Noname manuscript No.
(will be inserted by the editor)

On the Integrality Gap of Binary Integer Programs with Gaussian Data

Sander Borst · **Daniel Dadush** ·
Sophie Huiberts · **Samarth Tiwari**

the date of receipt and acceptance should be inserted later

Abstract For a binary integer program (IP) $\max c^\mathsf{T} x, Ax \leq b, x \in \{0,1\}^n$, where $A \in \mathbb{R}^{m\times n}$ and $c \in \mathbb{R}^n$ have independent Gaussian entries and the right-hand side $b \in \mathbb{R}^m$ satisfies that its negative coordinates have ℓ_2 norm at most $n/10$, we prove that the gap between the value of the linear programming relaxation and the IP is upper bounded by $\mathrm{poly}(m)(\log n)^2/n$ with probability at least $1-2/n^7-2^{-\mathrm{poly}(m)}$. Our results give a Gaussian analogue of the classical integrality gap result of Dyer and Frieze (Math. of O.R., 1989) in the case of random packing IPs. In constrast to the packing case, our integrality gap depends only polynomially on m instead of exponentially. Building upon recent breakthrough work of Dey, Dubey and Molinaro (SODA, 2021), we show that the integrality gap implies that branch-and-bound requires $n^{\mathrm{poly}(m)}$ time on random Gaussian IPs with good probability, which is polynomial when the number of constraints m is fixed. We derive this result via a novel meta-theorem, which relates the size of branch-and-bound trees and the integrality gap for random *logconcave* IPs.

Keywords Integer Programming · Integrality Gap · Branch-and-Bound

Mathematics Subject Classification (2010) 90C10

1 Introduction

Consider the following linear program with n variables and m constraints

$$\mathrm{val}_{\mathsf{LP}}(A,b,c) := \max_x \ \mathrm{val}(x) = c^\mathsf{T} x$$

A preliminary version of this work was published in the proceedings of the 22nd conference on Integer Programming and Combinatorial Optimization (IPCO 2021) [2].

This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement QIP–805241)

This work was done while the second and third named authors were participating in a program at the Simons Institute for the Theory of Computing.

Centrum Wiskunde & Informatica (CWI), Amsterdam, The Netherlands
E-mail: {sander.borst,dadush,s.huiberts,samarth.tiwari}@cwi.nl

$$\text{s.t. } Ax \le b \qquad \text{(Primal LP)}$$
$$x \in [0,1]^n$$

Let $\text{val}_{\mathsf{IP}}(A,b,c)$ be the value of the same optimization problem with the additional restriction that x is integral, i.e., $x \in \{0,1\}^n$. Now we define the integrality gap to be the quantity $\mathsf{IPGAP}(A,b,c) := \text{val}_{\mathsf{LP}}(A,b,c) - \text{val}_{\mathsf{IP}}(A,b,c)$.

The integrality gap of integer linear programs forms an important measure for the complexity of solving said problem in a number of works on the average-case complexity of integer programming [1,3,6,5,13,18].

So far, probabilistic analyses of the integrality gap have focussed on 0–1 packing IPs and the generalized assignment problem. In particular, the entries of $A \in \mathbb{R}^{m\times n}, b \in \mathbb{R}^m, c \in \mathbb{R}^n$ in these problems are all non-negative, and the entries of b were assumed to scale linearly with n.

In this paper, we analyze the integrality gap of (Primal LP) under the assumption that the entries of A and c are all independent Gaussian $\mathcal{N}(0,1)$ distributed, and that the negative part of b is small: $\|b^-\|_2 \le n/10$.

We prove that, with high probability, the integrality gap $\mathsf{IPGAP}(A,b,c)$ is small, i.e., (Primal LP) admits a solution $x \in \{0,1\}^n$ with value close to the optimum.

Theorem 1 *There exists an absolute constant $C \ge 200$, such that, for $m \ge 1$, $n \ge Cm^{4.5}$, $b \in \mathbb{R}^m$ with $\|b^-\|_2 \le n/10$, if A and c have i.i.d. $\mathcal{N}(0,1)$ entries, then*

$$\Pr\left(\mathsf{IPGAP}(A,b,c) \ge 10^{15} \cdot t \cdot \frac{m^{2.5}(m+\log n)^2}{n}\right) \le 4 \cdot \left(1 - \frac{1}{25}\right)^t + n^{-7},$$

for all $1 \le t \le \frac{n}{Cm^{2.5}(m+\log n)^2}$.

In the previous probabilistic analyses by [6,5,18], it is assumed that $b = \beta n$ for fixed $\beta \in (0,1/2)^m$ and the entries of (A,c) are independently distributed uniformly in the interval $[0,1]$. Those works prove a similar bound as above, except that in their results the dependence on m is exponential instead of polynomial. Namely, for $\beta_{\min} := \min_{i\in[n]} \beta_i$, they require $n \ge (1/\beta_{\min})^m \ge 2^m$ and the integrality gap scales like $O(1/\beta_{\min})^m \log^2 n/n$. We note that the integrality gap in Theorem 1 does not depend on the "shape" of b (other than requiring $\|b^-\|_2 \le n/10$). We give a high-level overview of the proof of Theorem 1 in subsection 1.2, describing the similarities and differences with the analysis of Dyer and Frieze [5]

Building on breakthrough work of Dey, Dubey and Molinaro [3], we show that the integrality gap above also implies that branch-and-bound applied to the above IP produces a tree of size at most $n^{\text{poly}(m)}$ with good probability. For this purpose, we give a novel meta-theorem relating the integrality gap and the complexity of branch-and-bound for random *logconcave* IPs. We detail this in the next subsection.

1.1 Relating the Integrality Gap to Branch-and-Bound

In recent breakthrough work, Dey, Dubey and Molinaro [3] provided a framework for deriving upper bounds on the size of branch-and-bound trees for random IPs with small integrality gaps. Their framework consists of two parts. In the first part, one deterministically relates the size any branch-and-bound tree using best-bound first node selection to the size of knapsack polytopes whose weights are induced by reduced costs and whose capacity is equal to the integrality gap. We recall that in the best-bound first rule, the next node to be processed is always the node whose LP relaxation value is the largest. This is formally encoded by the following theorem, which corresponds to a slightly adapted version of [3, Corollary 2].

Theorem 2 *Consider a binary integer program of the form*

$$\begin{aligned} &\max c^\mathsf{T} x \\ &\text{s.t. } Ax \le b \\ &\quad x \in \{0,1\}^m. \end{aligned} \tag{Primal IP}$$

Then, the best bound first branch-and-bound algorithm produces a tree of size

$$n^{O(m)} \cdot \max_{\lambda \in \mathbb{R}^m} |\{x \in \{0,1\}^n : \sum_{i=1}^{n} x_i |(A^\mathsf{T}\lambda - c)_i| \le \mathsf{IPGAP}(A,b,c)\}| + 1. \tag{1}$$

In the second part of the framework, one leverages the randomness in the coefficients of A, c to upper bound the maximum size of any knapsack in (1). In [3], they give such an upper bound for the specific packing instances studied by Dyer and Frieze [5]. In the present work, we generalize their probabilistic framework to random *logconcave* IPs. We now state our main meta-theorem, which we prove in Section 7.

Theorem 3 *Let* $n \ge 100(m+1)$*,* $b \in \mathbb{R}^m$*, and* $W := \begin{bmatrix} c^\mathsf{T} \\ A \end{bmatrix} \in \mathbb{R}^{n\times(m+1)}$ *be a matrix whose columns are independent logconcave random vectors with identity covariance. Then, for* $G \ge 0$*,* $\delta \in (0,1)$*, with probability at least*

$$1 - \Pr_{A,c}[\text{IPGAP}(A,b,c) \ge G] - \delta - e^{-n/5},$$

the best bound first branch-and-bound algorithm applied to (Primal IP) *produces a tree of size at most*

$$n^{O(m)} e^{2\sqrt{2nG}} / \delta. \tag{2}$$

The class of logconcave distributions is quite rich (see subsection 2.6 for a formal definition), e.g. the uniform distribution over any convex body as well as all of its marginals are logconcave. We are therefore hopeful that interesting bounds on the size of branch-and-bound trees can be obtained for a wide range of random logconcave IPs, which by Theorem 3 reduces to obtaining suitable bounds on the integrality gap.

When A, c have i.i.d. uniform $[0,1]$ coefficients and $b = \beta n$, $\beta \in (0,1/2)^m$ and $\beta_{\min} := \min_{i\in[n]} \beta_i$, Dyer and Frieze [5] proved that for n large enough

$$\Pr_{A,c}[\text{IPGAP}(A,b,c) \ge \alpha a_1 \log^2 n/n] \le 2^{-\alpha/a_2} + 1/(2n), \forall \alpha \ge 1,$$

where $a_1 = \Theta(1/\beta_{\min})^m$ and $a_2 = 2^{\Theta(m)}$. In [3], Dey, Dubey and Molinaro use this integrality gap result combined with a probabilistic analysis of the bound in Theorem 2 to show that the tree size is at most

$$n^{O(ma_1 \log a_1 + \alpha a_1 \log m)}$$

with probability $1 - 2^{-\alpha/a_2} - 1/n$. A stronger bound can be obtained from Theorem 3.

We first observe that $2\sqrt{3}\mathrm{IPGAP}(A,b,c) = \mathrm{IPGAP}(2\sqrt{3}A, 2\sqrt{3}b, 2\sqrt{3}c)$, noting that $W = 2\sqrt{3}\begin{bmatrix} c^\mathsf{T} \\ A \end{bmatrix}$ has identity covariance. Plugging $G = 2\sqrt{3}\alpha a_1 \log^2 n/n$ into Theorem 3 with $\delta = 1/(2n) - e^{-n/5}$, we get an improved tree-size bound of

$$n^{O(m)} e^{2\sqrt{4\sqrt{3}\alpha a_1 \log^2 n}} = n^{O(m) + 4\sqrt{\sqrt{3}\alpha a_1}}.$$

Proceeding in a similar fashion, we can easily derive a tree-size bound for Gaussian IPs by combining Theorem 1 and Theorem 3.

Corollary 1 *For $C \geq 200$ as in Theorem 1, $m \in \mathbb{N}$, $n \geq Cm^{4.5}$, $A \in \mathbb{R}^{m\times n}, c \in \mathbb{R}^n$ with i.i.d. $\mathcal{N}(0,1)$ entries and $b \in \mathbb{R}^m$, $\|b^-\|_2 \leq n/10$. Then, for $1 \leq t \leq \frac{n}{Cm^{2.5}(m+\log n)^2}$, with probability at least $1 - 4(1-\frac{1}{25})^t - 2/n^7$, the size of any best bound first branch-and-bound tree for solving* (Primal IP) *is at most $e^{O(\sqrt{t}m^{2.25})} n^{O(\sqrt{t}m^{1.25})}$.*

Proof Since A,b,c satisfy the conditions of Theorem 1, for $G = 10^{15} \cdot \frac{m^{2.5}(m+\log n)^2}{n}$, we have that $\mathrm{IPGAP}(A,b,c) \geq tG$ with probability at most $4(1-\frac{1}{25})^t + 1/n^7$.

Applying Theorem 3 to W with $\delta = 1/(2n^7)$, using the fact that $W \in \mathbb{R}^{m+1\times n}$ has i.i.d. $\mathcal{N}(0,1)$ entries, with probability at least

$$1 - (4(1-\frac{1}{25})^t + 1/n^7) - \delta - e^{-n/5} \geq 1 - 4(1-\frac{1}{25})^t - 2/n^7,$$

we get that the size of the branch-and-bound tree is at most

$$n^{O(m)} e^{2\sqrt{2tGn}}/\delta \leq n^{O(m)} e^{O(\sqrt{t}m^{1.25}(m+\log n))}(2n^7) = e^{O(\sqrt{t}m^{2.25})} n^{O(\sqrt{t}m^{1.25})}. \quad \square$$

1.2 Proof Overview for Theorem 1

Our proof strategy follows along similar lines to that of Dyer and Frieze [5], which we now describe. In their strategy, one first solves an auxiliary LP $\max c^\mathsf{T}x, Ax \leq b - \varepsilon 1_m$, for $\varepsilon > 0$ small, to get its optimal solution x^*, which is both feasible and nearly optimal for the starting LP (proved by a simple scaling argument), together with its optimal dual solution $u^* \geq 0$ (see subsection 2.2 for the formulation of the dual). From here, they round down the fractional components of x^* to get a feasible IP solution $x' := \lfloor x^* \rfloor$. We note that the feasibility of x' depends crucially on the packing structure of the LPs they work with, i.e., that A has non-negative entries (which does not hold in the Gaussian setting). Lastly, they construct a nearly optimal integer solution x'', by carefully choosing a subset of coordinates $T \subset \{i \in [n] : x'_i = 0\}$ of size $O(\mathrm{poly}(m)\log n)$, where they flip the coordinates of x' in T from 0 to 1 to get

x''. The coordinates of T are chosen accordingly the following criteria. Firstly, the coordinates should be *very cheap* to flip, which is measured by the absolute value of their *reduced costs*. Namely, they enforce that $|c_i - A_{\cdot,i}^{\mathsf{T}} u^*| = O(\log n/n)$, $\forall i \in T$. Secondly, T is chosen to make the *excess slack* $\|A(x^* - x'')\|_\infty \le 1/\mathrm{poly}(n)$, i.e., negligible. We note that guaranteeing the existence of T is highly non-trivial. Crucial to the analysis is that after conditioning on the exact value of x^* and u^*, the columns of $W := \begin{bmatrix} c^{\mathsf{T}} \\ A \end{bmatrix} \in \mathbb{R}^{(m+1)\times n}$ (the objective extended constraint matrix) that are indexed by $N_0 := \{i \in [n] : x_i^* = 0\}$ are independently distributed subject to having negative reduced cost, i.e., subject to $c_i - A_{\cdot,i}^{\mathsf{T}} u^* < 0$ for $i \in N_0$ (see Lemma 11). It is the large amount of left-over randomness in these columns that allowed Dyer and Frieze to show the existence of the subset T via a discrepancy argument (more on this below). Finally, given a suitable T, a simple sensitivity analysis is used to show the bound on the gap between $c^{\mathsf{T}} x''$ and the (Primal LP) value. This analysis uses the basic formula for the optimality gap between primal and dual solutions (see (Gap Formula) in subsection 2.2), and relies upon bounds on the size of the reduced costs of the flipped variables, the total excess slack and the norm of the dual optimal solution u^*.

Adapting to the Gaussian setting. As a first difference with the above strategy, we are able to work directly with the optimal solution x^* of the original LP without having to replace b by $b' := b - \varepsilon 1_m$. The necessity of working with this more conservative feasible region in the packing setting of [5] is that flipping 0 coordinates of x' to 1 can only *decrease* $b - Ax'$. In particular, if the coordinates of $b - Ax' \ge 0$ are too small, it becomes difficult to find a set T that doesn't force x'' to be infeasible. By working with b' instead of b, they can ensure that $b - Ax' \ge \varepsilon 1_m$, which avoids this problem. In the Gaussian setting, it turns out that we have equal power to both increase and decrease the slack of $b - Ax'$, due to the fact that the Gaussian distribution is symmetric about 0. We are in fact able to simultaneously fix both the feasibility and optimality error of x', which gives us more flexibility. In particular, we will be able to use randomized rounding when we move from x^* to x', which will allow us to start with a smaller initial slack error than is achievable by simply rounding x^* down.

The Discrepancy Lemma. Our main quantitative improvement – the reduction from an exponential to a polynomial dependence in m – arises from two main sources. The first source of improvement is a substantially improved version of a discrepancy lemma of Dyer and Frieze [5, Lemma 3.4]. This lemma posits that for any large enough set of “suitably random” columns in $\mathbb{R}^m$ and any not too big target vector $D \in \mathbb{R}^m$, then with non-negligible probability there exists a set containing half the columns whose sum is very close to D. This is the main lemma used to show the existence of the subset T, chosen from a suitably filtered subset of the columns of A in N_0, used to reduce the excess slack. The non-negligible probability in their lemma was of order $2^{-O(m)}$, which implied that one had to try $2^{O(m)}$ disjoint subsets of the filtered columns before having a constant probability of success of finding a suitable T. In our improved variant of the discrepancy lemma, we show that by sub-selecting a $1/(2\sqrt{m})$-fraction of the columns instead of 1/2-fraction, we can increase the success probability to constant, with the caveat of requiring a slightly larger set of initial columns. The formal statement of our improved discrepancy lemma is given below.

Lemma 1 *For $k,m \in \mathbb{N}$, let $a = \lceil 2\sqrt{m} \rceil$ and $\theta > 0$ satisfy $\left(\frac{2\theta}{\sqrt{2\pi k}}\right)^m \binom{ak}{k} = 1$. Let $Y_1, \dots, Y_{ak} \in \mathbb{R}^m$ be i.i.d. random vectors with independent coordinates. For $k_0 \in \mathbb{N}, \gamma \geq 0, M > 0$, assume that $\forall i \in [m]$, $Y_{1,i}$ is a (γ, k_0)-Gaussian convergent continuous random variable with maximum density at most M. Then, if*

$$k \geq \max\{(4\sqrt{m}+2)k_0, 144m^{\frac{3}{2}}(\log M + 3), 150\,000(\gamma+1)m^{\frac{7}{4}}\},$$

for any vector $D \in \mathbb{R}^m$ with $\|D\|_2 \leq \sqrt{k}$ the following holds:

$$\Pr\left[\exists K \subset [ak] : |K| = k, \|(\sum_{j \in K} Y_j) - D\|_\infty \leq \theta\right] \geq \frac{1}{25}. \tag{3}$$

The notion of Gaussian convergence used above (see Section 2.7 for a formal definition), quantifies the speed at which the density of normalized sums of i.i.d. random variables converges to the standard Gaussian density. This definition will in fact enforce that the entries of all the vectors in Lemma 1 have mean 0 and variance 1. Apart from the increased probability of success, we improve many other aspects of [5, Lemma 3.4]. In particular, we remove the restriction that the entries be bounded random variables, and we support targets of norm exactly $\sqrt{k}$ instead of k^α, for any $\alpha < 1/2$. Furthermore, [5, Lemma 3.4] is proved only in the asymptotic regime where $k \to \infty$, whereas we give explicit parameter dependencies, which are all polynomial in m. Taken together, these improvements make the lemma easier to use and more flexible, which should enable further applications. We refer the reader to Section 6 for more details.

Reduced cost filtering. The second source of improvement is the use of a much milder filtering step mentioned above. In both the uniform and Gaussian case, the subset T is chosen from a subset of N_0 associated with columns of A having reduced costs of absolute value at most some parameter $\Delta > 0$. The probability of finding a suitable T increases as Δ grows larger, since we have more columns to choose from, and the target integrality gap scales linearly with Δ, as the columns we choose from become more expensive as Δ grows. Depending on the distribution of c and A, the reduced cost filtering induces non-trivial correlations between the entries of the corresponding columns of A, which makes it difficult to use them within the context of the discrepancy lemma. To deal with this problem in the uniform setting, Dyer and Frieze filtered more aggressively, by additionally restricting to the columns of A lying in a sub-cube $[\alpha, \beta_{\min}]^m$, where $\alpha = \Omega(\log^3 n/n)$ and $\beta_{\min} := \min_{i \in [m]} \beta_i$ as above. This allowed them to ensure that the distribution of the filtered columns in A is uniform in $[\alpha, \beta_{\min}]^m$, thereby removing the unwanted correlations. In the packing setting, both aggressive and reduced cost filtering can have success probability $\Theta(\beta_{\min})^m \Delta$, so aggressive filtering is not much more expensive than reduced cost filtering. For an illustrative calculation, if $u^* = 1_m/(m\beta_{\min})$ and $\Delta \leq 1$, then for $i \in N_0$,

$A_i \in [0,1]^m, c_i \in [0,1]$ distributed uniformly, the reduced cost filtering probability is essentially equal to

$$\Pr[A_i^\mathsf{T} u^* - c_i \in [0,\Delta]] \leq \Delta \Pr[\sum_{j=1}^m A_{ij}/(m\beta_{\min}) \leq 2]$$
$$\leq \Delta \mathrm{vol}_m(\{a \geq 0 : \sum_{j=1}^m a_j \leq 2m\beta_{\min}\}) \leq \Delta(2e\beta_{\min})^m.$$

When A_i, c_i have $\mathcal{N}(0,1)$ entries however, we have $\Pr_{A_i,c_i}[A_i^\mathsf{T} u^* - c_i \in [0,\Delta]] = \Theta(\Delta/\|(1,u^*)\|_2)$ for $\Delta \in [0,1]$. Given this much larger success probability, we show how to work with only reduced cost filtering in the Gaussian setting. While the entries of the filtered columns of A do indeed correlate, using the rotational symmetry of the Gaussian distribution, we show that after applying a suitable rotation R, the coordinates of the filtered columns of RA are all independent (see Lemma 12). This allows us to apply the discrepancy lemma in a "rotated space", thereby completely avoiding the correlation issues in the uniform setting.

Sparsity of x^* and boundedness of u^*. As already mentioned, we are also able to substantially relax the rigid requirements on the right hand side b and to remove any stringent "shape-dependence" of the integrality gap on b. Specifically, the shape parameter $\beta_{\min}$ above is used to both lower bound $|N_0|$ by roughly $\Omega((1-2\beta_{\min})n)$, the number of zeros in x^*, as well as upper bound the ℓ_1 norm of the optimal dual solution u^* by $O(1/\beta_{\min})$ (this a main reason for the choice of the $[\alpha, \beta_{\min}]^m$ subcube above). These bounds are both crucial for determining the existence of T. In the Gaussian setting, we are able to establish $|N_0| = \Omega(n)$ and $\|u^*\|_2 = O(1)$, using only that $\|b^-\|_2 \leq n/10$. Due to the different nature of the distributions we work with, our arguments to establish these bounds are completely different from those used by Dyer and Frieze. Firstly, the lower bound on $|N_0|$, which is strongly based on the packing structure of the IP in [5], is replaced by a sub-optimality argument. Namely, we show that the objective value of any LP basic solution with too few zero coordinates must be sub-optimal, using the concentration properties of the Gaussian distribution (see Lemma 10). The upper bound on the ℓ_1 norm of u^* in [5] is deterministic and based on packing structure; namely, that the objective value of a Primal LP of packing-type is at most $\sum_{i=1}^n c_i \leq n$ (since $c_i \in [0,1], \forall i \in [m]$). In the Gaussian setting, we prove our bound on the norm of u^* by first establishing simple upper and lower bounds on the dual objective function, which hold with overwhelming probability, and optimizing over these simple approximations (see Lemma 8).

Future directions. Given the above, a first question is whether one can extend the integrality gap argument above to a larger class of logconcave IPs. An important technical difficulty is to understand whether Lemma 1 can be generalized to handle random columns whose entries are allowed to have non-trivial correlations and whose entries have non-zero means. A second question is whether one can improve the current parameter dependencies, both in terms of improving the integrality gap and relaxing the restrictions on b. For this purpose, one may try to leverage flipping both

0s to 1 and 1s to 0 in the rounding of x' to x''. The columns of W associated with the one coordinates of x^* are no longer independent however. A final open question is whether these techniques can be extended to handle discrete distributions on A and c.

1.3 Related Work

The worst-case complexity of solving $\max\{c^\top x : Ax = b, x \geq 0, x \in \mathbb{Z}^n\}$ scales as $n^{O(n)}$ times a polynomial factor in the bit complexity of the problem. This is a classical result due to Lenstra [16] and Kannan [15] which is based on lattice basis reduction techniques.

Beyond these worst-case bounds, the performance of basis reduction techniques for determining the feasibility of random integer programs has been analyzed. In this context, basis reduction is used to reformulate $Ax \leq b, x \in \mathbb{Z}^n$ as $AUw \leq b, w \in \mathbb{Z}^n$ for some unimodular matrix $U \in \mathbb{Z}^{n\times n}$, after which a simple variable branching scheme is applied (i.e., branching on integer hyperplanes in the original space). Furst and Kannan [11] showed that subset-sum instances of the form $\sum_{i=1}^n x_i a_i = b$, $x \in \{0,1\}^n$, where each a_i, $i \in [n]$, is chosen uniformly from $\{1,\ldots,M\}$ and $b \in \mathbb{Z}_+$, can be solved in polynomial time with high probability in this way if $M = 2^{\Omega(n^2)}$. Pataki, Tural and Wong [21] proved generalizations of this result for IPs of the form $f \leq Ax \leq g, l \leq x \leq u, x \in \mathbb{Z}^n$, where the coefficients of A are uniform in $\{1,\ldots,M\}$ and M is "large" compared to $\|(g-f, u-l)\|$. Apart from the different type of branching, compared to the present work, we note that the IPs analyzed in these models are either infeasible or have a unique feasible solution with high probability.

Another line of works has analyzed dynamic programming algorithm solving IPs with integer data [7,14,20]. For $A \in \mathbb{Z}^{m\times n}, b \in \mathbb{Z}^m$, [14] proved that $\max\{c^\top x : Ax = b, x \geq 0, x \in \mathbb{Z}^n\}$ can be solved in time $O(\sqrt{m}\Delta)^{2m}\log(\|b\|_\infty) + O(nm)$, where Δ is the largest absolute value of entries in the input matrix A . Integer programs of the form $\max\{c^\top x : Ax = b, 0 \leq x \leq u, x \in \mathbb{Z}^n\}$ can similarly be solved in time

$$n \cdot O(m)^{(m+1)^2} \cdot O(\Delta)^{m\cdot(m+1)} \log^2(m\cdot\Delta),$$

which was proved in [7]. Note that integer programs of the form $\max\{c^\top x : Ax \leq b, x \in \{0,1\}^n\}$ can be rewritten in this latter form by adding m slack variables.

The complexity of integer programming has also been studied from the perspective of *smoothed analysis*. In this context, Röglin and Vöcking [23] proved that a class of IPs satisfying some minor conditions has polynomial smoothed complexity if and only if that class admits a pseudopolynomial time algorithm. An algorithm has polynomial smoothed complexity if its running time is polynomial with high probability when its input has been perturbed by adding random noise, where the polynomial may depend on the inverse magnitude φ^{-1} of the noise as well as the dimensions n, m of the problem. An algorithm runs in pseudopolynomial time if the running time is polynomial when the numbers are written in unary, i.e., when the input data consists of integers of absolute value at most Δ and the running time is bounded by a polynomial $p(n,m,\Delta)$. In particular, they prove that solving the randomly perturbed problem requires only polynomially many calls to the pseudopolynomial time algorithm with numbers of size $(nm\varphi)^{O(1)}$ and considering only the first $O(\log(nm\varphi))$ bits of each of the perturbed entries.

One may in fact compare the complexity of dynamic programming and branch-and-bound for Gaussian IPs using the result of [23]. If we choose $A \in \mathbb{R}^{m\times n}, c \in \mathbb{R}^n$ as well as $b \in \mathbb{R}^m$ to have i.i.d. $N(0,1)$ entries, the result of [23] implies that with high probability, to solve (Primal IP) it is sufficient to solve polynomially many problems with integer entries of size $n^{O(1)}$. Since $\Delta = n^{O(m)}$ in this setting (by Hadamard's inequality), the result of [7] implies that (Primal IP) can be solved in time $n^{O(m^3)}$ with high probability. In comparison, by Corollary 1, for any fixed $\varepsilon \in (0,1)$, branch-and-bound solves (Primal IP) in time $n^{O(m^{1.25})}$, for $n \geq 2^m$, with probability $1-\varepsilon$.

1.4 Organization

In Section 2, we give preliminaries on probability theory, linear programming and integer rounding. In Section 3, we prove properties of the optimal primal and dual LP solutions x^* and u^*, and in Section 4, we characterize the distribution of the columns of the objective extended constraint matrix corresponding to the zero entries of x^*. In Section 5, we prove Theorem 1, using a discrepancy result that we prove in Section 6. In Section 7, we prove Theorem 3, our meta-theorem for random logconcave IPs.

2 Preliminaries

2.1 Basic Notation

We denote the reals and non-negative reals by $\mathbb{R}, \mathbb{R}^+$ respectively, and the integers and positive integers by $\mathbb{Z}, \mathbb{N}$ respectively. For $k \geq 1$ an integer, we let $[k] := \{1,\dots,k\}$. For $s \in \mathbb{R}$, we let $s^+ := \max\{s,0\}$ and $s^- := \min\{s,0\}$ denote the positive and negative part of s. We extend this to a vector $x \in \mathbb{R}^n$ by letting $x^{+(-)}$ correspond to applying the positive (negative) part operator coordinate-wise. We let $\|x\|_2 = \sqrt{\sum_{i=1}^n x_i^2}$ and $\|x\|_1 = \sum_{i=1}^n |x_i|$ denote the ℓ_2 and ℓ_1 norm respectively. We use $\log x$ to denote the base e natural logarithm. We use $0_m, 1_m \in \mathbb{R}^m$ to denote the all zeros and all ones vector respectively, and $e_1,\dots,e_m \in \mathbb{R}^m$ denote the standard coordinate basis. We write $\mathbb{R}^m_+ := [0,\infty)^m$.

For a random variable $X \in \mathbb{R}$, we let $\mathbb{E}[X]$ denote its expectation and $\mathrm{Var}[X] := \mathbb{E}[X^2] - \mathbb{E}[X]^2$ denote its variance. For a random vector $X \in \mathbb{R}^d$, we define its mean $\mathbb{E}[X] := (\mathbb{E}[X_1],\dots,\mathbb{E}[X_d])$ and its covariance matrix

$$\mathrm{Cov}(X) := \mathbb{E}[XX^\mathsf{T}] - \mathbb{E}[X]\,\mathbb{E}[X]^\mathsf{T} = (\mathbb{E}[X_iX_j] - \mathbb{E}[X_i]\,\mathbb{E}[X_j])_{i,j\in[d]}.$$

For any $u \in \mathbb{R}^d$, we note that $\mathrm{Var}[u^\mathsf{T}X] = \mathbb{E}[(u^\mathsf{T}X)^2] - \mathbb{E}[u^\mathsf{T}X]^2 = u^\mathsf{T}\,\mathrm{Cov}(X)u$. We say that X has identity covariance if $\mathrm{Cov}(X) = I_d$, the $d\times d$ identity matrix.

$X \in \mathbb{R}^d$ is a continuous random vector if it admits a probability density $f : \mathbb{R}^d \to \mathbb{R}_+$ satisfying $\Pr[X \in A] = \int_A f(x)dx$, for all measurable $A \subseteq \mathbb{R}^d$. We will say that a continuous random vector has maximum density at most $M > 0$ if its probability density f satisfies $\sup_{x\in\mathbb{R}^d} f(x) \le M$.

2.2 The Dual Program, Gap Formula and the Optimal Solutions

A convenient formulation of the dual of (Primal LP) is given by

$$\begin{aligned} &\min \operatorname{val}^\star(u) := b^\mathsf{T} u + \sum_{i=1}^n (c - A^\mathsf{T} u)_i^+ && \text{(Dual LP)} \\ &\text{s.t. } u \ge 0. \end{aligned}$$

To keep the notation concise, we will often use the identity $\|(c-A^t u)^+\|_1 = \sum_{i=1}^n (c - A^\mathsf{T} u)_i^+$.

For any primal solution x and dual solution u to the above pair of programs, we have the following standard formula for the primal-dual gap:

$$\begin{aligned} \operatorname{val}^\star(u) - \operatorname{val}(x) &:= b^\mathsf{T} u + \sum_{i=1}^n (c - A^\mathsf{T} u)_i^+ - c^\mathsf{T} x && \text{(Gap Formula)} \\ &= (b - Ax)^\mathsf{T} u + \left(\sum_{i=1}^n x_i (A^\mathsf{T} u - c)_i^+ + (1 - x_i)(c - A^\mathsf{T} u)_i^+ \right). \end{aligned}$$

Throughout the rest of the paper, we let x^* and u^* denote primal and dual optimal basic feasible solutions for (Primal LP) and (Dual LP) respectively, which we note are unique with probability 1. We use the notation

$$W := \begin{bmatrix} c^\mathsf{T} \\ A \end{bmatrix} \in \mathbb{R}^{(m+1)\times n}, \tag{4}$$

to denote the objective extended constraint matrix. We will frequently make use of the sets $N_b := \{i \in [n] : x_i^* = b\}$, $b \in \{0,1\}$, the 0 and 1 coordinates of x^*, and $S := \{i \in [n] : x_i^* \in (0,1)\}$, the fractional coordinates of x^*. We will also use the fact that $|S| \le m$, which follows since x^* is a basic solution to (Primal LP) and A has m rows.

2.3 Chernoff Bounds and Binomial Sums

Let $X_1, \dots, X_n$ independent $\{0,1\}$ random variables with $\mu = \mathbb{E}[\sum_{i=1}^n X_i]$. Then, the Chernoff bound gives [4, Corollary 1.10]

$$\begin{aligned} &\Pr[\sum_{i=1}^n X_i \le \mu(1-\varepsilon)] \le e^{-\frac{\varepsilon^2 \mu}{2}}, \varepsilon \in [0,1]. \\ &\Pr[\sum_{i=1}^n X_i \ge \mu(1+\varepsilon)] \le e^{-\frac{\varepsilon^2 \mu}{3}}, \varepsilon \in [0,1]. \end{aligned} \tag{5}$$

The same concentration holds for the size of the intersection of two random sets.

Lemma 2 *Let K, K' be two i.i.d. random subsets of $[ak]$, such that $|K| = |K'| = k$ and where $\Pr[i \in K] = \frac{1}{a}$ for every $1 \le i \le ak$. Then for every $\varepsilon \in (0,1)$ we have*

$$\Pr\left[|K \cap K'| \ge \frac{(1+\varepsilon)k}{a}\right] \le 2\exp\left(-\frac{k\varepsilon^2}{3a}\right),$$
$$\Pr\left[|K \cap K'| \le \frac{(1-\varepsilon)k}{a}\right] \le 2\exp\left(-\frac{k\varepsilon^2}{2a}\right).$$

Proof The set size $|K \cap K'|$ follows a hypergeometric distribution. To see this, we let $K \subset [ak]$ denote the set of successes, and we sample $|K'|$ elements from $[ak]$ without replacement. Then $|K \cap K'|$ counts the number of successes. The bound now follows directly from [4, Theorem 1.17]. □

We will need the following standard upper bound on binomial sums. For $n \ge 1$ and $n/2 \le k \le n$, we have that

$$|\{S \subseteq [n] : |S| \ge k\}| = \sum_{i=k}^{n} \binom{n}{i} \le e^{nH(k/n)}, \tag{6}$$

where $H(x) = -x\log(x) - (1-x)\log(1-x)$, $x \in [0,1]$, is the base e entropy function [12, Theorem 3.1]. We recall that $H(x)$ is concave in x and $H(x) = H(1-x)$, and hence is maximized at $H(1/2) = \log 2$.

2.4 Bounds on the Moment Generating Function

Lemma 3 *Let $Z \in \mathbb{R}$ be a random variable satisfying $\mathbb{E}[Z] = 0$ and $\mathbb{E}[e^{|Z|}] < \infty$. Then, $\mathbb{E}[e^Z] \le \mathbb{E}[\cosh(\sqrt{3/2}Z)]$, where $\cosh(x) := \frac{1}{2}(e^x + e^{-x}) = \sum_{k=0}^{\infty} \frac{x^{2k}}{(2k)!}$.*

Proof

$$\begin{aligned}
\mathbb{E}[e^Z] &= \sum_{k=0}^{\infty} \frac{\mathbb{E}[Z^k]}{k!} \quad \text{(by dominated convergence)}\\
&= 1 + \sum_{k=1}^{\infty} \frac{\mathbb{E}[Z^{2k}]}{(2k)!} + \sum_{k=1}^{\infty} \frac{\mathbb{E}[Z^{2k+1}]}{(2k+1)!} \quad \text{(} \mathbb{E}[Z] = 0 \text{)}\\
&\le 1 + \sum_{k=1}^{\infty} \frac{\mathbb{E}[Z^{2k}]}{(2k)!} + \sum_{k=1}^{\infty} \frac{(\mathbb{E}[Z^{2k}]\,\mathbb{E}[Z^{2k+2}])^{1/2}}{(2k+1)!} \quad \text{(by Hölder's inequality)}\\
&\le 1 + \sum_{k=1}^{\infty} \frac{\mathbb{E}[Z^{2k}]}{(2k)!} + \sum_{k=1}^{\infty} \frac{(2k+1)\,\mathbb{E}[Z^{2k}] + \mathbb{E}[Z^{2k+2}]/(2k+1)}{2(2k+1)!} \quad \text{(by AM-GM)}\\
&= 1 + \frac{3}{2}\frac{\mathbb{E}[Z^2]}{2} + \sum_{k=2}^{\infty} \frac{\mathbb{E}[Z^{2k}]}{(2k)!}\left(1 + 1/2 + \frac{2k}{2(2k-1)}\right)\\
&\le \sum_{k=0}^{\infty} \frac{\mathbb{E}[(\sqrt{3/2}Z)^{2k}]}{(2k)!} = \mathbb{E}[\cosh(\sqrt{3/2}Z)].
\end{aligned}$$

The last inequality follows by the fact that $(3/2)^k \ge 3/2 + \frac{k}{2k-1}$ whenever $k \ge 2$. □

2.5 Gaussian and Sub-Gaussian Random Variables

The standard, mean zero and variance 1, Gaussian $\mathcal{N}(0,1)$ has density function $\varphi(x) := \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$. A standard Gaussian vector in $\mathbb{R}^d$, denoted $\mathcal{N}(0, I_d)$, has probability density $\prod_{i=1}^d \varphi(x_i) = \frac{1}{\sqrt{2\pi}^d} e^{-\|x\|^2/2}$ for $x \in \mathbb{R}^d$. A random variable $Y \in \mathbb{R}$ is σ-sub-Gaussian if for all $\lambda \in \mathbb{R}$, we have

$$\mathbb{E}[e^{\lambda Y}] \leq e^{\sigma^2 \lambda^2/2}. \tag{7}$$

A standard normal random variable $X \sim \mathcal{N}(0,1)$ is 1-sub-Gaussian. If variables $Y_1, \ldots, Y_k \in \mathbb{R}$ are independent and respectively σ_i-sub-Gaussian, $i \in [k]$, then $\sum_{i=1}^k Y_i$ is $\sqrt{\sum_{i=1}^k \sigma_i^2}$-sub-Gaussian.
For a σ-sub-Gaussian random variable $Y \in \mathbb{R}$ we have the following standard tail-bound:

$$\max\{\Pr[Y \leq -\sigma s], \Pr[Y \geq \sigma s]\} \leq e^{-\frac{s^2}{2}}, s \geq 0. \tag{8}$$

For $X \sim \mathcal{N}(0, I_d)$, we will use the following higher dimensional analogue:

$$\Pr[\|X\|_2 \geq s\sqrt{d}] \leq e^{-\frac{d}{2}(s^2 - 2\log s - 1)} \leq e^{-\frac{d}{2}(s-1)^2}, s \geq 1. \tag{9}$$

We will use this bound to show that the columns of A corresponding to the fractional coordinates in the (almost surely unique) optimal solution x^* are bounded.

Lemma 4 *Letting $S := \{i \in [n] : x_i^* \in (0,1)\}$, we have that*

$$\Pr[\exists i \in S : \|A_{\cdot,i}\|_2 \geq (4\sqrt{\log(n)} + \sqrt{m})] \leq n^{-7}.$$

Proof Using the union bound over all n columns and Equation (9) we get that:

$$\begin{aligned}
\Pr[\exists i \in S : \|A_{\cdot,i}\|_2 \geq (4\sqrt{\log(n)} + \sqrt{m})] \\
&\leq \Pr[\exists i \in [n] : \|A_{\cdot,i}\|_2 \geq (4\sqrt{\log(n)} + \sqrt{m})] \\
&\leq n \Pr_{X \sim \mathcal{N}(0, I_d)}[\|X\|_2 \geq (4\sqrt{\log(n)} + \sqrt{m})] \\
&\leq n \exp\left(-\frac{m}{2}(4\sqrt{\log(n)/m})^2\right) = n^{-7}. \quad \square
\end{aligned}$$

We will need the following analogue of the Chernoff bound for truncated Gaussian sums.

Lemma 5 *Let $X_1, \ldots, X_n$ be i.i.d. $\mathcal{N}(0,1)$. Then*

$$\Pr\left[\left|\sum_{i=1}^n X_i^+ - \frac{n}{\sqrt{2\pi}}\right| \geq \sqrt{2n}s\right] \leq 2e^{-s^2/2}, s \geq 0.$$

Proof For $X \sim \mathcal{N}(0,1)$, a direct computation yields $\mu := \mathbb{E}[X^+] = \frac{1}{\sqrt{2\pi}}$. For $\lambda \in \mathbb{R}$, we get that

$$\begin{aligned}
\mathbb{E}[e^{\lambda(X^+-\mu)}] &\le \mathbb{E}[\cosh(\sqrt{3/2}\lambda(X^+-\mu))] \quad (\text{ by Lemma 3 }) \\
&\le \mathbb{E}[\cosh(\sqrt{3/2}\lambda(X-\mu))] \quad \left(\, |X^+-\mu| \le |X-\mu| \,\right) \\
&\le e^{\frac{3}{2}\lambda^2/2}\cosh(\sqrt{\frac{3}{2}}\lambda\mu) \quad (\text{ by (7) }) \\
&\le e^{(1+\mu^2)\frac{3}{4}\lambda^2} \quad \left(\cosh(x) \le e^{x^2/2} \right) \\
&\le e^{\lambda^2} \quad \left(\mu^2 = \frac{1}{2\pi} \le \frac{1}{3} \right).
\end{aligned}$$

By the above, we have that $X^+ - \frac{1}{2\pi}$ is $\sqrt{2}$-sub-Gaussian. Therefore, if $X_1,\ldots,X_n$ are i.i.d. $\mathcal{N}(0,1)$, the random variable $Y = (\sum_{i=1}^n X_i^+) - \frac{n}{\sqrt{2\pi}} = \sum_{i=1}^n (X_i^+ - \frac{1}{\sqrt{2\pi}})$ is $\sqrt{2n}$-sub-Gaussian. The desired result now follows directly from the sub-Gaussian tail bound (8) and the union bound. □

2.6 Logconcave Distributions

A probability measure μ on $\mathbb{R}^d$ is logconcave μ admits a probability density $f : \mathbb{R}^d \to \mathbb{R}_+$ such that $\log f : \mathbb{R}^d \to \mathbb{R} \cup \{-\infty\}$ is concave. We say that a random vector $X \in \mathbb{R}^d$ is logconcave if its distributed according to a logconcave probability measure. Important examples of logconcave probability distributions are the Gaussian distribution and the uniform distribution on a compact convex set.

Logconcave distributions have many useful analytical properties. In particular, the marginals of logconcave random vectors are also logconcave.

Theorem 4 **([22])** *Let $X \in \mathbb{R}^d$ be a logconcave random vector. Then, for any surjective linear transformation $T : \mathbb{R}^d \to \mathbb{R}^k$, TX is a logconcave random vector.*

The following theorem, which combines results from [17, 10], yields important properties of 1 dimensional logconcave distributions that we will need.

Theorem 5 *Let $\omega \in \mathbb{R}$ be a logconcave random variable with* $\mathrm{Var}[\omega] = 1$.

- *[17, Lemma 5.7]:* $\Pr[|\omega - \mathbb{E}[\omega]| \ge s] \le e^{-s+1}$, $\forall s \ge 0$.
- *[10, Theorem 4]: ω has maximum density at most* 1.

2.7 A Local Limit Theorem

We now introduce the formalization of Gaussian convergence used in Lemma 1.

Definition 1 Suppose $X_1, X_2, \dots$ is a sequence of i.i.d copies of a random variable X with density f. For $k_0 \in \mathbb{N}$, $\gamma \geq 0$, we define X to be (γ, k_0)-Gaussian convergent if the density f_n of $\sum_{i=1}^{n} X_i/\sqrt{n}$ satisfies:

$$|f_n(x) - \varphi(x)| \leq \frac{\gamma}{n} \quad \forall x \in \mathbb{R}, n \geq k_0,$$

where $\varphi := \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$ is the probability density function of the standard Gaussian.

The above definition quantifies the speed of convergence in the context of the central limit theorem. The rounding strategy used to obtain the main result utilizes random variables that are the weighted sum of a uniform and an independent normal variable. Crucially, the given convergence estimate will hold for these random variables:

Lemma 6 *Let U be uniform on $[-\sqrt{3}, \sqrt{3}]$ and let $Z \sim \mathcal{N}(0,1)$. Then there exists a universal constant $k_0 \geq 1$ such that $\forall \varepsilon \in [0,1]$, the random variable $\sqrt{\varepsilon}U + \sqrt{1-\varepsilon}Z$ is $(1/10, k_0)$-Gaussian convergent and has maximum density at most 1.*

We note that the $O(1/n)$-convergence rate to Gaussian achieved above is due to the first 3 moments of $X := \sqrt{\varepsilon}U + \sqrt{1-\varepsilon}Z$ matching those of the standard Gaussian, namely $\mathbb{E}[X] = \mathbb{E}[X^3] = 0$ and $\mathbb{E}[X^2] = 1$, and the fact that the characteristic function $c(s) := \mathbb{E}[e^{isX}]$ decays like $1/|s|$. More generally, if the first $l \geq 2$ moments match and the characteristic function decays quickly enough, the convergence rate is $O(1/n^{(l-1)/2})$. The proof follows from the following local limit theorem (a special case of [9, Theorem XVI.2.2]):

Theorem 6 *Let $X_1, X_2, \dots$ be a sequence of i.i.d. random variables with density having moments $\mathbb{E}[X_1] = 0, \mathbb{E}[X_1^2] = 1, \mathbb{E}[X_1^3] = 0$, $\mathbb{E}[X_1^4] = \mu_4 > 0$. Also, assume that we have $|\mathbb{E}[e^{isX}]| \leq \beta/(|s|+1)^{\alpha}, \forall s \in \mathbb{R}$ for some $\beta, \alpha > 0$. Define $S_n = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} X_i$ and $\varphi(x) = \frac{1}{\sqrt{2\pi}} \exp(-x^2/2)$. Then, for all $n \geq 1$, S_n admits a density f_n satisfying*

$$f_n(x) = \varphi(x) \left[1 + \frac{\mu_4 - 3}{24n}(x^4 - 6x^2 + 3)\right] + o\left(\frac{C_{\alpha,\beta}}{n}\right), \forall x \in \mathbb{R},$$

where $C_{\alpha,\beta}$ depends only on α, β.

Proof (Lemma 6) Letting $X = \sqrt{\varepsilon}U + \sqrt{1-\varepsilon}Z$, a straightforward calculation yields $\mathbb{E}[X] = \mathbb{E}[X^3] = 0$, $\mathbb{E}[X^2] = 1$ and $\mu_4 := \mathbb{E}[X^4] = \varepsilon^2 9/5 + 6(1-\varepsilon)\varepsilon + (1-\varepsilon)^2 3$. Since convolution does not increase the maximum density, as $\max\{\sqrt{\varepsilon}, \sqrt{1-\varepsilon}\} \geq 1/\sqrt{2}$, we see that the maximum density of $\sqrt{\varepsilon}U + \sqrt{1-\varepsilon}Z$ is upper bounded by $\min\{\frac{1}{2\sqrt{3\varepsilon}}, \frac{1}{\sqrt{2\pi(1-\varepsilon)}}\} \leq 1$, where the terms correspond to the maximum density of $\sqrt{\varepsilon}U$ and $\sqrt{1-\varepsilon}Z$ respectively. Further,

$$|\mathbb{E}[e^{is(\sqrt{\varepsilon}U + \sqrt{1-\varepsilon}Z)}]| = |\frac{\sin(\sqrt{3\varepsilon}s)}{\sqrt{3\varepsilon}s} e^{-(1-\varepsilon)s^2/2}| \leq \frac{2}{1+\sqrt{3\varepsilon}|s|} \cdot \frac{1}{1+(1-\varepsilon)s^2/2}$$

$$\leq \frac{10}{1+|s|}.$$

From here, $\max_{x\in\mathbb{R}} |\varphi(x)\frac{1}{24}(x^4-6x^2+3)|$ is maximized at $x=0$ attaining value $\varphi(0)\frac{3}{24} = \frac{1}{8\sqrt{2\pi}}$. Let $\gamma' := \frac{6}{40\sqrt{2\pi}} \geq |\frac{(\mu_4-3)}{8\sqrt{2\pi}}|$. Applying Theorem 6 with $\beta=10, \alpha=1$, one can choose $k_0 = O_{\alpha,\beta}(1)$ large enough so that the $o\left(C_{\alpha,\beta}/n\right)$ term is at most γ'/n for all $n \geq k_0$. The lemma now follows letting $\gamma := 1/10 \geq 2\gamma'$. □

2.8 Nets

Let $\mathbb{S}^{d-1} = \{x \in \mathbb{R}^d : \|x\|_2 = 1\}$ denote the unit sphere in $\mathbb{R}^d$. We say that $N \subseteq \mathbb{S}^{d-1}$ is an ε-net if for every $x \in \mathbb{S}^{d-1}$ there exists $y \in N$ such that $\|x-y\| \leq \varepsilon$. A classic result we will need is that S^{d-1} admits an ε-net N_ε of size $|N_\varepsilon| \leq (1+\frac{2}{\varepsilon})^d$. See, for example, [8, Chapter 5]. We note that the same bound holds if we wish to construct a net of some subset $A \subseteq \mathbb{S}^{d-1}$ and we wish to have $N_\varepsilon \subseteq A$.

2.9 Rounding to Binary Solutions

In the proof of Theorem 1, we will take our optimal solution x^* and round it to an integer solution x', by changing the fractional coordinates. Note that as x^* is a basic solution, it has at most m fractional coordinates. One could round to a integral solution by setting all of them to 0, i.e., $x' = \lfloor x^* \rfloor$. If we assume that the Euclidean norm of every column of A is bounded by C, then we have $\|A(x^*-x')\|_2 \leq mC$, since x^* has at most m fractional variables. However, by using randomized rounding we can make this bound smaller, as stated in the next lemma. We use this to obtain smaller polynomial dependence in Theorem 1.

Lemma 7 *Consider an $m \times n$ matrix A with $\|A_{\cdot,i}\|_2 \leq C$ for all $i \in [m]$ and $y \in [0,1]^n$. Let $S = \{i \in [n] : y_i \in (0,1)\}$. There exists a vector $y' \in \{0,1\}^n$ with $\|A(y-y')\|_2 \leq C\sqrt{|S|}/2$ and $y'_i = y_i$ for all $i \notin S$.*

Proof Let Y be the random variable in $\{0,1\}^n$ with independent components such that $\mathbb{E}(Y) = y$. Note that this implies that $\mathrm{Var}(Y_i) \leq 1/4$ for all i and $\mathrm{Var}(Y_i) = 0$ for $i \notin S$. Then:

$$\mathbb{E}(\|A(y-Y)\|_2)^2 \leq \mathbb{E}(\|A(y-Y)\|_2^2) = \sum_{i=1}^n \|A_{\cdot,i}\|_2^2 \,\mathrm{Var}(Y_i)$$
$$\leq \sum_{i\in S} C^2\,\mathrm{Var}(Y_i) \leq \frac{C^2|S|}{4}$$

So $\mathbb{E}(\|A(y-Y)\|_2) \leq C\sqrt{|S|}/2$, which directly implies the existence of a value $y' \in \{0,1\}^n$ with $\|A(y-y')\|_2 \leq C\sqrt{|S|}/2$. □

3 Properties of the Optimal Solutions

The following lemma is the main result of this section, which gives principal properties we will need of the optimal primal and dual LP solutions. Namely, we prove an upper bound on the norm of the optimal dual solution u^* and a lower bound on the number of zero coordinates of the optimal primal solution x^*.

Lemma 8 *Given $\delta := \frac{\sqrt{2\pi}}{n}\|b^-\|_2 \in [0,1/2)$, $\varepsilon \in (0,1/5)$, let x^*, u^* denote the optimal primal and dual LP solutions, and let $\alpha := \frac{1}{\sqrt{2\pi}}\sqrt{\left(\frac{1-3\varepsilon}{1-\varepsilon}\right)^2 - \delta^2}$ and choose $\beta \in [1/2,1]$ with $H(\beta) = \frac{\alpha^2}{4}$. Then, with probability at least $1 - 2\left(1+\frac{2}{\varepsilon}\right)^{m+1} e^{-\frac{\varepsilon^2 n}{8\pi}} - e^{-\frac{\alpha^2 n}{4}}$, the following holds:*

1. $c^\mathsf{T} x^* \geq \alpha n$.
2. $\|u^*\|_2 \leq \frac{1+\varepsilon}{1-3\varepsilon-(1-\varepsilon)\delta}$.
3. $|\{i \in [n] : x_i^* = 0\}| \geq (1-\beta)n - m$.

To prove Lemma 8, we will require two technical lemmas. The first, Lemma 9, shows that a random Gaussian matrix forms a good embedding from ℓ_2 into a "truncated" version of ℓ_1 (i.e., ℓ_1 restricted to non-negative coordinates). The second, Lemma 10, upper bounds the value of maximizing a Gaussian objective over the hypercube restricted to vectors having a large number of coordinates set to one.

Lemma 9 *Let $G \in \mathbb{R}^{n\times d}$ be a random matrix with independent $\mathcal{N}(0,1)$ entries. Then, for $\varepsilon \in (0,1)$,*

$$\Pr\left[\exists\, v \in S^{d-1}, \|(Gv)^+\|_1 \notin \left[\frac{1-3\varepsilon}{1-\varepsilon}, \frac{1+\varepsilon}{1-\varepsilon}\right]\cdot \frac{n}{\sqrt{2\pi}}\right] \leq 2\left(1+\frac{2}{\varepsilon}\right)^d e^{-\frac{\varepsilon^2 n}{8\pi}}. \quad (10)$$

Proof Let N_ε denote an ε-net of S^{d-1}. Let B denote the event in equation (10). Let E denote the event that there exists $v' \in N_\varepsilon$ such that $\|(Gv')^+\|_1 \notin [1-\varepsilon, 1+\varepsilon]\frac{n}{\sqrt{2\pi}}$.

We now show that $\Pr[B] \leq \Pr[E]$. For this purpose, it suffices to show that $\neg E \Rightarrow \neg B$. We thus condition G on the complement of E and show that B does not occur. For every $v \in S^{d-1}$, choose an $\tilde{v} \in N_\varepsilon$ satisfying $\|v - \tilde{v}\|_2 \leq \varepsilon$. Then, we have that

$$\begin{aligned}\max_{v\in S^{d-1}} \|(Gv)^+\|_1 &\leq \max_{v\in S^{d-1}} \|(G\tilde{v})^+\|_1 + \|(G(v-\tilde{v}))^+\|_1 \\ &\leq (1+\varepsilon)\frac{n}{\sqrt{2\pi}} + \varepsilon \max_{v\in S^{d-1}} \|(Gv)^+\|_1 \Rightarrow \\ \max_{v\in S^{d-1}} \|(Gv)^+\|_1 &\leq \frac{1+\varepsilon}{1-\varepsilon}\frac{n}{\sqrt{2\pi}}. \end{aligned} \quad (11)$$

To get lower bounds, we use that

$$\begin{aligned}\min_{v\in S^{d-1}} \|(Gv)^+\|_1 &\geq \min_{v\in S^{d-1}} \|(G\tilde{v})^+\|_1 - \|(G(v-\tilde{v}))^-\|_1 \\ &\geq (1-\varepsilon)\frac{n}{\sqrt{2\pi}} - \varepsilon \max_{v\in S^{d-1}} \|(Gv)^-\|_1\end{aligned}$$

$$= (1-\varepsilon)\frac{n}{\sqrt{2\pi}} - \varepsilon \max_{v\in S^{d-1}} \|(Gv)^+\|_1. \quad (12)$$

From this inequality, we deduce that

$$\min_{v\in S^{d-1}} \|(Gv)^+\|_1 \geq \left(1-\varepsilon\left(1+\frac{1+\varepsilon}{1-\varepsilon}\right)\right)\frac{n}{\sqrt{2\pi}} = \frac{1-3\varepsilon}{1-\varepsilon}\frac{n}{\sqrt{2\pi}}. \quad (13)$$

Thus $\neg E \Rightarrow \neg B$, as needed. Using that $Gv \sim \mathcal{N}(0, I_n)$ for $v \in S^{d-1}$, we have that

$$\begin{aligned}
\Pr[E] &\leq \sum_{\tilde{v}\in N_\varepsilon} \Pr\left[\|(G\tilde{v})^+\|_1 \notin [1-\varepsilon, 1+\varepsilon]\cdot\frac{n}{\sqrt{2\pi}}\right] \\
&= |N_\varepsilon| \Pr_{X\sim\mathcal{N}(0,I_n)}\left[\|X^+\|_1 \notin [1-\varepsilon, 1+\varepsilon]\frac{n}{\sqrt{2\pi}}\right] \\
&\leq 2\left(1+\frac{2}{\varepsilon}\right)^d e^{-\frac{\varepsilon^2 n}{8\pi}} \quad (\text{ by Lemma 5 }).
\end{aligned}$$

The lemma thus follows. □

Lemma 10 *Let $c \sim \mathcal{N}(0, I_n)$ and $\alpha \in [0, 2\sqrt{\log 2}]$. Choose $\beta \in [1/2, 1]$ satisfying $H(\beta) = \frac{\alpha^2}{4}$. Then*

$$\Pr\left[\max_{x\in\{0,1\}^n, \|x\|_1\geq\beta n} c^\mathsf{T} x \geq \alpha n\right] \leq e^{-\frac{\alpha^2 n}{4}}. \quad (14)$$

Proof For $x \in \{0,1\}^n$, we have that $c^\mathsf{T} x \sim \mathcal{N}(0, \|x\|_1)$. By the sub-Gaussian tail bound (8), we see that

$$\Pr[c^\mathsf{T} x \geq \alpha n] = \Pr_{z\in\mathcal{N}(0,1)}[\|x\|_2 z \geq \alpha n] \leq \Pr_{z\in\mathcal{N}(0,1)}[z \geq \alpha\sqrt{n}] \leq e^{-\frac{\alpha^2 n}{2}}.$$

By union bound and (6), we conclude that

$$\begin{aligned}
\Pr\left[\max_{x\in\{0,1\}^n, \|x\|_1\geq\beta n} c^\mathsf{T} x \geq \alpha n\right] &\leq |\{x \in \{0,1\}^n : \|x\|_1 \geq \beta n\}| \Pr_{z\in\mathcal{N}(0,1)}[z \geq \alpha\sqrt{n}] \\
&\leq e^{H(\beta)n} e^{-\frac{\alpha^2}{2}n} = e^{-\frac{\alpha^2 n}{4}}, \text{ as needed.} \quad \square
\end{aligned}$$

We are now ready to prove Lemma 8.

Proof (Lemma 8)
Let $W \in \mathbb{R}^{(m+1)\times n}$ be as in (4). Applying Lemma 9 to $(W^\mathsf{T}, \varepsilon)$ and Lemma 10 to (c, α, β), the events (E_1) $\|(W^\mathsf{T} u)^+\|_1 \in \left[\frac{1-3\varepsilon}{1-\varepsilon}, \frac{1+\varepsilon}{1-\varepsilon}\right]\cdot\frac{n}{\sqrt{2\pi}}$, $\forall \|u\|_2 = 1$, and (E_2) $\max_{x\in\{0,1\}^n, \|x\|_1\geq\beta n} c^\mathsf{T} x < \alpha n$ hold with probability at least $1 - 2\left(1+\frac{2}{\varepsilon}\right)^{m+1} e^{-\frac{\varepsilon^2 n}{8\pi}} - e^{-\frac{\alpha^2 n}{4}}$. It thus suffices to prove that properties 1-3 hold under E_1, E_2. We prove each in turn below.

Proof of 1. To lower bound $c^\mathsf{T} x^*$, i.e., the optimal LP value, it suffices to show that every dual solution has value at least αn. Take $u \in \mathbb{R}^m_+$, then we can bound the dual value $\mathrm{val}^*(u^*)$ using the inequalities

$$\begin{aligned} b^\mathsf{T} u + \|(c - A^\mathsf{T} u)^+\|_1 &\geq (b^-)^\mathsf{T} u + \|(W^\mathsf{T}(1,-u)^\mathsf{T})^+\|_1 \\ &\geq -\|b^-\|_2\|u\|_2 + \frac{1-3\varepsilon}{1-\varepsilon}\frac{n}{\sqrt{2\pi}}\sqrt{1+\|u\|_2^2} \quad (\text{ by } E_1 \text{ }) \\ &= \frac{n}{\sqrt{2\pi}} \cdot \left(-\delta\|u\|_2 + \frac{1-3\varepsilon}{1-\varepsilon}\sqrt{1+\|u\|_2^2}\right) \qquad (15) \\ &\geq \frac{n}{\sqrt{2\pi}}\sqrt{(\frac{1-3\varepsilon}{1-\varepsilon})^2 - \delta^2} = \alpha n, \end{aligned}$$

where the last inequality follows by minimizing the expression over $\|u\|$, which occurs at $\|u\| = \frac{\delta}{\sqrt{(\frac{1-3\varepsilon}{1-\varepsilon})^2 - \delta^2}}$.

Proof of 2. To bound $\|u^*\|_2$, we compare bounds for $\mathrm{val}^*(u^*)$:

$$\begin{aligned} \frac{1+\varepsilon}{1-\varepsilon}\frac{n}{\sqrt{2\pi}} &\geq \|(W^\mathsf{T}(1,0_m)^\mathsf{T})^+\|_1 = \|c^+\|_1 \quad (\text{by } E_1) \\ &= \mathrm{val}^*(0_m) \geq \mathrm{val}^*(u^*) \\ &\geq \frac{n}{\sqrt{2\pi}} \cdot \left(-\delta\|u^*\|_2 + \frac{1-3\varepsilon}{1-\varepsilon}\sqrt{1+\|u^*\|_2^2}\right) \quad (\text{ by (15) }) \\ &\geq \frac{n}{\sqrt{2\pi}} \cdot \left(\frac{1-3\varepsilon}{1-\varepsilon} - \delta\right)\|u^*\|_2 \\ &\Rightarrow \|u^*\|_2 \leq \frac{1+\varepsilon}{1-3\varepsilon-(1-\varepsilon)\delta}. \end{aligned}$$

Proof of 3. The optimal feasible solution x^* is unique almost surely, and as such it is a basic feasible solution. As such, we know that $x^* \in [0,1]^n$ has at most m fractional components. In particular, $|\{i \in [n] : x_i^* = 0\}| \geq n - m - |\{i \in [n] : x_i^* = 1\}|$. Thus, it suffices to show that $|\{i \in [n] : x_i^* = 1\}| \leq \beta n$. Define $\bar{x} \in \{0,1\}^n$ satisfying

$$\bar{x}_i = \begin{cases} x_i^* & : \ x_i^* \in \{0,1\} \\ 1 & : \ x_i^* \in (0,1), c_i \geq 0 \\ 0 & : \ x_i^* \in (0,1), c_i < 0 \end{cases}, \qquad \forall i \in [n].$$

Clearly, $c^\mathsf{T}\bar{x} \geq c^\mathsf{T} x^* \geq \alpha n$. Thus, by E_2 we have that $\beta n > |\{i \in [n] : \bar{x}_i = 1\}| \geq |\{i \in [n] : x_i^* = 1\}|$, as needed. □

4 Properties of the 0 Columns

For $Y := (c, a_1, \ldots, a_m) \sim \mathcal{N}(0, I_{m+1})$ and $u \in \mathbb{R}^m_+$, let Y^u denote the random variable Y conditioned the event $c - \sum_{i=1}^m u_i a_i \leq 0$. We will crucially use the following lemma directly adapted from Dyer and Frieze [5, Lemma 2.1], which shows that the columns of W associated with the 0 coordinates of x^* are independent subject to having negative reduced cost.

Recall that $N_b = \{i \in [n] : x_i^* = b\}$, that $S = \{i \in [n] : x_i^* \in (0,1)\}$ and that $W := \begin{bmatrix} c^\mathsf{T} \\ A \end{bmatrix} \in \mathbb{R}^{(m+1)\times n}$ is the objective extended constraint matrix.

Lemma 11 *Let $N_0' \subseteq [n]$. Conditioning on $N_0 = N_0'$, the submatrix $W_{\cdot,[n]\setminus N_0'}$ uniquely determines x^* and u^* almost surely. If we further condition on the exact value of $W_{\cdot,[n]\setminus N_0'}$, assuming x^* and u^* are uniquely defined, then any column $W_{\cdot,i}$ with $i \in N_0'$ is distributed according to Y^{u^*} and independent of $W_{\cdot,[n]\setminus\{i\}}$.*

Proof Knowing N_0', we solve the following program to obtain its primal and dual optimal feasible solutions $\bar{x}$ and $\bar{u}$.

$$\begin{aligned} \max\ & c^\mathsf{T} x \\ \text{s.t.} & \sum_{i \in [n]\setminus N_0} x_i a_{ji} \le b_j && \forall j \in [m] \\ & x_i = 0 && \forall i \in N_0' \\ & x \in [0,1]^n. \end{aligned}$$

This does not require knowledge of $W_{\cdot,N_0'}$, and the optimal feasible primal and dual solutions are unique almost surely.

If $N_0 = N_0'$, then $\bar{x} = x^*$ and $\bar{u} = u^*$. Since these solutions satisfy complementary slackness, this is equivalent to the following system of equations.

$$(1-\bar{x}_i)(c_i - \sum_{j=1}^m \bar{u}_i a_{ji})^+ = 0, \qquad \forall i \in [n]. \tag{16}$$

$$\bar{x}_i(\sum_{j=1}^m \bar{u}_i a_{ji} - c_i)^+ = 0, \qquad \forall i \in [n]. \tag{17}$$

$$\bar{u}_j(b - A\bar{x})_j = 0, \qquad \forall j \in [m]. \tag{18}$$

Note that for $i \in N_0$, eqs. (17) and (18) are trivially satisfied. By definition, the distribution of $W_{\cdot,i} := (c_i, a_{1i}, \dots, a_{mi})$ conditioned on eq. (16) for $\bar{x}_i = 0$ has the same law as $Y^{\bar{u}}$. Note that each of these conditions depends on only one $i \in N_0'$, so all columns of $W_{\cdot,N_0'}$ are independent.

We conditioned only on $N_0 = N_0'$, which has non-zero probability, and we have shown that for every possible realization of $W_{\cdot,[n]\setminus N_0'}$, the columns of $W_{\cdot,N_0'}$ are independently distributed as Y^{u^*}, which proves the lemma. □

To make the distribution of the columns $A_{\cdot,i}$ easier to analyze we rotate them.

Lemma 12 *Let R be a rotation that sends the vector u to the vector $\|u\|_2 e_m$. Suppose $(c,a) \sim Y^u$. Define $a' := Ra$. Then $(c, Ra) \sim (c', a')$, where (c', a') is the value of $(\bar{c}', \bar{a}') \sim \mathcal{N}(0, I_{m+1})$ conditioned on $\|u\|_2 \bar{a}'_m - \bar{c}' \ge 0$.*

Proof Recall that $(c,a) \sim Y^{u'}$ is generated by conditioning $(\bar{c},\bar{a}) \sim \mathcal{N}(0,I_{m+1})$ on $\bar{c} - u^\mathsf{T}\bar{a} \le 0$. The latter is equivalent to $\bar{c} - \|u\|_2(R\bar{a})_m = \bar{c} - u^\mathsf{T}R^{-1}(R\bar{a}) \le 0$, i.e. to $\|u\|_2(R\bar{a})_m - \bar{c} \ge 0$. Setting $(\bar{c}',\bar{a}') = (\bar{c},R\bar{a})$, we see that $(c,Ra) \sim (c',a')$, where (c',a') is the value of $(\bar{c}',\bar{a}') \sim \mathcal{N}(0,I_{m+1})$ conditioned on $\|u\|_2\bar{a}'_m - \bar{c}' \ge 0$. □

We will slightly change the distribution of the (c',a'_m) above using rejection sampling, as stated in the next lemma. This will make it easier to apply the discrepancy result of Lemma 1, which is used to round x^* to an integer solution of nearby value. In what follows, we denote the probability density function of a random variable X by f_X. In the following lemma, we use $\mathrm{unif}(0,\nu)$ to denote the uniform distribution on the interval $[0,\nu]$, for $\nu \ge 0$.

Lemma 13 *For any $\omega \ge 0$, $\nu > 0$, let $X,Y \sim \mathcal{N}(0,1)$ be independent random variables and let $Z = \omega Y - X$. Let X',Y',Z' be these variables conditioned on $Z \ge 0$. We apply rejection sampling on (X',Y',Z') with acceptance probability*

$$\Pr[accept|Z'=z] = \frac{2\varphi(\nu/\sqrt{1+\omega^2})\mathbf{1}_{z\in[0,\nu]}}{2\varphi(z/\sqrt{1+\omega^2})}.$$

Let X'',Y'',Z'' be the variables X',Y',Z' conditioned on acceptance. Then:

1. $\Pr[accept] = 2\nu\varphi(\nu/\sqrt{1+\omega^2})/\sqrt{1+\omega^2}$.
2. *$Y'' \sim W+V$ where $W \sim \mathcal{N}(0,\frac{1}{1+\omega^2})$, $V \sim \mathrm{unif}(0,\frac{\nu\omega}{1+\omega^2})$ and W,V are independent.*

Proof Because $Z \sim \mathcal{N}(0,1+\omega^2)$ and $Z' = Z \mid Z \ge 0$, for the density function $f_{Z'}$ we have $f_{Z'}(z) = 2\cdot\mathbf{1}_{z\ge 0}\varphi(z/\sqrt{1+\omega^2})/\sqrt{1+\omega^2}$.

$$\Pr[\mathrm{accept}] = \int_0^\nu \frac{2\varphi(z/\sqrt{1+\omega^2})}{\sqrt{1+\omega^2}}\frac{2\varphi(\nu/\sqrt{1+\omega^2})}{2\varphi(z/\sqrt{1+\omega^2})}dz = \frac{2\nu\varphi(\nu/\sqrt{1+\omega^2})}{\sqrt{1+\omega^2}}.$$

Now the probability density function of Z'', the variable Z' conditioned on acceptance, is:

$$f_{Z''}(z) = \frac{f_{Z'}(z)\cdot\Pr[\mathrm{accept}|Z'=z]}{\Pr[\mathrm{accept}]} = \mathbf{1}_{z\in[0,\nu]}/\nu.$$

So Z'' is uniformly distributed in $[0,\nu]$.

Let $W = Y - \frac{\omega}{1+\omega^2}Z$. By a direct calculation, one can check that $\mathbb{E}[W] = \mathbb{E}[Z] = \mathbb{E}[WZ] = 0$, $\mathbb{E}[Z^2] = 1+\omega^2$, $\mathbb{E}[W^2] = \frac{1}{1+\omega^2}$. Since W,Z are jointly Gaussian, this covariance structure implies that $W \sim \mathcal{N}(0,\frac{1}{1+\omega^2})$, $Z \sim \mathcal{N}(0,1+\omega^2)$ and that W,Z are independent. Noting that $Y = W + \frac{\omega}{1+\omega^2}Z$, we see that the law of Y'' is the same as that of $W + \frac{\omega}{1+\omega^2}Z''$, where W,Z'' are independent. Finally observe that $\frac{\omega}{1+\omega^2}Z'' \sim \mathrm{unif}(0,\frac{\nu\omega}{1+\omega^2})$. □

5 Proof of Theorem 1

Recall that $S = \{i \in [n] : x_i^* \in (0,1)\}$ and $N_0 = \{i \in [n] : x_i^* = 0\}$. To prove Theorem 1, we will need to condition on the following event, which we denote by E:

1. $\|A_{\cdot,i}\|_2 \le 4\sqrt{\log(n)} + \sqrt{m}$, $\forall i \in S$.
2. $\|u^*\| \le 3$.
3. $|N_0| \ge n/500$.

Using Lemmas 4 and 8 we can show that E hold with probability $1 - n^{-\Omega(1)}$. Now we take our optimal basic solution x^* and round it to an integral vector x' using Lemma 7. Then we can generate a new solution x'' from x' by flipping the values at indices $T \subseteq N_0$ to one. In Lemma 14 we show that with high probability there is such a set T, such that x'' is a feasible solution to our primal problem and that $\mathrm{val}(x^*) - \mathrm{val}(x'')$ is small.

We do this by looking at t disjoint subsets of N_0 with small reduced costs. Then we show for each of these sets that with constant probability it contains a subset T such that for x'' obtained from T, x'' is feasible and all constraints that are tight for x^* are close to being tight for x''. This argument relies on the Lemma 1 from the introduction, which we prove in Section 6.

If a suitable T exists, then using the gap formula we show that $\mathrm{val}(x^*) - \mathrm{val}(x'')$ is small. Because the t sets independent the probability of failure decreases exponentially with t. Hence, we can make the probability of failure arbitrarily small by increasing t. We know $\mathrm{val}(x^*) = \mathrm{val}_{\mathsf{LP}}(A,b,c)$ and because $x'' \in \{0,1\}^n$ we have $\mathrm{val}_{\mathsf{IP}}(A,b,c) \ge \mathrm{val}(x'')$, so $\mathsf{IPGAP}(A,b,c) = \mathrm{val}_{\mathsf{LP}}(A,b,c) - \mathrm{val}_{\mathsf{IP}}(A,b,c) \le \mathrm{val}(x^*) - \mathrm{val}(x'')$, which is small with high probability.

Lemma 14 *For $n \ge \exp(k_0)$, with k_0 as in Lemma 6, we have that*

$$\Pr\left[\mathsf{IPGAP}(A,b,c) > 10^{15} t \cdot \frac{m^{2.5}(\log n + m)^2}{n} \mid E\right] \le 2 \cdot \left(1 - \frac{1}{25}\right)^t \tag{19}$$

for $1 \le t \le \frac{n}{20000\sqrt{m}k^2}$, where $k := \lceil 165\,000 m(\log(n) + m)\rceil$.

Proof To prove the lemma, we show that the desired probability bound holds when we condition on the exact values of $N_0 \subseteq [n]$ and $W_{\cdot,[n]\setminus N_0}$ subject to 1-3 defining E. Since N_0 and $W_{\cdot,[n]\setminus N_0}$ determine E, this is clearly sufficient. By Lemma 11, we may further assume that this conditioning uniquely determines x^* and u^*.

Now let R be a rotation that sends the vector u^* to the vector $\|u^*\|_2 e_m$. Define:

$$\begin{aligned}
\Delta &:= 10\,000\sqrt{m}k/n = \Theta(m^{1.5}(m+\log n)/n), \\
B_i &:= RA_{\cdot,i}, && \text{for } i \in N_0, \\
Z_t &:= \{i \in N_0 : \|u^*\|_2 (B_i)_m - c_i \in [0, t\Delta]\}, && \text{for } 1 \le t \le \frac{1}{2\Delta k}.
\end{aligned}$$

We consider a (possibly infeasible) integral solution x' to the LP, generated by rounding the fractional coordinates from x^*. By Lemma 7 we can find such a solution with $\|A(x^* - x')\|_2 \le (4\sqrt{\log n} + \sqrt{m})\sqrt{|S|}/2 \le (4\sqrt{\log n} + \sqrt{m})\sqrt{m}/2$. We will select a subset $T \subseteq Z_t$ of size $|T| = k$ of coordinates to flip from 0 to 1 to obtain

$x'' \in \{0,1\}^n$ from x', so $x'' := x' + \sum_{i\in T} e_i$. By complementary slackness, we know for $i \in [n]$ that $x^*_i(A^\mathsf{T}u^* - c)^+_i = (1-x^*_i)(c - A^\mathsf{T}u^*)^+ = 0$ and that $x^*_i \notin \{0,1\}$ implies $(c - A^\mathsf{T}u^*)_i = 0$, and for $j \in [m]$ that $u^*_j > 0$ implies $b_j = (Ax^*)_j$. This observation allows us to prove the following key bound for the integrality gap of (Primal LP)

$$\begin{aligned}
&\mathrm{val}(x^*) - \mathrm{val}(x'') = \mathrm{val}^\star(u^*) - \mathrm{val}(x'') \\
&= (b - Ax'')^\mathsf{T}u^* \\
&+ \left(\sum_{i=1}^n x''_i(A^\mathsf{T}u^* - c)^+_i + (1-x''_i)(c - A^\mathsf{T}u^*)^+_i\right) && \text{(by (Gap Formula))} \\
&= (x^* - x'')^\mathsf{T}A^\mathsf{T}u^* + \sum_{i\in T}(A^\mathsf{T}u^* - c)_i && \text{(by complementary slackness)} \\
&\leq \sqrt{m}\|u^*\|_2 \left\|A(x'' - x^*)\right\|_\infty + t\Delta k && \text{(since } T \subseteq Z_t\text{).}
\end{aligned}$$

Condition 2 tells us that $\|u^*\|_2 \leq 3$, and by definition we have

$$t\Delta k \leq 27226 \cdot 10^{10} t \cdot \frac{m^{2.5}(\log(n)+m)^2}{n},$$

so the rest of this proof is dedicated to showing the existence of a set $T \subseteq Z_t$ such that $\|A(x'' - x^*)\|_\infty \leq O(1/n)$ and $Ax'' \leq b$.

By applying Lemma 11, we see that $\{(c_i, A_{\cdot,i})\}_{i\in N_0}$ are independent vectors, distributed as $\mathcal{N}(0, I_{m+1})$ conditioned on $c_i - A^\mathsf{T}_{\cdot,i}u^* \leq 0$. This implies that the vectors $\{(c_i, B_i)\}_{i\in N_0}$, are also independent. By Lemma 12, it follows that $(c_i, B_i) \sim \mathcal{N}(0, I_{m+1}) \mid \|u^*\|_2(B_i)_m - c_i \geq 0$. Note that the coordinates of B_i are therefore independent and $(B_i)_j \sim \mathcal{N}(0,1)$ for $j \in [m-1]$.

To simplify the upcoming calculations, we apply rejection sampling as specified in Lemma 13 with $\nu = \Delta t$ on $(c_i, (B_i)_m)$, for each $i \in N_0$. Let $Z'_t \subseteq N_0$ denote the indices which are accepted by the rejection sampling procedure. By the guarantees of Lemma 13, we have that $Z'_t \subseteq Z_t$ and

$$\Pr[i \in Z'_t \mid i \in N_0] = \frac{2\Delta t\varphi(\Delta t/\sqrt{1+\|u^*\|_2^2})}{\sqrt{1+\|u^*\|_2^2}} \geq \frac{2\Delta t\varphi(1/2)}{\sqrt{10}} \geq \Delta t/5.$$

Furthermore, for $i \in Z'_t$ we know that $(B_i)_m$ is distributed as a sum of independent $N(0, \frac{1}{1+\|u^*\|_2^2})$ and $\mathrm{unif}(0, t\Delta)$ random variables, and thus $(B_i)_m$ has mean and variance

$$\begin{aligned}
\mu_t &:= \mathbb{E}[(B_i)_m | i \in Z'_t] = \Delta t/2, \\
\sigma_t^2 &:= \mathrm{Var}[(B_i)_m | i \in Z'_t] = \frac{1}{1+\|u^*\|_2^2} + \frac{1}{12}\left(\frac{\|u^*\|_2\Delta t}{1+\|u^*\|_2^2}\right)^2 \in [1/10, 2].
\end{aligned}$$

Now define $\Sigma^{(t)}$ to be the diagonal matrix with $\Sigma^{(t)}_{j,j} = 1$, $j \in [m-1]$, and $\Sigma^{(t)}_{m,m} = \sigma_t$. Conditional on $i \in Z'_t$, define $B^{(t)}_i$ as the random variable

$$B^{(t)}_i := (\Sigma^{(t)})^{-1}(B_i - \mu_t e_m) \mid i \in Z'_t.$$

This ensures that all coordinates of $B^{(t)}$ are independent, mean zero and have variance one.

We have assumed that $|N_0| \geq n/500$ and we know $\Pr[i \in Z_t' | i \in N_0] \geq \Delta t/5$. Now, using the Chernoff bound (5) we find that:

$$\begin{aligned}\Pr[|Z_t'| < 2t\sqrt{m}k] &\leq \Pr\left[|Z_t'| < \frac{1}{5} t\Delta |N_0|/2\right] \\ &\leq \exp\left(-\frac{1}{8}\cdot\frac{1}{5} t\Delta |N_0|\right) \\ &\leq \left(1-\frac{1}{25}\right)^t. \end{aligned} \tag{20}$$

Now we define:

$$\begin{aligned} \theta &:= \frac{\sqrt{2\pi k}}{2}\binom{\lceil 2\sqrt{m}k\rceil}{k}^{-1/m}, & d &:= A(x^* - x'). \\ \theta' &:= 2\sqrt{m}\theta. & d' &:= d - 1_m\theta'. \end{aligned}$$

Observe that

$$\theta = \frac{\sqrt{2\pi k}}{2}\binom{\lceil 2\sqrt{m}k\rceil}{k}^{-1/m} \leq \frac{\sqrt{2\pi k}}{2}\left(2\sqrt{m}\right)^{-k/m} \leq \frac{1}{32m^2 n}.$$

So $\theta' \leq 1/8$.

If $|Z_t'| \geq \lceil 2\sqrt{m}\rceil kt$, then we can take t disjoint subsets $Z_t^{(1)}, \ldots Z_t^{(k)}$ of Z_t' of size $\lceil 2\sqrt{m}\rceil k$. Conditioning on this event, we wish to apply Lemma 1 to each $\{B_i^{(t)}\}_{i\in Z_t^{(l)}}$, $l \in [t]$, to help us find a candidate rounding of x' to a "good" integer solution x''.

Now we check that all conditions of Lemma 1 are satisfied. By definition we have $\left(\frac{2\theta}{\sqrt{2\pi k}}\right)^m \binom{ak}{k} = 1$, and we can bound

$$\begin{aligned} \left\|(\Sigma^{(t)})^{-1}(Rd' - e_m k\mu_t)\right\|_2 &\leq \max(1, 1/\sigma_t)(\|Rd\|_2 + \theta' + k\mu_t) \\ &\leq \sqrt{10}(\|RA(x^* - x')\|_2 + \theta' + k\Delta t/2) \\ &\leq \sqrt{10}\left(\sqrt{m}(4\sqrt{\log n} + \sqrt{m})/2 + \frac{1}{8} + \frac{1}{4}\right) \\ &\leq 4\sqrt{10m(\log n + m)} \leq \sqrt{k}. \end{aligned}$$

We now show that the conditions of Lemma 1 for $M = 1, \gamma = 1/10$, and k_0 specified below, are satisfied by $\{B_i^{(t)}\}_{i\in Z_t^{(l)}}$, $\forall l \in [t]$.

First, we observe that the $B_i^{(t)}$ are distributed as $(B_i^{(t)})_m \sim \sqrt{\varepsilon}V + \sqrt{1-\varepsilon}U$ for $\varepsilon = \frac{1}{(1+\|u^*\|_2^2)\sigma_t^2}$, where U is uniform on $[-\sqrt{3}, \sqrt{3}]$ and $V \sim \mathcal{N}(0,1)$. By Lemma 6, $(B_i^{(t)})_m$ is $(1/10, k_0)$-Gaussian convergent for some k_0 and has maximum density at most 1. Recalling that the coordinates of $B_i^{(t)}$, $i \in Z_t'$, are independent and $(B_i^{(t)})_j \sim$

$\mathcal{N}(0,1), \forall j \in [m-1]$, we see that $B_i^{(t)}$ has independent $(1/10, k_0)$-Gaussian convergent entries of maximum density at most 1. Lastly, we note that

$$\begin{aligned} k &= 165\,000m(\log(n)+m) \geq 165\,000(m^2+k_0 m) \\ &\geq \max\{(4\sqrt{m}+2)k_0, 144m^{\frac{3}{2}}(\log 1+3), 150\,000(\gamma+1)m^{\frac{7}{4}}\} \end{aligned}$$

as needed to apply Lemma 1, using that $n \geq \exp(k_0)$.

Therefore, applying Lemma 1, for each $l \in [t]$, with probability at least $1-1/25$, there exists a set $T_l \subseteq Z_t^{(l)}$ of size k such that:

$$\left\| \sum_{i \in T_l} B_i^{(t)} - (\Sigma^{(t)})^{-1}(Rd' - e_m k \mu_t) \right\|_\infty \leq \theta. \tag{21}$$

Call the event that (21) is valid for any of the t sets E_t. Because the success probabilities for each of the t sets are independent, we get:

$$\Pr[\neg E_t \mid |Z_t'| \geq \lceil 2\sqrt{m} \rceil tk] \leq \left(1 - \frac{1}{25}\right)^t.$$

Combining this with Equation (20), we see that $\Pr[\neg E_t] \leq 2 \cdot (1-\frac{1}{25})^t$. If E_t occurs, we choose $T \subseteq Z_t'$, $|T| = k$, satisfying (21). Then,

$$\begin{aligned} \left\| \sum_{i \in T} A_{\cdot,i} - d' \right\|_\infty &\leq \left\| \sum_{i \in T} A_{\cdot,i} - d' \right\|_2 = \left\| \sum_{i \in T} B_{\cdot,i} - Rd' \right\|_2 \\ &= \left\| \sum_{i \in T} (\Sigma^{(t)}) B_{\cdot,i}^{(t)} + k\mu_t e_m - Rd' \right\|_2 \\ &\leq \max(1, \sigma_t)\sqrt{m} \left\| \sum_{i \in T} B_{\cdot,i}^{(t)} - (\Sigma^{(t)})^{-1}(Rd' - e_m k \mu_t) \right\|_\infty \\ &\leq 2\sqrt{m}\theta = \theta'. \end{aligned}$$

Now we will show that when E_t occurs, x'' is feasible and $\|A(x''-x^*)\|_\infty = O(1/n)$. First we check feasibility:

$$\begin{aligned} \sum_{i=1}^m x_i'' a_{ji} &= (Ax')_j + \sum_{i \in T} a_{ji} \leq (Ax')_j + d_j' + \theta' \\ &= (Ax')_j + (A(x^* - x'))_j = (Ax^*)_j \leq b_j. \end{aligned}$$

Hence the solution is feasible for our LP. We also have

$$\begin{aligned} \|A(x''-x^*)\|_\infty &= \|Ax'' - Ax' - d\|_\infty \\ &= \|\sum_{i \in T} A_{\cdot,i} - d'\|_\infty \leq \|\sum_{i \in T} A_{\cdot,i} - d\|_\infty + \theta' \leq 2\theta'. \end{aligned}$$

Now we can finalize our initial computation:

$$\operatorname{val}(x^*) - \operatorname{val}(x'') \leq \sqrt{m}\|u^*\|_2 \left\|A(x''-x^*)\right\|_\infty + t\Delta k$$

$$\begin{aligned}
&\le 6\sqrt{m}\theta' + 10\,000 \cdot \frac{\sqrt{m}\cdot t \cdot k^2}{n} \\
&\le \frac{12}{32mn} + 27226 \cdot 10^{10} t \cdot \frac{m^{2.5}(\log n + m)^2}{n} \\
&\le 10^{15} t \cdot \frac{m^{2.5}(\log n + m)^2}{n}. \quad \square
\end{aligned}$$

Now we have all ingredients to prove Theorem 1.

Proof (Theorem 1) Substituting $\varepsilon = 1/9$ in Lemma 8 gives

$$\begin{aligned}
\delta &= \sqrt{2\pi}\|b^-\|_2/n \le \sqrt{2\pi}/10 \le 1/3 \\
\alpha &= \frac{1}{\sqrt{2\pi}}\sqrt{\frac{1-3\varepsilon}{1-\varepsilon} - \delta^2} \ge \frac{1}{\sqrt{2\pi}}\sqrt{\left(\frac{3}{4}\right)^2 - \delta^2} \\
\|u^*\|_2 &\le \frac{\varepsilon+1}{1-3\varepsilon-(1-\varepsilon)\delta} \le 3
\end{aligned}$$

Now note that $H(499/500) < \frac{\alpha^2}{4}$, so $\beta < 499/500$. If we choose n, in such a way that $(499/500-\beta)n \ge m$, then $(1-\beta)n - m \ge n/500$. So Lemma 8 yields that the probability that conditions 2 and 3 do not hold is at most $(20)^{m+1}\exp(-2^{-11}n) + \exp(-n/16\pi)$. We can still choose C such that $n \ge Cm$ and $t \le \frac{n}{Cm^{2.5}(m+\log n)^2}$, so by taking $C \ge 2^{12}$, this probability is smaller than $2(1-1/25)^t$. By Lemma 4 condition 1 holds with probability at least $1-n^{-7}$.

If 1, 2 and 3 hold Lemma 14 implies that $\mathsf{IPGAP}(A,b,c) \ge 10^{15} t \cdot \frac{m^{2.5}(\log n+m)^2}{n}$ holds with probability at most $2(1-1/25)^t$, as long as we set $C \ge \max(\exp(k_0), 10^{15})$ for k_0 as defined in Lemma 6. So:

$$\begin{aligned}
&\Pr\left[\mathsf{IPGAP}(A,b,c) > 10^{15} t \cdot \frac{m^{2.5}(\log n + m)^2}{n}\right] \\
&\le 2\cdot\left(1-\frac{1}{25}\right)^t + n^{-7} + 2\left(1-\frac{1}{25}\right)^t \\
&\le 4(1-1/25)^t + n^{-7}. \quad \square
\end{aligned}$$

6 Proof of the Improved Discrepancy Lemma

In this section, we prove Lemma 1, which is an improved variant of a discrepancy lemma of Dyer and Frieze [5, Lemma 3.4]. This lemma is the main tool used to restore feasibility and near-optimality of the rounding x' of x^* in the preceding section. Compared to Dyer and Frieze's original lemma, the main difference is that we achieve a constant probability of success instead of $(2/\sqrt{3})^{-m}$, which is crucial for reducing the exponential in m dependence of the integrality gap down to polynomial in m. For this purpose, we choose our subsets of size k from a slightly larger universe, i.e., $2\sqrt{mk}$ instead of $2k$, to reduce correlations. We also provide a tighter analysis of the lemma allowing us to show that the conclusion holds as long as $k = \Omega(m^{7/4})$, where the $\Omega(\cdot)$ hides a dependence on the parameters of the underlying coordinate distributions. We restate the lemma below.

Lemma 1 *For $k,m \in \mathbb{N}$, let $a = \lceil 2\sqrt{m} \rceil$ and $\theta > 0$ satisfy $\left(\frac{2\theta}{\sqrt{2\pi k}}\right)^m \binom{ak}{k} = 1$. Let $Y_1,\dots,Y_{ak} \in \mathbb{R}^m$ be i.i.d. random vectors with independent coordinates. For $k_0 \in \mathbb{N}, \gamma \geq 0, M > 0$, assume that $\forall i \in [m]$, $Y_{1,i}$ is a (γ,k_0)-Gaussian convergent continuous random variable with maximum density at most M. Then, if*

$$k \geq \max\{(4\sqrt{m}+2)k_0, 144m^{\frac{3}{2}}(\log M + 3), 150\,000(\gamma+1)m^{\frac{7}{4}}\},$$

for any vector $D \in \mathbb{R}^m$ with $\|D\|_2 \leq \sqrt{k}$ the following holds:

$$\Pr\left[\exists K \subset [ak] : |K| = k, \|(\sum_{j\in K} Y_j) - D\|_\infty \leq \theta\right] \geq \frac{1}{25}. \tag{3}$$

Our proof of Lemma 1 follows the same proof strategy as [5, Lemma 3.4]. Namely, we use the second moment method to lower bound the success probability by $\mathbb{E}[Z]^2/\mathbb{E}[Z^2]$, where Z counts the number of the k-subsets of $[ak]$ which satisfy (3). The value of θ is calibrated to ensure that

$$\mathbb{E}[Z] = \binom{ak}{k} \Pr\left[\|(\sum_{j=1}^{k} Y_j) - D\|_\infty \leq \theta\right] \approx 1.$$

This relies upon the fact that $\sum_{j=1}^k Y_j$ is very close in distribution to $\mathcal{N}(0,kI_m)$ and that the target D is close to the origin. To lower bound the ratio $\mathbb{E}[Z]^2/\mathbb{E}[Z^2]$, the key challenge from here is to show that for two *typical* k-subsets $K_1,K_2 \subset [ak]$, the joint success probability is close to that of the *independent* case, i.e., where $K_1 \cap K_2 = \emptyset$. Namely, we wish to show that

$$\Pr\left[\|(\sum_{j\in K_i} Y_j) - D\|_\infty \leq \theta, i \in \{1,2\}\right] \leq C \Pr\left[\|(\sum_{j=1}^{k} Y_j) - D\|_\infty \leq \theta\right]^2 \tag{22}$$

for some not too large $C \geq 1$. This turns out to yield a lower bound $\mathbb{E}[Z]^2/\mathbb{E}[Z]^2 = \Omega(1/C)$. To upper bound C, we rely upon the following key technical lemma, which upper bounds the probability that two correlated *nearly-Gaussian* random sums land in the same interval.

Lemma 15 *Let $r,k \in \mathbb{N}$ such that $\alpha := k/(k-r) \geq 4/3$. Let $Y_1, Y_2, \dots Y_{k+r}$ be i.i.d. copies of a (γ,k_0)-Gaussian convergent random variable Y. Let $U = \sum_{j=1}^{r} Y_j, V = \sum_{j=r+1}^{k} Y_j, W = \sum_{j=k+1}^{r+k} Y_j$. Assume that $k \geq 280(\gamma+1)\alpha^{\frac{7}{2}}$, $\min\{r, k-r\} \geq k_0$ and $\theta \in \left[0, \frac{1}{\sqrt{k}}\right]$. For any $D \in \mathbb{R}$, we have that*

$$\Pr[U+V \in [D-\theta, D+\theta], W+V \in [D-\theta, D+\theta]] \leq \frac{4\theta^2}{2\pi k}\left(1+\frac{16}{9\alpha^2}\right).$$

The correlation is quantified by $\alpha > 0$, which controls the number of shared terms k/α in both sums. As α increases, the probability of the joint event approaches the worst-case bound for the independent case. More precisely, if X_1 and X_2 are independent $\mathscr{N}(0,k)$ (i.e., sums of disjoint sets of k independent standard Gaussians), $D = 0$ and θ is very small, then

$$\Pr[X_1 \in [-\theta,\theta], X_2 \in [-\theta,\theta]] = \Pr_{X\sim\mathscr{N}(0,1)}\left[X \in \left[-\frac{\theta}{\sqrt{k}}, \frac{\theta}{\sqrt{k}}\right]\right]^2$$
$$\approx \left(\frac{2\theta}{\sqrt{k}}\varphi(0)\right)^2 = \frac{4\theta^2}{2\pi k}.$$

In their proof, Dyer and Frieze proved a special case of the above for $\alpha \approx 2$. In their setting, they pick $K_1, K_2 \subseteq [2k]$ of size k, where the *typical* intersection size is $|K_1 \cap K_2| \approx k/2$. To upper bound C in (22), they apply Lemma 15 to each of the coordinates $\sum_{j\in K_1} Y_{j,i}, \sum_{j\in K_2} Y_{j,i}$, for $i \in [m]$, with $\alpha \approx 2$, to deduce a $2^{O(m)}$ bound on C. By increasing the universe size from $[2k]$ to $[ak]$, where $a = \lceil 2\sqrt{m}\rceil$, we reduce the typical intersection size to k/a. This allows us to set $\alpha \approx 1/a$, which reduces C to roughly $(1+\frac{1}{a^2})^m = O(1)$.

Remark on Gaussian convergence. The attentive reader may have noticed that our definition of $(\gamma,k_0)-$Gaussian convergence requires the density of the normalized sum of k i.i.d. copies of a random variable $X \in \mathbb{R}$ to be within $\frac{\gamma}{k}$ of the density of the standard Gaussian, for $k \geq k_0$, which is a stronger requirement than the more conventional $\frac{\gamma}{\sqrt{k}}$-type convergence. Indeed, this faster rate of convergence can only be expected when the first three moments of X match those of the standard Gaussian (as is the case in our application), whereas the slower rate requires only that X have mean 0 and variance 1 and that the density of X be sufficiently "nice". One can show however that Lemma 1 still holds only assuming $\frac{\gamma}{\sqrt{k}}$-convergence, provided that the requirement $k \geq 150000(\gamma+1)m^{7/4}$ is replaced by $k \geq 150000^2(\gamma+1)^2 m^{7/2}$. It is in fact easy to verify that every inequality in the proof of Lemma 1 utilizing Gaussian convergence is preserved when replacing k by $\sqrt{k}$, and hence the required lower bound on k is replaced by the same lower bound for $\sqrt{k}$. The larger $m^{7/2}$ dependence on m would however increase the dependence on m in Theorem 1, which is why we chose to state Lemma 1 with the stronger Gaussian convergence requirement.

We now prove Lemma 1 using Lemma 15, deferring the proof of the latter to the end of the section.

Proof (Lemma 1) For every $K \subset [ak]$ let E_K denote the event that $\|\sum_{j\in K} Y_j - D\|_\infty \leq \theta$. Let $Z = \sum_{K\subset[ak],|K|=k} \mathbf{1}_{E_K}$.

$$\Pr\left[\exists K \subset [ak] : |K| = k, \|\sum_{j\in K} Y_j - D\|_\infty \leq \theta\right] = \Pr[Z > 0].$$

By an application of the second moment method, if Z is a positive-integer valued random variable of finite mean and variance then

$$\Pr[Z>0] \geq \frac{\mathbb{E}[Z]^2}{\mathbb{E}[Z^2]}.$$

Define $K_i = \{i+1, i+2, \ldots, i+k\}, 0 \leq i \leq k$.

$$\mathbb{E}[Z] = \sum_{K \subset [ak], |K|=k} \Pr[E_K] = \binom{ak}{k} \Pr[E_{K_0}],$$
$$\mathbb{E}[Z^2] = \sum_{K,K' \subset [ak], |K|=|K'|=k} \Pr[E_K \cap E_{K'}],$$

where the first equality follows since $Y_1, \ldots, Y_{ak}$ are independent and identically distributed. By the same reasoning, $\Pr[E_K \cap E_{K'}]$ depends only on $|K \cap K'|$. Noting that $|K_0 \cap K_r| = k - r$ for $0 \leq r \leq k$, we may rewrite $\mathbb{E}[Z^2]$ as

$$\mathbb{E}[Z^2] = \binom{ak}{k}^2 \sum_{r=0}^{k} \Pr\big[\big|K \cap K'\big| = k-r\big] \Pr[E_{K_0} \cap E_{K_r}],$$

where K, K' are independent uniformly random subsets of $[ak]$ of size k. Since we have $\Pr[|K \cap K'| = k] = \binom{ak}{k}^{-1}$,

$$\mathbb{E}[Z^2] = \mathbb{E}[Z] + \binom{ak}{k}^2 \sum_{r=1}^{k} \Pr\big[\big|K \cap K'\big| = k-r\big] \Pr[E_{K_0} \cap E_{K_r}].$$

In the rest of the proof, we show $\mathbb{E}[Z^2]/\mathbb{E}[Z]^2 \leq 25$, which implies the desired lower bound $\mathbb{E}[Z]^2/\mathbb{E}[Z^2] \geq 1/25$. This will be established for sufficiently large k and the required constraints for k will be collected at the end. Let

$$J := [k] \cap \left(k - \frac{(1+\varepsilon)k}{a}, k - \frac{(1-\varepsilon)k}{a}\right) \text{ for } \varepsilon = \frac{1}{2}.$$

We decompose

$$\mathbb{E}[Z^2]/\mathbb{E}[Z]^2 = \frac{1}{\mathbb{E}[Z]} + \sum_{r \in J} \Pr[|K \cap K'| = k-r] \frac{\Pr[E_{K_0} \cap E_{K_r}]}{\Pr[E_{K_0}]^2}$$
$$+ \sum_{r \in [k] \setminus J} \Pr[|K \cap K'| = k-r] \frac{\Pr[E_{K_0} \cap E_{K_r}]}{\Pr[E_{K_0}]^2}. \tag{23}$$

We upper bound $\mathbb{E}[Z^2]/\mathbb{E}[Z]^2$ in three steps. First, a lower bound on $\mathbb{E}[Z]$. Then, a worst-case upper bound on $\frac{\Pr[E_{K_0} \cap E_{K_r}]}{\Pr[E_{K_0}]^2}$ for every $1 \leq r \leq k$, and finally a superior average-case upper bound on $\frac{\Pr[E_{K_0} \cap E_{K_r}]}{\Pr[E_{K_0}]^2}$ for $r \in J$.

Let us prove a trivial upper bound for θ. Since $\binom{ak}{k} \geq a^k$, we have

$$\left(\frac{2\theta}{\sqrt{2\pi k}}\right)^m \binom{ak}{k} = 1 \Rightarrow \theta \leq a^{-k/m} \sqrt{\frac{\pi k}{2}}.$$

If $k \geq 2m\log m$, then $k/m \geq \log k$, so that

$$\theta \leq \left(2\sqrt{m}\right)^{-\frac{k}{m}} \sqrt{\frac{\pi k}{2}} \leq 2^{-\frac{k}{m}} \sqrt{\frac{\pi k}{2}} \leq \frac{1}{\sqrt{k}}.$$

Henceforth, we will assume $\theta \leq 1/\sqrt{k}$ and require $k \geq 2m\log m$.

Lower Bound for the Expectation. The goal of this subsection is to show that

$$\mathbb{E}[Z] = \binom{ak}{k} \Pr[E_{K_0}] \geq 1/e.$$

Since the individual coordinates of each vector Y_j, $j \in [k]$, are independent:

$$\Pr[E_{K_0}] = \prod_{i=1}^{m} \Pr\left[\sum_{j=1}^{k} Y_{j,i} \in [D_i - \theta, D_i + \theta]\right].$$

Each term in this product is at least

$$\Pr\left[\sum_{j=1}^{k} Y_{j,i} \in [D_i - \theta, D_i + \theta]\right] \geq \frac{2\theta}{\sqrt{2\pi k}} \exp\left(-\frac{D_i^2}{2k} - \frac{2e\sqrt{2\pi}(\gamma+1)}{k}\right), \forall 1 \leq i \leq m.$$

To prove this estimate, let $\bar{g}_i^k$ denote the density function of $\sum_{j=1}^{k} Y_{j,i}/\sqrt{k}$. Then,

$$\Pr\left[\sum_{j=1}^{k} Y_{j,i} \in [D_i - \theta, D_i + \theta]\right] = \Pr\left[\sum_{j=1}^{k} \frac{Y_{j,i}}{\sqrt{k}} \in \left[\frac{D_i - \theta}{\sqrt{k}}, \frac{D_i + \theta}{\sqrt{k}}\right]\right] = \int_{\frac{D_i-\theta}{\sqrt{k}}}^{\frac{D_i+\theta}{\sqrt{k}}} \bar{g}_i^k(x)\mathrm{d}x.$$

Recall that for $i \in [m]$, $Y_{1,i}, \ldots, Y_{k,i}$ are i.i.d. (γ, k_0)-Gaussian convergent random variables. If $k \geq k_0$, we have that

$$\bar{g}_i^k(x) \geq \varphi(x) - \frac{\gamma}{k}, \forall x \in \mathbb{R},$$

where $\varphi(x) := \frac{\exp(-x^2/2)}{\sqrt{2\pi}}$. Assuming that $k \geq 4\sqrt{2\pi}e(\gamma+1)m$, the desired estimate is derived as follows:

$$\begin{aligned}
\int_{\frac{D_i-\theta}{\sqrt{k}}}^{\frac{D_i+\theta}{\sqrt{k}}} \bar{g}_i^k(x)\mathrm{d}x &\geq \int_{\frac{D_i-\theta}{\sqrt{k}}}^{\frac{D_i+\theta}{\sqrt{k}}} \varphi(x) - \frac{\gamma}{k}\mathrm{d}x \\
&\geq \int_{\frac{D_i-\theta}{\sqrt{k}}}^{\frac{D_i+\theta}{\sqrt{k}}} \varphi\left(\frac{D_i}{\sqrt{k}}\right) - \frac{\theta}{\sqrt{k}} - \frac{\gamma}{k}\mathrm{d}x \quad (\ \varphi \text{ is 1-Lipschitz }) \\
&\geq \frac{2\theta}{\sqrt{2\pi k}} \left(\exp\left(-\frac{D_i^2}{2k}\right) - \frac{\sqrt{2\pi}\theta}{\sqrt{k}} - \frac{\sqrt{2\pi}\gamma}{k}\right) \\
&\geq \frac{2\theta}{\sqrt{2\pi k}} \left(\exp\left(-\frac{D_i^2}{2k}\right) - \frac{\sqrt{2\pi}(\gamma+1)}{k}\right) \quad \left(\text{since } \theta \leq \frac{1}{\sqrt{k}}\right)
\end{aligned}$$

$$\geq \frac{2\theta}{\sqrt{2\pi k}} \exp\left(-\frac{D_i^2}{2k}\right)\left(1 - \frac{2\sqrt{2\pi}(\gamma+1)}{k}\right)$$
$$\left(\|D\|_2 \leq \sqrt{k} \Rightarrow \exp\left(-\frac{D_i^2}{2k}\right) \geq 1/2\right)$$
$$\geq \frac{2\theta}{\sqrt{2\pi k}} \exp\left(-\frac{D_i^2}{2k} - \frac{2e\sqrt{2\pi}(\gamma+1)}{k}\right)$$
$$\left(\text{ since } 1 - x/e \geq e^{-x}, 0 \leq x \leq 1 \right).$$

Building on this estimate, we now consider all coordinates and get

$$\begin{aligned}\Pr[E_{K_0}] &\geq \left(\frac{2\theta}{\sqrt{2\pi k}}\right)^m \prod_{i=1}^m \exp\left(-\frac{D_i^2}{2k} - \frac{2e\sqrt{2\pi}(\gamma+1)}{k}\right) \\ &\geq \left(\frac{2\theta}{\sqrt{2\pi k}}\right)^m \exp\left(-\sum_{i=1}^m \frac{D_i^2}{2k} - \frac{1}{2}\right) \quad \left(\text{since } k \geq 4\sqrt{2\pi}e(\gamma+1)m \right) \\ &\geq \left(\frac{2\theta}{\sqrt{2\pi k}}\right)^m \exp(-1) \quad \left(\text{since } \|D_i\|_2 \leq \sqrt{k} \right).\end{aligned}$$

Since $\left(\frac{2\theta}{\sqrt{2\pi k}}\right)^m \binom{ak}{k} = 1$, we conclude that $\mathbb{E}[Z] \geq 1/e$, as needed.

Upper bound on $\frac{\Pr[E_{K_0} \cap E_{K_r}]}{\Pr[E_{K_0}]^2}$. To analyze $\Pr[E_{K_0} \cap E_{K_r}]$, for $r \in [k]$, we define

$$U_i := \sum_{j=1}^r Y_{j,i}, V_i := \sum_{j=r+1}^k Y_{j,i}, W_i := \sum_{j=k+1}^{r+k} Y_{j,i}, \text{ for } i \in [m].$$

The probability of $E_{K_0} \cap E_{K_r}$ can now be expressed as follows:

$$\Pr[E_{K_0} \cap E_{K_r}] = \prod_{i=1}^m \Pr[U_i + V_i \in [D_i - \theta, D_i + \theta], W_i + V_i \in [D_i - \theta, D_i + \theta]]. \quad (24)$$

For $i \in [m]$, let $g_i : \mathbb{R} \to \mathbb{R}_+$ denote the probability density of $Y_{1,i}$, which satisfies $\sup_{x \in \mathbb{R}} g_i(x) \leq M$ by assumption. Since $Y_{1,i}, \ldots, Y_{k,i}$ are i.i.d., we see that U_i, W_i, V_i are independent, that U_i, W_i both have density g_i^{*r} and that V_i has density $g_i^{*(k-r)}$, where $g_i^{*r}, g_i^{*(k-r)}$ are the r and $k-r$-fold convolutions of g_i.

We now split the analysis into a worst-case bound for any $r \in [k]$ and an average case bound for $r \in J$.

Worst-case Upper Bound. Since the convolution operation does not increase the maximum density, we note that $\max_{x\in\mathbb{R}} g_i^{*r}(x) \le M$, $i \in [m]$. From here, we see that

$$\begin{aligned}
&\Pr[U_i+V_i \in [D_i-\theta, D_i+\theta], W_i+V_i \in [D_i-\theta, D_i+\theta]] \\
&\quad = \int_{-\infty}^{\infty} \Pr[U_i \in [D_i-\theta-y, D_i+\theta-y]]^2 g_i^{*(k-r)}(y)dy \\
&\quad = \int_{-\infty}^{\infty} \left(\int_{-\theta}^{\theta} g_i^{*r}(D_i-y+x)dx\right)^2 g_i^{*(k-r)}(y)dy \\
&\quad \le \int_{-\infty}^{\infty} \left(\int_{-\theta}^{\theta} M dx\right)^2 g_i^{*(k-r)}(y)dy = (2M\theta)^2.
\end{aligned}$$

By (24), this gives $\Pr[E_{K_0} \cap E_{K_r}] \le (2\theta M)^{2m}$ for $r \in [k]$. The worst case upper bound is therefore

$$\frac{\Pr[E_{K_0} \cap E_{K_r}]}{\Pr[E_{K_0}]^2} \le e^2 \left(\frac{(2\theta M)\sqrt{2\pi k}}{2\theta}\right)^{2m} \le e^2 \left(M\sqrt{2\pi k}\right)^{2m}.$$

Before moving on to the average case bound for $\Pr[E_{K_0} \cap E_{K_r}]^2 / \Pr[E_{K_0}]^2$ for $r \in J$, we first upper bound the total contribution to (23) of the terms associated with $r \in [k] \setminus J$. Recall that $J = [k] \cap \left(k - \frac{(1+\varepsilon)k}{a}, k - \frac{(1-\varepsilon)k}{a}\right)$ for $\varepsilon = \frac{1}{2}$. For $r \in J$, note that $|K_0 \cap K_r| \in (\frac{k}{2a}, \frac{3k}{2a})$. That is, $|K_0 \cap K_r|$ is close to the average intersection size $\frac{k}{a}$ for two uniform k-subsets K, K' of $[ak]$. Applying Lemma 2 together with the worst-case upper bound for indices $r \in [k] \setminus J$, assuming that k is large enough (to be specified below), we get that

$$\begin{aligned}
&\sum_{r\in[k]\setminus J} \Pr[|K \cap K'| = k-r] \frac{\Pr[E_{K_0} \cap E_{K_r}]}{\Pr[E_{K_0}]^2} \\
&\qquad \le e^2 (M\sqrt{2\pi k})^{2m} \Pr\left[|K \cap K'| \notin (\frac{k}{2a}, \frac{3k}{2a})\right] \\
&\qquad \le 4e^2 (M\sqrt{2\pi k})^{2m} \exp(-\frac{k}{12a}) \le 1. \qquad (25)
\end{aligned}$$

To establish the last inequality, using that $a = \lceil 2\sqrt{m} \rceil \le 3\sqrt{m}$ and $x/c \ge \log x$ for $x \ge 2c\log c$ for $c \ge 1$, it suffices to show that the logarithm of the last expression is non-positive:

$$\begin{aligned}
&\log(4e^2) + 2m\left(\log M + \frac{\log(2\pi k)}{2}\right) - \frac{k}{12a} \\
&\le m\left(\log(8e^2\pi) + 2\log M + \log k - \frac{k}{36m^{3/2}}\right) \\
&\le m\left(\log(8e^2\pi) + 2\log M - \frac{k}{72m^{3/2}}\right) \quad \left(\text{if } k \ge 144m^{3/2}\log(72m^{3/2})\right), \\
&\le 0 \quad \left(\text{if } k \ge 72m^{3/2}(2\log M + \log(8e^2\pi))\right).
\end{aligned}$$

Simplifying the conditions above, (25) holds for

$$k \geq \max\{144m^{3/2}(\log M+3), 216m^{3/2}(\log m+3)\}.$$

Average-case Upper Bound. Observe that for $r \in J$, we have $r = k - \frac{k}{\alpha}$ for $\alpha \in \left[\frac{4}{3}\sqrt{m}, 4\sqrt{m}+2\right]$. To derive the average case bound, we will apply Lemma 15 to U_i, W_i, V_i, D_i and θ for each $i \in [m]$ with parameters k,r,α and γ, k_0. We first show that the requisite conditions are satisfied for k large enough. Firstly, recall that $\theta \in [0, 1/\sqrt{k}]$, $\alpha \geq 4/3\sqrt{m} \geq 4/3$, and that $Y_{1,i}, \ldots, Y_{k,i}$ are i.i.d. (γ, k_0)-Gaussian convergent random variables. Assuming $k \geq (4\sqrt{m}+2)k_0$, we have that $\min\{r, k-r\} = k\min\{\frac{1}{\alpha}, (1-\frac{1}{\alpha})\} \geq k_0$. Lastly, assuming $k \geq 280(\gamma+1)(4\sqrt{m}+2)^{7/2}$, we also have $k \geq 280(\gamma+1)\alpha^{7/2}$ by assumption on α. Therefore, for $r \in J$, we may apply Lemma 15 to (24) to conclude that

$$\begin{aligned}\frac{\Pr[E_{K_0} \cap E_{K_r}]}{\Pr[E_{K_0}]^2} &\leq \left(\left(1+\frac{16}{9\alpha^2}\right)\frac{4\theta^2}{2\pi k}\right)^m e^2 \left(\frac{\sqrt{2\pi k}}{2\theta}\right)^{2m} \leq e^2\left(1+\frac{16}{9\alpha^2}\right)^m \\ &\leq e^2\left(1+\frac{1}{m}\right)^m \leq e^3.\end{aligned}$$

Because $4\sqrt{m}+2 \leq 6\sqrt{m}$, we simplify the condition $k \geq 280(\gamma+1)(4\sqrt{m}+2)^{7/2}$ to the stronger condition $k \geq 150000(\gamma+1)m^{7/4}$.

In Conclusion: Combining the bounds from the previous sections, assuming that the constraints on k are all satisfied, we get that:

$$\begin{aligned}\mathbb{E}[Z^2]/\mathbb{E}[Z]^2 &= \frac{1}{\mathbb{E}[Z]} + \sum_{r=1}^{k} \Pr[|K \cap K'| = k-r]\frac{\Pr[E_{K_0} \cap E_{K_r}]}{\Pr[E_{K_0}]^2} \\ &\leq e + \sum_{r \in [k]\setminus J} \Pr[|K \cap K'| = k-r]\frac{\Pr[E_{K_0} \cap E_{K_r}]}{\Pr[E_{K_0}]^2} \\ &\qquad + \sum_{r \in J} \Pr[|K \cap K'| = k-r]\frac{\Pr[E_{K_0} \cap E_{K_r}]}{\Pr[E_{K_0}]^2} \\ &\leq e + \sum_{r \in [k]\setminus J} \Pr[|K \cap K'| = k-r] e^2 \left(M\sqrt{2\pi k}\right)^m \\ &\qquad + \sum_{r \in J} \Pr[|K \cap K'| = k-r] e^3 \\ &\leq e + 1 + e^3 \leq 25.\end{aligned}$$

Aggregating the constraints on k, the above gives a lower bound of $1/25$ for the success probability whenever k satisfies:

$$\begin{aligned}k \geq \max\{&2m\log m, k_0, 4\sqrt{2\pi}e(\gamma+1)m,\ 144m^{3/2}(\log M+3),\ 216m^{3/2}(\log m+3), \\ &(4\sqrt{m}+2)k_0,\ 150000(\gamma+1)m^{7/4}\}.\end{aligned}$$

Removing the dominated terms, it suffices for k to satisfy

$$k \geq \max\{(4\sqrt{m}+2)k_0,\ 144m^{\frac{3}{2}}(\log M+3),\ 150000(\gamma+1)m^{\frac{7}{4}}\},$$

as needed. □

Proof (Lemma 15) For convenience, let $P = \Pr[U+V \in [D-\theta, D+\theta], W+V \in [D-\theta, D+\theta]]$, and define the normalized sums: $\bar{U} = U/\sqrt{r}, \bar{V} = V/\sqrt{k-r}, \bar{W} = W/\sqrt{r}$. Let us denote the density of $\bar{U}$ and $\bar{W}$ by $\bar{g}$, and the density of $\bar{V}$ by $\bar{h}$. $\bar{U}, \bar{W}$ are independent and identically distributed, from which we see that

$$\begin{aligned}
P &= \Pr\left[\bar{U}\sqrt{r} + \bar{V}\sqrt{k-r} \in [D-\theta, D+\theta], \bar{W}\sqrt{r} + \bar{V}\sqrt{k-r} + \in [D-\theta, D+\theta]\right] \\
&= \int_{-\infty}^{\infty} \Pr\left[\bar{U} \in \left[\frac{D-\theta}{\sqrt{r}} - y\sqrt{\frac{k-r}{r}}, \frac{D+\theta}{\sqrt{r}} - y\sqrt{\frac{k-r}{r}}\right]\right]^2 \bar{h}(y) \\
&= \int_{-\infty}^{\infty} \left(\int_{-\frac{\theta}{\sqrt{r}}}^{\frac{\theta}{\sqrt{r}}} \bar{g}\left(\frac{D}{\sqrt{r}} - \frac{y}{\sqrt{\alpha-1}} + x\right) \mathrm{d}x \right)^2 \bar{h}(y).
\end{aligned}$$

Since Y is (γ, k_0)-Gaussian convergent and $\min\{r, k-r\} \geq k_0$, we have that

$$|\bar{g}(x) - \varphi(x)| \leq \frac{\gamma}{r} = \frac{\alpha}{\alpha-1}\frac{\gamma}{k}, \quad |\bar{h}(x) - \varphi(x)| \leq \frac{\gamma}{k-r} = \alpha\frac{\gamma}{k}, \quad \forall x \in \mathbb{R}.$$

Using the above, we upper bound P as follows:

$$\begin{aligned}
P &= \int_{-\infty}^{\infty} \left(\int_{-\frac{\theta}{\sqrt{r}}}^{\frac{\theta}{\sqrt{r}}} \bar{g}\left(\frac{D}{\sqrt{r}} - \frac{y}{\sqrt{\alpha-1}} + x\right) \mathrm{d}x \right)^2 \bar{h}(y)\mathrm{d}y \\
&\leq \int_{-\infty}^{\infty} \left(\int_{-\frac{\theta}{\sqrt{r}}}^{\frac{\theta}{\sqrt{r}}} \varphi\left(\frac{D}{\sqrt{r}} - \frac{y}{\sqrt{\alpha-1}} + x\right) + \frac{\gamma}{k}\frac{\alpha}{\alpha-1} \mathrm{d}x \right)^2 \bar{h}(y)\mathrm{d}y \\
&\leq \int_{-\infty}^{\infty} \left(\int_{-\frac{\theta}{\sqrt{r}}}^{\frac{\theta}{\sqrt{r}}} \varphi\left(\frac{D}{\sqrt{r}} - \frac{y}{\sqrt{\alpha-1}}\right) + \frac{2\gamma}{k} + \frac{\theta}{\sqrt{r}} \mathrm{d}x \right)^2 \bar{h}(y)\mathrm{d}y \quad (\varphi \text{ is 1-Lipschitz}) \\
&\leq \frac{4\theta^2}{r} \int_{-\infty}^{\infty} \left(\varphi\left(\frac{D}{\sqrt{r}} - \frac{y}{\sqrt{\alpha-1}}\right) + \frac{2\gamma}{k} + \frac{\theta}{\sqrt{r}} \right)^2 \bar{h}(y)\mathrm{d}y \\
&\leq \frac{4\theta^2}{r} \left(2\varphi(0)\left(\frac{2\gamma}{k} + \frac{\theta}{\sqrt{r}}\right) + \left(\frac{2\gamma}{k} + \frac{\theta}{\sqrt{r}}\right)^2 + \int_{-\infty}^{\infty} \varphi\left(\frac{D}{\sqrt{r}} - \frac{y}{\sqrt{\alpha-1}}\right)^2 \bar{h}(y)\mathrm{d}y \right)
\end{aligned}$$

$$\leq \frac{4\theta^2}{r}\left(\frac{2(\gamma+1)}{k}+\left(\frac{2(\gamma+1)}{k}\right)^2+\int_{-\infty}^{\infty}\varphi\left(\frac{D}{\sqrt{r}}-\frac{y}{\sqrt{\alpha-1}}\right)^2\bar{h}(y)\mathrm{d}y\right),$$

where the last inequality follows from $\theta \leq \frac{1}{\sqrt{k}}$, $r=(1-\frac{1}{\alpha})k \geq k/4$ for $\alpha \geq 4/3$ and $\varphi(0)\leq \frac{1}{2}$. Next, we upper bound the term $Q=\int_{-\infty}^{\infty}\varphi\left(\frac{D}{\sqrt{r}}-\frac{y}{\sqrt{\alpha-1}}\right)^2\bar{h}(y)\mathrm{d}y$.

$$\begin{aligned}
Q &= \int_{-\infty}^{\infty}\varphi\left(\frac{D}{\sqrt{r}}-\frac{y}{\sqrt{\alpha-1}}\right)^2\bar{h}(y)\mathrm{d}y\\
&\leq \int_{-\infty}^{\infty}\varphi\left(\frac{D}{\sqrt{r}}-\frac{y}{\sqrt{\alpha-1}}\right)^2\left(\varphi(y)+\frac{\alpha\gamma}{k}\right)\mathrm{d}y\\
&= \frac{\alpha\gamma}{k}\int_{-\infty}^{\infty}\varphi\left(\frac{D}{\sqrt{r}}-\frac{y}{\sqrt{\alpha-1}}\right)^2\mathrm{d}y+\int_{-\infty}^{\infty}\varphi\left(\frac{D}{\sqrt{r}}-\frac{y}{\sqrt{\alpha-1}}\right)^2\varphi(y)\mathrm{d}y\\
&= \frac{\alpha\sqrt{\alpha-1}\gamma}{\sqrt{2}k}\int_{-\infty}^{\infty}\varphi\left(\frac{y}{\sqrt{2}}\right)^2\mathrm{d}y+\int_{-\infty}^{\infty}\varphi\left(\frac{D}{\sqrt{r}}-\frac{y}{\sqrt{\alpha-1}}\right)^2\varphi(y)\mathrm{d}y\\
&= \frac{\alpha\sqrt{\alpha-1}\gamma}{2\sqrt{\pi}k}+\frac{1}{(2\pi)^{3/2}}\int_{-\infty}^{\infty}\exp\left(-\left(\frac{D}{\sqrt{r}}-\frac{y}{\sqrt{\alpha-1}}\right)^2-\frac{y^2}{2}\right)\mathrm{d}y\\
&= \frac{\alpha\sqrt{\alpha-1}\gamma}{2\sqrt{\pi}k}+\frac{1}{(2\pi)^{3/2}}\int_{-\infty}^{\infty}\exp\left(-\frac{y^2}{2}\left(1+\frac{2}{\alpha-1}\right)+\frac{2yD}{\sqrt{r(\alpha-1)}}-\frac{D^2}{r}\right)\mathrm{d}y\\
&\leq \frac{\alpha\sqrt{\alpha-1}\gamma}{2\sqrt{\pi}k}\\
&+\frac{1}{(2\pi)^{3/2}}\int_{-\infty}^{\infty}\exp\left(-\frac{1}{2}\left(y\sqrt{\frac{\alpha+1}{\alpha-1}}-\frac{2D}{\sqrt{(\alpha+1)r}}\right)^2+\frac{2D^2}{(\alpha+1)r}-\frac{D^2}{r}\right)\mathrm{d}y\\
&\leq \frac{\alpha\sqrt{\alpha-1}\gamma}{2\sqrt{\pi}k}+\frac{1}{(2\pi)^{3/2}}\int_{-\infty}^{\infty}\exp\left(-\frac{1}{2}\left(y\sqrt{\frac{\alpha+1}{\alpha-1}}\right)^2\right)\mathrm{d}y\\
&\quad\left(\text{because } \alpha\geq\frac{4}{3}\Rightarrow\frac{2D^2}{(\alpha+1)r}-\frac{D^2}{r}\leq 0\right)\\
&= \frac{\alpha\sqrt{\alpha-1}\gamma}{2\sqrt{\pi}k}+\frac{1}{2\pi}\sqrt{\frac{\alpha-1}{\alpha+1}}.
\end{aligned}$$

The final expression is:

$$P \leq \frac{4\theta^2}{r}\left(\frac{2(\gamma+1)}{k}+\left(\frac{2(\gamma+1)}{k}\right)^2+\frac{\alpha\sqrt{\alpha-1}\gamma}{2\sqrt{\pi}k}+\frac{1}{2\pi}\sqrt{\frac{\alpha-1}{\alpha+1}}\right).$$

Since $r = \frac{\alpha-1}{\alpha}k$, we have

$$P \leq \frac{4\theta^2}{k}\left(1+\frac{1}{\alpha-1}\right)\left(\frac{2(\gamma+1)}{k}+\left(\frac{2(\gamma+1)}{k}\right)^2+\frac{\alpha\sqrt{(\alpha-1)}\gamma}{2\sqrt{\pi}k}+\frac{1}{2\pi}\sqrt{\frac{\alpha-1}{\alpha+1}}\right).$$

We require $P \leq \frac{4\theta^2}{2\pi k}\left(1+\frac{16}{9\alpha^2}\right)$. Using $\alpha \geq 4/3$, observe that

$$\begin{aligned}\left(1+\frac{1}{\alpha-1}\right)\sqrt{\frac{\alpha-1}{\alpha+1}} = \sqrt{1+\frac{1}{\alpha^2-1}} &\leq 1+\frac{1}{2(\alpha^2-1)} \\ &\leq 1+\frac{16}{14\alpha^2} \quad (\text{ since } 14\alpha^2 \leq 32(\alpha^2-1)\ ).\end{aligned}$$

So all we need is

$$\left(1+\frac{1}{\alpha-1}\right)\left(\frac{2(\gamma+1)}{k}+\left(\frac{2(\gamma+1)}{k}\right)^2+\frac{\alpha\sqrt{(\alpha-1)}\gamma}{2\sqrt{\pi}k}\right) \leq \frac{4}{(2\pi)7\alpha^2},$$

because $\frac{16}{14}+\frac{8}{14} \leq \frac{16}{9}$. Using $\alpha \geq \frac{4}{3}$, we simplify the condition to

$$\frac{2(\gamma+1)}{k}+\left(\frac{2(\gamma+1)}{k}\right)^2+\frac{\alpha\sqrt{(\alpha-1)}\gamma}{2\sqrt{\pi}k} \leq \frac{1}{(2\pi)7\alpha^2}.$$

A stronger condition is $3\cdot\frac{\alpha^{3/2}2(\gamma+1)}{k} \leq \frac{1}{(2\pi)7\alpha^2}$, which is satisfied whenever $k \geq 280(\gamma+1)\alpha^{\frac{7}{2}}$. □

7 Bounding the Tree Size in Branch-and-Bound

The goal of this section is to prove Theorem 3 from the introduction. Our main technical lemma, which will allow us to upper bound (1), is given below.

Lemma 16 *Let $d \in \mathbb{N}, n \geq 100d$, $G \geq 0$, and $W = (W_1,\dots,W_n) \in \mathbb{R}^{d\times n}$ be a matrix whose columns are independent logconcave random vectors with identity covariance. Then, for $\delta \in (0,1)$, with probability at least $1-\delta-e^{-n/5}$, we have that*

$$\max_{\|u\|_2=1} |\{x \in \{0,1\}^n : \sum_{i=1}^n x_i|u^\mathsf{T}W_i| \leq G\}| \leq 60(357n^6)^{d+1}e^{2\sqrt{2nG}}/\delta.$$

Before proving Lemma 16, we first show how to derive Theorem 3 from Theorem 2 and Lemma 16.

Proof (of Theorem 3) For any $\lambda \in \mathbb{R}^m$, we see that

$$\begin{aligned}\{x \in \{0,1\}^n : \sum_{i=1}^n x_i|(c-A^\mathsf{T}\lambda)_i| \leq G\} &= \{x \in \{0,1\}^n : \sum_{i=1}^n x_i|(W^\mathsf{T}(1,-\lambda))_i| \leq tG\} \\ &\subseteq \{x \in \{0,1\}^n : \sum_{i=1}^n x_i|(\frac{W^\mathsf{T}(1,-\lambda)}{\|(1,-\lambda)\|_2})_i| \leq G\}.\end{aligned}$$

Given the above, we have that

$$\max_{\lambda \in \mathbb{R}^m} |\{x \in \{0,1\}^n : \sum_{i=1}^n x_i |(c - A^\mathsf{T}\lambda)_i| \leq G\}|$$
$$\leq \max_{\|u\|_2=1} |\{x \in \{0,1\}^n : \sum_{i=1}^n x_i |(W^\mathsf{T} u)_i| \leq G\}|. \tag{26}$$

Applying Lemma 16 to W, δ and $d = m+1$, with probability at least $1 - \delta - e^{-n/5}$, we get that

$$\max_{\|u\|_2=1} |\{x \in \{0,1\}^n : \sum_{i=1}^n x_i |(u^\mathsf{T} W)_i| \leq tG\}|$$
$$\leq 60(357n^6)^{m+2} e^{2\sqrt{2nG}}/\delta = n^{O(m)} e^{2\sqrt{2nG}}/\delta. \tag{27}$$

By Theorem 2 and the union bound, with probability at least

$$1 - \Pr_{A,c}[\mathrm{IPGAP}(A,b,c) \geq G] - \delta - e^{-n/5},$$

we have that the size of the branch-and-bound tree is at most

$$n^{O(m)} e^{2\sqrt{2Gn}}/\delta. \quad \square$$

We now sketch the high level ideas of the proof of Lemma 16. For $g \geq 0$, $u \in \mathbb{S}^{d-1}$, define the knapsack

$$K(u,g) := \{x \in \{0,1\}^n : \sum_{i=1}^n x_i |u^\mathsf{T} W_i| \leq g\}.$$

The proof of the lemma will proceed in two steps. In the first step, we show that the expected size of the knapsack polytope $K(u,G)$ satifies $\mathbb{E}_W[|K(u,G)|] \leq e^{2\sqrt{2nG}}$, for any fixed $u \in \mathbb{S}^{d-1}$ and $G \geq 0$ (see Lemma 17). In the second step, we extend this bound to all $u \in \mathbb{S}^{d-1}$ via the union bound applied to a carefully constructed net of knapsacks of the form $K(u, G+1/n)$ for $u \in \mathscr{K}$, where $\mathscr{K} \subseteq \mathbb{S}^{d-1}$ will have size $|N| = n^{O(d)}$. The slight increase in capacity $G \to G + 1/n$ is to ensure that every $K(u,G)$ knapsack is contained in some $K(u', G+1/n)$ knapsack, for some $u' \in \mathscr{K}$. To ensure this, we rely on Lemma 18 below to help us control the distance between knapsacks induced by nearby u's.

One complicating factor in the construction of $\mathscr{K}$ is the lack of any bound on the norms of the column means $\mu_i := \mathbb{E}[W_i]$, $i \in [n]$. To deal with arbitrary means, we will make use of a hyperplane arrangement $\mathscr{H}$ induced by the μ_i's, such that for any full-dimensional cell C of $\mathscr{H}$ and $i \in [n]$, we have that $|u^\mathsf{T}\mu_i|$ for $u \in C$ either lies in a small interval or is so large that $x_i = 0$ for any $x \in K(u,G)$.

We now give our main bound on the expected size of knapsack polytopes with random weights.

Lemma 17 *Let $\omega_1, \ldots, \omega_n \in \mathbb{R}$ be independent continuous random variables with maximum density at most* 1. *Then, for any $g \geq 0$, we have*

$$\mathbb{E}[|\{x \in \{0,1\}^n : \sum_{i=1}^n x_i |\omega_i| \leq g\}|] \leq e^{2\sqrt{2ng}}.$$

Proof Let $K := \{x \in \{0,1\}^n : \sum_{i=1}^n x_i|\omega_i| \le g\}$. For any $\gamma \ge 0$, we first note that

$$|K| \le e^{\gamma g} \prod_{i=1}^n (1 + e^{-\gamma \omega_i}). \tag{28}$$

To see this, note that each $x \in \{0,1\}^n$ can be associated with the term $e^{\gamma(g - \sum_{i=1}^n x_i \omega_i)}$ on the right hand side (after expanding out the product) and that each term with $x \in K$ contributes at least 1.

For each $i \in [n]$, letting $f_i : \mathbb{R} \to \mathbb{R}_+$ be the probability density of ω_i, we have that

$$\mathbb{E}[e^{-\gamma|\omega_i|}] = \int_0^\infty e^{-\gamma x}(f_i(x) + f_i(-x))dx \le 2\int_0^\infty e^{-\gamma x} dx = \frac{2}{\gamma}, \tag{29}$$

where we have used the assumption that $\max_{x\in\mathbb{R}} f_i(x) \le 1$, $\forall i \in [n]$.

Combining (28), (29), using that $\omega_1, \dots, \omega_n$ are independent, we get that

$$\mathbb{E}[|K|] \le e^{\gamma g} \prod_{i=1}^n \mathbb{E}[1 + e^{-\gamma\omega_i}] \le e^{\gamma g}(1 + \frac{2}{\gamma})^n \le e^{\gamma g + \frac{2n}{\gamma}}.$$

Setting $\gamma = \sqrt{\frac{2n}{g}}$, we get that $\mathbb{E}[|K|] \le e^{2\sqrt{2gn}}$, as claimed. □

We remark that an $e^{\Omega(\sqrt{nG})}$ dependence above is necessary. This holds for $\omega_1, \dots, \omega_n$ uniform in $[0,1]$ and $G \le n$. Letting $S = \{i \in [n] : \omega_i \le \sqrt{G/n}\}$, note that any subset of at most $\lfloor \sqrt{nG} \rfloor$ elements of S fits inside the knapsack K. It is easy to verify that $\mathbb{E}[|S|] = n(\sqrt{G/n}) = \sqrt{nG}$ and that $\Pr[|S| \ge \lfloor \sqrt{nG} \rfloor] \ge 1/2$. In particular,

$$\mathbb{E}[|K|] \ge \mathbb{E}[|\{T : T \subseteq S, |T| \le \lfloor \sqrt{nG} \rfloor\}|] \ge \frac{1}{2} 2^{\lfloor \sqrt{Gn} \rfloor} = e^{\Omega(\sqrt{nG})}, \text{ as needed.}$$

The next lemma give us control on the distance between knapsack weights induced by nearby u's. The proof follows along the same lines as the upper bound in Lemma 9, using Theorem 5 to give the requisite tailbounds.

Lemma 18 *Let $n \ge 100d$ and $W := (W_1, \dots, W_n) \in \mathbb{R}^{d\times n}$ be a matrix whose columns are independent logconcave random vectors with identity covariance. Then,*

$$\Pr[\max_{\|u\|_2=1} \|u^\mathsf{T}(W - \mathbb{E}[W])\|_1 \ge 4n] \le e^{-n/5}.$$

Proof Since the statement of the lemma is invariant to adding a fixed matrix to W, we assume without loss of generality that $\mathbb{E}[W] = 0$. For $i \in [n]$, $\|u\|_2 = 1$, by Theorem 4 we see that $u^\mathsf{T} W_i$ is logconcave. Furthermore, $\mathrm{Var}[u^\mathsf{T} W_i] = \mathbb{E}[(u^\mathsf{T}(W_i - \mathbb{E}[W_i]))^2] = \|u\|_2^2 = 1$. Therefore, for $\lambda \in [0,1)$, by Theorem 5 part 1 we have that:

$$\begin{aligned}
\mathbb{E}[e^{\lambda|u^\mathsf{T} W_i|}] &= \int_0^\infty \Pr[e^{\lambda|u^\mathsf{T} W_i|} \ge t]dt = \int_0^\infty \Pr[|u^\mathsf{T} W_i| \ge \log t/\lambda]dt \\
&\le \int_0^\infty \min\{1, e^{1-\log t/\lambda}\}dt = \int_0^\infty \min\{1, et^{-1/\lambda}\}dt \\
&= e^\lambda + e\int_{e^\lambda}^\infty t^{-1/\lambda} dt = e^\lambda + e\left[\frac{1}{1-1/\lambda} t^{1-1/\lambda}\right]_{e^\lambda}^\infty
\end{aligned}$$

$$= e^{\lambda} + \frac{\lambda}{1-\lambda}e^{\lambda} = \frac{e^{\lambda}}{1-\lambda}.$$

Therefore, for $\|u\|_2 = 1$ and $s \geq 2$, we have that

$$\begin{aligned}\Pr[\sum_{i=1}^{n} |u^{\mathsf{T}} W_i| \geq sn] &\leq \min_{\lambda \in [0,1)} e^{-\lambda sn}\, \mathbb{E}[e^{\sum_{i=1}^{n} \lambda |u^{\mathsf{T}} W_i|}] \\ &\leq \min_{\lambda \in [0,1)} e^{-\lambda sn} \frac{e^{\lambda n}}{(1-\lambda)^n} = e^{-n(s-2-\log(s-1))}, \qquad (30)\end{aligned}$$

where the minimum is attained at $\lambda = \frac{s-2}{s-1} \in [0,1)$. Letting N_ε be a minimal ε-net of $\mathbb{S}^{d-1}$, similar to the computation for (11), we get that

$$\max_{\|u\|_2=1} \|u^{\mathsf{T}} W\|_1 \leq \frac{1}{1-\varepsilon} \max_{u \in N_\varepsilon} \|u^{\mathsf{T}} W\|_1. \qquad (31)$$

Using the above for $\varepsilon = \frac{1}{4}$, we deduce the desired probability bound

$$\begin{aligned}\Pr[\max_{\|u\|_2=1} \|u^{\mathsf{T}} W\|_1 \geq 4n] &\underbrace{\leq}_{\text{by (31)}} \Pr[\max_{u \in N_{1/4}} \|u^{\mathsf{T}} W\|_1 \geq 3n] \\ &\underbrace{\leq}_{\text{by (30)}} |N_{1/4}| e^{-n(1-\log(2))} \leq 9^d e^{-n/4} \underbrace{\leq}_{n \geq 100d} e^{-n/5}. \qquad \square\end{aligned}$$

We now have all the ingredients needed to prove Lemma 16.

Proof (of Lemma 16) Let $K(u,g) := \{x \in \{0,1\}^n : \sum_{i=1}^n x_i |u^{\mathsf{T}} W_i| \leq g\}$ for $u \in \mathbb{S}^{d-1}$, $g \geq 0$. Noting that $|K(u,G)| \leq |\{0,1\}|^n \leq 2^n$, we may assume that $G \leq n$ since otherwise the bound of $n^{O(m)} e^{2\sqrt{2nG}}$ follows trivially.

We begin by constructing a suitable net of knapsacks as described at the beginning of the section. Let $\mu_i := \mathbb{E}[W_i]$, $i \in [n]$, and let $\mathscr{H}$ in denote the hyperplane arrangement on $\mathbb{R}^d$ induced by the hyperplanes $u^{\mathsf{T}} \mu_i = \frac{j}{2n^2}$, $i \in [n]$, $j \in \{-10n^3, \ldots, 10n^3\}$. Noting that this arrangement has $l := n(20n^3+1)$ hyperplanes, it is well-known that the number of d-dimensional cells of $\mathscr{H}$ is at most $\sum_{i=0}^{d} \binom{l}{i} \leq l^{d+1} \leq (21n^4)^{d+1}$ (see for example [19, Proposition 6.1.1]). Letting $\varepsilon = 1/(8n^2)$, for each d-cell C of $\mathscr{H}$, let N_ε^C denote a minimal ε-net of $\mathbb{S}^{d-1} \cap C$. Finally, we let $\mathscr{K} := \cup_C N_\varepsilon^C$, where C ranges over all d-cells of $\mathscr{H}$. The size of $\mathscr{K}$ is bounded by

$$|\mathscr{K}| \leq (1+2/\varepsilon)^d (21n^4)^{d+1} \leq (1+16n^2)^d (21n^4)^{d+1} \leq (357n^6)^{d+1}.$$

Claim Let E_1 denote the event that for all $u \in \mathbb{S}^{d-1}$, there exists $u' \in \mathscr{K}$ such that $K(u,G) \subseteq K(u', G+1/n)$. Then, E_1 holds with probability at least $1 - e^{-n/5}$.

Proof Let E_1' denote the event that $\|u^{\mathsf{T}}(W - \mathbb{E}[W])\|_1 < 4n$, for all $\|u\|_2 \in \mathbb{S}^{d-1}$. By Lemma 18, we see that E_1' holds with probability at least $1 - e^{-n/5}$. To prove the claim, we condition on E_1' and show that E_1 holds.

Take $u \in \mathbb{S}^{d-1}$. Let C denote a d-cell of $\mathscr{H}$ containing u, and let $u' \in \mathscr{K}$ denote the closest point in $N_\varepsilon^C \subseteq C \cap \mathbb{S}^{d-1}$ to u.

Let $B \subseteq [n]$ denote the (possibly empty) subset of indices such that either $\mu_i^\mathsf{T} v \leq -5n$ or $\mu_i^\mathsf{T} v \geq 5n$ is valid for all $v \in C$. Let $A = [n] \setminus B$. Since $5n = \frac{10n^3}{2n^2}$, for all $i \in A$, there exists $j_i \in \{-10n^3, \ldots, 10n^3 - 1\}$ such that $\frac{j_i}{2n^2} \leq \mu_i^\mathsf{T} v \leq \frac{j_i+1}{2n^2}$ is valid for all $v \in C$. In particular this implies that for all $i \in A$, $u, u' \in C$, we have

$$|(u-u')^\mathsf{T} \mu_i| \leq \frac{1}{2n^2}. \tag{32}$$

We first show that if $x \in K(u,G)$, then $x_i = 0, \forall i \in B$. For the sake of contradiction, assume $x \in K(u,G)$ and $x_i = 1$ for some $i \in B$. Then, since $G \leq n$, we have that

$$\sum_{j=1}^n x_j |u^\mathsf{T} W_j| \geq |u^\mathsf{T} W_i| \geq |u^\mathsf{T} \mu_i| - |u^\mathsf{T}(W_i - \mu_i)|$$
$$\underbrace{\geq}_{i \in B} 5n - \|u^\mathsf{T}(W - \mathbb{E}[W])\|_1 \underbrace{>}_{\text{by } E_1'} 5n - 4n \geq G,$$

a clear contradiction to the assumption that $x \in K(u,G)$.

Take $x \in K(u,G)$. We now show that $x \in K(u', G+1/n)$ as follows,

$$\begin{aligned} G &\geq \sum_{i=1}^n x_i |u^\mathsf{T} W_i| \geq \sum_{i=1}^n x_i (|(u')^\mathsf{T} W_i| - |(u-u')^\mathsf{T}(W-\mu_i)| - |(u-u')^\mathsf{T}\mu_i|) \\ &\geq \left(\sum_{i=1}^n x_i |(u')^\mathsf{T} W_i|\right) - \|(u-u')^\mathsf{T}(W - \mathbb{E}[W])\|_1 - \sum_{i \in A} 1/(2n^2) \\ &\quad (\text{ since } x_i = 0, \forall i \in B \text{ and } (32)) \\ &\geq \left(\sum_{i=1}^n x_i |(u')^\mathsf{T} W_i|\right) - 4n\varepsilon - |A|/(2n^2) \geq \left(\sum_{i=1}^n x_i |(u')^\mathsf{T} W_i|\right) - 1/n. \quad \square \end{aligned}$$

For $u \in \mathbb{S}^{d-1}$, by Theorem 4 we know that $u^\mathsf{T} W_i$, for $i \in [n]$, are independent and logconcave. Furthermore, $\mathrm{Var}[u^\mathsf{T} W_i] = \mathbb{E}[u^\mathsf{T}(W_i - \mu_i)^2] = \|u\|_2^2 = 1$, $\forall i \in [n]$. Therefore, by Theorem 5 part 2, the densities of $u^\mathsf{T} W_i$, $i \in [n]$, have maximum density at most 1. Applying Lemma 17 with $\omega_i = u^\mathsf{T} W_i$, $i \in [n]$ and $g = G + 1/n$, we get that $\mathbb{E}_W[|K(u, G+1/n)|] \leq e^{2\sqrt{2n(G+1/n)}} \leq e^{2\sqrt{2nG}+4}$.

Let E_2 denote the event $\forall u \in \mathscr{K}$, $|K(u, G+1/n)| \leq |\mathscr{K}| e^{2\sqrt{2nG}+4}/\delta$ for $\delta \in (0,1)$. By Markov's inequality, for $u \in \mathscr{K}$ we have that

$$\Pr[|K(u, G+1/n)| \geq |\mathscr{K}| e^{2\sqrt{2nG}+4}/\delta] \leq \delta/(|\mathscr{K}|).$$

Therefore, by the union bound, E_2 occurs with probability at least $1 - \delta$.

By the above claim, noting that

$$|\mathscr{K}| e^{2\sqrt{2nG}+4} \leq (357n^6)^{d+1} e^{2\sqrt{2nG}+4} \leq 60(357n^6)^{d+1} e^{2\sqrt{2nG}},$$

we see that

$$\begin{aligned} \Pr[\max_{u \in \mathbb{S}^{d-1}} |K(u,G)| \leq 60(357n^6)^{d+1} e^{2\sqrt{2nG}}/\delta] &\geq 1 - \Pr[\neg E_1] - \Pr[\neg E_2] \\ &\geq 1 - \delta - e^{-n/5}, \end{aligned}$$

as needed. $\square$

References

1. Beier, R., Vöcking, B.: Probabilistic analysis of knapsack core algorithms. In: J.I. Munro (ed.) Proceedings of the Fifteenth Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2004, New Orleans, Louisiana, USA, January 11-14, 2004, pp. 468–477. SIAM (2004)
2. Borst, S., Dadush, D., Huiberts, S., Tiwari, S.: On the Integrality Gap of Binary Integer Programs with Gaussian Data. In: M. Singh, D.P. Williamson (eds.) Integer Programming and Combinatorial Optimization, Lecture Notes in Computer Science, pp. 427–442. Springer International Publishing, Cham (2021). DOI 10.1007/978-3-030-73879-2_30
3. Dey, S.S., Dubey, Y., Molinaro, M.: Branch-and-bound solves random binary IPs in polytime. In: Proceedings of the 2021 ACM-SIAM Symposium on Discrete Algorithms (SODA), pp. 579–591. Society for Industrial and Applied Mathematics (2021). DOI 10.1137/1.9781611976465.35
4. Doerr, B.: Analyzing Randomized Search Heuristics: Tools from Probability Theory, pp. 1–20. DOI 10.1142/9789814282673_0001
5. Dyer, M., Frieze, A.: Probabilistic analysis of the multidimensional knapsack problem. Mathematics of OR **14**(1), 162–176 (1989). DOI 10.1287/moor.14.1.162
6. Dyer, M., Frieze, A.: Probabilistic analysis of the generalised assignment problem. Math. Program. **55**(1-3), 169–181 (1992). DOI 10.1007/bf01581197
7. Eisenbrand, F., Weismantel, R.: Proximity results and faster algorithms for integer programming using the steinitz lemma. ACM Trans. Algorithms **16**(1), 1–14 (2020). DOI 10.1145/3340322
8. Eldar, Y.C., Kutyniok, G. (eds.): Compressed Sensing. Cambridge University Press (2009). DOI 10.1017/cbo9780511794308
9. Feller, W.: An introduction to probability theory and its applications, vol 2. John Wiley & Sons (1991)
10. Fradelizi, M.: Hyperplane sections of convex bodies in isotropic position. Beiträge Algebra Geom **40**(1), 163–183 (1999)
11. Furst, M.L., Kannan, R.: Succinct certificates for almost all subset sum problems. SIAM Journal on Computing **18**(3), 550–558 (1989)
12. Galvin, D.: Three tutorial lectures on entropy and counting. arXiv:1406.7872 [math] (2014)
13. Goldberg, A., Marchetti-Spaccamela, A.: On finding the exact solution of a zero-one knapsack problem. In: Proceedings of the sixteenth annual ACM symposium on Theory of computing - STOC '84. ACM Press (1984). DOI 10.1145/800057.808701
14. Jansen, K., Rohwedder, L.: Integer programming (2019). DOI 10.1002/9781119454816.ch10
15. Kannan, R.: Minkowski's convex body theorem and integer programming. Mathematics of OR **12**(3), 415–440 (1987). DOI 10.1287/moor.12.3.415
16. Lenstra, H.: Integer programming with a fixed number of variables. Mathematics of OR **8**(4), 538–548 (1983). DOI 10.1287/moor.8.4.538
17. Lovász, L., Vempala, S.: The geometry of logconcave functions and sampling algorithms. Random Structures & Algorithms **30**(3), 307–358 (2007)
18. Lueker, G.S.: On the average difference between the solutions to linear and integer knapsack problems. In: Applied Probability-Computer Science: The Interface Volume 1, pp. 489–504. Birkhäuser Boston (1982). DOI 10.1007/978-1-4612-5791-2
19. Matousek, J.: Lectures on discrete geometry, vol. 212. Springer Science & Business Media (2013)
20. Papadimitriou, C.H.: On the complexity of integer programming. J. ACM **28**(4), 765–768 (1981). DOI 10.1145/322276.322287
21. Pataki, G., Tural, M., Wong, E.B.: Basis Reduction and the Complexity of Branch-and-Bound. In: Proceedings of the Twenty-First Annual ACM-SIAM Symposium on Discrete Algorithms, pp. 1254–1261. Society for Industrial and Applied Mathematics (2010). DOI 10.1137/1.9781611973075.100
22. Prékopa, A.: Logarithmic concave measures with application to stochastic programming. Acta Scientiarum Mathematicarum **32**, 301–316 (1971)
23. Röglin, H., Vöcking, B.: Smoothed analysis of integer programming. Math. Program. **110**(1), 21–56 (2007). DOI 10.1007/s10107-006-0055-7