Simple Lifted Cover Inequalities and Hard Knapsack Problems

Brady Hunsaker Craig A. Tovey

April 29, 2004

Abstract

We consider a class of random knapsack instances described by Chvátal, who showed that with probability going to 1, such instances require an exponential number of branch-and-bound nodes. We show that even with the use of simple lifted cover inequalities, an exponential number of nodes is required with probability going to 1.

It is not surprising that there exist integer programming (IP) instances for which solution by branch-and-bound requires an exponential number of nodes, since integer programming is an NP-complete problem while the linear programs solved at each branch-and-bound node are polynomially solvable. Examples of such instances were given by Jeroslow [5], who presented a set of simple instances of the knapsack problem which require an exponential number of branch-and-bound nodes when branching on variables, and by Chvátal [1], who considered a class of random instances of the knapsack problem and showed that with probability converging to 1, such a random instance requires exponentially many branch-and-bound nodes to solve.

Most modern IP solvers use branch-and-cut algorithms, which combine branch-and-bound with the use of cutting planes. Gu, Nemhauser, and Savelsbergh [4] considered solving the knapsack problem with branch-and-cut. They presented a set of instances that require an exponential number of branch-and-bound nodes even with the addition of simple lifted cover inequalities. More recent work in proving exponential worst-case bounds in the presence of various cutting planes has been done by Dash [2], who proved worst-case exponential bounds in the presence of lift-and-project cuts, Chvátal-Gomory inequalities, and matrix cuts as described by Lovász and Schrijver.

The work of Gu et al. and Dash is similar to Jeroslow's work in that specific "worst-case" examples are presented. In this paper we build on Chvátal's results, which are concerned with average-case performance over a class of random instances. We add all simple lifted cover inequalities to his formulation and show that an exponential number of branch-and-bound nodes is required with probability converging to 1. This result is not suggested by the NP-hardness of binary knapsack problems, because cover inequality separation for these problems is NP-hard [6].

1 Statement of the result

Following Chvátal [1], we consider the following class of knapsack instances:

$$\begin{array}{rrcll} \max & \sum_{i=1}^{n} a_i x_i & & & \\ s.t. & \sum_{i=1}^{n} a_i x_i & \leq & \lfloor \frac{\sum_{i=1}^{n} a_i}{2} \rfloor & \\ & x_i & \in & \{0,1\} & i = 1, \ldots, n, \end{array} \tag{1}$$

where the coefficients a_i are integers selected independently and uniformly such that $1 \leq a_i \leq 10^{n/2}$.

For ease of discussion, we denote the right-hand-side of the inequality by $r \equiv \lfloor \frac{\sum_{i=1}^{n} a_i}{2} \rfloor$ and the upper bound on coefficients by $B \equiv 10^{n/2}$.

Rather than a standard branch-and-bound framework, Chvátal considered a slight generalization, a class of algorithms that he called *recursive algorithms*. These have the capabilities of branching, fathoming, dominance, and improving the current solution. In particular, branching is performed on a single variable, though the selection of branching variable and the process of exploring nodes may be arbitrary. In terms of branch-and-bound, dominance allows the removal of a node if there is another node with the same set of fixed variables that has—considering only the fixed variables—at least as much slack in the constraint and at least as good an objective value. For a precise definition of this class of algorithms, see [1].

We will present our results using the language of branch-and-bound, though our results do apply to Chvátal's class of recursive algorithms.

Theorem 1 (Chvátal) *With probability converging to 1 as $n \rightarrow \infty$, every recursive algorithm (as described in the previous paragraph) operating on an instance of (1) will create at least $2^{n/10}$ nodes in the process of solving.*

For a knapsack problem with constraint $\sum_{i=1}^{n} a_i x_i \leq b$, a *cover* is a set $C \subseteq \{1, \ldots, n\}$ such that $\sum_{i \in C} a_i > b$. A *minimal cover* is a cover C such that no subsets of C are covers. A minimal cover C defines the following *cover inequality*, which is a valid inequality for the knapsack problem:

$$\sum_{i \in C} x_i \leq |C| - 1.$$

Although cover inequalities are not facet-defining in general, they can be strengthened to form facet-defining inequalities through a process called *lifting*. We will consider a special case of a lifted cover inequality called a *simple lifted cover*. Given a cover C, a simple lifted cover inequality has the form

$$\sum_{i \in C} x_i + \sum_{i \notin C} \alpha_i x_i \leq |C| - 1,$$

The values α_i are called *lifted coefficients* and are determined through a process called *sequential lifting*. See Gu et al.[3] or Wolsey [8] for discussions of lifted cover inequalities. Here we describe the process briefly for simple lifted cover inequalities.

Definition 2 *The sequential lifting process for simple cover inequalities is as follows. Let C be the cover. Let the indices not in C be ordered arbitrarily $i_1, i_2, \ldots, i_m$.*

1. *Initialize $K = \emptyset$, $a = 1$.*
2. *Let $j = i_a$.*
3. *Determine lifted coefficient α_j as follows:*

$$\alpha_j = |C| - 1 - \max\left\{\sum_{i \in C} x_i + \sum_{k \in K} \alpha_k x_k : x \in S, x_j = 1\right\}, \qquad (2)$$

   where S is the set of feasible integer solutions to the original knapsack problem.

4. *Set $K = K \cup \{j\}$, and $a = a + 1$.*
5. *If $a \leq m$, return to Step 2.*

Note that $\alpha_j \leq |C| - 1$. Also note that by induction, (2) shows that α_j is integer for all j.

Gu et al. [4] considered the use of simple lifted cover inequalities on knapsack problems. They showed that branch-and-cut using simple lifted cover inequalities requires an exponential number of nodes for the following set of knapsack instances, parametrized by scalar n and vectors δ and ξ:

$$\begin{array}{rlcl} \max & \sum_{j=1}^{12n}(2\theta - \xi_j)x_j + \sum_{12n+1}^{20n}(3\theta - \xi_j)x_j & & \\ s.t. & \sum_{j=1}^{12n}(2 \cdot 2^n - \delta_j)x_j + \sum_{j=12n+1}^{20n}(3 \cdot 2^n - \delta_j)x_j & \leq & 6n \cdot 2^n \\ & x & \in & \{0,1\}^{20n}, \end{array} \qquad (3)$$

where $n \geq 10$, $\theta = (60n \cdot 2^n)^{20n+1}$, $\delta_j \in \{1, \ldots, \lfloor 2^{n-1}/3n \rfloor\}$ for all $1 \leq j \leq 20n$, and $\xi_j \in \{1, \ldots, 2^n\}$ for all $1 \leq j \leq 20n$.

Like system (1), system (3) requires large coefficients. Note that (3) can be viewed as perturbations of the underlying instance given when $\delta_j = \xi_j = 0$ for all j.

We consider the same random instances as Chvátal but with the presence of simple lifted cover inequalities. We assume that all simple lifted cover inequalities are present, so our results represent the best possible performance of a branch-and-cut algorithm.

The central result of this paper is the following:

Theorem 3 *With probability going to 1, every branch-and-bound algorithm that branches on variables operating on an instance of (1) with the addition of all simple lifted cover inequalities will create at least $2^{n/30}$ branch-and-bound nodes in the process of solving.*

Section 2 presents several properties that an instance of (1) possesses with probability going to 1. The fact that they occur with probability going to 1 is proved in Section 3. Section 4 proves that any instance possessing the properties will require an exponential number of branch-and-bound nodes, which leads to the proof of Theorem 3. Conclusions appear in Section 5.

2 Properties of the random instances

For convenience in later discussion, let the knapsack coefficients be labeled so that $a_1 \leq a_2 \leq \cdots \leq a_n$. As before, we denote the upper bound of the distribution of coefficients by $B \equiv 10^{n/2}$ and the right-hand-side of the knapsack inequality by $r \equiv \lfloor \frac{1}{2} \sum a_i \rfloor$.

Let $\delta > 0$ be a constant that will be chosen later. We consider instances that possess the following properties:

1. For every q such that $\frac{n}{100} \leq q \leq \frac{99n}{100}$, the qth smallest coefficient, a_q, satisfies $a_q < \frac{q}{n+1} B(1+\delta)$.

2. The right-hand-side of the knapsack constraint, $r \equiv \lfloor \frac{1}{2} \sum a_i \rfloor$, satisfies

$$\frac{nB}{4}(1-\delta) < r < \frac{nB}{4}(1+\delta).$$

3. All covers include at least 7 variables with coefficients larger than $\frac{3}{5}B$.

We will refer to these as Properties 1, 2, and 3, respectively.

3 Instances possess the properties with probability going to 1

In this section we show that Properties 1, 2, and 3 are satisfied by an instance of (1) with probability going to 1 as n increases. We actually prove a slightly stronger type of convergence: with probability 1, the properties are eventually satisfied as n increases.

Let Y_n be a random variable following the empirical measure defined by the random sample coefficient values $a_1 \dots a_n$. (The empirical measure assigns mass $1/n$ to each value $a_i : i = 1 \dots n$.) Note that $E[Y_n]$ is the sample mean, $\sum_{i=1}^{n} a_i/n$, and more generally $E[Y_n | Y_n \leq v]$ is the average value of those coefficients less than or equal to v.

Lemma 4 *Let $\delta > 0$, $c > 0$, and $0 < t < 1$ be arbitrary constants. Let a_q be the qth smallest coefficient, and let Y_n be as defined just above. Almost surely as $n \to \infty$, for every q such that $cn < q < (1-c)n$, the following three relations hold:*

$$\frac{q}{n+1} B(1-\delta) < a_q \leq \frac{q}{n+1} B(1+\delta),$$

$$\frac{B}{2}(1-\delta) < E[Y_n] < \frac{B}{2}(1+\delta),$$

$$\frac{tB}{2}(1-\delta) < E[Y_n|Y_n \le tB] < \frac{tB}{2}(1+\delta).$$

Proof: Let X be a random variable with uniform distribution $U(0,1)$. Then a random variable for a coefficient in an instance of (1) can be generated by the transformation $\lceil BX \rceil$.

Let $U(t)$ be the cdf for X. Given a sample size n, consider the empirical measure cdf $U_n(t)$, which is the number of samples less than or equal to t divided by n. The Glivenko-Cantelli theorem states that $U_n(t)$ converges uniformly and almost surely to $U(t)$ [7]. That is, as $n \to \infty$,

$$\sup_t |U_n(t) - U(t)| \to 0 \ a.s.$$

By definition, and by the stated transformation, $a_q \le \frac{q}{n+1}B(1+\delta)$ iff $U_n(\lfloor \frac{q}{n+1}B(1+\delta)\rfloor/B) \ge q/n$. Similarly, $a_q > \frac{q}{n+1}B(1-\delta)$ iff $U_n(\lfloor \frac{q}{n+1}B(1-\delta)\rfloor/B) < q/n$. The first relation now follows from $U_n(t) \to U(t) = t \ \forall 0 < t < 1 \ a.s.$

For the second and third relations, let X_n be a random variable governed by U_n. Then $E[X_n]$ is the mean value of the sample. By the uniform convergence of Glivenko-Cantelli, $E[X_n] \to E[X]$ and for all $t > 0$, $E[X_n|X_n \le t] \to E[X|X \le t] = t/2$, almost surely. This follows from the continuity of the conditional mean functional because $E[X|X \le t] = (1/t)\int_0^t (1-U(t))dt$.

From the transformation that generates coefficients, $B \cdot E[X_n] \le \frac{\sum_i a_i}{n} = E[Y_n] \le 1 + B \cdot E[X_n]$. Since $E[X_n]$ converges to $\frac{1}{2}$ almost surely, this implies that with probability 1, for all $\delta > 0$, $(1-\delta)(B/2) \le E[Y_n] \le (1+\delta)(B/2)$ eventually as $n \to \infty$. This gives the second relation.

To generalize to the third relation, $E[Y_n|Y_n \le tB] = E[\lceil BX_n \rceil | \lceil BX_n \rceil \le tB] = E[\lceil BX_n \rceil | BX \le \lfloor tB \rfloor] \le 1 + E[BX_n|BX_n \le tB]$. Combining this with $E[X_n|X_n \le t] \to t/2 \ a.s.$ gives the right hand inequality, and the left hand inequality derivation is very similar. ∎

The first relation of Lemma 4 proves Property 1 by taking $c = \frac{1}{100}$.

Lemma 5 *Property 2: For any constant $\delta > 0$, the right-hand-side of the knapsack constraint, $r \equiv \lfloor \frac{1}{2}\sum a_i \rfloor$, almost surely satisfies*

$$\frac{nB}{4}(1-\delta) < r < \frac{nB}{4}(1+\delta),$$

eventually as $n \to \infty$.

Proof: From Lemma 4, we know that for arbitrary $\delta' > 0$, $\frac{B}{2}(1-\delta') < E[Y_n] < \frac{B}{2}(1+\delta')$ almost surely as $n \to \infty$. Since $r = \lfloor \frac{n}{2}E[Y] \rfloor$, this directly gives

$$\lfloor \frac{nB}{4}(1-\delta') \rfloor \le r < \frac{nB}{4}(1+\delta').$$

Given δ, we select $\delta' < \delta$ such that $\frac{nB}{4}(1-\delta) < \frac{nB}{4}(1-\delta') - 1$ holds eventually as $n \to \infty$. ■

The next lemma states that the sum of all the coefficients less than $\frac{3}{5}B$ is not enough to form a cover with probability going to 1. This is used in Lemma 7 to show that there are some large coefficients in any cover.

Lemma 6 *There exists a constant $\delta_1 > 0$ such that for all $0 < \delta < \delta_1$, the following relation holds almost surely as $n \to \infty$:*

$$\sum_{\{i: a_i \leq \frac{3}{5}B\}} a_i < \frac{nB}{4}(1-\delta).$$

Proof: The given sum is equivalent to $E[Y_n | Y_n \leq \frac{3}{5}]$ times the number of coefficients no more than $\frac{3}{5}$. The first relation of Lemma 4 shows that for all $\delta' > 0$, almost surely no more than $(\frac{3}{5} + \delta')n$ coefficients are in the summation. Set $t = \frac{3}{5}$ and use the third relation in Lemma 4 to conclude that for sufficiently large n and small δ',

$$\sum_{\{i: a_i \leq \frac{3}{5}B\}} a_i < (\frac{3}{5} + \delta')n \frac{3}{5}\frac{B}{2}(1+\delta') < \frac{10}{50}nB(1+\delta') = \frac{nB}{5}(1+\delta').$$

By choosing δ' small enough, clearly there exists $\delta_1 > 0$ such that $\frac{nB}{5}(1 + \delta') < \frac{nB}{4}(1-\delta_1)$, which proves the lemma. ■

Lemma 7 *Property 3: Eventually, all covers have at least 7 coefficients greater than 3/5, with probability 1.*

Proof: Using Lemma 6, we see that there must be at least one coefficient greater than 3/5. In fact, by considering the proof of Lemma 6, we see that the gap between the two quantities is almost surely

$$\frac{nB}{4}(1-\delta) - \frac{nB}{5}(1+\delta) = \Omega(nB),$$

if $0 < \delta < \frac{1}{9}$ is constant.

Since coefficients are bounded by B, this proves that $\Omega(n)$ coefficients greater than 3/5 are needed in order to form a cover. In particular, almost surely as $n \to \infty$ we have at least 7 such coefficients. ■

Theorem 8 *Let $0 < \delta < \delta_1$ be constant, where δ_1 is a constant satisfying Lemma 5. Then with probability 1 as $n \to \infty$, an instance of (1) eventually has Properties 1, 2, and 3.*

Proof: By Lemmas 4, 5, and 6, each of the three properties eventually holds with probability 1. The intersection of this finite number of events also eventually holds with probability 1. ■

4 Instances that satisfy the properties require exponential trees

In this section, we show that instances possessing properties 1, 2, and 3 will require an exponential number of branch-and-bound nodes to solve. Together with Theorem 8, this will prove Theorem 3.

This section is split into three stages. Section 4.1 presents results on the lifted coefficients in any cover. Section 4.2 uses these results to prove a central lemma about the form of lifted cover inequalities. Section 4.3 uses the lemmas to prove Theorem 3.

4.1 Lifted coefficients are small

In this section we prove that lifted coefficients have value 0, 1, or 2, with at most a small number of coefficients with value 2. For the first lemma, recall from (1) that a_j is the knapsack coefficient of variable x_j, which corresponds to lifted coefficient α_j.

Lemma 9 *Consider the lifting process of Definition 2. In step 3 of the process, if the lifted coefficient is α_j, then the corresponding knapsack coefficient a_j must be at least as large as the optimal objective value of the following IP.*

C is the cover set, and K is as defined in Definition 2. There are binary variables x_i and y_i for each $i \in C \cup K$ and a single continuous variable Δ. We continue to use $r \equiv \lfloor \frac{1}{2} \sum_{i=1}^{n} a_i \rfloor$.

$$\begin{array}{llcll} \max & \sum_{i \in C \cup K} a_i y_i + \Delta & & & \\ s.t. & \sum_{i \in C} x_i + \sum_{k \in K} \alpha_k x_k & = & |C| - 1 & \\ & \Delta & = & r - \sum_{i \in C \cup K} a_i x_i & \\ & \Delta & \geq & 0 & \\ & \sum_{i \in C} y_i + \sum_{k \in K} \alpha_k y_k & = & \alpha_j - 1 & \\ & y_i & \leq & x_i & \forall i \in C \cup K \\ & x_i, y_i & \in & \{0, 1\} & \forall i \in C \cup K \end{array} \tag{4}$$

Note that Δ is in fact constrained to be integer since all coefficients and variables are integer.

Proof: We will think of the binary variable vectors x and y as representing subsets of the original variables. For example, to represent the cover set C with x, we would set $x_i = 1$ for $i \in C$ and $x_i = 0$ otherwise.

First note that the IP is always feasible. To see this, let x represent the cover set C less any one element, so that the first equality of (4) is satisfied. Such a set x cannot violate the original knapsack constraint, so Δ as defined in the second equality must be nonnegative, satisfying the third constraint. Let y represent any subset of x of size $\alpha_j - 1$. Since $\alpha_j \leq |C| - 1$ by Definition 2, such subsets exist. This value for y satisfies the fourth and fifth constraints in (4), so the solution is feasible.

Assume the lemma is not true, so that a_j is smaller than the optimal objective value to (4). Let (x^*, y^*, Δ^*) be an optimal solution to (4) and let $z^* = x^* - y^*$. Note that $z^* \in \{0,1\}^n$, since $y_i^* = 1$ implies $x_i^* = 1$.

Define $\tilde{z}$ by $\tilde{z}_j = 1$, $\tilde{z}_i = z_i^*$ for all $i \in C \cup K$, and $\tilde{z}_i = 0$ otherwise.

Consider $\tilde{z}$ in light of the maximization in (2). We have

$$\sum_{i=1}^{n} a_i \tilde{z}_i = a_j + \sum_{i \in C \cup K} a_i z_i^* = a_j + \sum_{i \in C \cup K} a_i x_i^* - \sum_{i \in C \cup K} a_i y_i^* = a_j + r - \Delta^* - \sum_{i \in C \cup K} a_i y_i^*.$$

Since $a_j < \sum_{i \in C \cup K} a_i y_i^* + \Delta^*$ by assumption, this gives $\sum_{i=1}^{n} a_i \tilde{z}_i < r$. Therefore, $\tilde{z}$ satisfies the knapsack constraint and $\tilde{z} \in S$.

The value of the maximization in (2) is given by

$$\begin{aligned} \sum_{i \in C} \tilde{z}_i + \sum_{k \in K} \alpha_k \tilde{z}_k &= \sum_{i \in C} z_i^* + \sum_{k \in K} \alpha_k z_k^* = \sum_{i \in C} x_i^* + \sum_{k \in K} \alpha_k x_k^* - \sum_{i \in C} y_i^* - \sum_{k \in K} \alpha_k y_k^* \\ &= |C| - 1 - (\alpha_j - 1) = |C| - \alpha_j. \end{aligned}$$

This proves that $\alpha_j \leq |C| - 1 - (|C| - \alpha_j) = \alpha_j - 1$. This contradiction proves the lemma. ■

Lemma 10 *For instances that possess Property 3, all simple lifted cover inequalities will have $\alpha_i \leq 2$ for all i.*

Proof: Assume that $\alpha_j \geq 3$. Let x be defined by the original cover, and let y represent the set containing the two largest coefficients from the cover. Let these two coefficients be a_k and a_l. These values of x and y satisfy (4), so the objective value $a_k + a_l + \Delta$ gives a lower bound on the optimum. By Property 3 we know that a_k and a_l are greater than $\frac{3}{5}B$, so we conclude that $a_j > \frac{6}{5}B + \Delta > \frac{6}{5}B$. But this is impossible, so the lemma is proved by contradiction. ■

Lemma 11 *For instances that possess Property 3, the number of indices for which $\alpha_i = 2$ in any simple lifted cover inequality is no more than 3.*

Proof: Consider IP (4) of Lemma 9. We will consider lower bounds on the objective value. In this case the α_j value is 2, so y represents a set containing a single variable, either from the cover set or one with a lifted coefficient of 1.

Assume for a contradiction that there are at least four lifted coefficients with value 2 and that the first four are $\alpha_{i_1}, \alpha_{i_2}, \alpha_{i_3}$, and α_{i_4}.

For α_{i_1}, let $x^{(1)}$ be given by the variables in the cover C with the $|C| - 1$ largest coefficients a_i. Clearly this leads to a feasible solution $(x^{(1)}, y^{(1)}, \Delta^{(1)})$ to (4). Let the 7 largest coefficients in C be $k_1, k_2, \ldots, k_7$. These must exist by Property 3.

Consider α_{i_2}. We construct a feasible solution $x^{(2)}$ by "trading" indices k_1 and k_2 for i_1. Specifically, construct $x^{(2)}$ by setting $x_{i_1}^{(2)} = 1, x_{k_1}^{(2)} = 0, x_{k_2}^{(2)} = 0$, and $x_i^{(2)} = x_i^{(1)}$ for all other indices i. To see that $x^{(2)}$ satisfies the first equality in (4), note that we have replaced two variables from the cover with $\alpha_{i_1} = 2$. Any choice of a single variable for y from $x^{(2)}$ other than i_1 leads to a feasible system. The value for Δ is given by

$$\Delta^{(2)} = r - \sum a_i x_i^{(2)} = r - \sum a_i x_i^{(1)} + a_{k_1} + a_{k_2} - a_{i_1} = \Delta^{(1)} + a_{k_1} + a_{k_2} - a_{i_1}.$$

By Property 3, the two coefficients a_{k_1} and a_{k_2} are each over $\frac{3}{5}B$ while a_{i_1} is most B, so $\Delta^{(2)} > \Delta^{(1)} + \frac{3}{5}B + \frac{3}{5}B - B = \Delta^{(1)} + \frac{1}{5}B$.

For α_{i_3}, we construct a solution $x^{(3)}$ by "trading" the next two largest coefficients, k_3 and k_4, for α_{i_2}. That is, let $x_{i_2}^{(3)} = 1, x_{k_3}^{(3)} = 0, x_{k_4}^{(3)} = 0$, and $x_i^{(3)} = x_i^{(2)}$ for all other i. This leads to a feasible solution as before, and by Property 3, $\Delta^{(3)} > \Delta^{(2)} + \frac{1}{5}B > \Delta^{(1)} + \frac{2}{5}B$.

Finally, for α_{i_4}, we construct $x^{(4)}$ by removing k_5 and k_6 and adding i_3. By Property 3, $\Delta^{(4)} > \Delta^{(1)} + \frac{3}{5}B$. In the objective of (4), y represents a single variable in the set represented by x. We can choose x_{k_7}, which is at least $\frac{3}{5}B$ by Property 3. Therefore we have shown that $a_{i_4} > \Delta^{(1)} + \frac{3}{5} + \frac{3}{5} > \frac{6}{5}$. This is impossible, which proves that we cannot have four lifted coefficients with value 2. ■

4.2 The ratio of cover size to sum of coefficients

In this section, we present the key lemma leading to the proof of Theorem 3.

Lemma 12 *There exists $\delta_2 > 0$ such that for every instance that satisfies Properties 1, 2, and 3 with $0 < \delta < \delta_2$, every simple lifted cover inequality satisfies* $\frac{|C|-1}{|C|+\sum_{i \notin C} \alpha_i} > \frac{3}{5}$.

Proof: Based on Property 2 and the upper bound of B on coefficients, we need at least $\frac{n}{4}(1-\delta) > \frac{n}{5}$ variables to form a cover. Consider a simple lifted cover constraint and let T be the set of variables with non-zero coefficients. Let $t = |T|$, so we know $\frac{n}{5} \le t \le n$.

For a given t, we wish to consider the minimum value of $\frac{|C|-1}{|C|+\sum \alpha_i}$. The denominator is between t and $t+3$, since at most 3 variables have $\alpha_i = 2$ and no values of α_i are higher. For the numerator, we would like to know how small the cover itself can be.

Since the lifted cover inequality is a valid inequality, it must be the case that no feasible solution to the knapsack problem has more than $|C|-1$ variables from T, so any set $U \subset T, |U| \ge |C|$ must satisfy $\sum_{i \in U} a_i > r$. We get a lower bound on the size of $|C|$ by considering the variables in T with the smallest coefficients and determining the number of them that it takes to exceed r. For fixed t, we want the lowest of these lower bounds. This occurs when T contains the variables with the t highest coefficients overall.

We will show that even in this case, at least $3/5$ of the coefficients are needed in the cover. Assume that the coefficients are indexed so that $a_1 \le a_2 \le \cdots \le$

a_n. Then we are considering the coefficients $a_{n-t+1}, a_{n-t+2}, \ldots, a_{n-\lceil \frac{2}{5}t \rceil}$, and their sum,

$$X \equiv \sum_{q=n-t+1}^{n-\lceil \frac{2}{5}t \rceil} a_q. \tag{5}$$

By Property 1, the qth smallest coefficient is no more than $\frac{q}{n+1}B(1+\delta)$, for $\frac{n}{100} \le q \le \frac{99n}{100}$. Let Y be the contribution to the sum from values of q outside this range. The contribution to Y from values of $q < \frac{n}{100}$ is at most $\frac{n}{100}\left(\frac{B}{100}+1\right)(1+\delta)$, while the contribution from the values of $q > \frac{99n}{100}$ is at most $\frac{n}{100}B$. Thus,

$$Y < \frac{n}{100}B + \frac{n}{100}\left(\frac{B}{100}+1\right)(1+\delta) < \frac{n}{100}\left(\frac{101B}{100}+1\right)(1+\delta).$$

Using the bounds on a_q, we have

$$\begin{aligned}
X &\le Y + \sum_{q=n-t+1}^{n-\lceil \frac{2}{5}t \rceil} \frac{q}{n+1}B(1+\delta) \\
&= Y + \frac{B}{n+1}(1+\delta) \sum_{q=n-t+1}^{n-\lceil \frac{2}{5}t \rceil} q \\
&= Y + \frac{B}{n+1}(1+\delta)\frac{1}{2}(n-t+1+n-\lceil\frac{2}{5}t\rceil)(t-\lceil\frac{2}{5}t\rceil) \\
&\le Y + \frac{B}{2(n+1)}(1+\delta)(2n+1-\frac{7}{5}t)(\frac{3}{5}t)
\end{aligned}$$

By taking the derivative with respect to t, we find that this value is maximized when $t = \frac{5}{7}n + \frac{5}{14}$, at which point we have

$$\begin{aligned}
X &\le Y + \frac{B}{2(n+1)}(1+\delta)\frac{3}{7}(n+\frac{1}{2})^2 \\
&\le Y + \frac{3}{14}B(n+\frac{1}{2})(1+\delta) \\
&\le \frac{n}{100}\left(\frac{101B}{100}+1\right)(1+\delta) + \frac{3}{14}B(n+\frac{1}{2})(1+\delta).
\end{aligned}$$

We need to show that $X < r$. By Property 2, that is equivalent to

$$\frac{n}{100}\left(\frac{101B}{100}+1\right)(1+\delta) + \frac{3}{14}B(n+\frac{1}{2})(1+\delta) < \frac{nB}{4}(1-\delta). \tag{6}$$

Since $\frac{n}{100}\left(\frac{101B}{100}+1\right) + \frac{3}{14}B(n+\frac{1}{2}) < \frac{nB}{4}$ for sufficiently high n, we can choose $\delta' > 0$ such that (6) holds for all $0 < \delta < \delta'$.

The one other requirement on δ is that it be less than δ_1 from Theorem 8. Therefore, we choose $\delta_2 < \min\{\delta', \delta_1\}$ and have proved the lemma. ■

Lemma 12 has given the final condition on the choice of δ for Properties 1, 2, and 3 from Section 2. Specifically, we choose $0 < \delta < \delta_2$ so that all the previous lemmas will hold.

4.3 Chvátal's class of problems requires an exponential tree even in the presence of simple lifted cover inequalities

We are now ready to prove Theorem 3. In part this is based on Chvátal's proof. The key additional idea comes from Lemma 12.

Proof of Theorem 3: We claim that if no more than $n/30$ variables are fixed by branching, then the LP solution of the resulting node cannot be fathomed.

Chvátal [1] proved that with probability going to 1 the following properties hold:

4. $\sum_{i \in I} a_i \leq r$ whenever $|I| \leq n/10$.

5. There is no set $I \subset \{1, 2, \ldots, n\}$ such that $\sum_{i \in I} a_i = r$.

By Theorem 8, Properties 1, 2, and 3 also hold with probability going to 1.

By Property 4, fixing at most $n/30 < n/10$ variables leaves the LP relaxation feasible. We will show that the LP relaxation with probability going to 1 has optimal objective value r. Then by Property 5, the node cannot be fathomed.

We claim that by setting all unfixed variables to $\frac{11}{20}$, the left-hand-side sum of the knapsack constraint will exceed r and no simple lifted cover constraints will be violated. Reducing the value of some of the unfixed variables will then give a feasible fractional solution with left-hand-side sum—and therefore objective value—of r exactly.

First we check that we can exceed r in the objective. Let F be the set of fixed variables, so we are interested in $\sum_{i \notin F} \frac{11}{20} a_i = \frac{11}{20} \sum_{i \notin F} a_i$. Since the maximum coefficient value is B, we have

$$\begin{aligned} \sum_{i=1}^{n} a_i &\geq 2r \\ \sum_{i \notin F} a_i &\geq 2r - \sum_{i \in F} a_i \\ \sum_{i \notin F} a_i &\geq 2r - \frac{n}{30} B. \end{aligned}$$

Using the lower bound for r from Property 2 gives that for any constant $\delta_1 > 0$, the following holds with probability going to 1:

$$\frac{11}{20} \sum_{i \notin F} a_i \geq \frac{11}{20} \left(\frac{nB}{2}(1 - \delta_1) - \frac{nB}{30} \right) = \frac{nB}{4} \left(\frac{11}{10}(1 - \delta_1) - \frac{11}{150} \right).$$

Choose $\delta_1 > 0$ and $\delta_2 > 0$ such that $\frac{11}{10}\delta_1 + \delta_2 < \frac{4}{150}$. Then $\delta_1 < \frac{10}{11} \cdot \frac{4}{150} - \frac{10}{11}\delta_2$, and

$$\frac{11}{20} \sum_{i \notin F} a_i \geq \frac{nB}{4} \left(\frac{11}{10}(1 - \delta_1) - \frac{11}{150} \right)$$

$$
\begin{aligned}
&> \frac{nB}{4}\left(\frac{11}{10} - (\frac{4}{150} - \delta_2) - \frac{11}{150}\right) \\
&= \frac{nB}{4}(1+\delta_2) > r.
\end{aligned}
$$

This shows that with probability going to 1, our fractional solution has objective value r.

Next we check that no simple lifted cover inequalities are violated. Let an inequality be specified by the sets C and K. From Lemma 11, we know that with probability going to 1 at most 3 lifted coefficients have value 2. Then we need to verify that

$$\sum_{i\in C} x_i + \sum_{i\in K} \alpha_i x_i \le |C| - 1. \tag{7}$$

A worst-case assumption is that all fixed variables are fixed to 1 and that all of them appear in an inequality, so we have

$$\sum_{i\in C} x_i + \sum_{i\in K} \alpha_i x_i \le \sum_{i\in C\cup K} x_i + 3 \le \frac{n}{30} + 3 + (|C| + |K| - \frac{n}{30})\frac{11}{20}.$$

Lemma 12 showed that $\frac{|C|-1}{|C|+\sum_{k\in K}\alpha_k} > 3/5$ with probability going to 1 for all simple lifted cover inequalities, from which we will use $|K| < \frac{2|C|}{3} - \frac{5}{3}$. Using this in the previous equation, we wish to confirm

$$
\begin{aligned}
\frac{n}{30} + 3 + (\frac{5}{3}|C| - \frac{5}{3} - \frac{n}{30})\frac{11}{20} &\le |C| - 1 \\
\frac{n}{30}\cdot\frac{9}{20} + 4 - \frac{11}{12} &\le \frac{|C|}{12}.
\end{aligned}
$$

Finally, we use the fact that $|C| > \frac{n}{4}(1-\delta) > \frac{n}{5}$, with probability going to 1. This gives

$$
\begin{aligned}
\frac{n}{30}\cdot\frac{9}{20} + 4 - \frac{11}{12} &\le \frac{n}{60} \\
4 - \frac{11}{12} &\le \frac{n}{600}.
\end{aligned}
$$

This verifies that the original inequality (7) holds with probability going to 1. Therefore, with probability going to 1 there are no violated inequalities.

Note that the conclusion applies to all simple lifted cover inequalities simultaneously. That is, we relied on Lemma 12, which applies to every simple lifted cover inequality.

We have shown that with at most $n/30$ variables fixed, the LP relaxation has optimal objective value r. Therefore, no such node can be fathomed. Since we are branching on variables, there are at least $2^{n/30}$ such nodes in the branch-and-bound tree, with probability going to 1. ■

The main idea of the proof is that as long as not too many variables are fixed, the simple lifted cover inequalities do not prevent us from having all fractional variables set to values strictly greater than 1/2.

5 Conclusions

We have shown that for a particular class of random knapsack instances the branch-and-bound tree will with probability going to 1 have an exponential number of nodes even if every simple lifted cover inequality is applied. Since the knapsack problem is NP-hard, it is not surprising that there exist knapsack instances for which the branch-and-bound tree has exponential size. The fact that almost every random instance as considered in this paper requires an exponential number of nodes even with a large number of cuts present is more significant. Moreover, since cover inequality separation is NP-hard, even for binary knapsack [6], complexity considerations would not necessarily lead one to expect the result obtained here.

If this result is indicative of general branch-and-cut performance, it suggests that while cutting planes may help reduce the number of nodes, the number is still exponential most of the time. It is possible, however, that this result is specific to the type of instance being considered and that branch-and-cut algorithms perform better on other knapsack instances and other problems. Whether or not this is true is an important open question.

References

[1] Vasek Chvátal. Hard knapsack problems. *Operations Research*, 28(6):1402–1412, 1980.

[2] Sanjeeb Dash. Exponential lower bounds on the lengths of some classes of branch-and-cut proofs. Technical Report RC22575, IBM Research Division, Sep 2002.

[3] Zonghao Gu, George L. Nemhauser, and Martin W. P. Savelsbergh. Lifted cover inequalities for 0-1 integer programs: Computation. *INFORMS Journal on Computing*, 10(4):427–437, 1998.

[4] Zonghao Gu, George L. Nemhauser, and Martin W. P. Savelsbergh. Lifted cover inequalities for 0-1 integer programs: Complexity. *INFORMS Journal on Computing*, 11(1):117–123, 1999.

[5] R. G. Jeroslow. Trivial integer programs unsolvable by branch-and-bound. *Mathematical Programming*, 6:105–109, 1974.

[6] Diego Klabjan, George L. Nemhauser, and Craig Tovey. The complexity of cover inequality separation. *Operations Research Letters*, 23(1-2):35–40, 1998.

[7] David Pollard. *Convergence of Stochastic Processes.* Springer-Verlag, New York, 1984.

[8] Laurence A. Wolsey. *Integer Programming.* John Wiley & Sons, New York, 1998.