thus, we are left with at most one constraint per *vertex* i. Moreover, the fractional vector x' satisfies these constraints as seen above. So, we can invoke our approach for CIPs to efficiently construct a feasible integral solution of value $O(y^*(1+\log(\lceil n/y^* \rceil)))$.

The reader is referred to [18, 5] for other interesting approaches to the problem. We also mention that for the important special case where the distances $c(i,j)$ are symmetric (i.e., $c(i,j) = c(j,i)$) and satisfy the triangle inequality, work of Shmoys, Tardos & Aardal [33] and Guha & Khuller [16] has led to good *constant-factor* approximations for this problem.

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. In our case, the opposite is true! The randomized algorithm suggested by the existence proof has 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 dealing with approximate solutions anyway.)

The basic idea here has been that by using the correlations, we could improve on the Boole-Bonferroni inequality for our applications. There is much interesting work on approximating the probability of a union using the structure of the underlying events: see, e.g., Hunter [20] and Aldous [2]. 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 § 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* [35, 3]. 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 $O(\sqrt{n \log n})$ [35, 3], a classical *non-constructive* result of Spencer shows the existence of a ψ with $\mathrm{disc}(\psi) = O(\sqrt{n})$ [36]. 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.

The set discrepancy problem, an undirected multicommodity flow problem presented in [28], and certain other NP-hard problems in VLSI routing, are all modeled by a class of integer programs that are referred to as *minimax integer programs* (MIPs) in recent work of this author [37]. These are different from PIPs and CIPs; an improved integrality gap for sparse instances of such problems is proved in [37].

Yet another potential room for improvement lies in lower-bounding, in the context of (1.4), 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 (1.4); 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 § 4. Also, for some important covering problems such as various "cut" problems on graphs and some network design problems, the number of constraints is *exponential* in the number of variables. So, our approach will not give good approximation algorithms for such problems, though some of these problems do admit good approximations via other methods [5]. It would be interesting to apply our approach to get improved approximations for some specific important packing and covering problems; one such recent result due to the author is for edge-disjoint paths and related problems [38]. Furthermore, it will be interesting to study the correlations involved in other relaxation approaches such as semi-definite programming relaxations.

For the unweighted set cover problem, recent work of Slavík improves upon our result, showing that a natural greedy algorithm outputs a solution of value at most $(y^* - 1/2)\ln(n/y^*) + y^*$ [34]. Furthermore, improved approximation algorithms for the unweighted set cover problem in many geometric settings, have been obtained in [6]. If the dual set-system of a given set cover instance has VC-dimension d and if c is the size of the optimal set cover, then the work of [6] presents an algorithm that delivers a solution of value at most $O(c \log(cd))$. Please refer to [6] for the definitions of dual set systems and VC-dimension. We just mention here that $d = O(1)$ in many geometric settings; thus, the work of [6] leads to an approximation guarantee of $O(\log c)$ in such cases.

As we had seen before, our bounds are incomparable with known results for some *weighted* CIPs, e.g., those considered in [8, 5]. For instance, our approximation ratio of $O(\ln(n/y^*))$ for the weighted set cover problem is incomparable with the ratio of $O(\ln d)$ obtained in [8], where d denotes the maximum column sum in A; this is

because unlike the unweighted case, n/y^* is not necessarily at most d, in the weighted case. Recent work of the author improves on many current results, including some of the results on *weighted* PIPs and CIPs from the present paper [37]. The work of [37] presents improved approximation guarantees for PIPs, CIPs, and MIPs, if the coefficient matrix A does not have "many" nonzero entries in any column; in particular, it generalizes the above-seen result of [8].

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. We acknowledge the valuable feedback of the STOC 1995 program committee and referee(s). We also thank the journal referees for their very helpful comments and suggestions, which have substantially improved the quality of this paper.

REFERENCES

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

[2] D. Aldous, *The Harmonic Mean Formula for probabilities of unions: applications to sparse random graphs*, Discrete Mathematics, 76 (1989), pp. 167–176.

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

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

[5] 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.

[6] H. Brönnimann and M. T. Goodrich, *Almost optimal set covers in finite VC-dimensions*, Discrete & Computational Geometry, 14 (1995), pp. 463–479.

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

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

[9] 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.

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

[11] J. D. Esary and F. Proschan, *Coherent structures of non-identical components*, Technometrics, 5 (1963), pp. 191–209.

[12] U. Feige, *A threshold of* $\ln n$ *for approximating set cover*, Proc. ACM Symposium on Theory of Computing, 1996, pp. 314–318.

[13] 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 (1982), pp. 584–591.

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

[15] 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.

[16] S. Guha and S. Khuller, *Greedy strikes back: improved facility location algorithms*, Proc. ACM-SIAM Symposium on Discrete Algorithms, 1998, pp. 649–657.

[17] J. Håstad, *Clique is hard to approximate within* $n^{1-\epsilon}$, Proc. IEEE Symposium on Foundations of Computer Science, 1996, pp. 627–636.

[18] D. S. Hochbaum, *Heuristics for the fixed cost median problem*, Mathematical Programming, 22 (1982), pp. 148–162.

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

[20] D. Hunter, *An upper bound for the probability of a union*, Journal of Applied Probability, 13 (1976), pp. 597–603.

[21] D. S. Johnson, *Approximation algorithms for combinatorial problems*, Journal of Computer and System Sciences, 9 (1974), pp. 256–278.
[22] A. Kolen and A. Tamir, *Covering Problems*, In Discrete Location Theory (P. B. Mirchandani and R. L. Francis, eds.), Wiley–Interscience Series, John Wiley & Sons, Inc., New York, 1990, pp. 263–304.
[23] L. Lovász, *On the ratio of optimal integral and fractional covers*, Discrete Mathematics, 13 (1975), pp. 383–390.
[24] C. Lund and M. Yannakakis, *On the hardness of approximating minimization problems*, Journal of the ACM, 41 (1994), pp. 960–981.
[25] R. Motwani and P. Raghavan, *Randomized Algorithms*, Cambridge University Press, 1995.
[26] M. Naor and R. M. Roth, *Optimal file sharing in distributed networks*, SIAM Journal on Computing, 24 (1995), pp. 158–183.
[27] S. A. Plotkin, D. B. Shmoys, and É. Tardos, *Fast approximation algorithms for fractional packing and covering problems*, Mathematics of Operations Research, 20 (1995), pp. 257–301.
[28] P. Raghavan, *Probabilistic construction of deterministic algorithms: approximating packing integer programs*, Journal of Computer and System Sciences, 37 (1988), pp. 130–143.
[29] P. Raghavan, *Randomized approximation algorithms in combinatorial optimization*, Proc. FST & TCS Conference, 1994, pp. 300–317, Lecture Notes in Computer Science 880, Springer-Verlag, Berlin.
[30] P. Raghavan and C. D. Thompson, *Randomized rounding: a technique for provably good algorithms and algorithmic proofs*, Combinatorica, 7 (1987), pp. 365–374.
[31] T. K. Sarkar, *Some lower bounds of reliability*, Technical Report 124, Department of Operations Research and Statistics, Stanford University, 1969.
[32] D. R. Shier, *Network Reliability and Algebraic Structures*, Oxford Science Publications, Oxford University Press, 1991.
[33] D. B. Shmoys, É. Tardos, and K. Aardal, *Approximation algorithms for facility location problems*, Proc. ACM Symposium on Theory of Computing, 1997, pp. 265–274.
[34] P. Slavík, *A tight analysis of the greedy algorithm for set cover*, Proc. ACM Symposium on Theory of Computing, 1996, pp. 435–441.
[35] J. H. Spencer, *Ten Lectures on the Probabilistic Method*, SIAM, Philadelphia, 1987.
[36] J. H. Spencer, *Six standard deviations suffice*, Transactions of the American Mathematical Society, 289 (1985), pp. 679–706.
[37] A. Srinivasan, *An extension of the Lovász Local Lemma, and its applications to integer programming*, Proc. ACM-SIAM Symposium on Discrete Algorithms, 1996, pp. 6–15.
[38] A. Srinivasan, *Improved approximations for edge-disjoint paths, unsplittable flow, and related routing problems*, Proc. IEEE Symposium on Foundations of Computer Science, 1997, pp. 416–425.
[39] P. Turán, *On an extremal problem in graph theory*, Matematicko Fizicki Lapok, 48 (1941), pp. 436–452. (See, for instance, pages 81 and 82 in N. Alon, J. H. Spencer and P. Erdős, *The Probabilistic Method*, John Wiley & Sons, Inc., 1992.)

Appendix

Appendix A. Standard analysis of randomized rounding for PIPs. Choose $\alpha = e(5n)^{1/B}$. We first handle an easy case. If $y^* \leq 3\alpha$, we just choose any j such that $c_j = 1$; set $x_j = 1$ and $x_k = 0$ for all $k \neq j$. This is clearly an $O(n^{1/B})$-approximation algorithm if $y^* \leq 3\alpha$.

So suppose $y^* > 3\alpha$. Let the bad events $E_1, E_2, \ldots, E_{n+1}$ be as in (1.1), with $\beta = 3$. If we avoid all these $(n+1)$ events, we would have a feasible solution with objective function value at least $y^*/(3\alpha)$. The goal now is to show

$$\sum_{i \in [n+1]} \Pr(E_i) \leq K_4, \tag{A.1}$$

for say some constant $K_4 < 1$.

For the purpose of analysis, view each z_j as a sum of $\lceil x'_j \rceil$–many independent $\{0,1\}$ r.v.s, where the first $\lceil x'_j \rceil - 1$ are 1 with probability 1, and the last is 1 with

probability $1 - (\lceil x'_j \rceil - x'_j)$. Recalling that $A_{i,j} \in [0,1]$ for all i, j, we see, for each $i \in [n]$, that $(Az)_i$ is a sum of independent r.v.s each of which takes on values in $[0,1]$. Also, $\mathbf{E}[(Az)_i] = (Ax')_i/\alpha \leq b_i/\alpha$.

Thus, for $i \in [n]$, $\Pr(E_i)$ can be upper-bounded using part (a) of Fact 1:

$$\Pr((Az)_i > b_i) \leq (e/\alpha)^{b_i} = (5n)^{-b_i/B} \leq 1/(5n), \tag{A.2}$$

since $b_i \geq B$. To upper-bound $\Pr(E_{n+1})$, even Chebyshev's inequality will suffice here. If Y is a sum of independent r.v.s, each taking values in $[0,1]$, it is easy to check that the variance of Y is at most $\mathbf{E}[Y]$. Thus, Chebyshev's inequality shows that for such a Y and any $a > 0$, $\Pr(|Y - \mathbf{E}[Y]| \geq a) \leq \mathbf{E}[Y]/a^2$. Since $y^* > 3\alpha$ now by assumption, we have $\mu \doteq \mathbf{E}[c^T \cdot z] = y^*/\alpha \geq 3$. So by Chebyshev's inequality,

$$\Pr(c^T \cdot z < \mu/3) \leq \Pr(|c^T \cdot z - \mu| > 2\mu/3] \leq 9/(4\mu) \leq 3/4. \tag{A.3}$$

Bounds (A.2) and (A.3) show that

$$\sum_{i \in [n+1]} \Pr[E_i] \leq 1/5 + 3/4 = 0.95;$$

by repeating this algorithm an appropriate constant number of times, the probability of failure can be reduced to any desired positive constant.

Next suppose $A \in \{0,1\}^{n \times m}$. Recall from Definition 1.1 that we take each b_i to be an integer here. Thus, $E_i \equiv ((Az)_i > b_i)$ is equivalent to $((Az)_i \geq b_i + 1)$ now; so B essentially gets replaced by $B + 1$ in (A.2). We may thus take $\alpha = \Theta(n^{1/(B+1)})$ and get an $O(\alpha)$ approximation as above.