An Extension of the Lovász Local Lemma, and its Applications to Integer Programming*

Aravind Srinivasan†

Abstract

The Lovász Local Lemma due to Erdős and Lovász (in *Infinite and Finite Sets*, Colloq. Math. Soc. J. Bolyai 11, 1975, pp. 609–627) is a powerful tool in proving the existence of rare events. We present an extension of this lemma, which works well when the event to be shown to exist is a conjunction of individual events, each of which asserts that a random variable does not deviate much from its mean. As applications, we consider two classes of NP-hard integer programs: minimax and covering integer programs. A key technique, *randomized rounding of linear relaxations*, was developed by Raghavan & Thompson (*Combinatorica, 7 (1987), pp. 365–374*) to derive good approximation algorithms for such problems. We use our extension of the Local Lemma to prove that randomized rounding produces, with non-zero probability, much better feasible solutions than known before, if the constraint matrices of these integer programs are *column-sparse* (e.g., routing using short paths, problems on hypergraphs with small dimension/degree). This complements certain well-known results from discrepancy theory. We also generalize the *method of pessimistic estimators* due to Raghavan (*J. Computer and System Sciences, 37 (1988), pp. 130–143*), to obtain constructive (algorithmic) versions of our results for covering integer programs.

Key Words and Phrases. Lovász Local Lemma, column-sparse integer programs, approximation algorithms, randomized rounding, discrepancy

1 Introduction

The powerful Lovász Local Lemma (LLL) is often used to show the existence of *rare* combinatorial structures by showing that a random sample from a suitable sample space produces them with positive probability [14]; see Alon & Spencer [4] and Motwani & Raghavan [27] for several such applications. We present an extension of this lemma, and demonstrate applications to rounding fractional solutions for certain families of integer programs. A preliminary version of this work appeared in [35], with a sketch of the proof for minimax integer programs, and proofs omitted for our constructive results for covering integer programs. In this version, we provide all proofs, further generalize the main covering result of [35] to Theorem 5.9, and present applications thereof; in particular, the constructive approach to covering integer programs detailed in §5.3, requires a fair amount of work.

Let e denote the base of natural logarithms. The symmetric case of the LLL shows that all of a set of "bad" events E_i can be avoided under some conditions:

*A preliminary version of this work appeared as a paper of the same title in the *Proc. ACM-SIAM Symposium on Discrete Algorithms*, pages 6–15, 1996. Work done in parts at: (i) the University of Maryland (supported in part by NSF Award CCR-0208005 and NSF ITR Award CNS-0426683), (ii) the National University of Singapore, (iii) DIMACS (supported in part by NSF-STC91-19999 and by support from the N.J. Commission on Science and Technology), (iv) the Institute for Advanced Study, Princeton, NJ (supported in part by grant 93-6-6 of the Alfred P. Sloan Foundation), and (v) while visiting the Max-Planck-Institut für Informatik, 66123 Saarbrücken, Germany.

†Department of Computer Science and Institute for Advanced Computer Studies, University of Maryland, College Park, MD 20742, USA. E-mail: srin@cs.umd.edu.

Lemma 1.1. **([14])** *Let $E_1, E_2, \ldots, E_m$ be any events with $\Pr(E_i) \leq p\ \forall i$. If each E_i is mutually independent of all but at most d of the other events E_j and if $ep(d+1) \leq 1$, then $\Pr(\bigwedge_{i=1}^{m} \overline{E_i}) > 0$.*

Though the LLL is powerful, one problem is that the "dependency" d is high in some cases, precluding the use of the LLL if p is not small enough. We present a partial solution to this via an extension of the LLL (Theorem 3.1), which shows how to essentially reduce d for a class of events E_i; this works well when each E_i denotes some random variable deviating "much" from its mean. In a nutshell, we show that such events E_i can often be decomposed suitably into sub-events; although the sub-events may have a large dependency among themselves, we show that it suffices to have a small "bipartite dependency" between the set of events E_i and the set of sub-events. This, in combination with some other ideas, leads to the following applications in integer programming.

It is well-known that a large number of NP-hard combinatorial optimization problems can be cast as integer linear programming problems (ILPs). Due to their NP-hardness, good approximation algorithms are of much interest for such problems. Recall that a ρ-approximation algorithm for a minimization problem is a polynomial-time algorithm that delivers a solution whose objective function value is at most ρ times optimal; ρ is usually called the *approximation guarantee*, *approximation ratio*, or *performance guarantee* of the algorithm. Algorithmic work in this area typically focuses on achieving the smallest possible ρ in polynomial time. One powerful paradigm here is to start with the linear programming (LP) relaxation of the given ILP wherein the variables are allowed to be *reals* within their integer ranges; once an optimal solution is found for the LP, the main issue is how to *round* it to a good feasible solution for the ILP. Rounding results in this context often have the following strong property: they present an integral solution of value at most $y^* \cdot \rho$, where y^* will throughout denote the optimal solution value of the LP relaxation. Since the optimal solution value OPT of the ILP is easily seen to be lower-bounded by y^*, such rounding algorithms are also ρ-approximation algorithms. Furthermore, they provide an upper bound of ρ on the ratio OPT/y^*, which is usually called the *integrality gap* or *integrality ratio* of the relaxation; the smaller this value, the better the relaxation.

This work presents improved upper bounds on the integrality gap of the natural LP relaxation for two families of ILPs: minimax integer programs (MIPs) and covering integer programs (CIPs). (The precise definitions and results are presented in § 2.) For the latter, we also provide the corresponding polynomial-time rounding algorithms. Our main improvements are in the case where the coefficient matrix of the given ILP is *column-sparse*: i.e., the number of nonzero entries in every column is bounded by a given parameter a. There are classical rounding theorems for such column-sparse problems (e.g., Beck & Fiala [6], Karp, Leighton, Rivest, Thompson, Vazirani & Vazirani [18]). Our results complement, and are incomparable with, these results. Furthermore, the notion of column-sparsity, which denotes no variable occurring in "too many" constraints, occurs naturally in combinatorial optimization: e.g., routing using "short" paths, and problems on hypergraphs with "small" degree. These issues are discussed further in § 2.

A key technique, *randomized rounding of linear relaxations*, was developed by Raghavan & Thompson [32] to get approximation algorithms for such ILPs. We use Theorem 3.1 to prove that this technique produces, with non-zero probability, much better feasible solutions than known before, if the constraint matrix of the given MIP/CIP is column-sparse. (In the case of MIPs, our algorithm iterates randomized rounding several times with different choices of parameters, in order to achieve our result.) Such results cannot be got via Lemma 1.1, as the dependency d, in the sense of Lemma 1.1, can be as high as $\Theta(m)$ for these problems. Roughly speaking, Theorem 3.1 helps show that if no column in our given ILP has more than a nonzero entries, then the dependency can essentially be brought down to a polynomial in a; this is the key driver behind our improvements.

Theorem 3.1 works well in combination with an idea that has blossomed in the areas of derandomization and pseudorandomness, in the last two decades: (approximately) decomposing a function of several variables into a sum of terms, each of which depends on only a few of these variables. Concretely, suppose Z is a sum of random variables Z_i. Many tools have been developed to upper-bound $\Pr(Z - \mathbf{E}[Z] \geq z)$

and $\Pr(|Z - \mathbf{E}[Z]| \geq z)$ even if the Z_is are only (almost) k-wise independent for some "small" k, rather than completely independent. The idea is to bound the probabilities by considering $\mathbf{E}[(Z - \mathbf{E}[Z])^k]$ or similar expectations, which look at the Z_i *k or fewer at a time* (via linearity of expectation). The main application of this has been that the Z_i can then be sampled using "few" random bits, yielding a derandomization/pseudorandomness result (e.g., [3, 23, 8, 26, 28, 33]). Our results show that such ideas can in fact be used to show that some structures exist! This is one of our main contributions.

What about polynomial-time algorithms for our existential results? Typical applications of Lemma 1.1 are "nonconstructive" [i.e., do not directly imply (randomized) polynomial-time algorithmic versions], since the positive probability guaranteed by Lemma 1.1 can be exponentially small in the size of the input. However, certain algorithmic versions of the LLL have been developed starting with the seminal work of Beck [5]. These ideas do not seem to apply to our extension of the LLL, and hence our MIP result is nonconstructive. Following the preliminary version of this work [35], two main algorithmic versions related to our work have been obtained: (i) for a subclass of the MIPs [20], and (ii) for a somewhat different notion of approximation than the one we study, for certain families of MIPs [11].

Our main algorithmic contribution is for CIPs and multi-criteria versions thereof: we show, by a generalization of the *method of pessimistic estimators* [31], that we can efficiently construct the same structure as is guaranteed by our nonconstructive argument. We view this as interesting for two reasons. First, the generalized pessimistic estimator argument requires a quite delicate analysis, which we expect to be useful in other applications of developing constructive versions of existential arguments. Second, except for some of the algorithmic versions of the LLL developed in [24, 25], most current algorithmic versions minimally require something like "$pd^3 = O(1)$" (see, e.g., [5, 1]); the LLL only needs that $pd = O(1)$. While this issue does not matter much in many applications, it crucially does, in some others. A good example of this is the existentially-optimal integrality gap for the edge-disjoint paths problem with "short" paths, shown using the LLL in [21]. The above-seen "$pd^3 = O(1)$" requirement of currently-known algorithmic approaches to the LLL, leads to algorithms that will violate the edge-disjointness condition when applied in this context: specifically, they may route up to three paths on some edges of the graph. See [9] for a different – random-walk based – approach to low-congestion routing. An algorithmic version of this edge-disjoint paths result of [21] is still lacking. It is a very interesting open question whether there is an algorithmic version of the LLL that can construct the same structures as guaranteed to exist by the LLL. In particular, can one of the most successful derandomization tools – the method of conditional probabilities or its generalization, the pessimistic estimators method – be applied, fixing the underlying random choices of the probabilistic argument one-by-one? This intriguing question is open (and seems difficult) for now; it is elaborated upon further in § 6. As a step in this direction, we are able to show how such approaches can indeed be developed, in the context of CIPs.

Thus, our main contributions are as follows. (a) The LLL extension is of independent interest: it helps in certain settings where the "dependency" among the "bad" events is too high for the LLL to be directly applicable. We expect to see further applications/extensions of such ideas. (b) This work shows that certain classes of column-sparse ILPs have much better solutions than known before; such problems abound in practice (e.g., short paths are often desired/required in routing). (c) Our generalized method of pessimistic estimators should prove fruitful in other contexts also; it is a step toward complete algorithmic versions of the LLL.

A note on reading this paper. In order to get the main ideas of this work across, we summarize the main ideas of each section and some of the subsections, at their beginning. We also have paragraphs marked "Intuition" in some of the sections, in order to spell out the main ideas informally. Furthermore, we have moved simple-but-tedious calculations to the Appendix; it may help the reader to read the body of the paper first, and to then read the material in the Appendix if necessary. The rest of this paper is organized as follows. Our results are first presented in § 2, along with a discussion of related work. The extended LLL, and some large-deviation methods that will be seen to work well with it, are shown in § 3.

Sections 4 and 5 are devoted to our rounding applications. Section 6 concludes, and some of the routine calculations are presented in the Appendix.

2 Our Results and Related Work

Let Z_+ denote the set of non-negative integers; for any $k \in Z_+$, $[k] \doteq \{1, \ldots, k\}$. "Random variable" is abbreviated by "r.v.", and logarithms are to the base 2 unless specified otherwise.

Definition 2.1. ***(Minimax Integer Programs)*** *An MIP (minimax integer program) has the following variables: $\{x_{i,j} : i \in [n], j \in [\ell_i]\}$, for some integers $\{\ell_i\}$, and an extra variable W. Let $N = \sum_{i \in [n]} \ell_i$ and let x denote the N-dimensional vector of the variables $x_{i,j}$ (arranged in any fixed order). An MIP seeks to minimize W, an unconstrained real, subject to:*

(i) Equality constraints: $\forall i \in [n]$ $\sum_{j \in [\ell_i]} x_{i,j} = 1$;

(ii) a system of linear inequalities $Ax \leq \vec{W}$, where $A \in [0,1]^{m \times N}$ and $\vec{W}$ is the m-dimensional vector with the variable W in each component, and

(iii) Integrality constraints: $x_{i,j} \in \{0,1\}$ $\forall i, j$.

We let g denote the maximum column sum in any column of A, and a be the maximum number of non-zero entries in any column of A.

To see what problems MIPs model, note, from constraints (i) and (iii) of MIPs, that for all i, any feasible solution will make the set $\{x_{i,j} : j \in [\ell_i]\}$ have precisely one 1, with all other elements being 0; MIPs thus model many "choice" scenarios. Consider, e.g., global routing in VLSI gate arrays [32]. Given are an undirected graph $G = (V, E)$, a function $\rho : V \rightarrow V$, and $\forall i \in V$, a set P_i of paths in G, each connecting i to $\rho(i)$; we must connect each i with $\rho(i)$ using exactly one path from P_i, so that the maximum number of times that any edge in G is used for, is minimized–an MIP formulation is obvious, with $x_{i,j}$ being the indicator variable for picking the jth path in P_i. This problem, the vector-selection problem of [32], and the discrepancy-type problems of Section 4, are all modeled by MIPs; many MIP instances, e.g., global routing, are NP-hard.

We now introduce the next family of integer programs that we will work with.

Definition 2.2. ***(Covering Integer Programs)*** *Given $A \in [0,1]^{m \times n}$, $b \in [1, \infty)^m$ and $c \in [0,1]^n$ with max_j $c_j = 1$, a covering integer program (CIP) seeks to minimize $c^T \cdot x$ subject to $x \in Z_+^n$ and $Ax \geq b$. If $A \in \{0,1\}^{m \times n}$, each entry of b is assumed integral. We define $B = \min_i\ b_i$, and let a be the maximum number of non-zero entries in any column of A. A CIP is called unweighted if $c_j = 1$ $\forall j$, and weighted otherwise.*

Note the parameters g, a and B of definitions 2.1 and 2.2. Though there are usually no restrictions on the entries of A, b and c in CIPs aside of non-negativity, it is well-known and easy to check that the above restrictions are without loss of generality. CIPs again model many NP-hard problems in combinatorial optimization. Recall that a hypergraph $H = (V, E)$ is a family of subsets E (edges) of a set V (vertices). The classical *set cover* problem–covering V using the smallest number of edges in E (and its natural weighted version) is a standard example of a CIP. The parameter a here is the maximum number of vertices in any edge.

Next, there is growing interest in *multi-criteria* optimization, since different participating individuals and/or organizations may have different objective functions in a given problem instance; see, e.g., [29]. Motivated by this, we study multi-criteria optimization in the setting of covering problems:

Definition 2.3. ***(Multi-criteria CIPs; informal)*** *A multi-criteria CIP has a system of constraints "$Ax \geq b$" as in CIPs, and has several given non-negative vectors $c_1, c_2, \ldots, c_\ell$; the aim is to keep all the values $c_i^T \cdot x$ "low". (For instance, we may aim to minimize $\max_i c_i^T \cdot x$ subject to $Ax \geq b$.) As in Definition 2.2, we assume that $A \in [0,1]^{m\times n}$, $b \in [1,\infty)^m$ and for all i, $c_i \in [0,1]^n$ with $\max_j c_{i,j} = 1$.*

We now present a lemma to quantify our approximation results; its proof is given in §3.

Lemma 2.4. *Given independent r.v.s $X_1, \ldots, X_n \in [0,1]$, let $X = \sum_{i=1}^n X_i$ and $\mu = \boldsymbol{E}[X]$.*

a. For any $\delta > 0$, $\Pr(X \geq \mu(1+\delta)) \leq G(\mu,\delta)$, where $G(\mu,\delta) = \left(e^\delta/(1+\delta)^{1+\delta}\right)^\mu$.

b. $\forall \mu > 0\ \forall p \in (0,1)$, $\exists \delta = H(\mu,p) > 0$ such that $\lceil \mu\delta \rceil \cdot G(\mu,\delta) \leq p$ and such that

$$H(\mu,p) = \Theta\left(\frac{\log(p^{-1})}{\mu \log(\log(p^{-1})/\mu)}\right) \text{ if } \mu \leq \log(p^{-1})/2;\ \ H(\mu,p) = \Theta\left(\sqrt{\frac{\log(\mu + p^{-1})}{\mu}}\right) \text{ otherwise.}$$

Given an ILP, we can find an optimal solution x^* to its LP relaxation efficiently, but need to round fractional entries in x^* to integers. The idea of randomized rounding is: given a real $v > 0$, 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 mean outcome is v. Starting with this idea, the analysis of [32] produces an integral solution of value at most $y^* + O(\min\{y^*, m\} \cdot H(\min\{y^*, m\}, 1/m))$ for MIPs (though phrased a bit differently); this is derandomized in [31]. But this does not exploit the sparsity of A; the previously-mentioned result of [18] produces an integral solution of value at most $y^* + g + 1$.

For CIPs, the idea is to solve the LP relaxation, scale *up* the components of x^* suitably, and then perform randomized rounding; see Section 5 for the details. Starting with this idea, the work of [32] leads to certain approximation bounds; similar bounds are achieved through different means by Plotkin, Shmoys & Tardos [30]. Work of this author [36] improved upon these results by observing a "correlation" property of CIPs, getting an approximation ratio of $1 + O(\max\{\ln(mB/y^*)/B, \sqrt{\ln(mB/y^*)/B}\})$. Thus, while the work of [32] gives a general approximation bound for MIPs, the result of [18] gives good results for sparse MIPs. For CIPs, the current-best results are those of [36]; however, no better results were known for sparse CIPs.

2.1 Improvements achieved

For MIPs, we use the extended LLL and an idea of Éva Tardos that leads to a bootstrapping of the LLL extension, to show:

Theorem 2.5. *For any given MIP, there exists an integral solution of value at most $y^* + O(1) + O(\min\{y^*, m\} \cdot H(\min\{y^*, m\}, 1/a))$.*

Since $a \leq m$, this is always as good as the $y^* + O(\min\{y^*, m\} \cdot H(\min\{y^*, m\}, 1/m))$ bound of [32] and is a good improvement, if $a \ll m$. It also is an improvement over the additive g factor of [18] in cases where g is not small compared to y^*.

Consider, e.g., the global routing problem and its MIP formulation, sketched above; m here is the number of edges in G, and $g = a$ is the maximum length of any path in $\bigcup_i P_i$. To focus on a specific interesting case, suppose y^*, the fractional congestion, is at most one. Then while the previous results ([32] and [18], resp.) give bounds of $O(\log m/\log\log m)$ and $O(a)$ on an integral solution, we get the improved bound of $O(\log a/\log\log a)$. Similar improvements are easily seen for other ranges of y^* also; e.g., if $y^* = O(\log a)$, an integral solution of value $O(\log a)$ exists, improving on the previously known bounds of $O(\log m/\log(2\log m/\log a))$ and $O(a)$. Thus, routing along *short* paths (this is the notion of sparsity for the global routing problem) is very beneficial in keeping the congestion low. Section 4 presents a scenario where we get such improvements, for discrepancy-type problems [34, 4]. In particular, we generalize a hypergraph-partitioning result of Füredi & Kahn [16].

Recall the bounds of [36] for CIPs mentioned in the paragraph preceding this subsection; our bounds for CIPs depend only on the set of constraints $Ax \geq b$, i.e., they hold for any non-negative objective-function vector c. Our improvements over [36] get better as y^* decreases. We show:

Theorem 2.6. *For any given CIP, there exists a feasible solution of value at most* $y^* \cdot (1 + O(max\{\ln(a+1)/B, \sqrt{\ln(a+B)/B}\}))$.

This CIP bound is better than that of [36] if y^* is small enough. In particular, we generalize the result of Chvátal [10] on weighted set cover. Consider, e.g., a facility location problem on a directed graph $G = (V, A)$: given a cost $c_i \in [0, 1]$ for each $i \in V$, we want a min-cost assignment of facilities to the nodes such that each node sees at least B facilities in its out-neighborhood–multiple facilities at a node are allowed. If Δ_{in} is the maximum in-degree of G, Theorem 2.6 guarantees an integrality gap of $1 + O(\max\{\ln(\Delta_{in} + 1)/B, \sqrt{\ln(B(\Delta_{in} + 1))/B}\})$. This improves on [36] if $y^* \leq |V|B/\max\{\Delta_{in}, B\}$; it shows an $O(1)$ (resp., $1 + o(1)$) integrality gap if B grows as fast as (resp., strictly faster than) $\log \Delta_{in}$.

A key corollary of Theorem 2.6 is that for families of instances of CIPs, we get a good ($O(1)$ or $1 + o(1)$) integrality gap if B grows at least as fast as $\log a$. Bounds on the result of a greedy algorithm for CIPs relative to the optimal *integral* solution, are known [12, 13]. Our bound of Theorem 2.6 improves that of [12] and is incomparable with [13]; for any given A, c, and the unit vector $b/||b||_2$, our bound improves on [13] if B is more than a certain threshold. As it stands, randomized rounding produces such improved solutions for several CIPs only with a very low, sometimes exponentially small, probability. Thus, it does not imply a randomized algorithm, often. To this end, we generalize Raghavan's method of pessimistic estimators to derive an algorithmic (polynomial-time) version of Theorem 2.6, in § 5.3.

We also show via Theorem 5.9 and Corollary 5.10 that multi-criteria CIPs can be approximated well. In particular, Corollary 5.10 shows some interesting cases where the approximation guarantee for multi-criteria CIPs grows in a very much sub-linear fashion with the number ℓ of given vectors c_i: the approximation ratio is at most $O(\log \log \ell)$ times what we show for CIPs (which correspond to the case where $\ell = 1$). We are not aware of any such earlier work on multi-criteria CIPs. The constructive version of Corollary 5.10 that we present in § 5.3 requires $\text{poly}(n^{\log \ell}, m)$ time, though. It would be interesting to improve this to a polynomial-time algorithm.

2.2 Preliminary version of this work, and followup

A preliminary version of this work appeared in [35], with a sketch of the proof for minimax integer programs, and proofs omitted for our constructive results on covering integer programs. In this version, we provide all proofs, further generalize the main covering result of [35] to Theorem 5.9, and present a sample application of Theorem 5.9 in Corollary 5.10. As mentioned in § 1, two main algorithmic versions related to our work have been obtained following [35]. First, for a subclass of the MIPs where the nonzero entries of the matrix A are "reasonably large", constructive versions of our results have been obtained in [20]. Second, for a notion of approximation that is different from the one we study, algorithmic results have been developed for certain families of MIPs in [11]. Furthermore, our Theorem 2.6 for CIPs has been used in [19] to develop approximation algorithms for CIPs that have given upper bounds on the variables x_j.

3 The Extended LLL and an Approach to Large Deviations

We now present our LLL extension, Theorem 3.1. For any event E, define $\chi(E)$ to be its indicator r.v.: 1 if E holds and 0 otherwise. Suppose we have "bad" events $E_1, \ldots, E_m$ with a "dependency" d' (in the sense of Lemma 1.1) that is "large". Theorem 3.1 shows how to essentially replace d' by a possibly much-smaller d, if we are able to define appropriate non-negative valued random variables $\{C_{i,j}\}$ for each E_i. It generalizes Lemma 1.1 (define one r.v., $C_{i,1} = \chi(E_i)$, for each i, to get Lemma 1.1), and its proof is very similar to the classical proof of Lemma 1.1. The motivation for Theorem 3.1 will be clarified by the applications; in particular, given some additional preparations, Theorem 4.2 and Theorem 2.6 follow respectively, and with relatively less work, from Theorem 3.1 and its proof approach.

Theorem 3.1. *Given events $E_1, \ldots, E_m$ and any $I \subseteq [m]$, let $Z(I) \doteq \bigwedge_{i \in I} \overline{E_i}$. Suppose that for some positive integer d, we can define, for each $i \in [m]$, a finite number of r.v.s $C_{i,1}, C_{i,2}, \ldots$ each taking on only non-negative values such that:*

(i) any $C_{i,j}$ is mutually independent of all but at most d of the events E_k, $k \neq i$, and

(ii) $\forall I \subseteq ([m] - \{i\})$, $\Pr(E_i \mid Z(I)) \leq \sum_j \boldsymbol{E}[C_{i,j} \mid Z(I)]$.

Let p_i denote $\sum_j \boldsymbol{E}[C_{i,j}]$. Suppose that for all $i \in [m]$ we have $ep_i(d+1) \leq 1$. Then $\Pr(\bigwedge_i \overline{E_i}) \geq (d/(d+1))^m > 0$.

Remark 3.2. *Note, by setting $I = \phi$ in (ii), that $\Pr(E_i) \leq p_i$ for all i. Also, $C_{i,j}$ and $C_{i,j'}$ can "depend" on different subsets of $\{E_k | k \neq i\}$; the only restriction is that these subsets be of size at most d. Note that we have essentially reduced the dependency among the E_is, to just d: $ep_i(d+1) \leq 1$ suffices. Another important point is that the dependency among the r.v.s $C_{i,j}$ could be much higher than d: all we count is the number of E_k that any $C_{i,j}$ depends on.*

Proof of Theorem 3.1. We prove by induction on $|I|$ that if $i \notin I$ then $\Pr(E_i \mid Z(I)) \leq ep_i$, which suffices to prove the theorem since $\Pr(\bigwedge_i \overline{E_i}) = \prod_{i \in [m]} (1 - \Pr(E_i \mid Z([i-1])))$. For the base case where $I = \emptyset$, $\Pr(E_i \mid Z(I)) = \Pr(E_i) \leq p_i$. For the inductive step, let $S_{i,j,I} \doteq \{k \in I \mid C_{i,j}$ depends on $E_k\}$, and $S'_{i,j,I} = I - S_{i,j,I}$; note that $|S_{i,j,I}| \leq d$. If $S_{i,j,I} = \emptyset$, then $\mathbf{E}[C_{i,j} \mid Z(I)] = \mathbf{E}[C_{i,j}]$. Otherwise, letting $S_{i,j,I} = \{\ell_1, \ldots, \ell_r\}$, we have

$$\mathbf{E}[C_{i,j} \mid Z(I)] = \frac{\mathbf{E}[C_{i,j} \cdot \chi(Z(S_{i,j,I})) \mid Z(S'_{i,j,I})]}{\Pr(Z(S_{i,j,I}) \mid Z(S'_{i,j,I}))} \leq \frac{\mathbf{E}[C_{i,j} \mid Z(S'_{i,j,I})]}{\Pr(Z(S_{i,j,I}) \mid Z(S'_{i,j,I}))},$$

since $C_{i,j}$ is non-negative. The numerator of the last term is $\mathbf{E}[C_{i,j}]$, by assumption. The denominator can be lower-bounded as follows:

$$\prod_{s \in [r]} (1 - \Pr(E_{\ell_s} \mid Z(\{\ell_1, \ell_2, \ldots, \ell_{s-1}\} \cup S'_{i,j,I}))) \geq \prod_{s \in [r]} (1 - ep_{\ell_s}) \geq (1 - 1/(d+1))^r \geq (d/(d+1))^d > 1/e;$$

the first inequality follows from the induction hypothesis. Hence, $\mathbf{E}[C_{i,j} \mid Z(I)] \leq e\mathbf{E}[C_{i,j}]$ and thus, $\Pr(E_i \mid Z(I)) \leq \sum_j \mathbf{E}[C_{i,j} \mid Z(I)] \leq ep_i \leq 1/(d+1)$. □

The crucial point is that the events E_i could have a large dependency d', in the sense of the classical Lemma 1.1. The main utility of Theorem 3.1 is that if we can "decompose" each E_i into the r.v.s $C_{i,j}$ that satisfy the conditions of the theorem, then there is the possibility of effectively reducing the dependency by much (d' can be replaced by the value d). Concrete instances of this will be studied in later sections.

The tools behind our MIP application are our new LLL, and a result of [33]. Define, for $z = (z_1, \ldots, z_n) \in \Re^n$, a family of polynomials $S_j(z), j = 0, 1, \ldots, n$, where $S_0(z) \equiv 1$, and for $j \in [n]$,

$$S_j(z) \doteq \sum_{1 \leq i_1 < i_2 \cdots < i_j \leq n} z_{i_1} z_{i_2} \cdots z_{i_j}. \quad (1)$$

Remark 3.3. *For real x and non-negative integral r, we define $\binom{x}{r} \doteq x(x-1)\cdots(x-r+1)/r!$ as usual; this is the sense meant in Theorem 3.4 below.*

We define a nonempty event to be any event with a nonzero probability of occurrence. The relevant theorem of [33] is the following:

Theorem 3.4. ([33]) *Given r.v.s $X_1, \ldots, X_n \in [0,1]$, let $X = \sum_{i=1}^{n} X_i$ and $\mu = \boldsymbol{E}[X]$. Then,*

(a) For any $q > 0$, any nonempty event Z and any non-negative integer $k \leq q$,

$$\Pr(X \geq q \mid Z) \leq \boldsymbol{E}[Y_{k,q} \mid Z],$$

where $Y_{k,q} = S_k(X_1, \ldots, X_n)/\binom{q}{k}$.

(b) If the X_is are independent, $\delta > 0$, and $k = \lceil \mu\delta \rceil$, then $\Pr(X \geq \mu(1+\delta)) \leq \boldsymbol{E}[Y_{k,\mu(1+\delta)}] \leq G(\mu,\delta)$, where $G(\cdot,\cdot)$ is as in Lemma 2.4.

(c) If the X_is are independent, then $\boldsymbol{E}[S_k(X_1, \ldots, X_n)] \leq \binom{n}{k} \cdot (\mu/n)^k \leq \mu^k/k!$.

Proof. Suppose $r_1, r_2, \ldots r_n \in [0,1]$ satisfy $\sum_{i=1}^{n} r_i \geq q$. Then, a simple proof is given in [33], for the fact that for any non-negative integer $k \leq q$, $S_k(r_1, r_2, \ldots, r_n) \geq \binom{q}{k}$. This clearly holds even given the occurrence of any nonempty event Z. Thus we get $\Pr(X \geq q \mid Z) \leq \Pr(Y_{k,q} \geq 1 \mid Z) \leq \mathbf{E}[Y_{k,q} \mid Z]$, where the second inequality follows from Markov's inequality. The proofs of (b) and (c) are given in [33]. □

We next present the proof of Lemma 2.4:

Proof of Lemma 2.4. Part (a) is the Chernoff-Hoeffding bound (see, e.g., Appendix A of [4], or [27]). For (b), we proceed as follows. For any $\mu > 0$, it is easy to check that

$$G(\mu,\delta) = e^{-\Theta(\mu\delta^2)} \text{ if } \delta \in (0,1); \tag{2}$$
$$G(\mu,\delta) = e^{-\Theta(\mu(1+\delta)\ln(1+\delta))} \text{ if } \delta \geq 1. \tag{3}$$

Now if $\mu \leq \log(p^{-1})/2$, choose

$$\delta = C \cdot \frac{\log(p^{-1})}{\mu \log(\log(p^{-1})/\mu)}$$

for a suitably large constant C. Since $\mu \leq \log(p^{-1})/2$, we can make $\delta \geq 1$ and have (3) hold, by choosing, e.g., $C \geq 1$. Simple algebraic manipulation now shows that if C is large enough, then $\lceil \mu\delta \rceil \cdot G(\mu,\delta) \leq p$ holds. Similarly, if $\mu > \log(p^{-1})/2$, we set $\delta = C \cdot \sqrt{\frac{\log(\mu+p^{-1})}{\mu}}$ for a large enough constant C, and use (2). □

4 Approximating Minimax Integer Programs

We now employ Theorem 3.1 to develop improved results on integral solutions for MIPs. We broadly proceed in two steps. First, we define a useful parameter t; a relatively direct application of Theorem 3.1 leads to Theorem 4.2. Second, we "bootstrap" Theorem 4.2 to develop the stronger Theorem 2.5. We present the intuitions behind these two theorems in the paragraph preceding Theorem 4.2 and in the paragraph preceding the proof of Theorem 2.5.

Let us first define a parameter t. This is useful as a first step in developing Theorem 4.2, but we will dispense with it when we work up to the proof of Theorem 2.5. Suppose we are given an MIP conforming to Definition 2.1. Define t to be $\max_{i\in[m]} NZ_i$, where NZ_i is the number of rows of A which have a non-zero coefficient corresponding to at least one variable among $\{x_{i,j} : j \in [\ell_i]\}$. Note that

$$g \le a \le t \le \min\{m, a \cdot \max_{i\in[n]} \ell_i\}. \tag{4}$$

Theorem 4.2 now shows how Theorem 3.1 can help, for sparse MIPs–those where $t \ll m$. We will then bootstrap Theorem 4.2 to get the further-improved Theorem 2.5. We start with a proposition, whose proof is shown in the Appendix:

Proposition 4.1. *If $0 < \mu_1 \le \mu_2$, then for any $\delta > 0$, $G(\mu_1, \mu_2\delta/\mu_1) \le G(\mu_2, \delta)$.*

Intuition for Theorem 4.2. Given an MIP, let us conduct randomized rounding in a natural way: independently for each i, randomly round exactly one $x_{i,j}$ to 1, with $x_{i,j}$ getting rounded to 1 with probability $x^*_{i,j}$. Now, we have one bad event E_i for each row i of A, corresponding to its row-sum going noticeably above its expectation. We would like to define random variables $C_{i,j}$ corresponding to each E_i, in order to employ Theorem 3.1. If we can choose a suitable integer k, then Theorem 3.4(a) suggests a natural choice for the $C_{i,j}$: since E_i is an upper-tail event of the type captured by Theorem 3.4(a), we can try and bijectively map the $C_{i,j}$ to the $\binom{n}{k}$ r.v.s "$X_{i_1}X_{i_2}\cdots X_{i_k}/\binom{q}{k}$" that constitute the r.v. "$Y_{k,q}$" of Theorem 3.4(a). Condition (ii) of Theorem 3.1 is satisfied by Theorem 3.4(a); we will bound the parameter d in condition (i) of Theorem 3.1 by using the definition of t. These, in conjunction with some simple observations, lead to Theorem 4.2.

Theorem 4.2. *Given an MIP conforming to Definition 2.1, randomized rounding produces a feasible solution of value at most $y^* + \lceil \min\{y^*, m\} \cdot H(\min\{y^*, m\}, 1/(et))\rceil$, with non-zero probability.*

Proof. Conduct randomized rounding: independently for each i, randomly round exactly one $x_{i,j}$ to 1, guided by the "probabilities" $\{x^*_{i,j}\}$. Recall again that $x^* = \{x^*_{i,j}\}$ is the vector obtained by solving the LP relaxation. We may assume that $\{x^*_{i,j}\}$ is a *basic* feasible solution to this relaxation. Hence, at most m of the $\{x^*_{i,j}\}$ will be neither zero nor one, and only these variables will participate in the rounding. Since all the entries of A are in $[0, 1]$, we can assume without loss of generality from now on that $y^* \le m$ (and that $\max_{i\in[n]} \ell_i \le m$); this explains the "$\min\{y^*, m\}$" term in our stated bounds. Let $b_i = (Ax^*)_i$ denote the r.h.s. value that we get for the ith row in the optimal solution for the LP relaxation; so we have $b_i \le y^*$. If $z \in \{0, 1\}^N$ denotes the randomly rounded vector, then $\mathbf{E}[(Az)_i] = b_i \le y^*$ by the linearity of expectation. Defining

$$k = \lceil y^* H(y^*, 1/(et))\rceil \tag{5}$$

and events $E_1, E_2, \ldots, E_m$ by $E_i \equiv$ "$(Az)_i \ge b_i + k$", we now show that $\Pr(\bigwedge_{i\in[m]} \overline{E_i}) > 0$ using Theorem 3.1. Rewrite the ith constraint of the MIP as

$$\sum_{r\in[n]} X_{i,r} \le W, \text{ where } X_{i,r} = \sum_{s\in[\ell_r]} A_{i,(r,s)} x_{r,s};$$

the notation $A_{i,(r,s)}$ assumes that the pairs $\{(r, s) : r \in [n], s \in [\ell_r]\}$ have been mapped bijectively to $[N]$, in some fixed way. Defining the r.v.

$$Z_{i,r} = \sum_{s\in[\ell_r]} A_{i,(r,s)} z_{r,s},$$

we note that for each i, the r.v.s $\{Z_{i,r} : r \in [n]\}$ lie in $[0, 1]$ and are *independent*. Also, $E_i \equiv$ "$\sum_{r\in[n]} Z_{i,r} \ge b_i + k$".

Theorem 3.4 suggests a suitable choice for the crucial r.v.s $C_{i,j}$ (to apply Theorem 3.1). Let $u = \binom{n}{k}$; we now define the r.v.s $\{C_{i,j} : i \in [m], j \in [u]\}$ as follows. Fix any $i \in [m]$. Identify each $j \in [u]$ with some distinct k-element subset $S(j)$ of $[n]$, and let

$$C_{i,j} \doteq \frac{\prod_{v \in S(j)} Z_{i,v}}{\binom{b_i+k}{k}}. \tag{6}$$

We now need to show that the r.v.s $C_{i,j}$ satisfy the conditions of Theorem 3.1. For any $i \in [m]$, let $\delta_i = k/b_i$. Since $b_i \leq y^*$, we have, for each $i \in [m]$,

$$\begin{aligned} G(b_i, \delta_i) &\leq G(y^*, k/y^*) \text{ (by Proposition 4.1)} \\ &\leq G(y^*, H(y^*, 1/(et))) \text{ (by (5))} \\ &\leq 1/(ekt) \text{ (by the definition of } H). \end{aligned}$$

Now by Theorem 3.4, we get

Fact 4.3. *For all $i \in [m]$ and for all nonempty events Z, $\Pr(E_i \mid Z) \leq \sum_{j \in [u]} \boldsymbol{E}[C_{i,j} \mid Z]$. Also, $p_i \doteq \sum_{j \in [u]} \boldsymbol{E}[C_{i,j}] < G(b_i, \delta_i) \leq 1/(ekt)$.*

Next since any $C_{i,j}$ involves (a product of) k terms, each of which "depends" on at most $(t-1)$ of the events $\{E_v : v \in ([m] - \{i\})\}$ by definition of t, we see the important

Fact 4.4. *$\forall i \in [m]$ $\forall j \in [u]$, $C_{i,j} \in [0,1]$ and $C_{i,j}$ "depends" on at most $d = k(t-1)$ of the set of events $\{E_v : v \in ([m] - \{i\})\}$.*

From Facts 4.3 and 4.4 and by noting that $ep_i(d+1) \leq e(kt - k + 1)/(ekt) \leq 1$, we invoke Theorem 3.1, to see that $\Pr(\bigwedge_{i \in [m]} \overline{E_i}) > 0$, concluding the proof of Theorem 4.2. □

We are now ready to improve Theorem 4.2, to obtain Theorem 2.5.

Intuition for Theorem 2.5. Theorem 4.2 gives good results if $t \ll m$, but can we improve it further, say by replacing t by a $(\leq t)$ in it? As seen from (4), the key reason for $t \gg a^{\Theta(1)}$ is that $\max_{i \in [n]} \ell_i \gg a^{\Theta(1)}$. If we can essentially "bring down" $\max_{i \in [n]} \ell_i$ by forcing many $x^*_{i,j}$ to be zero for *each* i, then we effectively reduce t ($t \leq a \cdot max_i\ \ell_i$, see (4)); this is so since only those $x^*_{i,j}$ that are neither zero nor one take part in the rounding. A way of bootstrapping Theorem 4.2 to achieve this using a "slow rounding" technique that proceeds in $O(\log \log t)$ iterations, is shown by Theorem 2.5.

Proof. **(For Theorem 2.5)** Let $K_0 > 0$ be a sufficiently large absolute constant. Now if

$$(y^* \geq t^{1/7}) \text{ or } (t \leq \max\{K_0, 2\}) \text{ or } (t \leq a^4) \tag{7}$$

holds, then we will be done by Theorem 4.2. So we may assume that (7) is false. Also, if $y^* \leq t^{-1/7}$, Theorem 4.2 guarantees an integral solution of value $O(1)$; thus, we also suppose that $y^* > t^{-1/7}$. The basic idea now is, as sketched above, to set many $x^*_{i,j}$ to zero for each i (without losing too much on y^*), so that $\max_i \ell_i$ and hence, t, will essentially get reduced. Such an approach, whose performance will be validated by arguments similar to those of Theorem 4.2, is repeatedly applied until (7) holds, owing to the (continually reduced) t becoming small enough to satisfy (7). There are two cases:

Case I: $y^* \geq 1$. Solve the LP relaxation, and set $x'_{i,j} := (y^*)^2 (\log^5 t) x^*_{i,j}$. Conduct randomized rounding on the $x'_{i,j}$ now, rounding *each* $x'_{i,j}$ independently to $z_{i,j} \in \{\lfloor x'_{i,j} \rfloor, \lceil x'_{i,j} \rceil\}$. (Note the key difference from Theorem 4.2, where for each i, we round exactly one $x^*_{i,j}$ to 1.)

Let $K_1 > 0$ be a sufficiently large absolute constant. We now use ideas similar to those used in our proof of Theorem 4.2 to show that with nonzero probability, we have both of the following:

$$\forall i \in [m], \qquad (Az)_i \leq (y^*)^3 \log^5 t \cdot (1 + K_1/((y^*)^{1.5} \log^2 t)), \quad \textbf{and} \tag{8}$$
$$\forall i \in [n], \qquad |\sum_j z_{i,j} - (y^*)^2 \log^5 t| \leq K_1 y^* \log^3 t. \tag{9}$$

To show this, we proceed as follows. Let $E_1, E_2, \ldots, E_m$ be the "bad" events, one for each event in (8) not holding; similarly, let $E_{m+1}, E_{m+2}, \ldots, E_{m+n}$ be the "bad" events, one for each event in (9) not holding. We want to use our extended LLL to show that with positive probability, all these bad events can be avoided; specifically, we need a way of associating each E_i with some finite number of non-negative r.v.s $C_{i,j}$. (Loosely speaking, we will "decompose" each E_i into a finite number of such $C_{i,j}$.) We do this as follows. For each event $E_{m+\ell}$ where $\ell \geq 1$, we define just one r.v. $C_{m+\ell,1}$: this is the indicator variable for the occurrence of $E_{m+\ell}$. For the events E_i where $i \leq m$, we decompose E_i into r.v.s $C_{i,j}$ just as in (6): each such $C_{i,j}$ is now a scalar multiple of at most

$$O((y^*)^3 \log^5 t/((y^*)^{1.5} \log^2 t)) = O((y^*)^{1.5} \log^3 t) = O(t^{3/14} \log^3 t)$$

independent binary r.v.s that underlie our randomized rounding; the second equality (big-Oh bound) here follows since (7) has been assumed to not hold. Thus, it is easy to see that for all i, $1 \leq i \leq m+n$, and for any j, the r.v. $C_{i,j}$ depends on at most

$$O(t \cdot t^{3/14} \log^3 t) \tag{10}$$

events E_k, where $k \neq i$. Also, as in our proof of Theorem 4.2, Theorem 3.4 gives a direct proof of requirement (ii) of Theorem 3.1; part (b) of Theorem 3.4 shows that for any desired constant K, we can choose the constant K_1 large enough so that for all i, $\sum_j \mathbf{E}[C_{i,j}] \leq t^{-K}$. Thus, in view of (10), we see by Theorem 3.1 that $\Pr(\bigwedge_{i=1}^{m+n} \overline{E_i}) > 0$.

Fix a rounding z satisfying (8) and (9). For each $i \in [n]$ and $j \in [\ell_i]$, we renormalize as follows: $x''_{i,j} := z_{i,j}/\sum_u z_{i,u}$. Thus we have $\sum_u x''_{i,u} = 1$ for all i; we now see that we have two very useful properties. First, since $\sum_j z_{i,j} \geq (y^*)^2 \log^5 t \cdot \left(1 - O(\frac{1}{y^* \log^2 t})\right)$ for all i from (9), we have, $\forall i \in [m]$,

$$(Ax'')_i \leq \frac{y^*(1 + O(1/((y^*)^{1.5} \log^2 t)))}{1 - O(1/(y^* \log^2 t))} \leq y^*(1 + O(1/(y^* \log^2 t))). \tag{11}$$

Second, since the $z_{i,j}$ are non-negative integers summing to at most $(y^*)^2 \log^5 t(1 + O(1/(y^* \log^2 t)))$, *at most $O((y^*)^2 \log^5 t)$ values $x''_{i,j}$ are nonzero*, for each $i \in [n]$. Thus, by losing a little in y^* (see (11)), our "scaling up–rounding–scaling down" method has given a fractional solution x'' with a much-reduced ℓ_i for each i; ℓ_i is now $O((y^*)^2 \log^5 t)$, essentially. Thus, t has been reduced to $O(a(y^*)^2 \log^5 t)$; i.e., t has been reduced to at most

$$K_2 t^{1/4+2/7} \log^5 t \tag{12}$$

for some constant $K_2 > 0$ that is independent of K_0, since (7) was assumed false. Repeating this scheme $O(\log \log t)$ times makes t small enough to satisfy (7). More formally, define $t_0 = t$, and $t_{i+1} = K_2 t_i^{1/4+2/7} \log^5 t_i$ for $i \geq 0$. Stop this sequence at the first point where either $t = t_i$ satisfies (7), or $t_{i+1} \geq t_i$ holds. Thus, we finally have t small enough to satisfy (7) or to be bounded by some absolute constant. How much has $\max_{i \in [m]}(Ax)_i$ increased in the process? By (11), we see that at the end,

$$\max_{i \in [m]} (Ax)_i \leq y^* \cdot \prod_{j \geq 0} (1 + O(1/(y^* \log^2 t_j))) \leq y^* \cdot e^{O(\sum_{j \geq 0} 1/(y^* \log^2 t_j))} \leq y^* + O(1), \tag{13}$$

since the values $\log t_j$ decrease geometrically and are lower-bounded by some absolute positive constant. We may now apply Theorem 4.2.

Case II: $t^{-1/7} < y^* < 1$. The idea is the same here, with the scaling up of $x^*_{i,j}$ being by $(\log^5 t)/y^*$; the same "scaling up–rounding–scaling down" method works out. Since the ideas are very similar to Case I, we only give a proof sketch here. We now scale up all the $x^*_{i,j}$ first by $(\log^5 t)/y^*$ and do a randomized rounding. The analogs of (8) and (9) that we now want are

$$\forall i \in [m], \qquad (Az)_i \le \log^5 t \cdot (1 + K_1'/\log^2 t), \quad \textbf{and} \tag{14}$$

$$\forall i \in [n], \qquad |\sum_j z_{i,j} - \log^5 t/y^*| \le K_1' \log^3 t/\sqrt{y^*}. \tag{15}$$

Proceeding identically as in Case I, we can show that with positive probability, (14) and (15) hold simultaneously. Fix a rounding where these two properties hold, and renormalize as before: $x''_{i,j} := z_{i,j}/\sum_u z_{i,u}$. Since (14) and (15) hold, we get that the following analogs of (11) and (12) hold:

$$(Ax'')_i \le \frac{y^*(1 + O(1/\log^2 t))}{1 - O(\sqrt{y^*}/\log^2 t)} \le y^*(1 + O(1/\log^2 t)); \quad \text{and}$$

$$t \text{ has been reduced to } O(a \log^5 t/y^*), \text{ i.e., to } O(t^{1/4+1/7} \log^5 t).$$

We thus only need $O(\log \log t)$ iterations, again. Also, the analog of (13) now is that

$$\max_{i\in[m]} (Ax)_i \le y^* \cdot \prod_{j\ge 0} (1 + O(1/\log^2 t_j)) \le y^* \cdot e^{O(\sum_{j\ge 0} 1/\log^2 t_j)} \le y^* + O(1).$$

This completes the proof of Theorem 2.5. □

We now study our improvements for discrepancy-type problems, which are an important class of MIPs that, among other things, are useful in devising divide-and-conquer algorithms. Given is a set-system (X, F), where $X = [n]$ and $F = \{D_1, D_2, \ldots, D_M\} \subseteq 2^X$. Given a positive integer ℓ, the problem is to partition X into ℓ parts, so that *each* D_j is "split well": we want a function $f : X \to [\ell]$ which minimizes $\max_{j\in[M],k\in[\ell]} |\{i \in D_j : f(i) = k\}|$. (The case $\ell = 2$ is the standard set-discrepancy problem.) To motivate this problem, suppose we have a (di)graph (V, A); we want a partition of V into $V_1, \ldots, V_\ell$ such that $\forall v \in V$, $\{|\{j \in N(v) \cap V_k\}| : k \in [\ell]\}$ are "roughly the same", where $N(v)$ is the (out-)neighborhood of v. See, e.g., [2, 17] for how this helps construct divide-and-conquer approaches. This problem is naturally modeled by the above set-system problem.

Let Δ be the degree of (X, F), i.e., $\max_{i\in[n]} |\{j : i \in D_j\}|$, and let $\Delta' \doteq \max_{D_j\in F} |D_j|$. Our problem is naturally written as an MIP with $m = M\ell$, $\ell_i = \ell$ for each i, and $g = a = \Delta$, in the notation of Definition 2.1; $y^* = \Delta'/\ell$ here. The analysis of [32] gives an integral solution of value at most $y^*(1 + O(H(y^*, 1/(M\ell))))$, while [18] presents a solution of value at most $y^* + \Delta$. Also, since any $D_j \in F$ intersects at most $(\Delta - 1)\Delta'$ other elements of F, Lemma 1.1 shows that randomized rounding produces, with positive probability, a solution of value at most $y^*(1 + O(H(y^*, 1/(e\Delta'\Delta\ell))))$. This is the approach taken by [16] for their case of interest: $\Delta = \Delta'$, $\ell = \Delta/\log \Delta$.

Theorem 2.5 shows the existence of an integral solution of value $y^*(1+O(H(y^*, 1/\Delta)))+O(1)$, i.e., *removes the dependence of the approximation factor on* Δ'. This is an improvement on all the three results above. As a specific interesting case, suppose ℓ grows at most as fast as $\Delta'/\log \Delta$. Then we see that good integral solutions–those that grow at the rate of $O(y^*)$ or better–exist, and this was not known before. (The approach of [16] shows such a result for $\ell = O(\Delta'/\log(\max\{\Delta, \Delta'\}))$. Our bound of $O(\Delta'/\log \Delta)$ is always better than this, and especially so if $\Delta' \gg \Delta$.)

5 Approximating Covering Integer Programs

One of the main ideas behind Theorem 3.1 was to extend the basic inductive proof behind the LLL by decomposing the "bad" events E_i appropriately into the r.v.s $C_{i,j}$. We now use this general idea in a different context, that of (multi-criteria) covering integer programs, with an additional crucial ingredient being a useful correlation inequality, the FKG inequality [15]. The reader is asked to recall the discussion of (multi-criteria) CIPs from § 2. We start with a discussion of randomized rounding for CIPs, the Chernoff lower-tail bound, and the FKG inequality in § 5.1. These lead to our improved, but nonconstructive, approximation bound for column-sparse (multi-criteria) CIPs, in § 5.2. This is then made constructive in § 5.3; we also discuss there what we view as novel about this constructive approach. The two paragraphs marked "Intuition" in § 5.2, as well as the first two paragraphs § 5.3, describe some of our main ideas here at an intuitive level.

5.1 Preliminaries

Let us start with a simple and well-known approach to tail bounds. Suppose Y is a random variable and y is some value. Then, for any $0 \leq \delta < 1$, we have

$$\Pr(Y \leq y) \leq \Pr((1-\delta)^Y \geq (1-\delta)^y) \leq \frac{\mathbf{E}[(1-\delta)^Y]}{(1-\delta)^y}, \tag{16}$$

where the inequality is a consequence of Markov's inequality.

We next setup some basic notions related to approximation algorithms for (multi-criteria) CIPs. Recall that in such problems, we have ℓ given non-negative vectors $c_1, c_2, \ldots, c_\ell$ such that for all i, $c_i \in [0,1]^n$ with $\max_j c_{i,j} = 1$; $\ell = 1$ in the case of CIPs. Let $x = (x_1^*, x_2^*, \ldots, x_n^*)$ denote a given fractional solution that satisfies the system of constraints $Ax \geq b$. We are not concerned here with how x^* was found: typically, x^* would be an optimal solution to the LP relaxation of the problem. (The LP relaxation is obvious if, e.g., $\ell = 1$, or, say, if the given multi-criteria aims to minimize $\max_i c_i^T \cdot x^*$, or to keep each $c_i^T \cdot x^*$ bounded by some target value v_i.) We now consider how to round x^* to some integral z so that:

(P1) the constraints $Az \geq b$ hold, and

(P2) for all i, $c_i^T \cdot z$ is "not much bigger" than $c_i^T \cdot x^*$: our approximation bound will be a measure of how small a "not much bigger value" we can achieve in this sense.

Let us now discuss the "standard" randomized rounding scheme for (multi-criteria) CIPs. We assume a fixed instance as well as x^*, from now on. For an $\alpha > 1$ to be chosen suitably, set $x'_j = \alpha x_j^*$, for each $j \in [n]$. We then construct a random integral solution z by setting, independently for each $j \in [n]$, $z_j = \lfloor x'_j \rfloor + 1$ with probability $x'_j - \lfloor x'_j \rfloor$, and $z_j = \lfloor x'_j \rfloor$ with probability $1 - (x'_j - \lfloor x'_j \rfloor)$. The aim then is to show that with positive (hopefully high) probability, **(P1)** and **(P2)** happen simultaneously. We now introduce some useful notation. For every $j \in [n]$, let $s_j = \lfloor x'_j \rfloor$. Let A_i denote the ith row of A, and let $X_1, X_2, \ldots, X_n \in \{0,1\}$ be *independent* r.v.s with $\Pr(X_j = 1) = x'_j - s_j$ for all j. The bad event E_i that the ith constraint is violated by our randomized rounding is given by $E_i \equiv$ "$A_i \cdot X < \mu_i(1-\delta_i)$", where $\mu_i = \mathbf{E}[A_i \cdot X]$, s is the vector with entries s_j, and $\delta_i = 1 - (b_i - A_i \cdot s)/\mu_i$. We aim to bound $\Pr(E_i)$ for all i, when the standard randomized rounding is used. We assume without loss of generality that $A_i \cdot s < b_i$ for each i; otherwise, the bad event E_i cannot happen. So, we have $\delta_i \in (0,1)$ for all i.

We now bound $\Pr(E_i)$; the proof involves routine Chernoff bounds and calculations, and is shown in the Appendix.

Lemma 5.1. *Define* $g(B, \alpha) \doteq (\alpha \cdot e^{-(\alpha-1)})^B$*. Then for all* i*,*

$$\Pr(E_i) \le \frac{\boldsymbol{E}[(1-\delta_i)^{A_i \cdot X}]}{(1-\delta_i)^{(1-\delta_i)\mu_i}} \le g(B, \alpha) \le e^{-B(\alpha-1)^2/(2\alpha)}$$

under standard randomized rounding.

Next, we state a special case of the FKG inequality [15]. Given binary vectors $\vec{a} = (a_1, a_2, \ldots, a_\ell) \in \{0,1\}^\ell$ and $\vec{b} = (b_1, b_2, \ldots, b_\ell) \in \{0,1\}^\ell$, let us partially order them by coordinate-wise domination: $\vec{a} \preceq \vec{b}$ iff $a_i \le b_i$ for all i. Now suppose $Y_1, Y_2, \ldots, Y_\ell$ are *independent* r.v.s, each taking values in $\{0,1\}$. Let $\vec{Y}$ denote the vector $(Y_1, Y_2, \ldots, Y_\ell)$. Suppose an event $\mathcal{A}$ is completely defined by the value of $\vec{Y}$. Define $\mathcal{A}$ to be *increasing* iff: for all $\vec{a} \in \{0,1\}^\ell$ such that $\mathcal{A}$ holds when $\vec{Y} = \vec{a}$, $\mathcal{A}$ also holds when $\vec{Y} = \vec{b}$, for any $\vec{b}$ such that $\vec{a} \preceq \vec{b}$. Analogously, event $\mathcal{A}$ is *decreasing* iff: for all $\vec{a} \in \{0,1\}^\ell$ such that $\mathcal{A}$ holds when $\vec{Y} = \vec{a}$, $\mathcal{A}$ also holds when $\vec{Y} = \vec{b}$, for any $\vec{b} \preceq \vec{a}$. The FKG inequality proves certain intuitively appealing bounds:

Lemma 5.2. ***(FKG inequality)*** *Let* $I_1, I_2, \ldots, I_t$ *be any sequence of increasing events and* $D_1, D_2, \ldots, D_t$ *be any sequence of decreasing events (each* I_i *and* D_i *completely determined by* $\vec{Y}$*). Then for any* $i \in [t]$ *and any* $S \subseteq [t]$*,*

(i) $\Pr(I_i | \bigwedge_{j \in S} I_j) \ge \Pr(I_i)$ *and* $\Pr(D_i | \bigwedge_{j \in S} D_j) \ge \Pr(D_i)$*;*

(ii) $\Pr(I_i | \bigwedge_{j \in S} D_j) \le \Pr(I_i)$ *and* $\Pr(D_i | \bigwedge_{j \in S} I_j) \le \Pr(D_i)$*.*

Returning to our random variables X_j and events E_i, we get the following lemma as an easy consequence of the FKG inequality, since each event of the form "$\overline{E_i}$" or "$X_j = 1$" is an increasing event as a function of the vector $(X_1, X_2, \ldots, X_n)$:

Lemma 5.3. *For all* $B_1, B_2 \subseteq [m]$ *such that* $B_1 \cap B_2 = \emptyset$ *and for any* $B_3 \subseteq [n]$*,* $\Pr(\bigwedge_{i \in B_1} \overline{E_i} \mid ((\bigwedge_{j \in B_2} \overline{E_j}) \wedge (\bigwedge_{k \in B_3} (X_k = 1))) \ge \prod_{i \in B_1} \Pr(\overline{E_i})$*.*

5.2 Nonconstructive approximation bounds for (multi-criteria) CIPs

We now work up to our main approximation bound for multi-criteria CIPs in Theorem 5.9; this theorem is presented in an abstract manner, and one concrete corollary is shown by Corollary 5.10. We develop Theorem 5.9 by starting with the special case of CIPs (which have just one objective function) and proving Theorem 2.6; the basic ideas are then generalized to obtain Theorem 5.9. The randomized rounding approach underlying Theorem 5.9, as it stands, may only construct the solution guaranteed with very low (much less than inverse-polynomial) probability; algorithmic versions of Theorem 5.9, which in some cases involve super-polynomial time, are then developed in § 5.3.

Definition 5.4. ***(The function*** $\mathcal{R}$***)*** *For any* s *and any* $j_1 < j_2 < \cdots < j_s$*, let* $\mathcal{R}(j_1, j_2, \ldots, j_s)$ *be the set of indices* i *such that row* i *of the constraint system "*$Ax \ge b$*" has at least one of the variables* j_k*,* $1 \le k \le s$*, appearing with a nonzero coefficient. (Note from the definition of* a *in Defn. 2.2, that* $|\mathcal{R}(j_1, j_2, \ldots, j_s)| \le a \cdot s$*.)*

Let the vector $x^* = (x_1^*, x_2^*, \ldots, x_n^*)$, the parameter $\alpha > 1$, and the "standard" randomized rounding scheme, be as defined in § 5.1. The standard rounding scheme is sufficient for our (nonconstructive) purposes now; we generalize this scheme as follows, for later use in § 5.3.

Definition 5.5. ***(General randomized rounding)*** *Given a vector* $p = (p_1, p_2, \ldots, p_n) \in [0,1]^n$*, the general randomized rounding with parameter* p *generates independent random variables* $X_1, X_2, \ldots, X_n \in \{0,1\}$ *with* $\Pr(X_j = 1) = p_j$*; the rounded vector* z *is defined by* $z_j = \lfloor \alpha x_j^* \rfloor + X_j$ *for all* j*. (As in the standard rounding, we set each* z_j *to be either* $\lfloor \alpha x_j^* \rfloor$ *or* $\lceil \alpha x_j^* \rceil$*; the standard rounding is the special case in which* $\boldsymbol{E}[z_j] = \alpha x_j^*$ *for all* j*.)*

We now present an important lemma, Lemma 5.6, to get correlation inequalities which "point" in the "direction" opposite to FKG. Some ideas from the proof of Lemma 1.1 will play a crucial role in our proof this lemma.

Lemma 5.6. *Suppose we employ general randomized rounding with some parameter* p*, and that* $\Pr(\bigwedge_{i=1}^{m} \overline{E_i})$ *is nonzero under this rounding. The following hold for any* q *and any* $1 \leq j_1 < j_2 < \cdots < j_q \leq n$.

(i)

$$\Pr(X_{j_1} = X_{j_2} = \cdots = X_{j_q} = 1 \mid \bigwedge_{i=1}^{m} \overline{E_i}) \leq \frac{\prod_{t=1}^{q} p_{j_t}}{\prod_{i \in \mathcal{R}(j_1, j_2, \ldots, j_q)} (1 - \Pr(E_i))}; \tag{17}$$

the events $E_i \equiv ((Az)_i < b_i)$ *are defined here w.r.t. the general randomized rounding.*

(ii) In the special case of standard randomized rounding,

$$\prod_{i \in \mathcal{R}(j_1, j_2, \ldots, j_q)} (1 - \Pr(E_i)) \geq (1 - g(B, \alpha))^{aq}; \tag{18}$$

the function g *is as defined in Lemma 5.1.*

Proof. (i) Note first that if we wanted a *lower* bound on the l.h.s., the FKG inequality would immediately imply that the l.h.s. is at least $p_{j_1} p_{j_2} \cdots p_{j_q}$. We get around this "correlation problem" as follows. Let $Q = \mathcal{R}(j_1, j_2, \ldots, j_q)$; let $Q' = [m] - Q$. Let

$$Z_1 \equiv (\bigwedge_{i \in Q} \overline{E_i}), \text{ and } Z_2 \equiv (\bigwedge_{i \in Q'} \overline{E_i}).$$

Letting $Y = \prod_{t=1}^{q} X_{j_t}$, note that

$$|Q| \leq aq \text{ and} \tag{19}$$

$$Y \text{ is independent of } Z_2. \tag{20}$$

Now,

$$\begin{aligned}
\Pr(Y = 1 \mid (Z_1 \wedge Z_2)) &= \frac{\Pr(((Y = 1) \wedge Z_1) \mid Z_2)}{\Pr(Z_1 \mid Z_2)} \\
&\leq \frac{\Pr((Y = 1) \mid Z_2)}{\Pr(Z_1 \mid Z_2)} \\
&= \frac{\Pr(Y = 1)}{\Pr(Z_1 \mid Z_2)} \quad \text{(by (20))} \\
&= \frac{\prod_{t=1}^{q} \Pr(X_{j_t} = 1)}{\Pr(Z_1 \mid Z_2)} \qquad (21) \\
&\leq \frac{\prod_{t=1}^{q} \Pr(X_{j_t} = 1)}{\prod_{i \in \mathcal{R}(j_1, j_2, \ldots, j_q)} (1 - \Pr(E_i))} \quad \text{(by Lemma 5.3).}
\end{aligned}$$

(ii) We get (18) from Lemma 5.1 and (19). □

Intuition. Note that in the proof of part (i) of Lemma 5.6, we broadly proceed as in the proofs of Lemma 1.1 and Theorem 3.1 up to (21). The key difference in the next (and final) step is that instead of applying induction to lower-bound the denominator as in the proofs of Lemma 1.1 and Theorem 3.1, we are directly able to obtain a lower bound via Lemma 5.3. In addition to giving a better lower bound, this type of explicitly-available denominator will be of considerable value in § 5.3.

We will use Lemmas 5.3 and 5.6 to prove Theorem 5.9. As a warmup, we start with Theorem 2.6, which deals with the special case of CIPs; recall that y^* denotes $c^T \cdot x^*$. We present a simple proposition, whose proof is in the Appendix:

Proposition 5.7. *Suppose α and β are chosen as follows, for some sufficiently large absolute constant $K > 0$:*

$$\alpha = K \cdot \ln(a+1)/B \text{ and } \beta = 2, \text{ if } \ln(a+1) \geq B, \text{ and} \quad (22)$$
$$\alpha = \beta = 1 + K \cdot \sqrt{\ln(a+B)/B}, \text{ if } \ln(a+1) < B. \quad (23)$$

Then we have $\beta(1 - g(B,\alpha))^a > 1$.

We now prove Theorem 2.6, our result for CIPs:

Proof. **(For Theorem 2.6)** Let α and β be as in (22) and (23). Conduct standard randomized rounding, and let $\mathcal{E}$ be the event that $c^T \cdot z > y^*\alpha\beta$. Setting $Z \equiv \bigwedge_{i\in[m]} \overline{E_i}$ and $\mu \doteq \mathbf{E}[c^T \cdot z] = y^*\alpha$, we see by Markov's inequality that $\Pr(\mathcal{E} \mid Z)$ is at most $R = (\sum_{j=1}^n c_j \Pr(X_j = 1 \mid Z))/(\mu\beta)$. Note that $\Pr(Z) > 0$ since $\alpha > 1$; so, we need only prove that $R < 1$. Lemma 5.6 shows that

$$R \leq \frac{\sum_j c_j p_j}{\mu\beta \cdot (1 - g(B,\alpha))^a} = \frac{1}{\beta(1 - g(B,\alpha))^a};$$

thus, the condition $\beta(1 - g(B,\alpha))^a > 1$ suffices. Proposition 5.7 completes the proof. □

Intuition. The basic approach of our proof of Theorem 2.6 is to follow the main idea of Theorem 3.1, and to decompose the event "$\mathcal{E} \mid Z$" into a non-negative linear combination of events of the form "$X_j = 1 \mid Z$'; we then exploited the fact that each X_j depends on at most a of the events comprising Z. We now extend Theorem 2.6 and also generalize to multi-criteria CIPs. Instead of employing just a "first moment method" (Markov's inequality) as in the proof of Theorem 2.6, we will work with higher moments: the functions S_k defined in (1) and used in Theorem 3.4.

Suppose some parameters $\lambda_i > 0$ are given, and that our goal is to round x^* to z so that the event

$$\mathcal{A} \equiv \left[(Az \geq b) \ \wedge \ (\forall i, \ c_i^T \cdot z \leq \lambda_i)\right] \quad (24)$$

holds. We first give a sufficient condition for this to hold, in Theorem 5.9; we then derive some concrete consequences in Corollary 5.10. We need one further definition before presenting Theorem 5.9. Recall that A_i and b_i respectively denote the ith row of A and the ith component of b. Also, the vector s and values δ_i will throughout be as in the definition of **standard** randomized rounding.

Definition 5.8. ***(The functions* ch *and* ch′*)*** *Suppose we conduct general randomized rounding with some parameter p; i.e., let $X_1, X_2, \ldots, X_n$ be independent binary random variables such that $\Pr(X_j = 1) = p_j$. For each $i \in [m]$, define*

$$\mathrm{ch}_i(p) \doteq \frac{\boldsymbol{E}[(1-\delta_i)^{A_i \cdot X}]}{(1-\delta_i)^{b_i - A_i \cdot s}} = \frac{\prod_{j\in[n]} \boldsymbol{E}[(1-\delta_i)^{A_{i,j}X_j}]}{(1-\delta_i)^{b_i - A_i\cdot s}} \quad \textit{and} \quad \mathrm{ch}'_i(p) \doteq \min\{\mathrm{ch}_i(p), 1\}.$$

(Note from (16) that if we conduct general randomized rounding with parameter p, then $\Pr((Az)_i < b_i) \leq \mathrm{ch}'_i(p)$; also, "ch" stands for "Chernoff-Hoeffding".)

Theorem 5.9. *Suppose we are given a multi-criteria CIP, as well as some parameters $\lambda_1, \lambda_2, \ldots, \lambda_\ell > 0$. Let $\mathcal{A}$ be as in (24). Then, for any sequence of positive integers $(k_1, k_2, \ldots, k_\ell)$ such that $k_i \leq \lambda_i$, the following hold.*

(i) Suppose we employ general randomized rounding with parameter $p = (p_1, p_2, \ldots, p_n)$. Then, $\Pr(\mathcal{A})$ is at least

$$\Phi(p) \doteq \left(\prod_{r\in[m]} (1 - \mathrm{ch}'_r(p))\right) - \sum_{i=1}^{\ell} \frac{1}{\binom{\lambda_i}{k_i}} \cdot \sum_{j_1 < \cdots < j_{k_i}} \left(\prod_{t=1}^{k_i} c_{i,j_t} \cdot p_{j_t}\right) \cdot \prod_{r \notin \mathcal{R}(j_1,\ldots,j_{k_i})} (1 - \mathrm{ch}'_r(p)); \quad (25)$$

as in Defn. 2.3, $c_{i,j} \in [0,1]$ is the jth coordinate of the vector c_i.

(ii) Suppose we employ the standard randomized rounding to get a rounded vector z. Let $\lambda_i = \nu_i(1+\gamma_i)$ for each $i \in [\ell]$, where $\nu_i = \boldsymbol{E}[c_i^T \cdot z] = \alpha \cdot (c_i^T \cdot x^*)$ and $\gamma_i > 0$ is some parameter. Then,

$$\Phi(p) \geq (1 - g(B,\alpha))^m \cdot \left(1 - \sum_{i=1}^{\ell} \frac{\binom{n}{k_i} \cdot (\nu_i/n)^{k_i}}{\binom{\nu_i(1+\gamma_i)}{k_i}} \cdot (1 - g(B,\alpha))^{-a \cdot k_i}\right). \tag{26}$$

In particular, if the r.h.s. of (26) is positive, then $\Pr(\mathcal{A}) > 0$ for standard randomized rounding.

The proof of Theorem 5.9 is a simple generalization of that of Theorem 2.6 – basically, we use higher moments (the moment S_{k_i} for objective function c_i) and employ Theorem 3.4, instead of using the first moment and Markov's inequality. This proof is deferred to the Appendix. Theorem 2.6 is the special case of Theorem 5.9 corresponding to $\ell = k_1 = 1$. To make the general result of Theorem 5.9 more concrete, we now present an additional special case, Corollary 5.10. We provide this as one possible "proof of concept", rather than as an optimized one; e.g., the constant "3" in the bound "$c_i^T \cdot z \leq 3\nu_i$" of Corollary 5.10 can be improved. The proof of Corollary 5.10 requires routine calculations after setting $k_i = \lceil \ln(2\ell) \rceil$ and $\gamma_i = 2$ for all i in part (ii) of Theorem 5.9; its proof is given in the Appendix.

Corollary 5.10. *There is an absolute constant $K' > 0$ such that the following holds. Suppose we are given a multi-criteria CIP with notation as in part (ii) of Theorem 5.9. Define $\alpha = K' \cdot \max\{\frac{\ln(a)+\ln\ln(2\ell)}{B},\ 1\}$. Now if $\nu_i \geq \log^2(2\ell)$ for all $i \in [\ell]$, then standard randomized rounding produces a feasible solution z such that $c_i^T \cdot z \leq 3\nu_i$ for all i, with positive probability.*

In particular, this can be shown by setting $k_i = \lceil \ln(2\ell) \rceil$ and $\gamma_i = 2$ for all i, in part (ii) of Theorem 5.9.

5.3 Constructive version

It can be shown that for many problems, randomized rounding produces the solutions shown to exist by Theorem 2.6 and Theorem 5.9, with very low probability: e.g., probability almost exponentially small in the input size. Thus we need to obtain constructive versions of these theorems. Our method will be a deterministic procedure that makes $O(n)$ calls to the function $\Phi(\cdot)$, in addition to poly(n,m) work. Now, if k' denotes the maximum of all the k_i, we see that Φ can be evaluated in poly$(n^{k'}, m)$ time. Thus, our overall procedure runs in time poly$(n^{k'}, m)$ time. In particular, we get constructive versions of Theorem 2.6 and Corollary 5.10 that run in time poly(n,m) and poly$(n^{\log \ell}, m)$, respectively.

Our approach is as follows. We start with a vector p that corresponds to standard randomized rounding, for which we know (say, as argued in Corollary 5.10) that $\Phi(p) > 0$. In general, we have a vector of probabilities $p = (p_1, p_2, \ldots, p_n)$ such that $\Phi(p) > 0$. If $p \in \{0,1\}^n$, we are done. Otherwise suppose some p_j lies in $(0,1)$; by renaming the variables, we will assume without loss of generality that $j = n$. Define $p' = (p_1, p_2, \ldots, p_{n-1}, 0)$ and $p'' = (p_1, p_2, \ldots, p_{n-1}, 1)$. The main fact we wish to show is that $\Phi(p') > 0$ or $\Phi(p'') > 0$: we can then set p_n to 0 or 1 appropriately, and continue. (As mentioned in the previous paragraph, we thus have $O(n)$ calls to the function $\Phi(\cdot)$ in total.) Note that although some of the p_j will lie in $\{0,1\}$, we can crucially continue to view the X_j as *independent* random variables with $\Pr(X_j = 1) = p_j$.

So, our main goal is: assuming that $p_n \in (0,1)$ and that

$$\Phi(p) > 0, \tag{27}$$

to show that $\Phi(p') > 0$ or $\Phi(p'') > 0$. In order to do so, we make some observations and introduce some simplifying notation. Define, for each $i \in [m]$: $q_i = \mathrm{ch}'_i(p)$, $q'_i = \mathrm{ch}'_i(p')$, and $q''_i = \mathrm{ch}'_i(p'')$. Also define the vectors $q \doteq (q_1, q_2, \ldots, q_m)$, $q' \doteq (q'_1, q'_2, \ldots, q'_m)$, and $q'' \doteq (q''_1, q''_2, \ldots, q''_m)$. We now present a useful lemma about these vectors:

Lemma 5.11. *For all $i \in [m]$, we have*

$$0 \le q_i'' \le q_i' \le 1; \quad (28)$$

$$q_i \ge p_n q_i'' + (1 - p_n) q_i'; \text{ and} \quad (29)$$

$$q_i' = q_i'' = q_i \text{ if } i \notin \mathcal{R}(n). \quad (30)$$

Proof. It is directly seen from the definition of "ch_i" that $\mathrm{ch}_i(p'') \le \mathrm{ch}_i(p')$. Since $q_i'' = \min\{\mathrm{ch}_i(p''), 1\}$ and $q_i' = \min\{\mathrm{ch}_i(p'), 1\}$, we get $q_i'' \le q_i'$. The remaining inequalities of (28), as well as the equalities in (30), are straightforward. As for (29), we proceed as in [36]. First of all, if $q_i = 1$, then we are done, since $q_i'', \ q_i' \le 1$. So suppose $q_i < 1$; in this case, $q_i = \mathrm{ch}_i(p)$. Now, Definition 5.8 shows that

$$\mathrm{ch}_i(p) = p_n \mathrm{ch}_i(p'') + (1 - p_n)\mathrm{ch}_i(p').$$

Therefore, $q_i = \mathrm{ch}_i(p) = p_n \mathrm{ch}_i(p'') + (1 - p_n)\mathrm{ch}_i(p') \ge p_n \mathrm{ch}_i'(p'') + (1 - p_n)\mathrm{ch}_i'(p')$. □

Since we are mainly concerned with the vectors p, p' and p'' now, we will view the values $p_1, p_2, \ldots, p_{n-1}$ as arbitrary but *fixed*, subject to (27). The function $\Phi(\cdot)$ now has a simple form; to see this, we first define, for a vector $r = (r_1, r_2, \ldots, r_m)$ and a set $U \subseteq [m]$,

$$f(U, r) = \prod_{i \in U} (1 - r_i).$$

Recall that $p_1, p_2, \ldots, p_{n-1}$ are considered as constants now. Then, it is evident from (25) that there exist constants $u_1, u_2, \ldots, u_t$ and $v_1, v_2, \ldots, v_{t'}$, as well as subsets $U_1, U_2, \ldots, U_t$ and $V_1, V_2, \ldots, V_{t'}$ of $[m]$, such that

$$\Phi(p) = f([m], q) - (\sum_i u_i \cdot f(U_i, q)) - (p_n \cdot \sum_j v_j \cdot f(V_j, q)); \quad (31)$$

$$\Phi(p') = f([m], q') - (\sum_i u_i \cdot f(U_i, q')) - (0 \cdot \sum_j v_j \cdot f(V_j, q'))$$

$$= f([m], q') - \sum_i u_i \cdot f(U_i, q'); \quad (32)$$

$$\Phi(p'') = f([m], q'') - (\sum_i u_i \cdot f(U_i, q'')) - (1 \cdot \sum_j v_j \cdot f(V_j, q''))$$

$$= f([m], q'') - (\sum_i u_i \cdot f(U_i, q'')) - (\sum_j v_j \cdot f(V_j, q'')). \quad (33)$$

Importantly, we also have the following:

$$\text{the constants } u_i, v_j \text{ are non-negative}; \quad \forall j, \ V_j \cap \mathcal{R}(n) = \emptyset. \quad (34)$$

Recall that our goal is to show that $\Phi(p') > 0$ or $\Phi(p'') > 0$. We will do so by proving that

$$\Phi(p) \le p_n \Phi(p'') + (1 - p_n)\Phi(p'). \quad (35)$$

Let us use the equalities (31), (32), and (33). In view of (30) and (34), the term "$-p_n \cdot \sum_j v_j \cdot f(V_j, q)$" on both sides of the inequality (35) cancels; defining $\Delta(U) \doteq (1 - p_n) \cdot f(U, q') + p_n \cdot f(U, q'') - f(U, q)$, inequality (35) reduces to

$$\Delta([m]) - \sum_i u_i \cdot \Delta(U_i) \ge 0. \quad (36)$$

Before proving this, we pause to note a challenge we face. Suppose we only had to show that, say, $\Delta([m])$ is non-negative; this is exactly the issue faced in [36]. Then, we will immediately be done by part (i) of Lemma 5.12, which states that $\Delta(U) \geq 0$ for any set U. However, (36) also has terms such as "$u_i \cdot \Delta(U_i)$" with a *negative* sign in front. To deal with this, we need something more than just that $\Delta(U) \geq 0$ for all U; we handle this by part (ii) of Lemma 5.12. We view this as the main novelty in our constructive version here.

Lemma 5.12. *Suppose $U \subseteq V \subseteq [m]$. Then, (i) $\Delta(U) \geq 0$, and (ii) $\Delta(U)/f(U,q) \leq \Delta(V)/f(V,q)$. (Since $\Phi(p) > 0$ by (27), we have that $q_i < 1$ for each i. So, $1/f(U,q)$ and $1/f(V,q)$ are well-defined.)*

Assuming that Lemma 5.12 is true, we will now show (36); the proof of Lemma 5.12 is given below. We have

$$\begin{aligned}
\Delta([m]) - \sum_i u_i \cdot \Delta(U_i) &= (\Delta([m])/f([m],q)) \cdot f([m],q) - \sum_i (\Delta(U_i)/f(U_i,q)) \cdot u_i \cdot f(U_i,q) \\
&\geq (\Delta([m])/f([m],q)) \cdot \left[f([m],q) - \sum_i u_i \cdot f(U_i,q) \right] \quad \text{(by Lemma 5.12)} \\
&\geq 0 \quad \text{(by (27) and (31))}.
\end{aligned}$$

Thus we have (36).

Proof of Lemma 5.12. It suffices to show the following. Assume $U \neq [m]$; suppose $u \in ([m] - U)$ and that $U' = U \cup \{u\}$. Assuming by induction on $|U|$ that $\Delta(U) \geq 0$, we show that $\Delta(U') \geq 0$, and that $\Delta(U)/f(U,q) \leq \Delta(U')/f(U',q)$. It is easy to check that this way, we will prove both claims of the lemma. The base case of the induction is that $|U| \in \{0,1\}$, where $\Delta(U) \geq 0$ is directly seen by using (29). Suppose inductively that $\Delta(U) \geq 0$. Using the definition of $\Delta(U)$ and the fact that $f(U',q) = (1-q_u)f(U,q)$, we have

$$\begin{aligned}
f(U',q) &= (1-q_u) \cdot [(1-p_n)f(U,q') + p_n f(U,q'') - \Delta(U)] \\
&\leq (1-(1-p_n)q'_u - p_n q''_u) \cdot [(1-p_n)f(U,q') + p_n f(U,q'')] - (1-q_u) \cdot \Delta(U),
\end{aligned}$$

where this last inequality is a consequence of (29). Therefore, using the definition of $\Delta(U')$ and the facts $f(U',q') = (1-q'_u)f(U,q')$ and $f(U',q'') = (1-q''_u)f(U,q'')$,

$$\begin{aligned}
\Delta(U') &= (1-p_n)(1-q'_u)f(U,q') + p_n(1-q''_u)f(U,q'') - f(U',q) \\
&\geq (1-p_n)(1-q'_u)f(U,q') + p_n(1-q''_u)f(U,q'') + \\
&\quad\ (1-q_u) \cdot \Delta(U) - (1-(1-p_n)q'_u - p_n q''_u) \cdot [(1-p_n)f(U,q') + p_n f(U,q'')] \\
&= (1-q_u) \cdot \Delta(U) + p_n(1-p_n) \cdot (f(U,q'') - f(U,q')) \cdot (q'_u - q''_u) \\
&\geq (1-q_u) \cdot \Delta(U) \quad \text{(by (28))}.
\end{aligned}$$

So, since we assumed that $\Delta(U) \geq 0$, we get $\Delta(U') \geq 0$; furthermore, we get that $\Delta(U') \geq (1-q_u) \cdot \Delta(U)$, which implies that $\Delta(U')/f(U',q) \geq \Delta(U)/f(U,q)$. □

6 Conclusion

We have presented an extension of the LLL that basically helps reduce the "dependency" much in some settings; we have seen applications to two families of integer programming problems. It would be interesting to see how far these ideas can be pushed further. Two other open problems suggested by this work are: (i) developing a constructive version of our result for MIPs, and (ii) developing a poly(n, m)-time constructive version of Theorem 5.9, as opposed to the poly$(n^{k'}, m)$-time constructive version that we present in § 5.3.

Finally, a very interesting question is to develop a theory of applications of the LLL (Lemma 1.1) that can be made constructive with (essentially) no loss. Suppose we have bad events $E_1, E_2, \ldots, E_m$ in the setting of the LLL, which are functions of n independent binary random variables $X_1, X_2, \ldots, X_n$ where $\Pr(X_i = 1) = p_i$. In an attempt to mimic the approach described in the second paragraph of Section 5.3, suppose we proceed as follows. The standard proof of the LLL [14] presents a function Φ such that: (i) $\Pr(\bigwedge_{i=1}^{m} \overline{E_i}) > 0$ if $\Phi(p) > 0$, and (ii) if the conditions of the LLL hold, then $\Phi(p) > 0$. Can we now try to set values for the X_i one-by-one, as in Section 5.3? Unfortunately, there are two problems we face in this regard. First, the function Φ given by the proof of the LLL does not appear to be polynomial-time computable, basically due to the type of induction it uses; indeed, as briefly mentioned in the paragraph following the proof of Lemma 5.6, the type of "explicitly-available denominator" that we obtain in the proof of Lemma 5.6, is one crucial driver in obtaining an efficiently-computable Φ in § 5.3. Second, again appealing to the notation of the second paragraph of Section 5.3, it is unclear if we can prove here that "$\Phi(p') > 0$ or $\Phi(p'') > 0$". (In fact, Joel Spencer has suggested to us the possibility that such a disjunction may not hold for all applications of the LLL.) It would be of much interest to guarantee these for some interesting class of applications of the LLL, or to develop fresh approaches to obtain constructive versions of the LLL with (essentially) no loss.

Acknowledgments. This work started while visiting the Sandia National Laboratories in the summer of 1994; I thank Leslie Goldberg and Z Sweedyk who were involved in the early stages of this work. I would like to thank Éva Tardos for suggesting the key idea that helped bootstrap Theorem 4.2 to get Theorem 2.5. I also thank Noga Alon, Alan Frieze, Tom Leighton, Chi-Jen Lu, Alessandro Panconesi, Prabhakar Raghavan, Satish Rao, Joel Spencer, the SODA 1996 referees and journal referees for their very helpful comments and suggestions.

References

[1] N. Alon, *A parallel algorithmic version of the Local Lemma*, Random Structures & Algorithms, 2 (1991), pp. 367–378.

[2] N. Alon, *The strong chromatic number of a graph*, Random Structures & Algorithms, 3 (1992), pp. 1–7.

[3] N. Alon, L. Babai, and A. Itai, *A fast and simple randomized parallel algorithm for the maximal independent set problem*, Journal of Algorithms, 7 (1986), pp. 567–583.

[4] N. Alon and J. H. Spencer, *The Probabilistic Method*, Second Edition, John Wiley & Sons, 2000.

[5] J. Beck, *An algorithmic approach to the Lovász Local Lemma*, Random Structures & Algorithms, 2 (1991), pp. 343–365.

[6] J. Beck and T. Fiala, *"Integer-making" theorems*, Discrete Applied Mathematics, 3 (1981), pp. 1–8.

[7] J. Beck and J. H. Spencer, *Integral approximation sequences*, Mathematical Programming, 30 (1984), pp. 88–98.

[8] B. Berger and J. Rompel, *Simulating* $(\log^c n)$*-wise independence in NC*, JACM, 38 (1991), pp. 1026–1046.

[9] A. Z. Broder, A. M. Frieze, and E. Upfal, *Static and dynamic path selection on expander graphs: a random walk approach*, Random Structures & Algorithms 14 (1999), pp. 87–109.

[10] V. Chvátal, *A greedy heuristic for the set covering problem*, Math. of O.R., 4 (1979), pp. 233–235.

[11] A. Czumaj and C. Scheideler, *An algorithmic approach to the general Lovász Local Lemma with applications to scheduling and satisfiability problems*, In Proc. ACM Symposium on Theory of Computing, 2000, pp. 38–47.

[12] G. Dobson, *Worst-case analysis of greedy heuristics for integer programming with nonnegative data*, Math. of O.R., 7 (1982), pp. 515–531.

[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] P. Erdős and L. Lovász, *Problems and results on* 3*-chromatic hypergraphs and some related questions*, In *Infinite and Finite Sets*, A. Hajnal et al., eds., Colloq. Math. Soc. J. Bolyai 11, North Holland, Amsterdam, 1975, pp. 609–627.

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

[16] Z. Füredi and J. Kahn, *On the dimensions of ordered sets of bounded degree*, Order, 3 (1986), pp. 15–20.

[17] H. J. Karloff and D. B. Shmoys, *Efficient parallel algorithms for edge coloring problems*, Journal of Algorithms, 8 (1987), pp. 39–52.

[18] R. M. Karp, F. T. Leighton, R. L. Rivest, C. D. Thompson, U. V. Vazirani, and V. V. Vazirani, *Global wire routing in two-dimensional arrays*, Algorithmica, 2 (1987), pp. 113–129.

[19] S. G. Kolliopoulos and N. E. Young, *Tight approximation results for general covering integer programs*, In Proc. IEEE Symposium on Foundations of Computer Science, 2001, pp. 522–528.

[20] F. T. Leighton, C. J. Lu, S. B. Rao and A. Srinivasan, *New Algorithmic Aspects of the Local Lemma with Applications to Routing and Partitioning*, SIAM J. Comput., 31 (2001), pp. 626-641.

[21] F. T. Leighton, S. B. Rao and A. Srinivasan, *Multicommodity Flow and Circuit Switching*, In Proc. Hawaii International Conference on System Sciences, 1998, pp. 459–465.

[22] L. Lovász, *On the ratio of optimal integral and fractional covers*, Discrete Math., 13 (1975), pp. 383–390.

[23] M. Luby, *A simple parallel algorithm for the maximal independent set problem*, SIAM J. Comput., 15 (1986), pp. 1036–1053.

[24] M. Molloy and B. Reed, *Further algorithmic aspects of the Local Lemma*, In Proc. ACM Symposium on Theory of Computing, 1998, pp. 524–529.

[25] M. Molloy and B. Reed, *Graph Colouring and the Probabilistic Method*, Springer-Verlag, 2001.

[26] R. Motwani, J. Naor, and M. Naor, *The probabilistic method yields deterministic parallel algorithms*, Journal of Computer and System Sciences, 49 (1994), pp. 478–516.

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

[28] N. Nisan, *Pseudorandom generators for space–bounded computation*, Combinatorica, 12 (1992), pp. 449–461.

[29] C. H. Papadimitriou and M. Yannakakis, *On the approximability of trade-offs and optimal access of web sources*, In Proc. IEEE Symposium on Foundations of Computer Science, 2000, pp. 86–92.

[30] S. A. Plotkin, D. B. Shmoys, and É. Tardos, *Fast approximation algorithms for fractional packing and covering problems*, Math. of O.R., 20 (1995), pp. 257–301.

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

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

[33] J. P. Schmidt, A. Siegel, and A. Srinivasan, *Chernoff-Hoeffding bounds for applications with limited independence*, SIAM J. Discrete Math., 8 (1995), pp. 223–250.

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

[35] A. Srinivasan, *An Extension of the Lovász Local Lemma, and its Applications to Integer Programming*, in Proc. ACM-SIAM Symposium on Discrete Algorithms, 1996, pp. 6–15.

[36] A. Srinivasan, *Improved approximation guarantees for packing and covering integer programs*, SIAM J. Comput., 29 (1999), pp. 648–670.

Appendix

A Proofs and Calculation-Details

Proof. **(For Proposition 4.1)** Taking logarithms on both sides, we aim to show that

$$\mu_1 \cdot [\mu_2\delta/\mu_1 - (1 + \mu_2\delta/\mu_1)\ln(1 + \mu_2\delta/\mu_1)] \leq \mu_2 \cdot [\delta - (1+\delta)\ln(1+\delta)]\,.$$

This simplifies to showing

$$(1 + \psi\delta)\ln(1 + \psi\delta) \geq \psi \cdot (1+\delta)\ln(1+\delta),$$

where $\psi = \mu_2/\mu_1 \geq 1$. This inequality, in turn, follows from the fact that the function $t \mapsto (1+t)\ln(1+t)$ is convex for $t \geq 0$ (since its second derivative, $1/(1+t)$, is positive), and since this function equals 0 when $t = 0$. ☐

Proof. **(For Lemma 5.1)** The first inequality follows from (16). Next, the Chernoff-Hoeffding lower-tail approach [4, 27] shows that

$$\frac{\mathbf{E}[(1-\delta_i)^{A_i \cdot X}]}{(1-\delta_i)^{(1-\delta_i)\mu_i}} \leq \left(\frac{e^{-\delta_i}}{(1-\delta_i)^{1-\delta_i}}\right)^{\mu_i} = \frac{{\mu_i}^{b_i - A_i \cdot s} \cdot e^{b_i - \mu_i - A_i \cdot s}}{(b_i - A_i \cdot s)^{b_i - A_i \cdot s}}. \tag{37}$$

Recall that

$$A_i \cdot s < b_i; \quad A_i \cdot s + \mu_i \geq \alpha b_i; \quad \alpha > 1. \tag{38}$$

Let us now bound the last (third) term in (37), denoted ψ, say.

Fix all parameters other than μ_i; subject to (38), let us first observe that ψ is maximized when the second inequality in (38) is an equality. This is easily seen by recalling the definition of ψ (from (37)) and noting by differentiation that ψ is a decreasing function of μ_i when $\mu_i > b_i - A_i \cdot s$. Next, since $\mu_i = \alpha b_i - A_i \cdot s$,

$$\psi = \left(\frac{\alpha b_i - A_i \cdot s}{b_i - A_i \cdot s}\right)^{b_i - A_i \cdot s} \cdot e^{(1-\alpha)b_i} = \left(1 + \frac{(\alpha-1)b_i}{b_i - A_i \cdot s}\right)^{b_i - A_i \cdot s} \cdot e^{(1-\alpha)b_i}.$$

Let $x = (\alpha - 1)b_i/(b_i - A_i \cdot s)$, and note that $x \geq \alpha - 1$. We have

$$\ln \psi = (1-\alpha)b_i + (\alpha - 1)b_i \ln(1+x)/x. \tag{39}$$

Keeping b_i fixed,

$$\begin{aligned}
\frac{d(\ln \psi)}{dx} &= (\alpha-1)b_i \cdot \frac{x/(1+x) - \ln(1+x)}{x^2} \\
&= (\alpha-1)b_i \cdot \frac{x/(1+x) + \ln(1 - x/(1+x))}{x^2} \\
&\leq (\alpha-1)b_i \cdot \frac{x/(1+x) - x/(1+x)}{x^2} \\
&= 0.
\end{aligned}$$

So, ψ is maximized when $x = \alpha - 1$ (i.e., when $A_i \cdot s = 0$). From (39), $\ln \psi = b_i \cdot (1 - \alpha + \ln \alpha)$; the term multiplying b_i here is non-positive, since $\ln \alpha = \ln(1 + \alpha - 1) \leq \alpha - 1$. Therefore, ψ is maximized when b_i equals its minimum value of B, and so, $\psi \leq g(B, \alpha)$.

Finally, let us check that $g(B, \alpha) \leq e^{-B(\alpha-1)^2/(2\alpha)}$. Taking logarithms on both sides, we want

$$1 - \alpha + \ln \alpha \leq -(\alpha - 1)^2/(2\alpha) \tag{40}$$

for $\alpha \geq 1$. The l.h.s. and r.h.s. of (40) are equal when $\alpha = 1$; their respective derivatives are $-1 + 1/\alpha$ and $-1/2 + 1/(2\alpha^2)$, and so it suffices to prove that $1 - 1/\alpha \geq 1/2 - 1/(2\alpha^2)$ for $\alpha \geq 1$. That is, we need $2\alpha(\alpha - 1) \geq (\alpha - 1) \cdot (\alpha + 1)$, which is true since $\alpha \geq 1$. □

Proof. **(For Proposition 5.7)** Suppose first that $\ln(a+1) \geq B$. Then, $\alpha = K \cdot \ln(a+1)/B$ and so if K is large enough, then $\alpha/e^{\alpha-1} \leq e^{-\alpha/2}$. Therefore,

$$g(B, \alpha) = (\alpha \cdot e^{-(\alpha-1)})^B \leq e^{-\alpha B/2} = (a+1)^{-K/2}.$$

So, since $\beta = 2$, our goal of proving that $\beta(1 - g(B,\alpha))^a > 1$ reduces to proving that $2 \cdot (1 - (a+1)^{-K/2})^a > 1$, which is true since $a \geq 1$ and since K is chosen sufficiently large.

We next consider the case where $\ln(a+1) < B$. Recall from the statement of Lemma 5.1 that $g(B, \alpha) \leq e^{-B(\alpha-1)^2/(2\alpha)}$. Since $\alpha < K + 1$ in our case, we now have

$$g(B, \alpha) < e^{-B(\alpha-1)^2/(2(K+1))} = e^{-K^2 \ln(a+B)/(2(K+1))} = (a+B)^{-K^2/(2(K+1))}.$$

So, our goal of showing that $\beta(1 - g(B, \alpha))^a > 1$ reduces to proving that

$$(1 + K \cdot \sqrt{\ln(a+B)/B}) \cdot (1 - (a+B)^{-K^2/(2(K+1))})^a \geq 1,$$

which in turn holds if $(1 + K \cdot \sqrt{\ln(a+B)/B}) \cdot (1 - a \cdot (a+B)^{-K^2/(2(K+1))}) \geq 1$; this final inequality is true if we choose K to be a sufficiently large constant, since $a \cdot (a+B)^{-K^2/(2(K+1))} \ll K \cdot \sqrt{\ln(a+B)/B}$ under such a choice. □

Proof. **(For Theorem 5.9)**

(i) Let $E_r \equiv ((Az)_r < b_r)$ be defined w.r.t. general randomized rounding with parameter p; as observed in Definition 5.8, $\Pr(E_r) \leq \mathrm{ch}'_r(p)$. Now if $\mathrm{ch}'_r(p) = 1$ for some r, then part (i) is trivially true; so we assume that $\Pr(E_r) \leq \mathrm{ch}'_r(p) < 1$ for all $r \in [m]$. Defining $Z \equiv (Az \geq b) \equiv \bigwedge_{r \in [m]} \overline{E_r}$, we get by the FKG inequality that

$$\Pr(Z) \geq \prod_{r \in [m]} (1 - \Pr(E_r)).$$

Define, for $i = 1, 2, \ldots, \ell$, the "bad" event $\mathcal{E}_i \equiv (c_i^T \cdot z > \lambda_i)$. Fix any i. Our plan is to show that

$$\Pr(\mathcal{E}_i \mid Z) \le \frac{1}{\binom{\lambda_i}{k_i}} \cdot \sum_{j_1 < j_2 < \cdots < j_{k_i}} \left(\prod_{t=1}^{k_i} c_{i,j_t} \cdot p_{j_t} \right) \cdot \left(\prod_{r \in \mathcal{R}(j_1, j_2, \ldots, j_{k_i})} (1 - \Pr(E_r))^{-1} \right). \tag{41}$$

If we prove (41), then we will be done as follows. We have

$$\Pr(\mathcal{A}) \ge \Pr(Z) \cdot \left(1 - \sum_i \Pr(\mathcal{E}_i \mid Z) \right) \ge \left(\prod_{r \in [m]} (1 - \Pr(E_r)) \right) \cdot \left(1 - \sum_i \Pr(\mathcal{E}_i \mid Z) \right). \tag{42}$$

Now, the term "$(\prod_{r \in [m]} (1 - \Pr(E_r)))$" is a decreasing function of each of the values $\Pr(E_r)$; so is the lower bound on "$-\Pr(\mathcal{E}_i \mid Z)$" obtained from (41). Hence, bounds (41) and (42), along with the bound $\Pr(E_r) \le \mathrm{ch}'_r(p)$, will complete the proof of part (i).

We now prove (41) using Theorem 3.4(a) and Lemma 5.6. Recall the symmetric polynomials S_k from (1). Define $Y = S_{k_i}(c_{i,1}X_1, c_{i,2}X_2, \ldots, c_{i,n}X_n)/\binom{\lambda_i}{k_i}$. By Theorem 3.4(a), $\Pr(\mathcal{E}_i \mid Z) \le \mathbf{E}[Y \mid Z]$. Next, the typical term in $\mathbf{E}[Y \mid Z]$ can be upper bounded using Lemma 5.6:

$$\mathbf{E}\left[\left(\prod_{t=1}^{k_i} c_{i,j_t} \cdot X_{j_t}\right) \mid \bigwedge_{i=1}^{m} \overline{E_i} \right] \le \frac{\prod_{t=1}^{k_i} c_{i,j_t} \cdot p_{j_t}}{\prod_{r \in \mathcal{R}(j_1, j_2, \ldots, j_{k_i})} (1 - \Pr(E_r))}.$$

Thus we have (41), and the proof of part (i) is complete.

(ii) We have

$$\Phi(p) = \left[\prod_{r \in [m]} (1 - \mathrm{ch}'_r(p)) \right] \cdot \left(1 - \sum_{i=1}^{\ell} \frac{1}{\binom{\lambda_i}{k_i}} \cdot \sum_{j_1 < \cdots < j_{k_i}} \left[\prod_{t=1}^{k_i} c_{i,j_t} \cdot p_{j_t}\right] \cdot \left(\prod_{r \in \mathcal{R}(j_1, \ldots, j_{k_i})} \frac{1}{1 - \mathrm{ch}'_r(p)} \right) \right). \tag{43}$$

Lemma 5.1 shows that under standard randomized rounding, $\mathrm{ch}'_r(p) \le g(B, \alpha) < 1$ for all r. So, the r.h.s. κ of (43) gets lower-bounded as follows:

$$\begin{aligned}
\kappa &\ge (1 - g(B,\alpha))^m \cdot \left(1 - \sum_{i=1}^{\ell} \frac{1}{\binom{\nu_i(1+\gamma_i)}{k_i}} \cdot \sum_{j_1 < \cdots < j_{k_i}} \left(\prod_{t=1}^{k_i} c_{i,j_t} \cdot p_{j_t} \right) \cdot \left[\prod_{r \in \mathcal{R}(j_1, \ldots, j_{k_i})} (1 - g(B,\alpha)) \right]^{-1} \right) \\
&\ge (1 - g(B,\alpha))^m \cdot \left(1 - \sum_{i=1}^{\ell} \frac{1}{\binom{\nu_i(1+\gamma_i)}{k_i}} \cdot \sum_{j_1 < \cdots < j_{k_i}} \left(\prod_{t=1}^{k_i} c_{i,j_t} \cdot p_{j_t} \right) \cdot (1 - g(B,\alpha))^{-ak_i} \right) \\
&\ge (1 - g(B,\alpha))^m \cdot \left(1 - \sum_{i=1}^{\ell} \frac{\binom{n}{k_i} \cdot (\nu_i/n)^{k_i}}{\binom{\nu_i(1+\gamma_i)}{k_i}} \cdot (1 - g(B,\alpha))^{-ak_i} \right),
\end{aligned}$$

where the last line follows from Theorem 3.4(c). □

Proof. **(For Corollary 5.10)** Let us employ Theorem 5.9(ii) with $k_i = \lceil \ln(2\ell) \rceil$ and $\gamma_i = 2$ for all i. We just need to establish that the r.h.s. of (26) is positive. We need to show that

$$\sum_{i=1}^{\ell} \frac{\binom{n}{k_i} \cdot (\nu_i/n)^{k_i}}{\binom{3\nu_i}{k_i}} \cdot (1 - g(B,\alpha))^{-a \cdot k_i} < 1;$$

it is sufficient to prove that for all i,

$$\frac{\nu_i^{k_i}/k_i!}{\binom{3\nu_i}{k_i}} \cdot (1 - g(B,\alpha))^{-a \cdot k_i} < 1/\ell. \tag{44}$$

We make two observations now.

- Since $k_i \sim \ln \ell$ and $\nu_i \geq \log^2(2\ell)$,

$$\binom{3\nu_i}{k_i} = (1/k_i!) \cdot \prod_{j=0}^{k_i-1} (3\nu_i - j) = (1/k_i!) \cdot (3\nu_i)^{k_i} \cdot e^{-\Theta(\sum_{j=0}^{k_i-1} j/\nu_i)} = \Theta((1/k_i!) \cdot (3\nu_i)^{k_i}).$$

- $(1 - g(B, \alpha))^{-a \cdot k_i}$ can be made arbitrarily close to 1 by choosing the constant K' large enough.

These two observations establish (44). □