

Integrality Gaps for Random Integer Programs via Discrepancy

Sander Borst *

Daniel Dadush *

Dan Mikulincer †

Abstract

In this work, we prove new bounds on the additive gap between the value of a random integer program $\max c^\top x, Ax \leq b, x \in \{0, 1\}^n$ with m constraints and that of its linear programming relaxation for a wide range of distributions on (A, b, c) . Our investigation is motivated by the work of Dey, Dubey, and Molinaro (SODA '21), who gave a framework for relating the size of Branch-and-Bound (B&B) trees to additive integrality gaps.

Dyer and Frieze (MOR '89) and Borst et al. (Mathematical Programming '22), respectively, showed that for certain random packing and Gaussian IPs, where the entries of A, c are independently distributed according to either the uniform distribution on $[0, 1]$ or the Gaussian distribution $\mathcal{N}(0, 1)$, the integrality gap is bounded by $O_m(\log^2 n/n)$ with probability at least $1 - 1/n - e^{-\Omega_m(1)}$. In this paper, we generalize these results to the cases where the entries of A are uniformly distributed on an integer interval (e.g., entries in $\{-1, 0, 1\}$), and where the columns of A are distributed according to an isotropic logconcave distribution. Second, we substantially improve the success probability to $1 - 1/\text{poly}(n)$, compared to constant probability in prior works (depending on m). Leveraging the connection to Branch-and-Bound, our gap results imply that for these IPs B&B trees have size $n^{\text{poly}(m)}$ with high probability (i.e., polynomial for fixed m), which significantly extends the class of IPs for which B&B is known to be polynomial.

Our main technical contribution and the key to achieving the above results is a new linear discrepancy theorem for random matrices. Our theorem gives general conditions under which a target vector is equal to or very close to a $\{0, 1\}$ combination of the columns of a random matrix A . Compared to prior results, our theorem handles a much wider range of distributions on A , both continuous and discrete, and achieves success probability exponentially close to 1, as opposed to the constant probability shown in earlier results. Our proof uses a Fourier analytic approach, building on the work of Hoberg and Rothvoss (SODA '19) and Franks and Saks (RSA '20) who studied the discrepancy of random set systems and matrices respectively.

1 Introduction

Consider an integer program (IP) in n variables and a fixed number m of constraints of the form:

$$\begin{aligned} \text{val}_{\text{IP}}(A, b, c) &:= \max_x c^\top x \\ \text{(Primal IP)} \quad &\text{s.t. } Ax \leq b, x \in \{0, 1\}^n. \end{aligned}$$

An important factor controlling our ability to solve IPs is the tightness of the linear programming (LP) relaxation. A natural way to measure tightness is the size of the gap

$$\text{IPGAP}(A, b, c) := \text{val}_{\text{LP}}(A, b, c) - \text{val}_{\text{IP}}(A, b, c),$$

where $\text{val}_{\text{LP}}(A, b, c)$ relaxes $x \in \{0, 1\}^n$ to $x \in [0, 1]^n$.

In practice, automated methods for tightening LP relaxations such as presolving and cutting planes are crucial to the performance of modern IP solvers [1, 2]. Presolving refers to simple inference rules applied to constraints in sequence, which among other things are used to find implied variable fixings, tighten variable bounds and strengthen the coefficients of inequalities [3]. Cutting planes refers to additional valid linear inequalities for the IP, which are often generated from the optimal simplex tableau [4]. The effectiveness of cutting planes is very often measured in terms of the fraction of the integrality gap they close, which helps justify the integrality gap as a key metric in practice [5, 6, 7].

*Centrum Wiskunde & Informatica (CWI), Amsterdam, The Netherlands. 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).

†Massachusetts Institute of Technology (MIT), Cambridge, USA.

1.1 Complexity of Branch-and-Bound Recently, Dey, Dubey, and Molinaro [8] provided theoretical evidence for the importance of the gap in IP solving, in the context of solving *random packing* IPs via Branch-and-Bound (B&B). We state a generalization of their result, from [9], to the case of random logconcave IPs:

THEOREM 1.1. ([8, 9]) *Let $n \geq \Omega(m)$, $b \in \mathbb{R}^m$, and $\begin{pmatrix} c \\ A \end{pmatrix} \in \mathbb{R}^{(m+1) \times n}$ be a matrix whose columns are independent logconcave random vectors with identity covariance. Then, for $G \geq 0$, with probability at least $1 - \Pr_{A,c}[\text{IPGAP}(A, b, c) \geq G] - 1/\text{poly}(n)$, the best bound first Branch-and-Bound algorithm applied to (Primal IP) produces a tree of size at most $n^{O(m)} e^{2\sqrt{2nG}}$.*

We recall that B&B proceeds by building a binary search tree whose leaves represent a partition of the solution space. Each node of the tree is associated with a subinstance, corresponding to a partial fixing of the variables to 0 or 1, and is labelled with the optimal solution to the LP relaxation consistent with this fixing. The tree is initialized with a single node corresponding to the empty fixing which is labelled with the optimal solution to $\text{val}_{\text{LP}}(A, b, c)$. B&B keeps track of the best integer solution found in the tree, picked from the nodes whose LP solutions are integral (or possibly found via heuristics), and removes from consideration all leaf nodes whose LP values are at most the value of this best found solution (a procedure known as node pruning). At every iteration, B&B chooses an unpruned leaf node in the current tree to process using a *node selection rule*. To process the node, B&B examines the optimal LP solution x at the node and chooses a fractional coordinate $x_i \in (0, 1)$ using a *branching rule*. B&B then creates two children, each one corresponding to fixing x_i to 0 or 1, computes the corresponding LP solutions and checks whether these nodes can be pruned. Once there are no more nodes left to process, the best found integer solution is the optimal solution to the IP. For more background on B&B, the interested reader may consult the following standard textbooks [10, 11].

As B&B is still the backbone of all modern IP solvers such as CPLEX, Gurobi, SCIP and Mosek, it is natural to try and understand under which circumstances the complexity of B&B can be meaningfully upper bounded. Theorem 1.1 above gives probabilistic upper bounds on the size of B&B trees for random logconcave IPs in terms of the integrality gap, where the branching rule can be arbitrary as long as 'best bound first' node selection rule is used. Best bound first corresponds to always processing leaf nodes with the largest LP value first. This rule ensures that we never process a node whose value is worse than that of the optimal IP solution, which is crucial for relating the size of the tree to the integrality gap.

While Theorem 1.1 is relatively general, it is of course only meaningful for distributions over which strong upper bounds on the integrality gap are known. Note that the class of logconcave distributions is very rich and includes, among others, the uniform distribution on any convex body, and unbounded distributions such the Gaussian and Laplace distributions. In terms of applications, due to the $n^{O(m)}$ term in the complexity estimate, they are mainly restricted to the case where the number of constraints m is constant, which includes fundamental problems such as knapsack. We discuss these applications together with our contributions in the next section.

1.2 Integrality Gap Bounds and Their Applications For models directly captured by Theorem 1.1, strong integrality have been proven for random packing [12, 13, 14] and Gaussian IPs [9], where the entries of (A, c) are either independent uniform $[0, 1]$ or $\mathcal{N}(0, 1)$. For random packing IPs, when $b \in (n/4, n/2)^m$, Dyer and Frieze [12] proved that $\text{IPGAP}(A, b, c) \leq 2^{O(m)} \log^2(n)/n$ with probability at least $1 - 1/\text{poly}(n) - 2^{-\text{poly}(m)}$. For Gaussian IPs, when $\|b^-\|_2 \leq n/10$ (b^- is the negative part of b), Borst et al [9] proved that $\text{IPGAP}(A, b, c) \leq \text{poly}(m) \log^2(n)/n$ with probability at least $1 - 1/\text{poly}(n) - 2^{-\text{poly}(m)}$. For these models, Theorem 1.1 implies that B&B is polynomial for fixed m with good probability. More precisely, letting $G = O_m(\log^2 n/n)$ be the gap for these models, the complexity of B&B grows as $e^{\sqrt{nmG}} = e^{O_m(\log n)} = n^{O_m(1)}$.

We remark that the works [13, 12] also introduced a specialized algorithm for solving random packing IPs, with the same complexity¹. At a high level, this algorithm simply examines all possible ways to round the optimal LP solution without violating the target gap bound. It is worthwhile to note that [8] (implicitly) bounds the complexity of B&B by showing that its work never exceed the work of the algorithm in [13, 12].

The above results yield some of the very few cases where meaningful estimates are known on the size of B&B trees (additional examples will be discussed in the related works section). Motivated by the scarcity of results on B&B, the main focus of this work is to extend the range of distributions over (A, b, c) for which strong gap upper bounds hold.

Our first result is a substantial generalization of the Gaussian gap bound of Borst et al [9]. The result applies to a wide range of constraint matrices whose columns are independent mean-zero random vectors, which we refer to as *Centered IPs*.

¹[13] examined the case $m = 1$ while [12] handled general m .

THEOREM 1.2. (GAP BOUND FOR CENTERED IPS) For $m \geq 1$, $n \geq \text{poly}(m)$, $b \in \mathbb{R}^m$ with $\|b^-\|_2 \leq O(n)$, if c has i.i.d. $\mathcal{N}(0,1)$ entries and the columns of A are independent isotropic, logconcave random vectors whose support is contained in a ball of radius $O(\sqrt{\log n} + \sqrt{m})$, then

$$\Pr\left(\text{IPGAP}(A, b, c) \geq \frac{\text{poly}(m)(\log n)^2}{n}\right) \leq n^{-\text{poly}(m)}.$$

Furthermore, the same result holds if the entries of A are distributed independently and uniformly in $\{0, \pm 1, \dots, \pm k\}$ and $b \in \mathbb{Z}^m$ with $\|b^-\|_2 \leq O(kn)$, for any $1 \leq k \leq n$.

We recall that a random vector $X \in \mathbb{R}^m$ is isotropic if $\mathbb{E}[X] = 0$ (mean zero) and $\mathbb{E}[(X - \mathbb{E}[X])(X - \mathbb{E}[X])^\top] = I_m$ (identity covariance). Since one can apply an affine transformation to any full dimensional random variable to make it isotropic, this condition may be regarded as a useful normalization. When A has iid $\mathcal{N}(0,1)$ entries, we remark that the columns of A have norm bounded by $O(\sqrt{m} + \sqrt{\log n})$ with probability $1 - 1/\text{poly}(n)$ by standard tail bounds. Thus, the above theorem yields a strict generalization of [9].

Our second result is a discrete variant of the random packing IP gap bound of [12].

THEOREM 1.3. (DISCRETE PACKING IPS) For $m \geq 1$, $k \geq 3$, $\beta \in (0, 1/4)$, $n \geq \text{poly}(m, k) \exp(\Omega(1/\beta))$, $b \in ((kn\beta, kn(1/2 - \beta)) \cap \mathbb{Z})^m$, if c has i.i.d. exponential entries and the entries of A are independent and uniform in $\{1, \dots, k\}$, then

$$\Pr\left(\text{IPGAP}(A, b, c) \geq \frac{\exp(O(1/\beta)) \text{poly}(m)(\log n)^2}{n}\right) \leq n^{-\text{poly}(m)}.$$

Compared to [12] we require exponentially distributed objective coefficients, i.e., with density e^{-x} , $x \geq 0$, instead of uniform $[0, 1]$. The additional smoothness of the distribution makes the arguments much cleaner while preserving the qualitative nature of the results. We remark that extending the above gap bound to the setting of non-negative logconcave random vectors is also possible, though we do not pursue this here.

As a consequence of these gap bounds, we derive the following generalized complexity bounds for B&B.

COROLLARY 1.1. With probability $1 - 1/\text{poly}(n)$, the best-bound first Branch-and-Bound algorithm applied to (Primal IP) produces a tree of size at most $n^{\text{poly}(m)}$ in the Centered IP model of Theorem 1.2 and of size $n^{\exp(O(1/\beta)) \text{poly}(m)}$ in the Discrete Packing IP model from Theorem 1.3.

In the discrete case, we note that the above result does not follow directly from Theorem 1.1, since it requires the constraint matrix to have a logconcave distribution (a suitable adaptation of the proof still applies however, see Section 4).

The above results extend our understanding of gap bounds in two significant ways. Firstly, they show that strong gap bounds are obtainable when the entries of the constraint matrix take only small integer values. This is well motivated by the fact that the constraints for practical IPs are often combinatorial and thus are expressed using small integer coefficients. Secondly, in the case of centered IPs, we show that non-trivial correlations between the variables in a column of the constraint (induced by logconcavity) can be handled. Thus, we establish a limited form of universality for these gap bounds.

We would like to stress that neither of these extensions is a priori obvious. To highlight the subtleties in the discrete setting, we note that even for $m = 1$, the gap bound does not hold when the right hand side b is non-integral. As a simple example, it is easy to see that for n odd, the integrality gap of the knapsack program $\max \sum_{i=1}^n x_i$ subject to $\sum_{i=1}^n x_i \leq n/2$, $x \in \{0, 1\}^n$ is $1/2$. It is not hard to check that a constant gap is preserved with overwhelming probability even if the profit vector c is exponentially distributed and weights a_1, \dots, a_n are drawn uniformly from $\{1, \dots, k\}$, for k fixed.

To circumvent the above issue, a crucial step in our proof in the discrete setting is to show that we can round the optimal LP solution x^* to an integer solution x' that is tight on the same set of constraints. That is, $(Ax^*)_i = b_i \Rightarrow (Ax')_i = b_i, \forall i \in [m]$ (which is only possible if b is integral). Indeed, one of our main technical contributions, described in the next section, is to give general conditions under which such exact roundings are possible, using Fourier analytic techniques from discrepancy theory. We remark that Dyer and Frieze [15] believed that a discrete extension of their gap bounds should be possible provided the range of the discrete distribution (i.e., k above) grows with n , which we confirm here without this assumption².

²We note that our gap bounds seemingly require that n be somewhat larger than the range k . This is an artificial restriction. For (very) large k , one must treat the discrete distribution as if it were continuous, which requires a slight adaptation of the proofs.

Beyond generalizing previous work, our results also improve the probabilistic guarantees. Compared to prior works, our gap bounds hold with high probability $1 - 1/\text{poly}(n)$ instead of $1 - 1/\text{poly}(n) - 2^{-\text{poly}(m)}$ (which is constant for fixed m). Achieving a high probability bound via prior estimates increases the gap by a $O(\log n)$ factor, which makes the corresponding B&B complexity quasi-polynomial instead of polynomial.

Is it natural to ask whether there are natural limitations to extending these gap bounds to much wider classes of IPs. For logconcave random IPs, one may observe that any worst-case instance can be encoded as the mean $\mathbb{E}[(A, b, c)]$ of the random instance. In this way, one can view the distribution (A, b, c) as a smoothed version of worst-case IP instance $\mathbb{E}[(A, b, c)]$. By appropriately scaling up the means, or equivalently scaling down the variance of the entries of (A, b, c) , the instance (A, b, c) converges to the worst-case instance $\mathbb{E}[(A, b, c)]$. It is not hard to check that for many hard combinatorial problems such as SET COVER or MAX 3-SAT, adding small random perturbations (of appropriate sign) to the instance data does not change the optimal solution, and thus we cannot expect strong integrality gap results in these settings.

We note that smoothed analysis of integer programming has indeed been studied by various works [16, 17]. In particular, Röglin and Vöcking [17] showed that under mild conditions, the class of IPs that are easy to solve on average over suitable random perturbations of the instance data (i.e., that can be solved in polynomial time with high probability over the perturbations) correspond exactly to the class of IPs solvable in pseudopolynomial time.

Gap Bounds in Related Settings. Strong gap bounds have also been proven for random instances of combinatorial optimization problems, though this has not translated into good upper bounds on the size of B&B trees. Dyer and Frieze [15], proved an $O_m(\log^2 n/n)$ gap bound for random instances of a generalized assignment problem on a bipartite graph with m vertices on the small side and n vertices on the big side. They further gave an $n^{O_m(1)}$ algorithm to solve this problem with high probability. Frieze and Sorkin [18] showed that the cycle cover relaxation for the asymmetric TSP has an expected $O(\log^2 n/n)$ additive integrality gap, where the edge weights are chosen uniformly from $[0, 1]$ for a complete digraph on n vertices and gave an $2^{\tilde{O}(\sqrt{n})}$ algorithm which solves these instances with high probability. Recently, Frieze [19] showed that any B&B tree using the cycle cover relaxation to prune nodes³ will have expected size $\Omega(2^{n^\alpha})$, for some $0 < \alpha < 1/20$. It is an interesting open question to understand if B&B can recover the same running times as the specialized algorithms above for these problems.

1.3 Techniques We now explain the high level strategy behind the obtaining integrality gap bounds of [12, 9]. We will then extract a natural discrepancy theoretic problem which will be the key to our generalizations.

Proving Gap Bounds. Given integer program indexed by (A, b, c) , we recall that the goal is to bound the gap between the value of the integer program and its linear programming relaxation. The strategy to bound the integrality gap is to suitably round an optimal basic primal solution $x^* \in [0, 1]^n$ (for possibly a slightly perturbed right-hand side b) to a nearly optimal solution $x'' \in \{0, 1\}^n$ to the IP. The rounding procedure will make use of the dual optimal solution $u^* \in \mathbb{R}_+^m$ (see Section 3.1 for the dual program).

To round x^* , we first round down its fractional components to 0 to get x' (there are at most m such components). From here, one selects a small subset of variables $T \subseteq \{i : x_i^* = 0\}$, $|T| = O_m(\log n)$, with tiny reduced costs (which are “cheap” to flip to 1), namely $|c_i - u^{*\top} a_i| = O_m(\log n/n)$, for $i \in T$. The subset T is carefully chosen such that $x'' := x' + 1_T \in \{0, 1\}^n$, obtained by flipping the variables in T from 0 to 1, is the desired feasible and nearly optimal IP solution. For feasibility, we need that $A(x' + 1_T) \leq b$. For near-optimality, we additionally need $x' + 1_T$ to be nearly (or even exactly) tight on the constraints that x^* is tight on. That is, we need that $(Ax^*)_i = b_i \Rightarrow (A(x' + 1_T))_i \approx b_i$, $\forall i \in [m]$. Simplifying slightly, a sufficient condition is $A1_T \approx A(x^* - x')$, thus we may think of T as a “slack repairing” set.

A crucial property, in both the packing and Gaussian setting, is that x^* has at least $\Omega(n)$ zero entries and that the corresponding columns of A are independent subject to having negative reduced cost (see Lemma 3.2 and Lemma 3.3). Furthermore, there are at least $\Omega_m(\log n)$ columns with suitably small reduced costs, and so there is large set of columns from which to select T .

To generalize the above strategy to other distributions, the crucial difficulty is understanding under what conditions the “slack repairing” set T above exists.

The Discrepancy Problem. Stated abstractly, the existence of T can be phrased as a natural discrepancy theoretic problem. Let $t \in \mathbb{R}^m$ be a target vector (e.g., the difference of slack vectors) and let $A \in \mathbb{R}^{m \times n}$ be a “nice” random matrix with independent columns. When can we ensure with high probability that t is equal to or very close to a $\{0, 1\}$ combination of the columns of A ?

³The cycle cover relaxation is, in fact, integral, but its solutions are not feasible (i.e., they may be a union of many disjoint directed cycles). B&B must therefore branch on variables that are integral in the current relaxation.

A general answer to this question was given in [9, Lemma 1], improving upon [12, Lemma 3.4]. However, it enforced very strict conditions on the entries of A . Specifically, the entries of A needed to be independent, mean zero with unit variance, absolutely continuous random variables of bounded density which “converge quickly enough” to a Gaussian when averaged. Furthermore, for the targets t in the “range” of A , the probability of successfully hitting t was only $\Theta(1)$.

As our main technical contribution, which yields the key ingredient for extending the IP gap bounds, we give a much more general and powerful discrepancy theorem. We state it below, restricted to special cases relevant for our applications (see Theorem 2.1 for the general result).

THEOREM 1.4. (LINEAR DISCREPANCY THEOREM FOR RANDOM MATRICES) *Let $A = (a_1, \dots, a_{\bar{n}}) \in \mathbb{R}^{m \times \bar{n}}$, $\bar{n} \geq \text{poly}(m)$, be a random matrix with independent columns with same mean $\mu \in \mathbb{R}^m$. Let $p \in (0, 1)$ satisfy $\frac{\text{poly}(m) \log \bar{n}}{\bar{n}} \leq p \leq \frac{1}{\text{poly}(m)}$. Then, the following holds:*

- *Discrete case: suppose that the entries A_{ij} are i.i.d. uniform on an integer interval $\{l, l+1, \dots, l+k\}$ for $l, k \in \mathbb{Z}$, $\bar{n} \geq k \geq \max\{2, |l|\}$. Then, with probability at least $1 - e^{-\Omega(p\bar{n})}$, for every $t \in \mathbb{Z}^m$ satisfying*

$$\|t - p\bar{n}\mu\|_2 \leq O\left(\frac{k\sqrt{p\bar{n}}}{m \log m}\right)$$

there exists $x \in \{0, 1\}^{\bar{n}}$ with $\|x\|_1 = \Theta(p\bar{n})$ such that $Ax = t$.

- *Continuous case: suppose that $a_1, \dots, a_{\bar{n}}$ are logconcave with identity covariance and that $\|\mu\| \leq \text{poly}(m)$. Then, with probability at least $1 - e^{-\Omega(p\bar{n})}$, for every $t \in \mathbb{R}^m$ satisfying,*

$$\|t - p\bar{n}\mu\|_2 \leq O\left(\frac{\sqrt{p\bar{n}}}{m \log m}\right),$$

there exists $x \in \{0, 1\}^{\bar{n}}$, with $\|x\|_1 = \Theta(p\bar{n})$ such that $\|t - Ax\|_2 \leq e^{-\Omega(p\bar{n}/m)}$.

When applying the above to find the slack repairing set T , \bar{n} will roughly be $O(\text{poly}(m) \log n)$ and $p = 1/\text{poly}(m)$. The corresponding success probability will now be $1 - n^{-\text{poly}(m)}$ as opposed to $\Theta(1)$, which is what allows us to much better tail bounds for the integrality gap.

Gap bounds via Discrepancy. Given the discrepancy theorem, the proof of the gap bounds mirrors the proofs in [12, 9] as described in the last section, though with non-trivial technical adaptations as well as some simplifications.

As in prior work, the relevant properties of the optimal primal x^* and dual solution u^* must be established in these generalized settings (see Lemma 3.3 and Lemma 3.10). In particular, one must show that the norm of u^* is suitably bounded, and that there are sufficiently many columns of A with small reduced costs indexed by the zeros of x^* . These properties can be derived using standard tools for concentration and anti-concentration of logconcave and uniform discrete random variables.

Providing “nice enough” columns for our discrepancy theorem can however be challenging. For centered IP where A has independent logconcave columns, conditioning on the size of the reduced costs can significantly perturb the mean of each column (possibly in different directions when the columns are not i.i.d.). We overcome this problem using sophisticated forms of rejection sampling, which allows to “virtually recenter” each logconcave column (see Lemma 3.8). Interestingly, the rejection sampling procedure does not even preserve logconcavity, however the properties required for our more general discrepancy theorem persist (see Theorem 2.1).

In terms of simplifications, compared to earlier proofs we no longer require repeated trials on disjoint subsets of columns of A to find a suitable slack repairing set T . In particular, due to our new discrepancy theorem, using all the small reduced cost columns together both exponentially decreases the probability of failure and increases the size of the targets one can hit. Furthermore, since the discrepancy theorem directly applies to columns with non-zero means, one can avoid ad-hoc reductions to the mean-zero case as in the proof of [12] for the packing case.

Relations to Discrepancy Theory. We first explain the connection to linear discrepancy. As defined by Lovász, Spencer, Vesztegombi [20], the linear discrepancy of a matrix $A \in \mathbb{R}^{m \times n}$ is $\text{lindisc}(A) := \max_{\lambda \in [0, 1]^n} \min_{x \in \{0, 1\}^m} \|A(x - \lambda)\|_\infty$. That is, it is the maximum “rounding error” one must incur to round a $[0, 1]$ combination of the columns to a $\{0, 1\}$ combination (the specific choice of ℓ_∞ vs ℓ_2 norm is not important in our context). The discrepancy $\text{disc}(A)$ of A is linear discrepancy restricted to $\lambda = \mathbf{1}_n/2$ (the all $1/2$ vector). It is more common to expressed discrepancy by

$\min_{x \in \{-1, 1\}^n} \|Ax\|_\infty = 2\text{disc}(A)$, in which case $x \in \{-1, 1\}^n$ is interpreted as a 2-coloring of the columns of A . While linear discrepancy is always larger than discrepancy, [20] showed the maximum discrepancy of any subset of the columns of A , known as hereditary discrepancy, upper bounds $\text{lindisc}(A)$ up to a factor of 2.

Bounds on the discrepancy of various matrix classes, often induced by the incidence matrix of set system, have found many applications in computational geometry and complexity (see [21, 22]). Over the last decade or so, efficient algorithms for producing low-discrepancy colorings have been developed [23, 24, 25, 26, 27] and have found many applications in the context of approximation algorithms [28, 29, 30, 31].

With the above perspective, Theorem 1.4 can be interpreted as bounding the linear discrepancy of the random matrix A for combinations $\lambda \in [0, 1]$ which are very close to $p\mathbf{1}_n$, where $p \in (0, 1)$ is as above. As the columns of A are independent with mean μ , one can expect that $A(p\mathbf{1}_n) \approx pn\mu$. Perhaps slightly less clear is that every $t \in \mathbb{R}^m$, which is close to $pn\mu$, will in fact be exactly expressible as $t = A\lambda$, where $\lambda \approx p\mathbf{1}_n$ with high probability (this requires an analysis of the singular values of A). Theorem 1.4 now shows that λ can be replaced by $x \in \{0, 1\}^n$ with $\|x\|_1 = \Theta(pn)$ incurring either no error in the discrete case (assuming $t \in \mathbb{Z}^n$) or exponentially small error in the continuous case, thereby bounding the linear discrepancy of the combination λ . At least in the continuous case, we note that one can in fact adapt the proof of Theorem 1.4 to directly bound the linear discrepancy of any “reasonable” combination λ (i.e., without the detour through the mean μ). This would be somewhat less useful for the integrality gap application we consider here, since it would give significantly less precise guarantees on the targets we can expect to hit. Interestingly, while the application pursued here is algorithmic, i.e. bounding the complexity of B&B, we are not aware of any efficient algorithm to compute the rounding x guaranteed by Theorem 1.4.

Crucial to the linear discrepancy bounds we achieve in Theorem 1.4, i.e., either exponential small or zero, is that the matrix A has many more columns than the dimension m . In particular, reducing to bounds on hereditary discrepancy become useless in this setting. The study of discrepancy in the “over complete” setting has been become very active somewhat more recently, with works focusing on the discrepancy of large random matrices and set systems [32, 33, 34, 35, 36, 37, 38]. Many of these works were motivated by the Beck-Fiala conjecture, which posits that discrepancy of any $\{0, 1\}$ matrix A in which every column has at most t ones is bounded by $O(\sqrt{t})$ (here A can be interpreted as the incidence matrix of a set system). Variants of this conjecture for random set systems and matrices were established in [32, 34, 35, 36, 37, 39, 38], where in the case $n \gg m$, it was shown that discrepancy quickly drops to 1 [35, 36, 37, 39] or even exponentially close to zero [32, 36] depending on whether the columns of the matrix A are discretely or continuously distributed.

Discrepancy via Fourier Analysis. To prove our linear discrepancy theorem, we rely on a Fourier analytic approach, which is completely different from the second-moment counting based proofs in [12, 9]. The high level approach was pioneered by Kuperberg, Peled and Lovett [40], who applied Fourier analytic techniques to show the existence of rigid combinatorial objects, such as orthogonal arrays, Steiner systems and regular hypergraphs. At a technical level, for a matrix $A \in \{0, 1\}^{m \times n}$, they were interested in understanding the minimum $p \in (0, 1]$ such that $pA\mathbf{1}_n = Ax$, where $x \in \{0, 1\}^n$ $\|x\|_1 = pn$. They examined this question for highly symmetric deterministic matrices coming from the above applications. The Fourier analytic approach was later applied by Hoberg and Rothvoss [35] and independently by Frank and Saks [36] to show very strong upper bounds on the discrepancy $\min_{x \in \{-1, 1\}^n} \|Ax\|_\infty$ of a random matrix A when $n \gg m$, where the columns of A were drawn iid from various distributions.

From a comparative perspective, our Theorem 1.4 sits in between the work of [40] and [35, 36]. We work with random matrices A as in [35, 36], though the question we attack is more in the spirit of [40]. We note that unlike [40], it is not sufficient for us to show that the existence a rounding $x \in \{0, 1\}^n$ such that $A(x - p\mathbf{1}_n)$ is small or zero. For our applications, we require this to be true with $Ap\mathbf{1}_n$ replaced by $pn\mu = \mathbb{E}[Ap\mathbf{1}_n]$. We further require the existence of x to hold for any target t close enough to $pn\mu$. This last requirement however generally comes for “free” with Fourier analytic techniques (the moment you can hit $pn\mu$ you can hit everything close to it as well). A more significant difficulty is that the concentration of $Ap\mathbf{1}_n$ around its mean $pn\mu$ is rather weak. That is, the probability that $Ap\mathbf{1}_n$ is close to its mean scales relative to the ambient dimension m (which is constant) instead of n . This makes achieving the $1 - e^{-\Omega(pn)}$ success probability more challenging.

We now explain the high-level approach. Let $A \in \mathbb{R}^{m \times n}$ be our random matrix with columns having mean $\mu \in \mathbb{R}^m$ and covariance matrix $\sigma^2 I_m$, and let $p \in (0, 1)$ be our parameter. The strategy is to directly analyze the probability mass function of the random variable $Y = AX$, where X_1, \dots, X_n are i.i.d. Bernoulli’s with probability p . [40] also use this distribution, whereas [35, 36] choose X_1, \dots, X_n to be uniform $\{-1, 1\}$ random variables. Restricting attention to the discrete case, where $Y \in \mathbb{Z}^n$, we show that $\Pr[Y = t] \gg 0$ for $t \in \mathbb{Z}^n$, when $\|np\mu - t\| = O_m(\sigma\sqrt{pn})$, where $O_m(\sigma^2 pn)$ roughly measures the “available variance” of AX in all directions. Obtaining a bound for $\Pr[Y = t]$ is done by applying the Fourier inversion formula, showing that the Fourier coefficients are close to those of a Gaussian and integrating (see Section 2 for an overview).

For $\theta \in [-1/2, 1/2]^m$, the corresponding of Fourier coefficient AX is expressed by $\mathbb{E}_X[e^{2\pi i \langle \theta, AX \rangle}]$. Similar to [35, 36], we control the magnitude of these coefficients using the anti-concentration properties of the columns of A . Specifically, for

our choice of column distributions, we need to show the the probability that $\langle \theta, A_i \rangle$, $i \in [n]$, is “close” to an integer decays predictably as a function of $\sigma \|\theta\|$. We note that the exact expression for the Fourier coefficients indeed differs depending on whether the entries of X are Rademacher or Bernoulli distributed, where the former is more prone to parity issues (e.g., if $A \in \{0, 1\}^{m \times n}$ and $x \in \{-1, 1\}^n$, the parity of Ax is fixed). Such parity issues do not arise in our setting. As mentioned above, we face a different difficulty, which is being able to pinpoint the exact targets whose probabilities we can accurately estimate due to the poor concentration of $\mathbb{E}_X[AX] = A(p\mathbf{1}_n)$ around $\mathbb{E}_{X,A}[AX] = np\mu$.

To deal with this issue, we first carefully subsample a set $S \subseteq [n]$ of columns from A , whose sum is close to the mean, and then generate Y from these subsampled columns. To construct S , we iterate through the columns one by one, adding A_i to S if $\langle \sum_{j \in S} (A_j - \mu), A_i - \mu \rangle \leq 0$ and $\|A_i - \mu\| \leq 2\mathbb{E}[\|A_i - \mu\|_2^2]^{1/2}$. This subsampling deterministically ensures that $\|\sum_{i \in S} (A_i - \mu)\|_2 \leq 2\sqrt{\sum_{i \in S} \mathbb{E}[\|A_i - \mu\|_2^2]}$, suitably biasing the sum towards the mean. For the distributions we work with, it is easy to show that $|S| = \Omega(n)$ with probability $1 - e^{-\Omega(n)}$, so we always have a constant fraction of the columns left over. Note that this subsampling crucially uses our flexibility to drop columns of A , a distinguishing feature of $\{0, 1\}$ combinations versus $\{-1, 1\}$ combinations. This subsampling process, however, causes non-trivial dependencies among the columns of A . That is, the submatrix we use to generate Y no longer has independent columns. Fortunately, we show that even with the conditioning induced by subsampling, the columns still retain enough of their anti-concentration properties to allow the Fourier analytic estimates to go through.

Related Work. Another success story in analysis of LP based B&B comes from fixed parameter tractability (FPT). In this context, it was shown that important combinatorial problems such as vertex cover [41] and multiway cut [41] are FPT when parameterized by the integrality gap of the natural LP relaxation. Since then, multiple general frameworks for capturing such problems have been developed [42, 43]. These results crucially rely on the use of half-integral and persistent LP relaxations, which are known only for very structured problems. The results presented here thus mainly provide complementary evidence for the effectiveness of B&B in the unstructured setting.

While variable based B&B is most commonly used in practice, there is a theoretical line of work examining the effectiveness of branching schemes for solving random integer programs based on general integer disjunctions. These schemes rely on sophisticated lattice basis reduction methods [44] and follow the template of Lenstra’s celebrated polynomial time algorithm for integer programming in fixed dimension [45]. In particular, Furst and Kannan [46] showed that certain random subset-sum instances can be solved in polynomial time via basis reduction, and Pataki, Tural and Wong [47] extended these results to certain IPs with multiple constraints. Apart from the different type of branching, compared to the present work, the IPs analyzed in these models are either infeasible or have a unique feasible solution with high probability.

Organization The preliminaries and necessary background to our results can be found in appendix A. Section 2 contains the proof of the discrepancy result, Theorem 1.4, and the bounds on the integrality gap, given in Theorems 1.2 and 1.3, are proven in Section 3. Section 4 is devoted to the proof of Corollary 1.1. Finally, in Section 5 we prove several anti-concentration results required by our method.

2 Discrepancy

In this section we describe our main discrepancy results. Let $A \in \mathbb{R}^{n \times m}$ be a random matrix with independent columns. Throughout the section we assume that the columns of A have the same mean, denoted as μ . We also make the assumption that all columns of A have the same period, as defined in (A.1), and denote it by $\text{period}(A)$. We will deal with two different cases: either $\text{period}(A) = \{0\}$, which means that the columns of A are absolutely continuous, or $\text{period}(A) = \mathbb{Z}^m$, in which case the columns of A are supported on the lattice \mathbb{Z}^m .

Notation and assumptions: Before stating the result, let us introduce some definitions. We first describe the distributions captured by our result; distributions with a non-negligible mass on every half-space passing through their mean.

DEFINITION 1. (APPROXIMATELY SYMMETRIC DISTRIBUTIONS) A probability distribution \mathcal{D} on \mathbb{R}^m , with mean μ , is called approximately symmetric if, for any $v \in \mathbb{R}^m$,

$$\Pr_{X \sim \mathcal{D}} (\langle X, v \rangle \geq \langle \mu, v \rangle) \geq \frac{1}{4e^2}.$$

REMARK 1. The constant $\frac{1}{4e^2}$ in the definition is somewhat arbitrary and could be relaxed to any smaller constant. It is immediately clear that any distribution which is symmetric around its mean is also approximately symmetric. Moreover, Grünbaum’s inequality in Lemma A.6 shows that logconcave measures are approximately symmetric.

Suppose that the columns of A are approximately symmetric. The main idea will be to choose $S \subseteq [n]$ randomly with $\Pr[i \in S] = p$ for $p \leq \frac{1}{\text{poly}(m)}$, independently for all $i \in [n]$. We then show that, with high probability over A and positive probability over S , $A\mathbf{1}_S$ is close to a target vector t . Thus, let us define the random vector $D := A\mathbf{1}_S$.

To understand the distribution of D we will consider its Fourier transform, denoted $\hat{D}(\theta) := \mathbb{E}[\exp(2\pi i \langle D, \theta \rangle)]$. Note, that under the assumptions above, we have $\text{period}(A) = \text{period}(D)$, and so we shall use V to denote the fundamental domain of D , as in (A.2).

The next definition quantifies an appropriate notion of anti-concentration properties for the columns of A .

DEFINITION 2. (ANTI-CONCENTRATION) *Let $\sigma \geq 0$ and $\kappa \in (0, 1)$. We say the measure \mathcal{D} is (σ, κ) -anti-concentrated if $\text{Cov}(\mathcal{D}) \preceq \sigma^2 \mathbf{I}_m$, and for any $\mathbf{v} \in \mathbb{R}^m$ and any $\theta \in V$,*

$$\text{(anti-concentration)} \quad \Pr_{X \sim \mathcal{D}} [d(\theta^\top X, \mathbb{Z}) \geq \kappa \min(1, \|\theta\|_\infty \sigma) \mid \langle \mathbf{v}, X \rangle \leq \langle \mathbf{v}, \mu \rangle] \geq \kappa,$$

where $d(\theta^\top X, \mathbb{Z}) := \inf_{z \in \mathbb{Z}} |\theta^\top X - z|$. When σ is clear from the context, we will sometimes omit the dependence on σ from the definition.

Without further details, the definition might seem opaque. Below we explain the rationale for considering this notion of anti-concentration and demonstrate some examples of distributions that satisfy Definition 2. For now, it shall suffice to say that the (anti-concentration) property appears naturally when trying to establish bounds on Fourier transforms.

In the following assumption, which we shall enforce throughout the section, we detail the possible dependencies between the different parameters. These dependencies, in turn, dictate the possible regimes in which our results hold.

ASSUMPTION 1. *We say that the random matrix $A \in \mathbb{R}^{n \times m}$ satisfies our assumptions if its columns are independent with a common period and a common mean $\mu \in \mathbb{R}^m$, are approximately symmetric, and (σ, κ) -anti-concentrated, with constants $\sigma, \kappa > 0$. Further, let $p \in [0, 1]$ and assume that the following technical condition is met:*

$$(2.1) \quad \frac{\left(10m \left(1 + \frac{\|\mu\|}{\sigma}\right)\right)^{20}}{\kappa^{16} n} \leq p \leq \frac{\kappa^8}{10^{15} m^6}.$$

Main discrepancy results: With the above notation, the main result of this section is:

THEOREM 2.1. *Let $p \in [0, 1]$ and suppose that A satisfies our assumptions, as in Assumption 1. Then, when $\text{period}(A) = \mathbb{R}^m$, with probability $1 - e^{-\Omega(\kappa pn)}$ we have:*

For all $t \in \mathbb{R}^m$ such that

$$\|t - pn\mu\| \leq \frac{\sigma \kappa^{\frac{3}{2}}}{10^4 \ln \left(\frac{10^3 m}{\kappa^3} \left(1 + \frac{\|\mu\|^2}{\sigma^2}\right)^{\frac{1}{m}} \right)} \frac{\sqrt{pn}}{m},$$

there is a set S of size $|S| \in [\frac{1}{16}pn, \frac{3}{16}pn]$ such that

$$\|A\mathbf{1}_S - t\| \leq \exp\left(-\frac{\kappa^3 pn}{80m}\right) \sigma n^3.$$

Theorem 2.1, as stated, only deals with distributions which are absolutely continuous with respect to the Lebesgue measure. However, the argument also applies to measures with singularities. In particular, the result also holds for distributions supported on the lattice \mathbb{Z}^m , the case which is most relevant to our work. Moreover, in the lattice case, if one finds a subset $S \subset [n]$, such that $A\mathbf{1}_S$ is very close to some target vector $t \in \mathbb{R}^m$, then, since $A\mathbf{1}_S$ belongs to the lattice as well, one can actually deduce $A\mathbf{1}_S = t$. We prove this statement as a part of the proof of Theorem 2.1.

THEOREM 2.2. *Suppose that $\text{period}(A) = \mathbb{Z}^m$. Then, under the same conditions of Theorem 2.1 together with $\exp\left(\frac{\kappa^3 pn}{80m}\right) \geq 2\sigma^2 n^2$, with probability $1 - e^{-\Omega(\kappa pn)}$, there is a set S of size $|S| \in [\frac{1}{16}pn, \frac{3}{16}pn]$ such that $A\mathbf{1}_S = t$, provided that $t \in \mathbb{Z}^m$.*

Remark that since, by (2.1), $\kappa^3 pn \geq \text{poly}(m)$, the condition $\exp\left(\frac{\kappa^3 pn}{80m}\right) \geq 2\sigma^2 n^2$ almost does not restrict generality.

Road map for the proof: Recall Fourier's inversion formula, Theorem A.1. In light of the formula, it will be enough to show that the integral in Theorem A.1 is positive for appropriate t , for most choices of A . In this case we will get that $\Pr[D = t] > 0$, which implies the existence of an appropriate subset of columns. At a very high level, we will show that most of the mass lies next to the origin and that the integrand has an exponential decay far from the origin. The following sequence of steps will achieve this:

1. Our first step will be to subsample the columns of A . This will result in appropriate concentration bounds that improve as n increases, irregardless of the value of m (Lemma 2.2).
2. We then show that with high probability $\arg(\hat{D}(\theta)) \in [-\frac{1}{8}\pi, \frac{1}{8}\pi]$ for all θ with $\|\theta\| = O(\frac{\text{poly}(m)}{\sigma\sqrt{n}})$. This will allow to establish that the real part of $\hat{D}(\theta) \exp(-2\pi i \langle \theta, t \rangle)$ is large when $\|t - pn\mu\| \|\theta\|$ is small (Lemma 2.4).
3. Next, we show that $\int_{\|\theta\| \leq r} |\hat{D}(\theta)| d\theta$ is large for some $r = O(\frac{\text{poly}(m)}{\sigma\sqrt{n}})$. In combination with Lemma 2.4, we shall conclude that $\int_{\|\theta\| \leq r} \hat{D}(\theta) \exp(-2\pi i \langle \theta, t \rangle) d\theta$ is large (Lemma 2.6).
4. Finally, we will show that for $\|\theta\| \geq r$, $|\hat{D}(\theta)|$ is rapidly decreasing. So the integral over these θ can only have a small negative contribution (Lemma 2.7).

The first bottleneck is Step 1, which will change the distribution and effective size of the random matrix A . Below we show that approximately symmetric distributions maintain many desirable properties after our subsampling procedure. The second bottleneck is Step 4. To establish a rapid enough decay of the Fourier spectrum, we require that the columns of A satisfy the (anti-concentration) property from Definition 2.

To gain a bit of intuition about Definition 2, recall that we are working in the Fourier domain. If X is a column of A , it is natural to require that $\langle \theta, X \rangle$ be bounded away from integer points. Otherwise, $\langle \theta, D \rangle$ could be close to an integer point with high probability, making $|\hat{D}(\theta)|$ large. An extra component in the definition says that the anti-concentration continues to hold after conditioning on an arbitrary half-space, passing through the mean. As will become apparent, this is a consequence of the subsampling scheme from Step 1.

Let us just note that the (anti-concentration) property is not vacuous. In fact Theorem 1.4 is a direct consequence of the following lemma (see the proof in Section 5) and Theorems 2.1 and 2.2 (note that by Remark 1 both cases are approximately symmetric).

LEMMA 2.1. *Suppose that for $X = (X_1, \dots, X_m) \sim \mathcal{D}$, one of the following holds,*

1. X is logconcave and isotropic.
2. X_i are i.i.d. uniformly on an integer interval $\{a, a+1, \dots, a+k\}$, with $k > 1$.

Then, \mathcal{D} satisfies (anti-concentration) with constant $\kappa > \frac{1}{50}$.

One may wonder about the necessity of the condition $k > 1$ in Case 2 of Lemma 2.1. A moment of reflection reveals that, if X is uniform on $\{0, 1\}$, then X is not anti-concentrated, for any $\kappa > 0$, and thus our framework does not directly apply to this case. However, by taking account of the possible bad cases, our analysis can be refined to also handle such distributions. We do not pursue this direction here.

Step 1 - subsampling: We will generate a sub-matrix of A by selecting a subset of the columns. This will ensure that the norm of the columns is bounded, as well as that the norm of their sum is small. Suppose that the columns of A satisfy (anti-concentration) with $\kappa, \sigma > 0$, for $i = 1, \dots, n$ we define random variables $Y_i \in \{0, 1\}$:

$$\Pr[Y_{k+1} = 1 | A_1, Y_1, \dots, A_k, Y_k] = \begin{cases} 1 & \text{if } \langle \sum_{j=1}^k Y_j (A_j - \mu), A_{k+1} - \mu \rangle < 0 \text{ and } \|A_{k+1} - \mu\| \leq 10\sigma \sqrt{\frac{m}{k}} \\ \frac{1}{2} & \text{if } \langle \sum_{j=1}^k Y_j (A_j - \mu), A_{k+1} - \mu \rangle = 0 \text{ and } \|A_{k+1} - \mu\| \leq 10\sigma \sqrt{\frac{m}{k}} \\ 0 & \text{else} \end{cases}$$

We then select all columns of A for which $Y_i = 1$. For now, let A'_i have the law of column A_i , conditional on being selected, and denote the selected set $S_A = \{i \in [n] | Y_i = 1\}$.

LEMMA 2.2. *Suppose that the columns of A satisfy (anti-concentration) with $\kappa, \sigma > 0$ and that it is an approximately symmetric distribution, in the sense of Definition 1. Then, if $n \gg \frac{m^2}{\kappa}$:*

1.

(norm concentration) $\|A'_i - \mu\| \leq 10\sigma\sqrt{\frac{m}{\kappa}}.$

2. With probability $1 - e^{-\Omega(n)}$, $|S_A| \geq \frac{n}{200}$,

(matrix concentration) $\sum_{i \in S_A} (A_i - \mu)(A_i - \mu)^\top \preceq 2n\sigma^2 \mathbf{I}_m$, and

(concentration) $\left\| \sum_{i \in S_A} A_i - \mu \right\|^2 \leq 2nm\sigma^2.$

3. $\Pr(d(\theta^\top A'_i, \mathbb{Z}) \geq \kappa \min(1, \|\theta\|_\infty \sigma) |A_1 Y_1, \dots, A_{i-1} Y_{i-1}) \geq \frac{\kappa}{2}$, for every $\theta \in V$, and $i \in [n']$.

Proof. [Proof] The first claim is immediate since we have conditioned the columns on the event $\{\|A_i - \mu\| \leq 10\sigma\sqrt{\frac{m}{\kappa}}\}$.

For (matrix concentration), let $Z_i \stackrel{\text{law}}{=} A_i | (\|A_i - \mu\| \leq 10\sigma\sqrt{\frac{m}{\kappa}})$ and note,

$$\sum_{i \in S_A} (A_i - \mu)(A_i - \mu)^\top = \sum_{i=1}^n Y_i (Z_i - \mu)(Z_i - \mu)^\top \leq \sum_{i=1}^n (Z_i - \mu)(Z_i - \mu)^\top.$$

As the random vectors $\{Z_i - \mu\}_{i=1}^n$ are mutually independent and $(Z_i - \mu)(Z_i - \mu)^\top \preceq 10\sigma\sqrt{\frac{m}{\kappa}} \mathbf{I}_m$ almost surely, (matrix concentration) follows from the matrix Bernstein inequality [48, Theorem 1.6.2].

For (concentration), since $\langle \sum_{j=1}^{i-1} Y_j (A_j - \mu), A_i - \mu \rangle \leq 0$,

$$\left\| \sum_{i \in S_A} A_i - \mu \right\|^2 \leq \sum_{i \in S_A} \|A_i - \mu\|^2 = \text{Tr} \left(\sum_{i \in S_A} (A_i - \mu)(A_i - \mu)^\top \right) \leq 2n \text{Tr}(\sigma^2 \mathbf{I}_m),$$

where the last inequality is monotonicity of the trace. Also, recalling that $\text{Cov}(A_i) \preceq \sigma^2 \mathbf{I}_d$, the fact that with high probability $|S_A| \geq \frac{n}{8}$ follows from Azuma's inequality. Indeed, for fixed $i \in [n]$, by Chebyshev's inequality,

$$\Pr \left(\|A_i - \mu\| > 10\sigma\sqrt{\frac{m}{\kappa}} \right) \leq \frac{\kappa \mathbb{E}[\|A_i - \mu\|^2]}{100\sigma^2 m} = \frac{\kappa \text{Tr}(\text{Cov}(A_i))}{100\sigma^2 m} \leq \frac{\kappa}{100}.$$

Since A_i has an approximately symmetric law, then, since A_i is independent from $\{Y_j, A_j\}_{j=1}^{i-1}$, by definition,

$$\Pr \left(\left\langle \sum_{j=1}^{i-1} Y_j (A_j - \mu), A_i - \mu \right\rangle \leq 0 \right) \geq \frac{1}{4e^2}.$$

Taken together, the above displays imply

$$\Pr(Y_i = 1 | A_1, \dots, A_{i-1}) \geq \frac{1}{100}.$$

Applying Azuma's inequality, as in (A.4), we get

$$\Pr \left(|S_A| \geq \frac{n}{200} \right) = 1 - e^{-\Omega(n)}.$$

Finally, we address the (anti-concentration) property. For fixed $i \in [n']$, let us define $v = \sum_{j=1}^{i-1} A_j Y_j$. So,

$$\begin{aligned} & \Pr(d(\theta^\top A'_i, \mathbb{Z}) \geq \kappa \min(1, \|\theta\|_\infty \sigma) |A_1 Y_1, \dots, A_{i-1} Y_{i-1}) \\ &= \Pr \left(d(\theta^\top A_i, \mathbb{Z}) \geq \kappa \min(1, \|\theta\|_\infty \sigma) | \langle v, A_i - \mu \rangle \leq 0 \text{ and } \|A_i - \mu\| \leq 10\sigma\sqrt{\frac{m}{\kappa}} \right) \\ &\geq \Pr(d(\theta^\top A_i, \mathbb{Z}) \geq \kappa \min(1, \|\theta\|_\infty \sigma) | \langle v, A_i - \mu \rangle \leq 0) - \frac{\kappa}{100} \geq \frac{\kappa}{2}. \end{aligned}$$

Here, the last inequality follows from Definition 2, while the first inequality is a union bound on the anti-concentration event and $\{\|A_i - \mu\| \leq 10\sigma\sqrt{\frac{m}{\kappa}}\}$. \square

In light of the lemma, in the sequel, all computations will be made conditioned on the high-probability event defined by Lemma 2.2, and we will only consider the selected columns. Thus, with a slight abuse of notation, from now on, the random variables D, \hat{D}, A_i , etc., will only be considered with respect to the selected columns. In particular, we will write n for $|S_A|$.

Step II - bounding the argument: We will now show that for small θ , the argument of $\hat{D}(\theta)$ is close to $2\pi n p \langle \theta, \mu \rangle$.

Before proceeding, we introduce an auxiliary parameter $\beta = \sqrt{\frac{10^3 m^2}{\kappa^3} \ln \left(\frac{10^3 m}{\kappa^3} \left(1 + \frac{\|\mu\|^2}{\sigma^2} \right)^{\frac{1}{m}} \right)}$, parametrizing the region in which we control $\hat{D}(\theta)$. We record here some facts which will be useful later on:

LEMMA 2.3. Assume that Assumption 1 is satisfied. If $\beta = \sqrt{\frac{10^3 m^2}{\kappa^3} \ln \left(\frac{10^3 m}{\kappa^3} \left(1 + \frac{\|\mu\|^2}{\sigma^2} \right)^{\frac{1}{m}} \right)}$, then for large enough n ,

$$\begin{aligned} \frac{1}{\sqrt{24\pi^2}} &\leq \beta \leq \frac{1}{80\sqrt{pm}} \\ \left(\frac{m}{\kappa}\right)^{\frac{3}{2}} \beta^3 &\leq \frac{\sqrt{pn}}{50000} \\ \left(1 + \frac{\|\mu\|}{\sigma}\right)^3 \beta^3 &\leq \frac{\sqrt{pn}}{50000} \end{aligned}$$

Proof. [Proof] The lower bound in the first inequality is trivial. The upper bound follows from the upper bound in (2.1), $p \leq \frac{\kappa^8}{10^{15} m^6} \frac{1}{\ln \left(10 \left(1 + \frac{\|\mu\|^2}{\sigma^2} \right) \right)}$, because

$$\begin{aligned} \beta &= \sqrt{\frac{10^3 m^2}{\kappa^3} \left(\ln \left(\frac{100m}{\kappa^3} \right) + \ln \left(10 \left(1 + \frac{\|\mu\|^2}{\sigma^2} \right)^{\frac{1}{m}} \right) \right)} \\ &\leq \sqrt{\frac{10^3 m^2}{\kappa^3} \left(3 \left(\frac{100m}{\kappa^3} \right)^{\frac{1}{3}} + \ln \left(10 \left(1 + \frac{\|\mu\|^2}{\sigma^2} \right)^{\frac{1}{m}} \right) \right)} \\ &\leq \sqrt{\frac{10^5 m^{2.5}}{\kappa^4} \ln \left(10 \left(1 + \frac{\|\mu\|^2}{\sigma^2} \right)^{\frac{1}{m}} \right)} \leq \frac{1}{80\sqrt{pm}}. \end{aligned}$$

The first inequality uses $\ln(x) = 3 \ln(x^{\frac{1}{3}}) \leq 3x^{\frac{1}{3}}$ and the second $x + y \leq xy$, valid when $x, y \geq 1$.

For the first inequality concerning β^3 , we use the above bound, which implies $\beta^3 \leq 10^9 \frac{m^{15}}{\kappa^6}$. We thus compute,

$$\left(\frac{m}{\kappa}\right)^{\frac{3}{2}} \beta^3 \leq 10^9 \frac{m^9}{\kappa^8} \left(1 + \frac{\|\mu\|}{\sigma}\right)^2 \leq \frac{\sqrt{pn}}{50000}$$

where the second inequality follows, again, from (2.1). Using a similar argument we also get,

$$\left(1 + \frac{\|\mu\|}{\sigma}\right)^5 \beta^3 \leq 10^9 \frac{m^8}{\kappa^6} \left(1 + \frac{\|\mu\|}{\sigma}\right)^2 \leq \frac{\sqrt{pn}}{50000}$$

which completes the proof. \square

The main observation in this step, is that, when we fix the columns $\{A_j\}_{j=1}^n$, a Taylor approximation implies the bound,

$$\left| \sum_{j=1}^n \arg(\mathbb{E}_S[\exp(\mathbf{1}_{i \in S} 2\pi i A_j)]) - 2\pi n p \langle \theta, \mu \rangle \right| \leq 2\pi p \left| \sum_{j=1}^n \langle \theta, A_j \rangle - n \langle \theta, \mu \rangle \right| + 50p \sum_{j=1}^n |\langle \theta, A_j \rangle|^3.$$

The term on the LHS controls $\arg(\hat{D}(\theta))$ and the two terms on the RHS can be bounded with (concentration) and (norm concentration) respectively. We then prove:

LEMMA 2.4. With probability $1 - e^{-\Omega(n)}$ over A , for $\|\theta\| \leq \frac{\beta}{\sigma\sqrt{pn}}$,

$$|\arg(\hat{D}(\theta)) - 2\pi n p \langle \theta, \mu \rangle| \leq 4\pi\beta\sqrt{pm} + \frac{5000((m/\kappa)^{\frac{3}{2}} + \|\mu\|^3/\sigma^3)\beta^3}{\sqrt{pn}}.$$

Proof. [Proof] Let $f(x) = \arg(p \cdot \exp(2\pi i \cdot x) + (1-p))$ and observe that $f(x) = \arctan\left(\frac{p \sin(2\pi x)}{p \cos(2\pi x) + (1-p)}\right)$. A calculation shows $f(0) = 0$, $f'(0) = 2\pi p$ and $f''(0) = 0$. Hence, we set $g(x) = f(x) - 2\pi p x$ and, with a second-order Taylor approximation of $f(x)$ around $x = 0$, we see that for every $x \geq 0$ there is some $x' \in [0, x]$ such that

$$g(x) = \frac{d^3 f}{dx^3}(x')x^3.$$

Another calculation shows that, as long as $p \leq 0.1$, we have $\frac{d^3 f}{dx^3}(x') \leq 50p$ and hence $|g(x)| \leq 50|x|^3 p$. Now, note,

$$\arg(\mathbb{E}[\exp(\mathbf{1}_{j \in S} 2\pi i x)]) = \arg(p \cdot \exp(2\pi i \cdot x) + (1-p)) = f(x).$$

Now,

$$\begin{aligned} |\arg(\hat{D}(\theta)) - 2\pi n p \langle \theta, \mu \rangle| &= \left| \arg\left(\prod_{j=1}^n \mathbb{E}[\exp(\mathbf{1}_{j \in S} 2\pi i \langle \theta, A_j \rangle)]\right) - 2\pi n p \langle \theta, \mu \rangle \right| \\ &= \left| \sum_{j=1}^n \arg(\exp(\mathbf{1}_{j \in S} 2\pi i \langle \theta, A_j \rangle)) - 2\pi n p \langle \theta, \mu \rangle \right| \\ &= \left| \sum_{j=1}^n f(\langle \theta, A_j \rangle) - 2\pi n p \langle \theta, \mu \rangle \right|, \end{aligned}$$

where we understand $|\cdot|$ as referring to distance on the circle. The Taylor approximation given above shows that, when $\|\theta\| \leq \frac{\beta}{\sigma\sqrt{pn}}$, we can bound this distance (where now the bound will be expressed as a distance between real numbers),

$$\begin{aligned} \left| \sum_{j=1}^n f(\langle \theta, A_j \rangle) - 2\pi n p \langle \theta, \mu \rangle \right| &\leq 2\pi p \left| \sum_{j=1}^n \langle \theta, A_j \rangle - n \langle \theta, \mu \rangle \right| + 50p \sum_{j=1}^n |\langle \theta, A_j \rangle|^3 \\ &\leq 2\pi p \left| \left\langle \theta, \sum_{j=1}^n (A_j - \mu) \right\rangle \right| + 50p \|\theta\|^3 \sum_{j=1}^n \|A_j\|_2^3 \\ &\leq 2\pi p \|\theta\| \left\| \sum_{j=1}^n (A_j - \mu) \right\| + 50p \|\theta\|^3 n \left(10\sigma\sqrt{m/\kappa} + \|\mu\|\right)^3 \\ &\leq 4\pi p \beta \sqrt{\frac{m\sigma^2}{\sigma^2 p}} + \frac{50(10\sigma\sqrt{m/\kappa} + \|\mu\|)^3 \beta^3}{\sigma^3 \sqrt{np}} \\ &= 4\pi\beta\sqrt{pm} + \frac{50\left(10\sqrt{m/\kappa} + \frac{\|\mu\|}{\sigma}\right)^3 \beta^3}{\sqrt{pn}}, \end{aligned}$$

where the third inequality follows from (norm concentration) and the penultimate inequality being a consequence of (concentration). \square

The following corollary is now immediate.

COROLLARY 2.1. With probability $1 - e^{-\Omega(n)}$, for all $\|\theta\| \leq \frac{\beta}{\sigma\sqrt{pn}}$,

$$|\arg(\hat{D}(\theta)) - 2\pi n p \langle \theta, \mu \rangle| \leq \frac{1}{8} \pi.$$

Proof. Lemma 2.3 implies $\beta\sqrt{pm} \leq \frac{1}{80}$, $\beta^3 \left(\frac{\|\mu\|}{\sigma}\right)^3 \leq \frac{\sqrt{pm}}{50000}$ and, $\beta^3(m/\kappa)^{\frac{3}{2}} \leq \frac{\sqrt{pm}}{50000}$. Thus,

$$4\pi\beta\sqrt{pm} + \frac{50 \left(10\sqrt{m/\kappa} + \frac{\|\mu\|}{\sigma}\right)^3 \beta^3}{\sqrt{pn}} \leq \frac{1}{8}\pi,$$

and the Corollary follows from Lemma 2.4. \square

Step III - bounding the integral from below, near the origin: Next, we prove that the modulus of \hat{D} is bounded, near the origin.

LEMMA 2.5. *The following holds,*

$$\int_{\|\theta\| \leq \frac{\beta}{\sigma\sqrt{pm}}} |\hat{D}(\theta)| d\theta \geq \left(\frac{1}{200\pi^3 nmp}\right)^{\frac{m}{2}} \frac{1}{\sigma^{m-1} (\sigma^2 + \|\mu\|^2)^{\frac{1}{2}}}.$$

Proof. [Proof] We have:

$$\begin{aligned} |\hat{D}(\theta)| &= \prod_{j=1}^n |\mathbb{E}[\exp(2\pi i \langle \theta, A_j \rangle)]| = \prod_{j=1}^n |(1-p) \cdot 1 + p \exp(2\pi i \langle \theta, A_j \rangle)| \\ &= \prod_{j=1}^n \sqrt{(1-p + p \cos(2\pi \langle \theta, A_j \rangle))^2 + p^2 \sin(2\pi \langle \theta, A_j \rangle)^2} \\ &\geq \prod_{j=1}^n (1-p + p \cos(2\pi \langle \theta, A_j \rangle)) \geq \exp(-6\pi^2 p \sum_{i=1}^n \langle \theta, A_i \rangle^2). \end{aligned}$$

Here the last inequality follows, as long as $p \leq 0.01$, from the elementary inequalities,

$$\begin{aligned} \cos(x) &\geq 1 - x^2 \\ \ln(1-x) &\geq -\frac{3}{2}x \quad \text{when } |x| \leq \frac{1}{2}. \end{aligned}$$

Indeed, it's enough to consider $\langle \theta, A_j \rangle \in [-1, 1]$, for which,

$$\ln(1-p + p \cos(2\pi \langle \theta, A_j \rangle)) \geq \ln(1 - p4\pi^2 \langle \theta, A_j \rangle^2) \geq -6\pi^2 p \langle \theta, A_j \rangle^2.$$

Hence, by applying the (matrix concentration) property to the obtained bound, we get,

$$\begin{aligned} |\hat{D}(\theta)| &\geq \exp\left(-12\pi^2 p \left(n\langle \theta, \mu \rangle^2 + \sum_{i=1}^n \langle \theta, A_i - \mu \rangle^2\right)\right) \geq \exp(-24\pi^2 np (\langle \theta, \mu \rangle^2 + \sigma^2 \theta \theta^\top)) \\ &\geq \exp(-24\pi^2 \theta (\mu^\top \mu + \sigma^2 \mathbf{I}_m) \theta^\top np). \end{aligned}$$

Let $Y \sim \mathcal{N}\left(0, \frac{1}{48\pi^2 np} (\mu^\top \mu + \sigma^2 \mathbf{I}_m)^{-1}\right)$, then,

$$\begin{aligned} \int_{\|\theta\| \leq \frac{\beta}{\sigma\sqrt{pm}}} |\hat{D}(\theta)| d\theta &\geq \int_{\|\theta\| \leq \frac{\beta}{\sigma\sqrt{pm}}} \exp(-24\pi^2 \theta (\mu^\top \mu + \sigma^2 \mathbf{I}_m) \theta^\top np) d\theta \\ &= \frac{1}{\sqrt{\det(96\pi^3 np (\mu^\top \mu + \sigma^2 \mathbf{I}_m))}} \Pr\left(\|Y\| \leq \frac{\beta}{\sigma\sqrt{pn}}\right) \\ (2.2) \quad &= \frac{1}{(96\pi^3 np)^{\frac{m}{2}} (\|\mu\|^2 + \sigma^2)^{\frac{1}{2}} \sigma^{m-1}} \Pr\left(\|Y\| \leq \frac{\beta}{\sigma\sqrt{pn}}\right). \end{aligned}$$

By Chebyshev's inequality,

$$\Pr\left(\|Y\| \geq \frac{\beta}{\sigma\sqrt{pn}}\right) \leq \frac{\sigma^2 pn}{\beta^2} \text{Tr}\left(\frac{1}{48\pi^2 np} (\mu^\top \mu + \sigma^2 \mathbf{I}_m)^{-1}\right) \leq \frac{m}{48\pi^2 \beta^2} \leq \frac{1}{2},$$

where in the last inequality, we have used $\beta \geq \sqrt{\frac{10^3 m^2}{\kappa^3}} \geq \sqrt{2 \frac{m}{48\pi^2}}$. The proof concludes by plugging this estimate into (2.2). \square

To handle the integral in Theorem A.1, we note that when $\|t - pn\mu\|$ is small enough, as dictated by Theorem 2.1, Corollary 2.1 implies,

$$|\arg(\hat{D}(\theta) \exp(2\pi i \langle \theta, t \rangle))| \leq |\arg(\hat{D}(\theta)) - 2\pi n p \langle \theta, \mu \rangle| + |\arg(\exp(2\pi i \langle \theta, t - pn\mu \rangle))| \leq \frac{1}{4} \pi,$$

whenever $\|\theta\| \leq \frac{\beta}{\sigma\sqrt{pn}}$ (recall $\beta = \sqrt{\frac{10^3 m^2}{\kappa^3} \ln\left(\frac{10^3 m}{\kappa^3} \left(1 + \frac{\|\mu\|^2}{\sigma^2}\right)^{\frac{1}{m}}\right)}$). For a complex number z , we have $\Re(z) = \cos(\arg(z))|z|$. Thus, from Lemma 2.5, we deduce the following result.

LEMMA 2.6. *Let $t \in \mathbb{R}^m$ such that*

$$\|t - pn\mu\| \leq \frac{\sigma \kappa^{\frac{3}{2}}}{10^4 \ln\left(\frac{10^3 m}{\kappa^3} \left(1 + \frac{\|\mu\|^2}{\sigma^2}\right)^{\frac{1}{m}}\right)} \frac{\sqrt{pn}}{m}.$$

Then, for $\|\theta\| \leq \frac{\beta}{\sigma\sqrt{pn}}$,

$$\Re[\hat{D}(\theta) \exp(-2\pi i \langle \theta, t \rangle)] \geq 0,$$

and

$$\Re \left[\int_{\|\theta\| \leq \frac{\beta}{\sigma\sqrt{pn}}} \hat{D}(\theta) \exp(-2\pi i \langle \theta, t \rangle) d\theta \right] \geq \cos\left(\frac{\pi}{4}\right) \left(\frac{1}{200\pi^3 nmp}\right)^{\frac{m}{2}} \frac{1}{\sigma^{m-1} (\sigma^2 + \|\mu\|^2)^{\frac{1}{2}}}.$$

Step IV - exponential decay of the Fourier spectrum: To show that the Fourier spectrum decays rapidly, we employ an ε -net argument over a very large box. The main difficulty is that $\langle \theta, D \rangle$ can be close to an integer, irregardless of the value $\|\theta\|$. Our anti-concentration assumption allows us to avoid this. Specifically, Item 3 in Lemma 2.2 implies that for any given θ in a dense enough net, we can expect many columns to satisfy that $\langle \theta, A_i \rangle$ is far from any integer point. Formally, we prove:

LEMMA 2.7. *Assume that Assumption 1 is satisfied. Then, with probability $1 - e^{-\Omega(\kappa n)}$, we have*

$$|\hat{D}(\theta)| \leq \exp\left(-\frac{1}{80} \kappa^3 n (1-p) p \pi^2 \min(1, \|\theta\|_\infty^2 \sigma^2)\right),$$

for $\theta \in \tilde{V}$, where V is the fundamental domain, as in (A.2), and

$$\tilde{V} := \left[-e^{\frac{\kappa^3 pn}{80m}} \frac{1}{4\sigma n^2}, e^{\frac{\kappa^3 pn}{80m}} \frac{1}{4\sigma n^2}\right]^m \cap V.$$

Proof. [Proof] We begin with a technical calculation which will help control the size of the net. Observe that (2.1) implies

$$160m \leq \sqrt{\kappa n} \implies \sqrt{\frac{m}{\kappa}} \leq \frac{1}{2} \sqrt{\frac{\kappa n}{80m}} \implies \sqrt{\frac{m}{\kappa}} \leq \frac{1}{2} e^{\frac{\kappa n}{80m}},$$

and similarly,

$$\frac{\|\mu\|}{\sigma} \leq \frac{1}{2} e^{\frac{\kappa n}{80m}}.$$

Now, let N be an ε -net of \tilde{V} for $\varepsilon = \frac{1}{2n^2(\sigma\sqrt{m/\kappa} + \|\mu\|)}$. Standard arguments show that one can take $|N| \leq \left(\sqrt{\frac{m}{\kappa}} + \frac{\|\mu\|}{\sigma}\right)^m e^{\frac{\kappa n}{80m}} \leq e^{\frac{\kappa n}{40m}}$, when n is large enough. Here we have used the bound, $\sqrt{\frac{m}{\kappa}} + \frac{\|\mu\|}{\sigma} \leq e^{\frac{\kappa n}{80m}}$, established above.

For $\theta \in N$, define

$$\varphi(\theta) = \frac{\kappa}{4} \min(1, \|\theta\|_\infty \sigma) \text{ and } E(\theta) = \{j : d(\langle \theta, A_j \rangle, \mathbb{Z}) \geq \varphi(\theta)\}.$$

If we fix θ and set $X_i := \mathbf{1}_{i \in E(\theta)}$ then $|E(\theta)| = \sum_{i=1}^n X_i$. The statement of Item 3 in Lemma 2.2 is $\mathbb{E}[X_i | X_1, \dots, X_{i-1}] \geq \frac{\kappa}{2}$. Applying Azuma's inequality (A.4), we get:

$$\Pr \left[|E(\theta)| \leq \frac{\kappa}{4} n \right] \leq \exp \left(-\frac{\kappa}{8} n \right).$$

In this case, by the union bound,

$$\mathbb{P} \left(\exists \theta \in N : |E(\theta)| \leq \frac{\kappa}{4} n \right) \leq e^{-\frac{\kappa}{16} n}.$$

If $j \in E(\theta)$, since $\varphi(\theta) \leq \frac{1}{4}$, we have:

$$\begin{aligned} |\mathbb{E}_S[\exp(\mathbf{1}_{j \in S} \cdot 2\pi i \langle \theta, A_j \rangle)]| &= \sqrt{(1-p + p \cos(2\pi \langle \theta, A_j \rangle))^2 + p^2 \sin(2\pi \langle \theta, A_j \rangle)^2} \\ &= \sqrt{1 + 2p^2 - 2p + 2(1-p)p \cos(2\pi \langle \theta, A_j \rangle)} \\ &\leq \sqrt{1 + 2p^2 - 2p + 2(1-p)p \cos(2\pi \varphi(\theta))} \\ &\leq \sqrt{1 - 2(1-p)p(2\pi \varphi(\theta))^2/5} \\ &\leq 1 - (1-p)p(2\pi \varphi(\theta))^2/5 \\ &\leq 1 - \frac{4}{5}(1-p)p\pi^2 \frac{\kappa^2}{16} \min(1, \|\theta\|_\infty^2 \sigma^2). \end{aligned}$$

Observe that $|\mathbb{E}_S[\exp(\mathbf{1}_{j \in S} 2\pi i x)]| = \sqrt{(1-p + p \cos(2\pi x))^2 + p^2 \sin(2\pi x)^2}$ is $4\pi p$ -Lipschitz in x , as long as $p \leq \frac{1}{4}$. Take an arbitrary $\theta \in \tilde{V}$ and let θ' be the closest point in N . Recall that $\max_i \|A_i\| \leq 10\sigma \sqrt{\frac{m}{\kappa}} + \|\mu\|$, which follows from eq. (norm concentration). So, by our choice of ε and with the Cauchy-Schwartz inequality,

$$|\langle \theta - \theta', A_j \rangle| \leq \frac{1}{2n^2}.$$

Thus:

$$\begin{aligned} |\hat{D}(\theta)| &= \prod_{j=1}^n |\mathbb{E}_S[\exp(\mathbf{1}_{j \in S} 2\pi i \langle \theta, A_j \rangle)]| \\ &\leq \prod_{j \in E(\theta')} (|\mathbb{E}_S[\exp(\mathbf{1}_{j \in S} 2\pi i \langle \theta', A_j \rangle)]| + 4\pi p |\langle \theta - \theta', A_j \rangle|) \\ &\leq \prod_{j \in E(\theta')} \left(1 - \frac{4}{5}(1-p)p\pi^2 \frac{\kappa^2}{16} \min(1, \|\theta'\|_\infty^2 \sigma^2) + 4\pi p |\langle \theta - \theta', A_j \rangle| \right) \\ &\leq \exp \left(-\frac{4}{5}(1-p)p\pi^2 \frac{\kappa^2}{16} |E(\theta')| \min(1, \|\theta'\|_\infty^2 \sigma^2) + 4\pi p \sum_{j=1}^n |\langle \theta - \theta', A_j \rangle| \right) \\ &\leq e^{\frac{2\pi}{n}} \exp \left(-\frac{1}{5} \frac{\kappa^3}{16} n(1-p)p\pi^2 \min(1, \|\theta'\|_\infty^2 \sigma^2) \right) \\ &\leq e^{\frac{2\pi}{n}} \exp \left(-\frac{1}{5} \frac{\kappa^3}{16} n(1-p)p\pi^2 \min(1, (1-\varepsilon)^2 \|\theta\|_\infty^2 \sigma^2) \right) \\ &\leq \exp \left(-\frac{1}{80} \kappa^3 n(1-p)p\pi^2 \min(1, \|\theta\|_\infty^2 \sigma^2) \right), \end{aligned}$$

where the last inequality holds, since by (2.1), $\kappa^3 pn \geq 10^{20}$. \square

By properly integrating the inequality, we have thus obtained:

LEMMA 2.8. *With probability $1 - e^{-\Omega(\kappa n)}$, the following inequality holds:*

$$\int_B |\hat{D}(\theta)| d\theta \leq 2 \left(\frac{1}{10^5 p m n \sigma^2 \left(1 + \frac{\|\mu\|^2}{\sigma^2}\right)^{\frac{1}{m}}} \right)^{\frac{m}{2}},$$

where

$$B = V \cap \left\{ \|\theta\|_\infty < e^{\frac{\kappa^3 p n}{80m}} \frac{1}{4\sigma n^2} \right\} \cap \left\{ \|\theta\| \geq \frac{\beta}{\sigma \sqrt{p n}} \right\}.$$

Proof. [Proof] From Lemma 2.7, when $p < 0.1$, and $\sigma \|\theta\|_\infty \leq \sigma \|\theta\| \leq \sqrt{m}$, we have

$$|D(\theta)| \leq \exp\left(-\frac{\kappa^3 p n \sigma^2 \|\theta\|^2}{10m}\right).$$

Now, if $Y \sim \mathcal{N}(0, \mathbf{I}_m)$ and $n \geq 100m^6$:

$$\begin{aligned} \int_{\sqrt{m} \geq \|\theta\| \geq \frac{\beta}{\sigma \sqrt{p n}}} |D(\theta)| d\theta &\leq \int_{\|\theta\| \geq \frac{\beta}{\sigma \sqrt{p n}}} \exp\left(-\frac{\kappa^3 p n \sigma^2 \|\theta\|^2}{10m}\right) d\theta \\ &\leq \left(\frac{5m}{\kappa^3 p n \sigma^2}\right)^{\frac{m}{2}} \mathbb{P}\left(\|Y\| \geq \frac{\beta}{\sigma \sqrt{p n}} \sqrt{\frac{\kappa^3 p n \sigma^2}{5m}}\right) \\ &= \left(\frac{5m}{\kappa^3 p n \sigma^2}\right)^{\frac{m}{2}} \mathbb{P}\left(\|Y\|^2 \geq \beta^2 \frac{\kappa^3}{5m}\right) \\ &= \left(\frac{5m}{\kappa^3 p n \sigma^2}\right)^{\frac{m}{2}} \mathbb{P}\left(\|Y\|^2 \geq 200m \ln\left(\frac{10^3 m}{\kappa^3} \left(1 + \frac{\|\mu\|^2}{\sigma^2}\right)^{\frac{1}{m}}\right)\right). \end{aligned}$$

The last equality holds since $\beta = \sqrt{\frac{10^3 m^2}{\kappa^3} \ln\left(\frac{10^3 m}{\kappa^3} \left(1 + \frac{\|\mu\|^2}{\sigma^2}\right)^{\frac{1}{m}}\right)}$. By Lemma A.2,

$$\begin{aligned} \mathbb{P}\left(\|Y\|^2 \geq 200m \ln\left(\frac{10^3 m}{\kappa^3} \left(1 + \frac{\|\mu\|^2}{\sigma^2}\right)^{\frac{1}{m}}\right)\right) &\leq \exp\left(-50m \ln\left(\frac{10^3 m}{\kappa^3} \left(1 + \frac{\|\mu\|^2}{\sigma^2}\right)^{\frac{1}{m}}\right)\right) \\ &\leq \left(\frac{\kappa^3}{10^6 m}\right)^{\frac{m}{2}} \frac{1}{\left(1 + \frac{\|\mu\|^2}{\sigma^2}\right)^{\frac{1}{2}}}. \end{aligned}$$

Hence,

$$\int_{\sqrt{m} \geq \|\theta\| \geq \frac{\beta}{\sigma \sqrt{p n}}} |D(\theta)| d\theta \leq \left(\frac{1}{10^5 p m n \sigma^2}\right)^{\frac{m}{2}} \frac{1}{\left(1 + \frac{\|\mu\|^2}{\sigma^2}\right)^{\frac{1}{2}}}.$$

Furthermore, for all $\theta \in \left[-\frac{\kappa^3 p n}{4\sigma n^2}, \frac{\kappa^3 p n}{4\sigma n^2}\right]^m$ with $\|\theta\|_\infty \sigma \geq 1$, Lemma 2.7 also implies,

$$|D(\theta)| \leq \exp\left(-\frac{\kappa^3 p n}{40m}\right).$$

So:

$$\int_{\frac{e^{\frac{\kappa^3 pn}{80m}}}{4\sigma n^2} \geq \|\theta\|_\infty \geq \frac{1}{4}} |D(\theta)| d\theta \leq \frac{1}{\sigma^m n^{2m}} \exp\left(\frac{\kappa^3 pn}{80m}\right)^m \cdot \exp\left(-\frac{\kappa^3 pn}{40}\right) = \frac{1}{\sigma^m n^{2m}} \exp\left(-\frac{\kappa^3 pn}{80}\right).$$

Summing the previous two inequalities yields

$$\int_B |\hat{D}(\theta)| d\theta \leq \left(\frac{1}{10^5 pmn \sigma^2 \left(1 + \frac{\|\mu\|^2}{\sigma^2}\right)^{\frac{1}{m}}} \right)^{\frac{m}{2}} + \left(\frac{1}{\sigma^2 n^4} \right)^{\frac{m}{2}} \exp\left(-\frac{\kappa^3 pn}{80}\right).$$

The proof concludes by recalling the assumption from (2.1), which implies a lower bound on $\kappa^3 pn$, and hence affords the bound,

$$\begin{aligned} & \left(\frac{1}{10^5 pmn \sigma^2 \left(1 + \frac{\|\mu\|^2}{\sigma^2}\right)^{\frac{1}{m}}} \right)^{\frac{m}{2}} + \left(\frac{1}{\sigma^2 n^4} \right)^{\frac{m}{2}} \exp\left(-\frac{\kappa^3 pn}{80}\right) \\ & \leq 2 \left(\frac{1}{10^5 pmn \sigma^2 \left(1 + \frac{\|\mu\|^2}{\sigma^2}\right)^{\frac{1}{m}}} \right)^{\frac{m}{2}}. \end{aligned}$$

□

Proving Theorem 2.1 and 2.2: If the support of \mathcal{D} is contained in \mathbb{Z}^m , then Lemma 2.6 and Lemma 2.8 are enough to prove Theorem 2.2.

Proof. [Proof of Theorem 2.2] Let B be defined as in Lemma 2.8 and note that since $2\sigma n^2 \leq e^{\frac{\kappa^3 pn}{80m}}$,

$$\left[-\frac{1}{2}, \frac{1}{2}\right]^m = V \subset \left\{ \|\theta\|_\infty < e^{\frac{\kappa^3 pn}{80m}} \frac{1}{4\sigma n^2} \right\}.$$

Thus,

$$B = \left[-\frac{1}{2}, \frac{1}{2}\right]^m \cap \left\{ \|\theta\| \geq \frac{\beta}{\sigma\sqrt{pn}} \right\}.$$

By Lemma 2.8,

$$(2.3) \quad \int_B |\hat{D}(\theta)| d\theta \leq 0.2 \left(\frac{1}{200\pi^3 nmp \sigma^2 \left(1 + \frac{\|\mu\|^2}{\sigma^2}\right)^{\frac{1}{m}}} \right)^{\frac{m}{2}}.$$

Moreover, by Lemma 2.6,

$$\begin{aligned} (2.4) \quad \Re \left[\int_{\|\theta\| \leq \frac{\beta}{\sigma\sqrt{pn}}} \hat{D}(\theta) \exp(-2\pi i \langle \theta, t \rangle) d\theta \right] & \geq \cos\left(\frac{\pi}{4}\right) \left(\frac{1}{200\pi^3 nmp} \right)^{\frac{m}{2}} \frac{1}{\sigma^{m-1} (\sigma^2 + \|\mu\|^2)^{\frac{1}{2}}} \\ & = \cos\left(\frac{\pi}{4}\right) \left(\frac{1}{200\pi^3 nmp \sigma^2 \left(1 + \frac{\|\mu\|^2}{\sigma^2}\right)^{\frac{1}{m}}} \right)^{\frac{m}{2}}. \end{aligned}$$

So,

$$\begin{aligned} \Pr[D = t] &= \Re \left[\int_{\theta \in [-\frac{1}{2}, \frac{1}{2}]^m} \hat{D}(\theta) \exp(-2\pi i \langle \theta, t \rangle) d\theta \right] \\ &\geq \Re \left[\int_{\|\theta\| \leq \frac{\beta}{\sigma\sqrt{pn}}} \hat{D}(\theta) \exp(-2\pi i \langle \theta, t \rangle) d\theta \right] - \int_B |\hat{D}(\theta)| d\theta \\ &\geq \frac{1}{2} \left(\frac{1}{200\pi^3 nmp\sigma^2 \left(1 + \frac{\|\mu\|^2}{\sigma^2}\right)^{\frac{1}{m}}} \right)^{\frac{m}{2}}. \end{aligned}$$

Finally, using the multiplicative Chernoff bound we can see that:

$$\Pr[|S| \notin [0.5pn, 1.5pn]] = \Pr \left[\left| |S| - \mathbb{E}[|S|] \right| \geq \frac{1}{2} \mathbb{E}[|S|] \right] \leq 2\exp(-pn/12).$$

By (2.1), $pn \geq \left(10m \left(1 + \frac{\|\mu\|}{\sigma}\right)\right)^{20}$. So, we have $\Pr[D = t] > 2\exp(-pn/12)$, which implies the existence of a suitable S with $|S| \in [0.5pn, 1.5pn]$ and $A\mathbf{1}_S = t$. \square

The proof of the continuous case requires an extra step to control the Fourier transform outside the domain of Lemma 2.8. To deal with continuous distributions, we define $R \sim \mathcal{N}(0, \gamma I_m)$ for γ to be determined later and $H = D + R$. If f_H is the density of H , we will show that, for appropriate t , $f_H(t)$ is positive, from which it will follow that with high probability there exists a suitable set S such that $\|A\mathbf{1}_S - t\|$ is small.

Proof. [Proof of Theorem 2.1] As in the proof of Theorem 2.2, let B be defined as in Lemma 2.8. Since the columns of A are now absolutely continuous we have $V = \mathbb{R}^m$, and so,

$$B = \left\{ \|\theta\|_\infty < e^{\frac{\kappa^3 pn}{80m}} \frac{1}{4\sigma n^2} \right\} \cap \left\{ \|\theta\| \geq \frac{\beta}{\sigma\sqrt{pn}} \right\}.$$

Another difference from the discrete case is that now, by the multiplication-convolution theorem (Theorem A.2), we have $\hat{H}(\theta) = \hat{D}(\theta) \cdot \hat{R}(\theta) = e^{-\frac{\|\theta\|^2 \gamma}{2}} \hat{D}(\theta)$. So, if $\theta_0 := \exp\left(\frac{\kappa^3 pn}{80m}\right) \frac{1}{\sigma n^2}$, the estimates in (2.3) and (2.4) imply,

$$\begin{aligned} &\Re \left[\int_{[-\theta_0, \theta_0]^m} \hat{H}(\theta) \exp(-2\pi i \langle \theta, t \rangle) d\theta \right] \\ &\geq e^{-\frac{\beta^2 \gamma}{\sigma^2 pn}} \left(\Re \left[\int_{\|\theta\| \leq \frac{\beta}{\sigma\sqrt{pn}}} \hat{D}(\theta) \exp(-2\pi i \langle \theta, t \rangle) d\theta \right] - \int_B |\hat{D}(\theta)| d\theta \right) \\ &\geq \frac{1}{4} \left(\frac{1}{200\pi^3 nmp\sigma^2 \left(1 + \frac{\|\mu\|^2}{\sigma^2}\right)^{\frac{1}{m}}} \right)^{\frac{m}{2}}, \end{aligned}$$

where the last inequality holds as long as $\gamma < \frac{\sigma^2 pn}{\beta^2}$. Since we wish to invoke Theorem A.1, we need to bound $\int_{\mathbb{R}^m} |\hat{H}(\theta)| \exp(-2\pi i \langle \theta, t \rangle) d\theta$ from below. In light of the above computation, it will be enough to choose γ such that,

$$(2.5) \quad \int_{\|\theta\|_\infty \geq \theta_0} |\hat{H}(\theta)| d\theta \leq \frac{1}{8} \left(\frac{1}{200\pi^3 nmp\sigma^2 \left(1 + \frac{\|\mu\|^2}{\sigma^2}\right)^{\frac{1}{m}}} \right)^{\frac{m}{2}}.$$

Towards finding an appropriate γ , let Y stand for the standard Gaussian in \mathbb{R}^m , and compute,

$$\begin{aligned} \int_{\|\theta\|_\infty \geq \theta_0} |\hat{H}(\theta)| d\theta &\leq \int_{\|\theta\| \geq \theta_0} e^{-\frac{\gamma\|\theta\|^2}{2}} d\theta \\ &= \left(\frac{2\pi}{\gamma}\right)^{\frac{m}{2}} \Pr(\|Y\|^2 \geq \gamma\theta_0^2) \\ &\leq \left(\frac{2\pi}{\gamma}\right)^{\frac{m}{2}} e^{-\frac{\gamma\theta_0^2}{3}}, \end{aligned}$$

where the last inequality is Lemma A.2. Let us choose now

$$\gamma = \exp\left(-\kappa^3 \frac{pn}{40m}\right) \sigma^2 n^5 = \frac{n}{\theta_0^2},$$

for which (2.5) holds. Also, $\gamma < \frac{n\sigma^2}{\beta^2}$ as required earlier and $\gamma\theta_0^2 \geq 7m$, as required by Lemma A.2. If $f_H(t)$ is the density of H at t , Theorem A.1 along with (2.5) give,

$$\begin{aligned} f_H(t) &= \Re \left[\int_{\mathbb{R}^m} \hat{H}(\theta) \exp(-2\pi i \langle \theta, t \rangle) d\theta \right] \\ &\geq \Re \left[\int_{\|\theta\|_\infty \leq \theta_0} \hat{H}(\theta) \exp(-2\pi i \langle \theta, t \rangle) d\theta \right] - \int_{\|\theta\|_\infty > \theta_0} |\hat{H}(\theta)| d\theta \\ &\geq \frac{1}{8} \left(\frac{1}{200\pi^3 n m p \sigma^2 \left(1 + \frac{\|\mu\|^2}{\sigma^2}\right)^{\frac{1}{m}}} \right)^{\frac{m}{2}} > 0. \end{aligned}$$

Now, recall that $H = D + R$, where $R \sim \mathcal{N}(0, \gamma \mathbf{I}_m)$. By (2.1), $n > \left(10m \left(1 + \frac{\|\mu\|}{\sigma}\right)\right)^{20}$. Hence, by applying Lemma A.2 again,

$$\Pr(\|R\| \geq \sqrt{\gamma n}) \leq e^{-\frac{n}{3}} \leq \frac{1}{2} f_H(t).$$

We conclude that with probability $1 - e^{-\Omega(\kappa n)}$ over A , there exists some $T \subset [n]$ and some $v \in \mathbb{R}^m$, with $\|v\| \leq \exp\left(-\frac{\kappa^3 pn}{80m}\right) \sigma n^2$, such that

$$A \mathbf{1}_T - v = t,$$

Or, in other words, $\|A \mathbf{1}_T - t\| = \|v\| \leq \exp\left(-\frac{\kappa^3 pn}{80m}\right) \sigma n^2$. Finally, we finish, as in the proof of the discrete case, with the multiplicative Chernoff bound:

$$\Pr[|S| \notin [0.5pn, 1.5pn]] = \Pr\left[||S| - \mathbb{E}[|S|]| \geq \frac{1}{2} \mathbb{E}[|S|]\right] \leq 2 \exp(-pn/12).$$

Again, since $\left(\left(1 + \frac{\|\mu\|}{\sigma}\right) 10m\right)^{20} \leq pn$, we have,

$$\Pr\left[\|D - t\| < \exp\left(-\frac{\kappa^3 pn}{80m}\right) \sigma n^2\right] > 2 \exp(-pn/12),$$

which implies, by a union bound, the existence of a suitable T with $|T| \in [0.5pn, 1.5pn]$ and

$$\|A \mathbf{1}_T - t\| \leq \exp\left(-\frac{\kappa^3 pn}{80m}\right) \sigma n^2.$$

□

3 Integrality Gap Bounds

3.1 Linear Programs and their Duals We begin with the basic linear programs relevant to this work. We will examine the integrality gap with respect to primal LP defined as follows:

$$\begin{aligned} \text{val}_{\text{LP}}(A, b, c) &:= \max_x \text{val}_c(x) = c^\top x \\ \text{(Primal LP)} \quad &\text{s.t. } Ax \leq b, x \in [0, 1]^n, \end{aligned}$$

This LP has the corresponding dual linear program, which we can express in the following convenient form:

$$\begin{aligned} \text{val}_{\text{LP}}^*(A, b, c) &:= \min_u \text{val}_b^*(u) = b^\top u + \left\| (c - A^\top u)^+ \right\|_1 \\ \text{(Dual LP)} \quad &\text{s.t. } u \geq 0. \end{aligned}$$

By strong duality, assuming (Primal LP) is bounded and feasible, we have that $\text{val}_{\text{LP}}(A, b, c) = \text{val}_{\text{LP}}^*(A, b, c)$.

For any primal solution x and dual solution u to the above pair of programs, we will make heavy use of the standard formula for the primal-dual gap:

$$\begin{aligned} \text{val}_b^*(u) - \text{val}_c(x) &= b^\top u + \left\| (c - A^\top u)^+ \right\|_1 - c^\top x \\ \text{(Gap Formula)} \quad &= (b - Ax)^\top u + (\langle x, (A^\top u - c)^+ \rangle + \langle \mathbf{1}_n - x, (c - A^\top u)^+ \rangle). \end{aligned}$$

In the sequel, we will let x^* denote the optimal solution to Primal LP and u^* denote the optimal solution to Dual LP. For all the LP distributions we work with, the objective c is continuously distributed (either Gaussian or exponentially distributed), from which it can be verified that conditioned on the feasibility of Primal LP (which depends only on A and b) both x^* and u^* are uniquely defined almost surely. Moreover, if $i \in [n]$, we shall use $A_{\cdot,i}$ to refer to the i^{th} column of A and extend this definition to other matrices as well.

Once the optimal solution is found for Primal LP, one can round its fractional coordinates to an integral vector. While the rounded vector may not be a feasible solution, we shall use the fact that, as long as the $A_{\cdot,i}$ are sufficiently bounded, it cannot be very far from a feasible solution.

LEMMA 3.1. ([9, LEMMA 7]) *There exists $x' \in \{0, 1\}^n$, such that,*

$$\|A(x^* - x')\| \leq \sqrt{m} \cdot \max_{i \in [n]} \|A_{\cdot,i}\|.$$

For the optimal solution x^* , define,

$$(3.6) \quad N_0 := \{i \in [n] | x_i^* = 0\}, \text{ and } N_1 := \{i \in [n] | x_i^* = 1\}.$$

Let W be the matrix with columns $W_{\cdot,i} = [c_i \ A_{\cdot,i}]^\top$. The distribution of the columns of W with indices in N_0 plays an important role in the proofs of Theorems 1.2 and 1.3. The following lemma essentially says that conditioning on the set N_0 and on the values of the non-0-columns preserves the mutual independence of the 0-columns. The conditional distribution of the the 0-columns is also identified. The reader is referred to [9, Lemma 5] for the proof.

LEMMA 3.2. *Let $N \subset [n]$. Conditional on $N_0 = N$ and on the values of sub-matrix $W_{\cdot,[n] \setminus N}$, x^* and u^* are almost surely well defined. Moreover, if $i \in N$, then $W_{\cdot,i}$ is independent from $W_{\cdot, N \setminus \{i\}}$ and the conditional law $W_{\cdot,i} | i \in N$ is the same as $W_{\cdot,i} | u^{*\top} A_{\cdot,i} - c_i > 0$.*

3.2 The Gap Bound for Centered IPs In this subsection we will prove Theorem 1.2. In the setting of Theorem 1.2, the objective $c \in \mathbb{R}^m$ has independent standard Gaussian entries, and the $m \times n$ constraint matrix A has independent columns which are distributed as either one of the following two possibilities:

- (LI) Isotropic logconcave distributions with support bounded by $O(\sqrt{\log n} + \sqrt{m})$.
- (DSU) Vectors with independent entries, uniform on a discrete symmetric interval of size $k \geq 3$.

To simplify the notation in the discrete case, we divide the constraint matrix A and the right hand side b by k (which clearly does not restrict generality). Thus, in the discrete case (DSU), we will assume that entries of A are uniformly distributed in $\{0, \pm 1/k, \dots, \pm 1\}$ and that the right hand side $b \in \mathbb{Z}^m/k$ satisfies $\|b^-\|_2 \leq O(n)$. In this way, the discrete case is usefully viewed as a discrete approximation of the continuous setting where the entries of A are uniformly distributed in $[-1, 1]$ (note that the covariance matrix of each column here is $I_m/3$, and thus essentially isotropic).

With the above setup, our goal is to show that $\text{IPGAP}(A, b, c) = O\left(\frac{\text{poly}(m)(\log n)^2}{n}\right)$ with probability $1 - n^{-\text{poly}(m)}$.

3.2.1 Properties of the Optimal Solutions To obtain the gap bound, we will need to show $|N_0| = \Omega(n)$ and that u^* , the optimal dual solution, has small norm. This is given by the following lemma, which is a technical adaptation of [9, Lemma 8].

LEMMA 3.3. *For $A \in \mathbb{R}^{m \times n}$, $n \geq 10^5 m$, distributed as (LI) or (DSU), $c \sim \mathcal{N}(0, I_m)$, $\|b^-\| \leq \frac{n}{12\sqrt{2}}$ with probability at least $1 - e^{-\Omega(n)}$, we have $\|u^*\| \leq 32$ and $|N_0| \geq \frac{n}{10^5}$.*

To prove this, we need two key lemmas. The first lemma will provide a good approximation for the value of any dual solution.

LEMMA 3.4. *Let $W^\top := (c, A^\top)$ where $c \sim \mathcal{N}(0, I_n)$ and $A \in \mathbb{R}^{m \times n}$ is distributed as (LI) or (DSU). Then, for $n = \Omega(m)$, we have that*

$$\Pr \left[\exists v \in \mathbb{S}^m : \|(v^\top W)^+\|_1 \notin \left[\frac{n}{12}, \frac{3n}{4} \right] \right] \leq e^{-\Omega(n)}.$$

Proof. [Proof] Fix $v \in \mathbb{S}^m$. We wish to understand $\Pr \left[\|(v^\top W)^+\|_1 \notin \left[\frac{n}{8}, \frac{5n}{8} \right] \right]$. Let $i \in [n]$, we first claim

$$(3.7) \quad \frac{1}{6} \leq \mathbb{E} [(v^\top W)_i^+] \leq \frac{1}{2}.$$

To see the right inequality, by Proposition A.1, $\mathbb{E} [(v^\top W)_i^+] = \frac{1}{2} \mathbb{E} [|v^\top W|_i^+]$. Now observe that every entry has variance at most 1, so with Jensen's inequality

$$\mathbb{E} [(v^\top W)_i^+] = \frac{1}{2} \mathbb{E} [|v^\top W|_i^+] \leq \frac{1}{2} \sqrt{\text{Var}(|v^\top W|_i^+)} \leq \frac{1}{2}.$$

For the left inequality, if the columns of W are isotropic log-concave (recall that the standard Gaussian is also log-concave) Lemma A.8 to get,

$$\frac{1}{6} \leq \frac{1}{2\sqrt{e}} \leq \frac{1}{2} \mathbb{E} [|v^\top W|_i^+] = \mathbb{E} [(v^\top W)_i^+].$$

If the columns of W are discrete, by Proposition A.2 every entry satisfies, $\text{Var}(W_{ij}) \geq \frac{1}{3}$. Hence, $\text{Var}((v^\top W)_i) \geq \frac{1}{3}$. Moreover, Proposition A.2 also implies, $\mathbb{E}[W_{ji}^4] \leq 3\mathbb{E}[W_{ji}^2]^2$. So, by Khinchine's inequality in Lemma A.11,

$$\frac{1}{6} \leq \frac{\sqrt{\text{Var}((v^\top W)_i)}}{2\sqrt{3}} \leq \frac{1}{2} \mathbb{E} [|v^\top W|_i^+] = \mathbb{E} [(v^\top W)_i^+].$$

Now, having established (3.7), we can bound $\Pr \left[\|(v^\top W)^+\|_1 \notin \left[\frac{n}{8}, \frac{5n}{8} \right] \right]$. In the log-concave case, Lemma A.10 immediately gives,

$$\Pr \left[\|(v^\top W)^+\|_1 \notin \left[\frac{n}{8}, \frac{5n}{8} \right] \right] \leq e^{-\Omega(n)}.$$

In the discrete case, by Lemma A.3, every entry of W is 1-sub-Gaussian, and Lemma A.4 shows that $(v^\top W)_i^+ - \mathbb{E} [(v^\top W)_i^+]$ is $\sqrt{2}$ -sub-Gaussian. After summing the coordinates we get that $\|(v^\top W)^+\|_1 - \mathbb{E} [\|(v^\top W)^+\|_1]$ is $\sqrt{2n}$ -sub-Gaussian. Applying (A.6), we can thus conclude a corresponding probability bound, as in the previous display.

We now turn to consider the entire sphere. Fix ε to be a small constant and let $N_\varepsilon \subset \mathbb{S}^{m-1}$ be an ε -net. It is standard to show that one may take $|N_\varepsilon| \leq \left(\frac{3}{\varepsilon}\right)^m$. Hence, by applying a union bound,

$$\Pr \left(\exists v \in N_\varepsilon : \|(v^\top W)^+\|_1 \notin \left[\frac{n}{8}, \frac{5n}{8} \right] \right) \leq \left(\frac{3}{\varepsilon}\right)^m e^{-\Omega(n)} \leq e^{-\Omega(n)},$$

where the last inequality holds when $n = \Omega(m)$.

Let us denote by E the event considered above and for $u \in \mathbb{S}^{m-1}$ let $\tilde{u} \in N_\varepsilon$, with $\|u - \tilde{u}\|_2 \leq \varepsilon$. Under E , we have,

$$\begin{aligned} \max_{u \in \mathbb{S}^{m-1}} \|(u^\top W)^+\|_1 &\leq \min_{v \in N_\varepsilon} \|(v^\top W)^+\|_1 + \|((u - \tilde{u})^\top W)^+\|_1 \\ &\leq \frac{5}{8}n + \varepsilon \max_{u \in \mathbb{S}^{m-1}} \|(u^\top W)^+\|_1, \end{aligned}$$

which is equivalent to,

$$\max_{u \in \mathbb{S}^{m-1}} \|(u^\top W)^+\|_1 \leq \frac{5}{8(1-\varepsilon)}n.$$

On the other hand,

$$\begin{aligned} \min_{u \in \mathbb{S}^{m-1}} \|(u^\top W)^+\|_1 &\geq \min_{v \in N_\varepsilon} \|(v^\top W)^+\|_1 - \|((u - \tilde{u})^\top W)^-\|_1 \\ &\geq \frac{n}{8} - \varepsilon \max_{u \in \mathbb{S}^{m-1}} \|(u^\top W)^+\|_1 \\ &\geq \frac{n}{8} - \varepsilon \frac{5}{8(1-\varepsilon)}n. \end{aligned}$$

Choose now $\varepsilon = \frac{5}{212}$ to conclude,

$$\frac{n}{12} \leq \min_{u \in \mathbb{S}^{m-1}} \|(u^\top W)^+\|_1 \leq \max_{u \in \mathbb{S}^{m-1}} \|(u^\top W)^+\|_1 \leq \frac{3n}{4}.$$

□

The second lemma will imply that any LP solution with large support must have small objective value.

LEMMA 3.5. *Let $c \sim \mathcal{N}(0, I_n)$. Then, for every $\alpha \in [0, 2\sqrt{\log(2)}]$,*

$$\Pr \left[\max_{x \in \{0,1\}^n, \|x\|_1 \geq \beta n} c^\top x \geq \alpha n \right] \leq e^{-\frac{\alpha^2 n}{2}},$$

where $\beta \in [1/2, 1]$ is such that $H(\beta) \leq \frac{\alpha^2}{4}$, where $H(p) = -p \log p - (1-p) \log(1-p)$, $p \in [0, 1]$, is base e entropy.

Proof. [Proof] For any $x \in \{0, 1\}^n$, $c^\top x \sim \mathcal{N}(0, \|x\|_2^2)$ and thus, by (A.6),

$$\Pr(c^\top x \geq \alpha n) \leq e^{-\frac{\alpha^2 n^2}{2\|x\|_2^2}} \leq e^{-\frac{\alpha^2 n}{2}}.$$

We now apply a union bound,

$$\Pr \left(\max_{x \in \{0,1\}^n, \|x\|_1 \geq \beta n} c^\top x \geq \alpha n \right) \leq |\{x \in \{0, 1\}^n, \|x\|_1 \geq \beta n\}| e^{-\frac{\alpha^2 n}{2}} \leq e^{H(\beta)n} e^{-\frac{\alpha^2 n}{2}} \leq e^{-\frac{\alpha^2 n}{4}}.$$

□

We now have the ingredients to prove the main lemma.

Proof. [Proof of Lemma 3.3] For the proof, we will consider the extended matrix $W^\top := (c, A^\top)$. We begin by showing that, for the optimal solution, $c^\top x^*$ is large. Let $u \geq 0$ be any dual solution. Then, under the complement of the event defined in Lemma 3.4 for W , using (Dual LP),

$$\begin{aligned} \text{val}_b^*(u) = b^\top u + \|(c - A^\top u)^+\|_1 &\geq -\|b^-\| \|u\| + \|((1, -u)^\top W)^+\|_1 \\ (3.8) \quad &\geq -\|b^-\| \|u\| + \sqrt{1 + \|u\|^2} \frac{n}{12} \geq \frac{n}{12} \left(-\frac{\|u\|}{\sqrt{2}} + \sqrt{1 + \|u\|^2} \right) \geq \frac{n}{12\sqrt{2}}. \end{aligned}$$

The second inequality is the lower bound in Lemma 3.4 and the last inequality follows since the function $\sqrt{1+t^2} - \frac{t}{\sqrt{2}}$ is minimized at $t = 1$. A lower bound on $c^\top x^*$ follows by noting,

$$c^\top x^* = \text{val}_c(x^*) = \text{val}_b^*(u^*).$$

We now prove that $\|u^*\|$ cannot be too large. Again, under the complement of the event in Lemma 3.4, but using the upper bound this time,

$$\begin{aligned} \frac{3n}{4} &\geq \|((1,0)^\top W)^+\|_1 = \|c^+\|_1 = \text{val}_b^*(0) \geq \text{val}_b^*(u^*) \\ &\geq \frac{n}{12} \left(-\frac{\|u^*\|}{\sqrt{2}} + \sqrt{1 + \|u^*\|^2} \right) \geq \frac{n}{12} \left(1 - \frac{1}{\sqrt{2}} \right) \|u^*\|, \end{aligned}$$

where in the third inequality we have applied (3.8) to v^* . Thus, rearranging we get $\|u^*\|_2 \leq \frac{9\sqrt{2}}{\sqrt{2}-1} \leq 32$. Finally, we show that the optimal solution has many 0 coordinates. Since x^* has at most m fractional coordinates,

$$|\{i \in [n] \mid x_i^* = 0\}| \geq n - m - |\{i \in [n] \mid x_i^* = 1\}|.$$

Since, by assumption, $n \geq 10^5 m$, to finish the proof it will suffice to show $|\{i \in [n] \mid x_i^* = 1\}| \leq \left(1 - \frac{2}{10^5}\right)n$. Define \bar{x} by,

$$\bar{x}_i := \begin{cases} x_i^* & \text{if } x_i^* \in \{0, 1\} \\ 1 & \text{if } x_i^* \notin \{0, 1\} \text{ and } c_i \geq 0. \\ 0 & \text{if } x_i^* \notin \{0, 1\} \text{ and } c_i < 0 \end{cases}$$

Letting $\alpha = \frac{1}{12\sqrt{2}}$ and $\beta = 1 - \frac{2}{10^5}$, a calculation reveals that $H(\beta) \leq \frac{1}{4\alpha^2}$. By (3.8), we have

$$c^\top \bar{x} \geq c^\top x^* \geq \alpha n,$$

and by conditioning on the complement of the event in Lemma 3.5 with β and α as above,

$$\beta n \geq |\{i \in [n] \mid \bar{x}_i = 1\}| \geq |\{i \in [n] \mid x_i^* = 1\}|.$$

The proof concludes by applying the union bound to the events in Lemmas 3.5 and 3.4. \square

3.2.2 Conditional Distribution of 0-columns of IP Let B be a random variable with the same distribution as the columns of A . By Lemma 2.1, B satisfies (anti-concentration) with constant $\kappa \leq 1$. Define $C := \frac{\sqrt{150}\|u^*\|}{\sqrt{\kappa}}$. We first show that the anti-concentration property is unaffected if we condition B on a strip of width $2C$.

LEMMA 3.6. *Let B' have the law of B , conditioned on $|u^{*\top} B| \leq C$. Then,*

1. $\Pr[|u^{*\top} B| \leq C] \geq 1 - \frac{\kappa}{150}$.
2. We have $\frac{1}{10}I_m \preceq \text{Cov}(B') \preceq 2I_m$.
3. If B is (DSU), then B' is symmetric and anti-concentrated with parameter $\kappa/2$.
4. If B is (LI), B' is logconcave.

Proof. [Proof] Let $E = \{a \in \mathbb{R}^m : |u^{*\top} a| \leq C\}$. From Chebyshev's inequality, and since distributions we consider satisfy $\text{Cov}(B) \preceq I_m$,

$$\Pr(B \in E) \geq 1 - \frac{\mathbb{E}[(u^{*\top} B)^2]}{C^2} \geq 1 - \frac{\|u^*\|^2}{C^2} \geq 1 - \frac{\kappa}{150},$$

which is the first claim.

If $w \in \mathbb{R}^m$, then

$$\mathbb{E}[|w^\top B'|^2] = \frac{\mathbb{E}[|w^\top B|^2 \mathbf{1}_E]}{\Pr(B \in E)} \leq 2\mathbb{E}[|w^\top B|^2] \leq 2\|w\|^2.$$

In the (DSU) case, to lower bound $\text{Cov}(B')$, we note that by Proposition A.2, $\mathbb{E}[W_{ji}^4] \leq 3\mathbb{E}[W_{ji}^2]^2$. As a consequence, Lemma A.11 implies $\sqrt{\mathbb{E}[|w^\top B|^4]} \leq \sqrt{3}\mathbb{E}[|w^\top B|^2]$. By the Cauchy-Schwarz inequality, $\mathbb{E}[|w^\top B \cdot \mathbf{1}_{B \notin E}|^2] \leq \sqrt{\mathbb{E}[|w^\top B|^4] \cdot \mathbb{E}[\mathbf{1}_{B \notin E}^4]} \leq \sqrt{3\Pr[B \notin E]}\mathbb{E}[|w^\top B|^2] \leq \sqrt{\frac{\kappa}{50}}\|w\|^2$. By Proposition A.2 we have $\mathbb{E}[|w^\top B|^2] \geq \frac{1}{3}\|w\|^2$. So:

$$\mathbb{E}[|w^\top B'|^2] = \frac{\mathbb{E}[|w^\top B|^2] - \mathbb{E}[|w^\top B|^2 \cdot \mathbf{1}_{B \notin E}]}{\Pr[B \in E]} \geq \frac{\frac{1}{3}\|w\|^2 - \sqrt{\frac{\kappa}{50}}\|w\|^2}{1 - \frac{\kappa}{50}} \geq \frac{1}{10}\|w\|^2,$$

proving the second claim for the (DSU) case.

In the (LI) case, $\langle w, B \rangle$ is logconcave. By Lemma A.5, $f_{(B,v)} \leq \frac{1}{\sqrt{\text{Var}\langle B,v \rangle}} = 1$. So, $f_{(v,B')} \leq \frac{1}{1-1/150} f_{(B,v)} = \frac{150}{149}$. Now,

Lemma A.1 implies that $\text{Var}(\langle B', v \rangle) \geq \frac{1}{13}$.

Now, if B is (DSU), it is symmetric, and because conditioning on a symmetric set preserves symmetry, so is B' . Consequently, in this case, $\mathbb{E}[B'] = 0$. Now set $\sigma' := \sqrt{\|\text{Cov}(B')\|_{\text{op}}} \leq \sqrt{2}$. For the third claim, let $I(\theta) = \{a \in \mathbb{R}^m : d(\theta^\top a, \mathbb{Z}) \geq \frac{\kappa}{2} \min(1, \sigma' \|\theta\|_\infty)\}$. Choose an arbitrary $v \in \mathbb{R}^n$. By the symmetry of B' we have:

$$\begin{aligned} \Pr[B' \in I(\theta) \mid \langle B, v \rangle \leq 0] &= \frac{\Pr[B \in I(\theta) \cap E \mid \langle B, v \rangle \leq 0]}{\Pr[B \in E \mid \langle B, v \rangle \leq 0]} \\ &\geq \Pr[B \in I(\theta) \mid \langle B, v \rangle \leq 0] - 2\Pr[B \notin E] \geq \frac{\kappa}{2}. \end{aligned}$$

The last inequality follows from (anti-concentration),

$$\Pr[B \in I(\theta) \mid \langle B, v \rangle \leq 0] \geq \Pr\left[d(\theta^\top B, \mathbb{Z}) \geq \frac{\kappa}{2} \min(1, \sigma' \|\theta\|_\infty) \mid \langle B, v \rangle \leq 0\right] \geq \kappa.$$

This shows anti-concentration when B is (DSU).

If B is (LI), then so is B' because it is a restriction to a convex set, proving the last claim.

□

In the proof of Theorem 1.2, we work with the columns of A that have negative reduced cost. We show that we can convert their distribution into the distribution of B' , by using rejection sampling.

LEMMA 3.7. *Let $B = A_{\cdot i}$ and let B', c' have $\text{Law}(B', c') = \text{Law}(B, c_i \mid u^{*\top} B - c_i \geq 0)$. When $\delta \leq C$, there exist a rejection sampling procedure ψ such that:*

- $\text{Law}(B' \mid \psi(B', c') = \text{accept}) = \text{Law}(B \mid |u^{*\top} A_{\cdot i}| \leq C)$.
- $\Pr[\psi(B', c') = \text{accept}] \geq \delta \exp\left(-\frac{10^4}{\kappa}\right)$.

Proof. [Proof]

Let c'' and B'' be independent random variables with $c'' \sim \text{Unif}(0, \delta)$ and let $\text{Law}(B'') = \text{Law}(B \mid |u^{*\top} B| \leq C)$. Now we will apply Lemma A.12 to transform (B', c') into (B'', c'') using rejection sampling. Observe that for any $\bar{B} \in \mathbb{R}^m$ with $u^{*\top} \bar{B} - \bar{c} \in [0, \delta]$:

$$\begin{aligned} \frac{f_{(B'', c'')}(\bar{B}, \bar{c})}{f_{(B', c')}(\bar{B}, \bar{c})} &= \frac{f_B(\bar{B}) / \Pr[|u^{*\top} B| \leq C] / \delta}{f_B(\bar{B}) f_{c_i}(\bar{c}) / \Pr[u^{*\top} B - c \geq 0]} = \frac{\Pr[u^{*\top} B - c \geq 0]}{\delta f_{c_i}(\bar{c}) \Pr[|u^{*\top} B| \leq C]} \\ &\leq \frac{1 - \frac{\kappa}{150}}{\delta \exp(-\frac{1}{2}(C + \delta)^2) / \sqrt{2\pi}} \leq \frac{3 \exp(2C^2)}{\delta}. \end{aligned}$$

For the last inequality we assumed that $\delta \leq C$. By Lemma A.12 we, using rejection sampling we can turn (B', c') into (B'', c'') with success probability $\frac{\delta}{3\exp(\frac{1}{2}(C+\delta)^2)}$. Since $C = \frac{\sqrt{2}\|u^*\|}{\sqrt{\kappa}}$, by Lemma 3.3, the success probability is at least $\frac{\delta}{3\exp(2C^2)} \leq \delta \exp\left(-\frac{10^4}{\kappa}\right)$. \square

When the columns of A are continuously distributed, we have to be more careful, because the distribution that we obtain from rejection sampling is not necessarily symmetric. As a result, B will not necessarily be mean-zero. We apply another step of rejection sampling to handle this case.

LEMMA 3.8. *Let B' denote the random variable $B \mid |u^{*\top} B| \leq C$. If B' is logconcave, then there exists a rejection sampling procedure ψ , such that $\Pr[\psi(B') = \text{accept}] = \Omega(1)$, and such that the random variable B'' with $\text{Law}(B'') = \text{Law}(B' \mid \psi(B') = \text{accept})$ satisfies:*

1. $\mathbb{E}[B''] = 0$.
2. $\text{Cov}(B'') \succeq \frac{1}{768} \mathbf{I}_m$.
3. The law of B'' is an admissible distribution, in the sense of Definition 1.
4. B'' satisfies (anti-concentration) with an $\Omega(1)$ constant.

Proof. [Proof] Let $\mu' := \mathbb{E}[B']$. By Hölder's inequality, we have any unit vector $v \in \mathbb{R}^m$,

$$\begin{aligned} \mathbb{E}[\langle B', v \rangle] &= \frac{\mathbb{E}[\langle B, v \rangle \cdot \mathbf{1}_{|u^{*\top} B| \leq C}]}{\Pr[|u^{*\top} B| \leq C]} = -\frac{\mathbb{E}[\langle B, v \rangle \cdot \mathbf{1}_{|u^{*\top} B| > C}]}{\Pr[|u^{*\top} B| \leq C]} \leq \frac{\sqrt{\mathbb{E}[\langle B, v \rangle^2]} \sqrt{\Pr[|\langle B, u^* \rangle| > C]}}{\Pr[|u^{*\top} B| \leq C]} \\ &\leq \frac{\sqrt{\frac{1}{150}}}{1 - 1/150} \leq \frac{1}{12} \end{aligned}$$

where the second equality follows since B is isotropic and from

$$\mathbb{E}[\langle B, v \rangle \cdot \mathbf{1}_{|u^{*\top} B| \leq C}] + \mathbb{E}[\langle B, v \rangle \cdot \mathbf{1}_{|u^{*\top} B| > C}] = \mathbb{E}[\langle B, v \rangle] = 0.$$

Recall $C \geq \sqrt{150}\|u^*\|$. Hence, the concentration bound in [49, Lemma 5.7], coupled with the fact that B is isotropic, gives,

$$\frac{\sqrt{\mathbb{E}[\langle B, v \rangle^2]} \sqrt{\Pr[|\langle B, u^* \rangle| > C]}}{\Pr[|u^{*\top} B| \leq C]} \leq \frac{\sqrt{\Pr[|\langle B, \frac{u^*}{\|u^*\|} \rangle| > \sqrt{150}]}}{\Pr[|u^{*\top} B| \leq C]} \leq 2e^{-5} \leq \frac{1}{12}.$$

So,

$$\|\mu'\| = \sup_{v \in \mathbb{R}^m, \|v\|=1} \mathbb{E}[\langle B', v \rangle] \leq \frac{1}{12}.$$

Define a convex subset of \mathbb{R}^m by

$$M := \left\{ \frac{\mathbb{E}[f(B')B']}{1/4} : f \in L_\infty(\mathbb{R}^m, \mathbb{R}), 0 \leq f(x) \leq 1 \text{ for every } x \in \mathbb{R}^m \text{ and } \mathbb{E}[f(B')] = \frac{1}{4} \right\},$$

and consider an arbitrary halfspace H that contains $-\mu'$. By Lemma A.6 we have,

$$\Pr[B' \in H] \geq \Pr[B \in H] - \Pr[|u^{*\top} B| > C] \geq \frac{1}{e} - \frac{1}{12} - \frac{1}{150} = \frac{1}{4}.$$

Therefore, there exists $S \subseteq H$ with $\Pr[B' \in S] = \frac{1}{4}$. Then, $\mathbb{E}[\mathbf{1}_S(B')] = \frac{1}{4}$ and, $\mathbb{E}[B' \mid B' \in S] = \mathbb{E}[\frac{\mathbf{1}_S(B')B'}{1/4}] \in M$, which implies that $M \cap H \neq \emptyset$. Because this holds for any H with $-\mu' \in H$, by the convexity of M we have $-\mu' \in M$. Indeed, suppose not,

then there is a hyperplane passing at $-\mu'$ which separates it from M , which cannot happen. We conclude that there exists $f: \mathbb{R}^m \rightarrow \mathbb{R}$, with $\|f\|_\infty \leq 1$ and $\mathbb{E}[f(B')] = \frac{1}{4}$, such that $\frac{\mathbb{E}[f(B')B']}{1/4} = -\mu'$. Let $g(x) = \frac{f(x)+4}{5}$.

Let B'' be a random variable with $f_{B''}(x) \propto g(x)f_{B'}(x)$. Note that $f_{B''}(x)/f_{B'}(x) = g(x)/\mathbb{E}_{B'}[g(B')] \leq 1/\mathbb{E}_{B'}[g(B')] \leq 2$. Now, by Lemma A.12 there exists a rejection sampling procedure ψ with $\Pr[\psi(B') = \text{accept}] = \frac{1}{2}$ and $\text{Law}(B' | \psi(B') = \text{accept}) = B''$. Now we will show that B'' satisfies the required properties.

Firstly, we have $\mathbb{E}[B''] = \frac{\mathbb{E}[g(B')B']}{\mathbb{E}[g(B')]} = \frac{\mathbb{E}[f(B')B'] + 4\mathbb{E}[B']}{5\mathbb{E}[g(B')]} = \frac{-4\mu' + 4\mu'}{5\mathbb{E}[g(B')]} = 0$, proving the first stated property. Secondly, for every halfspace H containing the origin, by the Grünbaum inequality, $\Pr[B \in H] \geq \frac{1}{e}$. So $\Pr[B' \in H] \geq \Pr[B \in H] - \Pr[|u^{*\top}B| > C] \geq \frac{1}{e} - \frac{1}{150} \geq \frac{1}{4}$. Because $f_{B''} \geq g \cdot f_{B'} \geq \frac{4}{5}f_{B'}$, we have $\Pr[B'' \in H] \geq \frac{4}{5} \cdot \Pr[B' \in H] \geq \frac{1}{5} \geq \frac{1}{4e^2}$. This proves that B'' has an admissible distribution.

For all unit vectors $v \in \mathbb{R}^m$ we have:

$$f_{\langle v, B'' \rangle}(x) \leq \frac{f_{\langle v, B' \rangle}(x)}{\Pr[\psi(B') = \text{accept}]} \leq 2f_{\langle v, B' \rangle}(x).$$

This implies $\text{Var}(\langle v, B'' \rangle) \preceq 2\text{Var}(\langle v, B' \rangle) \preceq 2I_m$. Because $\text{Cov}(B') \geq \frac{1}{10}I_m$ and the fact that $\langle B', v \rangle$ is logconcave, by Lemma A.5, $f_{\langle v, B' \rangle} \leq \frac{1}{\sqrt{\text{Var}(\langle v, B' \rangle)}} \leq 4$. Hence, $f_{\langle v, B'' \rangle} \leq 2 \cdot f_{\langle v, B' \rangle} \leq 768$. Now, Lemma A.1 implies that $\text{Var}(\langle B'', v \rangle) \geq \frac{1}{768}$. Hence $\frac{1}{768}I_m \preceq \text{Cov}(B'') \preceq 2I_m$. Now by Lemma 5.2 anti-concentration holds with a constant parameter. \square

3.2.3 Proof of Theorem 1.2

Proof. [Proof of Theorem 1.2] Consider the optimal solutions x^* and u^* to respectively (Primal LP) and (Dual LP). We condition on the event $|N_0| \geq n/10^5$ and $\|u^*\|_2 \leq 32$, where $N_0 := \{i \in [n] : x_i^* = 0\}$. By Lemma 3.3, this event occurs with probability at least $1 - e^{-\Omega(n)}$. Subject to this, we further condition on the exact values of x^* , u^* . We will show that for every such conditioning, the integrality gap is small with high probability over the randomness of A_{N_0} .

Set $\delta := \frac{\text{poly}(m) \log n}{n}$, where the polynomial factor is the same one as dictated by Theorem 2.1. We now show that we can construct a large subset $Z \subseteq N_0$, such that the reduced costs of the variables indexed by Z are small and the columns $A_{\cdot, i}, i \in Z$, are independent and satisfy the necessary conditions in order to apply Theorem 2.1 to round x^* to a near optimal solution. By Lemma 3.2, first note that $(c_i, A_{\cdot, i}), i \in N_0$ are independent and distributed according to $u^{*\top}A_{\cdot, i} - c_i > 0$.

By Lemma 3.7, using rejection sampling we can sample a set $Z \subseteq N_0$, such that $\text{Law}(A_{\cdot, i} | i \in Z) = \text{Law}(A_{\cdot, i} | |u^{*\top}A_{\cdot, i}| \leq C)$, $\Pr[i \in Z | i \in N_0] = \Omega(\delta)$ and such that $u^{*\top}A_{\cdot, i} - c_i \in [0, \delta]$ for all $i \in Z$. If columns of A have a discrete symmetric distribution (DSU), then $\mathbb{E}[A_{\cdot, i} | i \in Z] = 0$. Moreover, by Lemma 3.6, the distribution of the columns in Z is admissible (since it is symmetric), as per Definition 1, and satisfies the (anti-concentration) property with parameter $\kappa = \Omega(1)$. In the logconcave case, (LI), we apply a second round of rejection sampling to Z , as described in Lemma 3.8, which achieves that the law of $A_{\cdot, i}, i \in Z$ is mean-zero, admissible, and anti-concentrated with parameter $\kappa = \Omega(1)$. Furthermore, this second step of rejection sampling only decreases the probability that $i \in Z$ by at most a constant factor.

In both cases, we see that $\mathbb{E}[Z] \geq \Omega(\delta|N_0|) = \Omega(\text{poly}(m) \log n)$. Thus, by the Chernoff bound (A.3), $|Z| \geq \Omega(\text{poly}(m) \log n)$ with probability at $1 - n^{-\text{poly}(m)}$. We now condition on the exact set $Z \subseteq N_0$ subject to this size lower bound. Note that $A_{\cdot, i}, i \in Z$, are independent admissible, anti-concentrated with parameter $\kappa = \Omega(1)$ and mean-zero random vectors.

Set $p = \frac{\varepsilon \cdot \kappa^4}{1000m^5}$, where $\varepsilon > 0$ is chosen small enough to have $\kappa^3 \exp\left(\frac{\kappa^3}{3 \cdot 80^2 p m^5}\right) \geq 50000m^2$. We consider the rounded vector x' , from Lemma 3.1, and define the target, $t := A(x^* - x') - n^4 \exp(-p\kappa^3|Z|/m)\mathbf{1}_m$ in (LI) setting and $t := \lfloor kA(x^* - x') \rfloor / k$ in the (DSU) setting. We will now apply Theorem 2.1 to obtain a set $T \subseteq Z$ such that $\|\sum_{i \in T} A_i - t\|_2 \leq n^4 \exp(-\kappa^3 p|Z|/m)$ in the (LI) setting or $\sum_{i \in T} A_i = t$ in the (DSU) setting. This will help us both fix the slack introduced by the rounding as well as enforce that the resulting solution to be feasible.

We now invoke Lemma 3.1, which coupled with $|Z|p = \Omega(\text{poly}(m) \log(n))$ and the fact $\max_{i \in [n]} \|A_{\cdot, i}\| = O(\sqrt{\log(n)} + \sqrt{m})$, shows that, as long as the degree of the polynomial in δ is large enough,

$$\begin{aligned} \|t\| &\leq \|A(x^* - x')\| + m(n^4 \exp(-\kappa^3 p|Z|/m) + 1) \\ &\leq O(\sqrt{m \log n} + m) = o(\sqrt{p|Z|m}). \end{aligned}$$

Thus, for large n , Theorem 2.1 applies to the matrix $A_{\cdot, Z}$ and t in the (LI) setting and the matrix $kA_{\cdot, Z}, kt$ in the

(DSU) setting. Thus, with probability $1 - e^{-\Omega(p|Z|)} = 1 - n^{-\text{poly}(m)}$ there exists a set $T \subseteq Z$ such that $|T| \leq \frac{3}{2}p|Z|$, and $\|\sum_{i \in T} A_i - t\| \leq 32n^4 \exp(-\kappa^3 p|Z|/80m)$ in the (LI) setting and $\sum_{i \in T} A_i = t$ in the (DSU) setting.

Now we let $x'' = x' + \mathbf{1}_T$. We now show that

$$(3.9) \quad Ax'' \leq b \quad \text{and} \quad u^{*\top}(b - Ax'') \leq 1/\text{poly}(n).$$

Firstly, in the (DSU) setting, we have

$$Ax'' = Ax' + \lfloor k(Ax^* - Ax') \rfloor / k \leq Ax^* \leq b,$$

so x'' is a feasible integer solution. Take $j \in [m]$ such that $u_j^* > 0$. By complementary slackness we have that $(Ax^*)_j = b_j \in \mathbb{Z}/k$. Since $A \in \mathbb{Z}^{m \times n}/k$ and $x' \in \mathbb{Z}^n$, we have that $Ax' \in \mathbb{Z}^m/k$. In particular,

$$(Ax'')_j = (Ax')_j + \lfloor k(Ax^* - Ax') \rfloor_j / k = (Ax')_j + \lfloor k(b_j - Ax')_j \rfloor / k = (Ax')_j + k(b_j - Ax')_j / k = b_j.$$

We conclude that $u^{*\top}(b - Ax'') = 0$ as needed.

In the (LI) setting, we first note that

$$\|Ax'' - (Ax^* - n^4 \exp(-\kappa^3 p|Z|/m) \mathbf{1}_m)\| = \|\sum_{i \in T} A_i - t\| \leq n^4 \exp(-\kappa^3 p|Z|/m).$$

So, we must have $Ax'' \leq Ax^* \leq b$, and hence x'' is a feasible. Furthermore, by complementary slackness

$$\begin{aligned} u^{*\top}(b - Ax'') &= u^{*\top}(Ax^* - Ax'') \leq \|u^*\| \|Ax^* - Ax''\| \leq 32(\|Ax^* - t\| + \|t - Ax''\|) \\ &\leq 32(m+1)n^4 \exp(-\kappa^3 p|Z|/m) \leq 1/\text{poly}(n). \end{aligned}$$

To conclude, we use x'' to bound the integrality gap with the (Gap Formula) applied to x'' and u^* :

$$\begin{aligned} \text{IPGAP}(A, b, c) &= u^{*\top}(b - Ax'') + \left(\sum_{i=1}^n x''_i (A^\top u^* - c)_i^+ + (1 - x''_i) (c - A^\top u^*)_i^+ \right) \\ &= u^{*\top}(b - Ax'') + \sum_{i \in T} (A^\top u^* - c)_i \quad (\text{by complementary slackness}) \\ &\leq 1/\text{poly}(n) + |T| \cdot \delta \quad (\text{by (3.9) and } T \subseteq Z) \\ &\leq O\left(\frac{\text{poly}(m) \log(n)^2}{n}\right). \end{aligned}$$

□

3.3 The Gap Bound for Packing IPs In this section we will prove Theorem 1.3. Here, the objective $c \in \mathbb{R}^m$ has independent entries that are exponentially distributed with parameter $\lambda = 1$. The $m \times n$ constraint matrix A has independent columns which are distributed with (DU) independent entries which are uniform on the interval $\{1, \dots, k\}$, $k \geq 3$.

As in the centered case, we divide the constraint matrix A and the right hand side b by k . So, we will assume that the entries of A are uniformly distributed in $\{\frac{1}{k}, \dots, 1\}$ and that the right hand side b lies in $((n\beta, n(1/2 - \beta)) \cap \frac{\mathbb{Z}}{k})^m$. This way, we can see this setting as a discrete approximation of the continuous setting where the entries of A are uniformly distributed in $[0, 1]$, like in [12].

We want to show $\text{IPGAP}(A, b, c) \leq \frac{\exp(O(1/\beta)) \text{poly}(m)(\log n)^2}{n}$ with probability at least $1 - n^{-\text{poly}(m)}$. We will do this by first solving a slightly modified version of the LP-relaxation. We choose a $b' < b$. Now we let x^* be the minimizer of (Primal LP), where b is replaced by b' and let u^* be the optimal solution to the corresponding (Dual LP). We round down the solution, setting $x'_i := \lfloor x_i^* \rfloor$. Note that $\|A(x^* - x')\| \leq \sum_{i: x_i^* \in (0,1)} \|A_{\cdot, i}\| \leq m\sqrt{m}$.

Similar to the proof of Theorem 1.2, our proof proceeds by flipping x'_i to 1 for a subset of indices for which $x_i^* = 0$. By duality, these are columns with $A_{\cdot, i} - c_i \geq 0$. To be able to apply Theorem 1.4, we convert the conditional distribution of the columns of A back into their original distribution using rejection sampling:

LEMMA 3.9. *Let $B = A_{\cdot, i}$ and let B', c' have $\text{Law}(B', c'_i) = \text{Law}(B, c_i \mid u^{*\top} B - c_i \geq 0)$. When $\delta \leq 1$, here exists a rejection sampling procedure ψ such that:*

- $\Pr[\psi(B', c'_i) = \text{accept}] \geq \frac{1}{11} \delta \exp(-\|u^*\|_1)$.
- $\text{Law}(B' \mid \psi(B', c'_i) = \text{accept}) = \text{unif}(\left(\{\frac{1}{k}, \dots, 1\} \cap [\frac{1}{3m}, 1]^m\right))$.
- $u^{*\top} B' - c'_i \in [0, \delta]$ whenever $\psi(B', c'_i) = \text{accept}$.

Proof. [Proof] Let c'' and B'' be independent random variables with $c'' \sim \text{Unif}(0, \delta)$ and let $B'' \sim \text{Unif}(\left(\{\frac{1}{k}, \dots, 1\} \cap [\frac{1}{3m}, 1]^m\right))$. Note that $f_{B''}(x) \leq f_B(x)/(1 - \frac{1}{m})^m$. Now we will apply Lemma A.12 to transform (B', c') into (B'', c'') using rejection sampling. Observe that for any $\bar{B} \in \mathbb{R}^m, \bar{c} \in \mathbb{R}$ with $u^{*\top} B - c \in [0, \delta]$:

$$\begin{aligned} \frac{f_{(B'', c'')}(\bar{B}, \bar{c})}{f_{(B', c')}(\bar{B}, \bar{c})} &= \frac{f_{B''}(\bar{B})/\delta}{f_B(\bar{B})f_c(\bar{c})/\Pr_c[u^{*\top} B - \bar{c} \geq 0]} \leq \frac{f_B(\bar{B})/(1 - \frac{1}{m})^m/\delta}{f_B(\bar{B})f_c(\bar{c})} \leq \frac{4}{\delta f_c(\bar{c})} \\ &\leq \frac{4}{\delta \exp(-u^{*\top} B - \delta)} \leq \frac{11}{\delta \exp(-\|u^*\|_1)}. \end{aligned}$$

The stated result now follows directly by applying Lemma A.12 on (B', c'_i) to get a rejection sampling procedure that satisfies $\text{Law}((B', c'_i) \mid \psi(B', c'_i) = \text{accept}) = \text{Law}(B'', c'')$. \square

In the previous lemma, both the acceptance probability and the maximal size of δ depend on $\|u^*\|_1$. To prevent this from affecting the proof, we will show that with high probability $\Omega(\beta^4) \leq \|u^*\|_1 \leq O(\frac{1}{\beta})$. Because our proof of Theorem 1.3 will rely on flipping the columns for which $x_i^* = 0$, we will also show that with high probability the number of these columns is at least proportional to n .

LEMMA 3.10. *Consider the packing setting, with $\beta \in (0, 1/4)$ and $b' \in ((n\beta/2, n(1/2 - \beta)) \cap \frac{1}{k}\mathbb{Z})^m$. Then, with probability at least $1 - e^{-\Omega(\beta^2 n)}$, we have $\Omega(\beta^4) \leq \|u^*\|_1 \leq O(\frac{1}{\beta})$ and $|N_0| \geq \Omega(\beta^4 n)$.*

Proof. [Proof] Note that the distribution of the c_i 's is exponential and therefore logconcave with $\mathbb{E}[c_i] = 1$. By Lemma A.9 we now see that with probability $1 - e^{-\Omega(n)}$, we have $3n \geq \sum_{i=1}^n c_i$ and consequently,

$$3n \geq \sum_{i=1}^n c_i = c^\top \mathbf{1}_n \geq \text{val}_{\text{LP}}(x^*) = \text{val}^*(u^*) \geq \sum_{i=1}^m b'_i u_i^* \geq \frac{n \cdot \beta}{2} \|u^*\|_1.$$

Hence, we have $\|u^*\|_1 \leq \frac{6}{\beta}$, with high probability.

For the second claim, let $H : (0, \frac{1}{2}] \rightarrow (0, -\log(\frac{1}{2})]$ be defined with $H(x) = -x \log x - (1-x) \log(1-x)$. Set $\alpha := \min(\frac{1}{2}\beta, H^{-1}(\frac{1}{8}\beta^2))$. As $H(x) \leq 2\sqrt{x}$, we have $H^{-1}(x) \geq \frac{x^2}{4}$ and hence $\alpha \geq \frac{1}{256}\beta^4$. Let $x \in \{0, 1\}^n$ and suppose that $K := |\{i : x_i = 1\}| \geq (1 - \alpha)n$. By first using $b_1 \leq (\frac{1}{2} - \beta)n$, and $\mathbb{E}[(Ax)_1] = \frac{K}{2}$, and then applying Hoeffding's inequality we see,

$$\begin{aligned} \Pr[(Ax)_1 \leq b'_1] &\leq \Pr\left[(Ax)_1 \leq \frac{1 - \beta}{2} n\right] = \Pr\left[(Ax)_1 - \frac{1 - \alpha}{2} n \leq -\frac{\beta - \alpha}{2} n\right] \\ &\leq \Pr\left[(Ax)_1 - \frac{K}{2} \leq -\frac{\beta - \alpha}{2} n\right] = \Pr\left[(Ax)_1 - \mathbb{E}[(Ax)_1] \leq -\frac{\beta - \alpha}{2} n\right] \\ &\leq \exp(-(\beta - \alpha)^2 n) \leq \exp\left(-\frac{1}{4}\beta^2 n\right). \end{aligned}$$

Let $S = \{x \in \{0, 1\}^n : |\{i : x_i = 1\}| \geq (1 - \alpha)n\}$. Note that by Lemma A.13, $|S| \leq \sum_{i=0}^{\lfloor \alpha n \rfloor} \binom{n}{i} \leq \exp(H(\alpha)n)$. Taking the union bound over all $x \in S$, we see that

$$\Pr[\exists x \in S : (Ax)'_1 \leq b'_1] \leq |S| \exp(-\frac{1}{4}\beta^2 n) \leq \exp(H(\alpha)n - \frac{1}{4}\beta^2 n) \leq \exp(-\frac{1}{8}\beta^2 n).$$

So, with probability at least $1 - e^{-\Omega(\beta^2 n)}$ all feasible values $x \in \{0, 1\}^n$ have $|\{i : x_i = 0\}| \geq \alpha n$ and in particular $|N_0| \geq \alpha n \geq \frac{1}{256}\beta^4 n$.

At the same time, observe that when $i \in N_0$, we must have $c_i - u^{*\top} A_{\cdot,i} \leq 0$, so in particular $\|u^*\|_1 \geq c_i$. We have $\Pr[c_i \leq \log(\frac{1}{1-\alpha/2})] \leq 1 - \exp(-\log(\frac{1}{1-\alpha/2})) = \frac{1}{2}\alpha$. By the Chernoff bound (A.3) this implies that with probability at least $1 - \exp(-\Omega(n))$ we have $|\{i \in [n] : c_i \leq \log(\frac{1}{1-\alpha/2})\}| \leq \frac{3}{4}\alpha n$. If this event holds and at the same time we have $|N_0| \geq \alpha n$, then this implies $\|u^*\|_1 \geq \log(\frac{1}{1-\alpha/2})$ because otherwise $N_0 \subseteq \{i \in [n] : c_i \leq \log(\frac{1}{1-\alpha/2})\}$, contradicting the bounds on their size. So, we can conclude that with high probability we have $\|u^*\|_1 \geq \log(\frac{1}{1-\alpha/2}) \geq -\log(1 - 2^{-8}\beta^4) \geq 2^{-8}\beta^4$.

□

Proof. [Proof of Theorem 1.3]

Let $r = \lceil \frac{10^6 m^{12} \log(n)}{s^2} \rceil$, where s is a constant that we will choose later. Let $\mu = \frac{k + \lceil \frac{k}{3m} \rceil}{2k} = \mathbb{E}[U]$, where $U \sim \text{Uniform}(\{\frac{1}{k}, \dots, 1\} \cap [\frac{1}{3m}, 1])$. Now we define

$$(3.10) \quad \gamma = \frac{r}{1000m^5} \mu \text{ and } b' = b - \gamma \mathbf{1}.$$

Let x^* and u^* be the optimal solutions of (Primal LP) and (Dual LP) where b is replaced by b' . We will assume that $\frac{\beta^4}{C_1} \leq \|u^*\|_1 \leq \frac{C_1}{\beta}$ and $|N_0| \geq C_2 \cdot \beta^4 \cdot n$, for some constants C_1, C_2 . By Lemma 3.10 this happens probability $1 - e^{-\Omega(\beta^2 n)}$. Subject to this, we condition on the exact values of x^*, u^* .

Let $\delta := \frac{11 \exp(C_1/\beta) r}{C_2 \beta^4 n}$. Note that by our assumption that $n \geq \text{poly}(m) \exp(\Omega(1/\beta))$, we may assume that $\delta \leq \beta^4 / (C_1 m) \leq \frac{\|u^*\|_1}{3m}$. Thus, by Lemma 3.9 we can sample a set $Z \subseteq N_0$ such that for $\text{Law}(A_{\cdot,i} | i \in Z) = \text{Uniform}(\{\frac{1}{k}, \dots, 1\} \cap [\frac{1}{3m}, 1]^m)$ and that $\Pr[i \in Z | i \in N_0] = \frac{1}{|N_0|} \delta \exp(-\|u^*\|_1)$. Noting that $\mathbb{E}[|Z|] = 2r$, by Chernoff's inequality (A.3), with probability at least $1 - n^{-\text{poly}(m)}$, we have $|Z| \geq r$. Now we restrict Z to its first s elements, to get $|Z| = r$. Observe that $\mathbb{E}[A_{\cdot,i} | i \in Z] = \mu \mathbf{1}$.

We consider the target vector $t \in \mathbb{R}^m$, defined by:

$$t_i := \begin{cases} b_i - (Ax')_i : & u_i^* > 0 \\ \lfloor \gamma \rfloor : & \text{otherwise} \end{cases},$$

which satisfies $(b - Ax' - t)^\top u^* = 0$. Our next step will be to apply Theorem 1.4 on $kA_{\cdot,Z} \in \mathbb{Z}^{m \times r}$ and $kt \in \mathbb{Z}^m$ with parameter $p = \frac{\gamma}{\mu r} = \frac{1}{1000m^5}$, to get a set $T \subseteq Z$ such that $\sum_{i \in T} A_i = t$. Note that we have chosen γ and r to have $p^4 = \omega(\frac{m^3}{r})$.

To verify that t is indeed covered by Theorem 1.4, we first note that by Lemma 3.2 the columns $kA_{\cdot,i}$ for $i \in Z$ are independent with entries uniformly distributed in $\{\lceil k/m \rceil, \dots, k\}$. We now show that t is sufficiently close to the mean $|Z|p\mu$:

$$\begin{aligned} \|t - |Z|p\mu\| &= \|t - \gamma \mathbf{1}\| = \sqrt{\sum_{j: u_j^* > 0} (b_j - (Ax')_j)^2 + |\{j : u_j^* = 0\}| (\gamma - \lfloor \gamma \rfloor)^2} \\ &\leq \sqrt{\sum_{j: u_j^* > 0} (A(x^* - x')_j)^2 + |\{j : u_j^* = 0\}|} \leq \sqrt{\sum_{j: u_j^* > 0} \|x^* - x'\|_1^2 + |\{j : u_j^* = 0\}|} \\ &\leq m^{1.5} \leq \frac{s}{1000m^5} \sqrt{rm} = sp \sqrt{|Z|m} \leq s \sqrt{p|Z|m}. \end{aligned}$$

Now choose the constant s such that the previous inequality implies the condition from Theorem 1.4. As a result, with probability $1 - \exp(-p|Z|) \geq 1 - n^{-\text{poly}(m)}$ there exists a set $T \subseteq Z$, such that $\sum_{i \in T} A_i = t$.

Let $x'' = x' + \mathbf{1}_T$. Noting that x' was obtained from x^* , for i with $u_i^* = 0$ we have

$$(Ax'')_i = (Ax')_i + t_i \leq b'_i + \gamma = b_i.$$

For, i with $u_i^* > 0$ we have:

$$(Ax'')_i = (Ax')_i + t_i = b_i,$$

which means that x'' is a feasible solution to the integer program.

Using (Gap Formula) for x'' and u^* , we now get:

$$\begin{aligned}
 \text{IPGAP}(A, b, c) &= \text{val}_{\text{LP}}(A, b, c) - \text{val}_{\text{IP}}(A, b, c) \leq \text{val}_b^*(u^*) - \text{val}_c(x'') \\
 &= b^\top u^* + \sum_{i=1}^n (c - A^\top u^*)_i^+ - c^\top x'' \\
 &= (b - Ax'')^\top u^* + \left(\sum_{i=1}^n x''_i (A^\top u^* - c)_i^+ + (1 - x''_i)(c - A^\top u^*)_i^+ \right) \\
 &= (b - Ax' - t)^\top u^* + \sum_{i \in T} (A^\top u^* - c)_i \quad (\text{by complementary slackness}) \\
 &= \sum_{i \in T} (A^\top u^* - c)_i \quad (\text{since } (b - Ax' - t)^\top u^* = 0) \\
 &\leq \delta |T| \leq \frac{(\exp(C_1/\beta) \text{poly}(m) \log n)^2}{n}. \quad (\text{by Lemma 3.9})
 \end{aligned}$$

□

4 The Size of the Branch-and-Bound Tree

In this section, we prove Corollary 1.1 in the discrete setting. This will follow by adapting the proof of Theorem 1.1 from [9].

The first ingredient is the key theorem from [50, Theorem 3], which relates the Branch-and-Bound tree size to the size of a certain knapsack.

THEOREM 4.1. *Let $A \in \mathbb{R}^{m \times n}$, $b \in \mathbb{R}^m$, $c \in \mathbb{R}^n$. Then, the best bound first Branch-and-Bound algorithm applied to Primal IP with data A, b, c produces a tree of size*

$$(4.11) \quad 2n \cdot |\{x \in \{0, 1\}^n : \sum_{i=1}^n x_i |(A^\top u^* - c)_i| \leq \text{IPGAP}(A, b, c)\}| + 1,$$

where u^* is an optimal solution to (Dual LP).

Given the above, we must show how to upper bound the size of the knapsack in (4.11) in the setting of Corollary 1.1. Namely, when the entries of $c \in \mathbb{R}^n$ are either standard Gaussian or exponential and the entries of A are discrete.

For this purpose, we will require the following bound on the size on random knapsack polytopes.

LEMMA 4.1. ([9, LEMMA 11]) *Let $\omega_1, \dots, \omega_n \in \mathbb{R}$ be independent continuous random variables with maximum density at most 1. Then, for any $G \geq 0$, we have*

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

To get a bound on (4.11), we will use the above lemma for G being an upper bound on $\text{IPGAP}(A, b, c)$, using a union bound over all dual solutions of suitably bounded norm. In particular, we will require an upper bound on the norm of u^* . We note that the proof of Theorem 1.1 in [9] did not require a bound on the norm of the dual solution. In the logconcave setting, one could get around needing such a bound by using the anti-concentration properties of the columns of the objective extended constraint matrix $W^\top := (c, A^\top)$. In the present setting, we will only be able to rely on the anti-concentration properties of the objective c .

In both the centered and packing case, the coefficients of $c \in \mathbb{R}^n$ are independent and have maximum density at most 1. In particular, for any $u \in \mathbb{R}^m$, the entries of $(A^\top u - c)$ are also independent and have maximum density 1. By applying an appropriate union bound (e.g., see the proof of [9, Lemma 16]), one can derive the following bound over a family of knapsacks:

LEMMA 4.2. *Let $c \in \mathbb{R}^n$ have independent coordinates with maximum density at most 1 and let $A \in [-1, 1]^{m \times n}$ have independent entries. Then, for any $n \geq m, R \geq 2, G \geq 1/n$, we have that*

$$\Pr_{\|u\|_2 \leq R} \left[\max_{\|u\|_2 \leq R} |\{x \in \{0, 1\}^n : \sum_{i=1}^n x_i |(A^\top u - c)_i| \leq G\}| \geq (nR)^{\Theta(m)} e^{2\sqrt{2nG}} \right] \leq n^{-\Omega(m)}.$$

Proof. Let $K(u, G) := \{x \in \{0, 1\}^n : \sum_{i=1}^n x_i |(A^T u - c)_i| \leq G\}$. For $i \in [n]$, let N_i be an ε -net of $\{u \in \mathbb{R}^m : \|u\|_2 \leq R\}$ for $\varepsilon = \frac{1}{n^3}$. Note that we can choose N with $|N| = O(R^m n^{3m})$.

Lemma 4.1 implies that for any $u \in N'$, $\mathbb{E}[|K(u, G)|] \leq e^{2\sqrt{2nG}}$. So, by the Markov bound, we see:

$$\Pr[|K(u, G)| \geq n^m |N| e^{2\sqrt{2nG}}] \leq \frac{1}{n^m |N|}.$$

Taking the union bound over all $u' \in N$, we see:

$$\Pr[\exists u' \in N : |K(u, G)| \geq n^m |N| e^{2\sqrt{2nG}}] \leq |N| \frac{1}{n^m |N|} = n^{-m}$$

Now suppose that the event is true. Then for each u with $\|u\| \leq R$ there will be a $u' \in N$ with $\|u - u'\| \leq \varepsilon$. This implies that for all $x \in K(u, G)$ we have:

$$\begin{aligned} \sum_{i=1}^n x_i |(A^T u' - c)_i| &\leq \sum_{i=1}^n (x_i |(A^T u - c)_i| + x_i |(A^T (u' - u))_i|) \leq \sum_{i=1}^n x_i |(A^T u - c)_i| + \|A^T (u' - u)\|_1 \\ &\leq \sum_{i=1}^n x_i |(A^T u - c)_i| + \sqrt{mn} \varepsilon \leq G + \frac{1}{n}. \end{aligned}$$

So:

$$\Pr[\exists u' \in N : |K(u, G + \frac{1}{n})| \geq n^m |N| e^{2\sqrt{2nG}}] \leq n^{-m}.$$

Setting $G' := G - \frac{1}{n}$, we see:

$$\Pr[\exists u' \in N : |K(u, G')| \geq n^m |N| e^{2\sqrt{2nG'+2}}] \leq n^{-m}.$$

This proves the lemma because $n^m |N| e^{2\sqrt{2nG'+2}} \leq (nR)^{\Theta(m)} e^{2\sqrt{2nG'}}$. \square

We now have all the required ingredients to prove the bound on Branch-and-Bound trees in the discrete case.

Proof. [Proof of Corollary 1.1] As in section 3, in both the discrete centered and packing case, we divide the constraint matrix A by k . Thus, the entries of A are either uniform in $\{0, \pm 1/k, \pm 2/k, \dots, 1\}$ in the centered case or in $\{1/k, 2/k, \dots, 1\}$ in the packing case, and hence all contained in $[-1, 1]$.

Let $K = \{x \in \{0, 1\}^n : \sum_{i=1}^n x_i |(A^T u^* - c)_i| \leq \text{IPGAP}(A, b, c)\}$. By Theorem 4.1, to bound the size of the Branch-and-Bound tree, it suffices to prove a high probability upper bound on $|K|$.

In the centered case, by Theorem 1.2 we have that $\text{IPGAP}(A, b, c)$ is at most $G := \text{poly}(m)(\log n)^2/n$ with probability $1 - n^{-\text{poly}(m)}$, by Lemma 3.3 that $\|u^*\| \leq R := 32$ with probability $1 - e^{-\Omega(n)}$. Applying Lemma 4.2 with G, R together with the union bound, we conclude that $|K| \leq n^{\text{poly}(m)}$ with probability $1 - n^{-\Omega(m)}$, as needed.

In the packing case, by Theorem 1.3 the integrality gap $\text{IPGAP}(A, b, c)$ is upper bound by $G := \exp(O(1/\beta)) \text{poly}(m)(\log n)^2/n$ with probability $1 - n^{-\text{poly}(m)}$, and by Lemma 3.10 we have that $\|u^*\|_2 \leq \|u^*\|_1 \leq R := O(1/\beta)$ with probability $1 - e^{-\Omega(n)}$. Applying Lemma 4.2 with G, R together with the union bound, we conclude that $|K| \leq n^{\exp(O(1/\beta)) \text{poly}(m)}$ with probability $1 - n^{-\Omega(m)}$, as needed. \square

5 Anti-Concentration Results

Throughout this section we use the notation,

$$d(x, \mathbb{Z}) := \min_{z \in \mathbb{Z}} |x - z|.$$

Our goal in this section is to prove that Definition 2 is valid for a large family of distributions and prove Lemma 2.1.

Distributions with bounded densities: We begin with the following simple 1-dimensional lemma.

LEMMA 5.1. Let X be a random variable with $\mathbb{E}[X] = \mu$ and $\text{Var}(X) = \sigma^2$. Suppose that X has a density ρ , which satisfies,

$$\rho(x) \leq \frac{C}{\sigma},$$

for some $C > 0$. Then, for every $\varepsilon > 0$, then for $\delta = \frac{\varepsilon^2}{12C}$,

$$\Pr(d(X, \mathbb{Z}) \geq \delta \min(1, \sigma)) \geq 1 - \varepsilon.$$

Proof. By Chebyshev's inequality $\Pr(|X - \mu| \geq \frac{2}{\varepsilon}\sigma) \leq \frac{\varepsilon^2}{4}$. Define $\sigma' = \min(1, \sigma)$ and note that if $\mathbb{Z} + [-\delta\sigma', \delta\sigma'] := \bigcap_{z \in \mathbb{Z}} [z - \delta\sigma', z + \delta\sigma']$, then, for any $\delta > 0$,

$$\Pr\left(X \in \left[\mu - \frac{2}{\varepsilon}\sigma, \mu + \frac{2}{\varepsilon}\sigma\right] \cap (\mathbb{Z} + [-\delta\sigma', \delta\sigma'])\right) \leq \sum_{z \in \mathbb{Z}, |z - \mu| < \frac{2}{\varepsilon}\sigma - \delta\sigma'} \int_{z - \delta\sigma'}^{z + \delta\sigma'} \rho(x) dx.$$

If $\sigma \geq 1$, since $\rho(x) \leq \frac{C}{\sigma}$,

$$\sum_{z \in \mathbb{Z}, |z - \mu| < \frac{2}{\varepsilon}\sigma - \delta} \int_{z - \delta}^{z + \delta} \rho(x) dx \leq \frac{6}{\varepsilon} \sigma \cdot 2\delta \cdot \frac{C}{\sigma}.$$

We now choose $\delta = \frac{\varepsilon^2}{12C}$, so the right hand side becomes smaller than $\frac{\varepsilon}{2}$, and

$$\begin{aligned} \Pr(d(X, \mathbb{Z}) \geq \delta \min(1, \sigma)) \\ \geq \Pr\left(|X - \mu| < \frac{\varepsilon}{2}\sigma\right) - \Pr\left(X \in [\mu - \delta, \mu + \delta] \cap (\mathbb{Z} + [-\delta, \delta])\right) \geq 1 - \frac{\varepsilon^2}{4} - \frac{\varepsilon}{2} > 1 - \varepsilon. \end{aligned}$$

If $\sigma < 1$, then $\sigma' = \sigma$ and

$$\sum_{z \in \mathbb{Z}, |z - \mu| < \frac{2}{\varepsilon}\sigma - \delta\sigma} \int_{z - \delta\sigma}^{z + \delta\sigma} \rho(x) dx \leq \frac{6}{\varepsilon} \cdot 2\delta\sigma \cdot \frac{C}{\sigma}.$$

We then arrive at the same conclusion. \square

We now prove our anti-concentration result measures with an appropriate density bound.

LEMMA 5.2. Let X be a random vector in \mathbb{R}^m with $\Sigma := \text{Cov}(X)$. For $\theta \in \mathbb{R}^m$ let ρ_θ stand for the density of $\langle X, \theta \rangle$. Assume that there is a constant $C > 1$, satisfying the following three conditions:

- For every $\theta \in \mathbb{R}^m$, $x \in \mathbb{R}$, $\rho_\theta(x) \leq \frac{C}{\sqrt{\text{Var}(\langle X, \theta \rangle)}}$.
- For every $v \in \mathbb{R}^m$, $\Pr(\langle v, X \rangle \leq \langle v, \mu \rangle) \geq \frac{1}{C}$.
- $\|\Sigma\|_{\text{op}} \|\Sigma^{-1}\|_{\text{op}} \leq C$.

Then, for any $\theta, v \in \mathbb{R}^m$,

$$\Pr\left[d(\theta^\top X, \mathbb{Z}) \geq \frac{1}{48C^3} \min(1, \|\theta\|_\infty \sigma) \mid \langle v, X \rangle \leq \langle v, \mu \rangle\right] \geq \frac{1}{2C},$$

where $\sigma := \|\Sigma\|_{\text{op}}$. In other words, X satisfies (anti-concentration) with σ and $\kappa = \frac{1}{48C^3}$.

Before proving the result, we just note that by Lemmas A.6 and A.5, Lemma 5.2 applies to isotropic logconcave distributions and hence proves the first half of Lemma 2.1.

Proof. Let us denote $\eta^2 := \text{Var}(\theta^\top X)$ and observe

$$\frac{\|\theta\|^2}{\|\Sigma^{-1}\|_{\text{op}}} \leq \eta^2 \leq \|\theta\|^2 \|\Sigma\|_{\text{op}}.$$

With this in mind, we will actually show the seemingly stronger result,

$$\Pr \left[d(\theta^\top X, \mathbb{Z}) \geq \frac{\kappa}{C} \min(1, \sigma) \mid \langle \mathbf{v}, X \rangle \leq \langle \mathbf{v}, \mu \rangle \right] \geq \frac{\kappa}{C}.$$

However, since the class of measures considered by the lemma is preserved under rotations this is an actual equivalent statement.

By assumption, $\frac{\eta}{\sigma\|\theta\|} \geq \frac{1}{\sqrt{C}}$, and if we choose $\varepsilon = \frac{1}{2C}$ in Lemma 5.1, then

$$\Pr \left[d(\theta^\top X, \mathbb{Z}) \geq \frac{\delta}{\sqrt{C}} \min(1, \|\theta\|\sigma) \right] \geq 1 - \frac{1}{2C},$$

where $\delta = \frac{1}{48C^3}$. To complete the proof, by assumption \mathbf{v} ,

$$\Pr(\langle \mathbf{v}, X \rangle \leq \langle \mathbf{v}, \mu \rangle) \geq \frac{1}{C}.$$

Thus, with a union bound

$$\begin{aligned} & \Pr \left[d(\theta^\top X, \mathbb{Z}) \geq \frac{\delta}{\sqrt{C}} \min(1, \|\theta\|\sigma) \mid \langle \mathbf{v}, X \rangle \leq \langle \mathbf{v}, \mu \rangle \right] \\ &= \frac{\Pr \left[d(\theta^\top X, \mathbb{Z}) \geq \frac{\delta}{\sqrt{C}} \min(1, \|\theta\|\sigma) \text{ and } \langle \mathbf{v}, X \rangle \leq \langle \mathbf{v}, \mu \rangle \right]}{\Pr(\langle \mathbf{v}, X \rangle \leq \langle \mathbf{v}, \mu \rangle)} \\ &\geq \Pr \left[d(\theta^\top X, \mathbb{Z}) \geq \frac{\delta}{\sqrt{C}} \min(1, \|\theta\|\sigma) \right] + \Pr(\langle \mathbf{v}, X \rangle \leq \langle \mathbf{v}, \mu \rangle) - 1 \\ &\geq \left(\frac{1}{2C} + 1 - \frac{1}{C} \right) + \frac{1}{C} - 1 \geq \frac{1}{2C}. \end{aligned}$$

□

Discrete distributions: We now prove anti-concentration results for discrete distributions supported on \mathbb{Z}^m . Our result pertains to random variables which are uniform on intervals of length at least 3.

LEMMA 5.3. *Let $X = (X_1, \dots, X_m)$ be a random vector in \mathbb{R}^m , such that $\{X_i\}_{i=1}^m$ are i.i.d. uniformly on $\{a, a+1, \dots, a+k\}$, for some $a, k \in \mathbb{N}$, with $k > 1$. Set $\mu = \mathbb{E}[X_1]$ and $\sigma = \sqrt{\text{Var}(X_1)}$. Then, for every $\theta \in [-\frac{1}{2}, \frac{1}{2}]^m$, and every $\mathbf{v} \in \mathbb{R}^m$,*

$$\Pr \left(d(\theta^\top X, \mathbb{Z}) \geq \frac{1}{10} \min(\|\theta\|\sigma, 1) \mid \langle \mathbf{v}, X \rangle \leq \langle \mathbf{v}, \mu \mathbf{1} \rangle \right) \geq \frac{1}{40}.$$

In other words, X satisfies (anti-concentration) with σ and $\kappa = \frac{1}{40}$.

Proof. Observe that, as X is symmetric around its mean,

$$\begin{aligned} & \Pr \left(d(\theta^\top X, \mathbb{Z}) \geq \frac{1}{10} \min(\|\theta\|\sigma, 1) \mid \langle \mathbf{v}, X \rangle \leq \langle \mathbf{v}, \mu \mathbf{1} \rangle \right) \\ &= \frac{\Pr \left(d(\theta^\top X, \mathbb{Z}) \geq \frac{1}{10} \min(\|\theta\|\sigma, 1) \text{ and } \langle \mathbf{v}, X \rangle \leq \langle \mathbf{v}, \mu \mathbf{1} \rangle \right)}{\Pr(\langle \mathbf{v}, X \rangle \leq \langle \mathbf{v}, \mu \mathbf{1} \rangle)} \\ &\geq \Pr \left(d(\theta^\top X, \mathbb{Z}) \geq \frac{1}{10} \min(\|\theta\|\sigma, 1) \text{ and } \langle \mathbf{v}, X \rangle \leq \langle \mathbf{v}, \mu \mathbf{1} \rangle \right), \end{aligned}$$

With no loss of generality, let us assume $|\theta_m| = \|\theta\|_\infty$ and consider the event,

$$E = \left\{ \sum_{i=1}^{m-1} v_i X_i \leq \mu \sum_{i=1}^{m-1} v_i \text{ and } v_m X_m \leq \mu v_m \right\}.$$

Clearly, $E \subset \{\langle v, X \rangle \leq \langle v, \mu \mathbf{1} \rangle\}$, and by symmetry and independence, $\Pr(E) \geq \frac{1}{4}$. With the previous display,

$$\begin{aligned} \Pr\left(d(\theta^\top X, \mathbb{Z}) \geq \frac{1}{10} \min(\|\theta\|_\infty \sigma, 1) \mid \langle v, X \rangle \leq \langle v, \mu \rangle\right) \\ \geq \Pr\left(d(\theta^\top X, \mathbb{Z}) \geq \frac{1}{10} \min(\|\theta\|_\infty \sigma, 1), X \in E\right) \\ = \frac{1}{4} \Pr\left(d(\theta^\top X, \mathbb{Z}) \geq \frac{1}{10} \min(\|\theta\|_\infty \sigma, 1) \mid X \in E\right). \end{aligned}$$

Now, denote $r := \sum_{i=1}^{m-1} \theta_i X_i$, and rewrite,

$$d(\theta^\top X, \mathbb{Z}) = d(\theta_m X_m, (\mathbb{Z} - r)).$$

We observe that under the conditioning on E , depending on $\text{sign}(v_m)$, $\theta_m X_m$ is either uniform on $\{\theta_m a, \theta_m(a+1) \dots \theta_m \lfloor \mu \rfloor\}$, or on $\{\theta_m \lceil \mu \rceil \dots \theta_m(a+k)\}$. We are then interested in the size of the set

$$F = \{\theta_m \cdot x : d(\theta_m x, (\mathbb{Z} - r)) \geq \frac{1}{10} \min(\theta_m \sigma, 1) \text{ and } x \in \text{support}(X_m|E)\}.$$

Since $|\theta_m| \leq \frac{1}{2}$, $\sigma = \sqrt{\frac{k^2+2k}{12}}$ and $|\text{Support}(X_m|E)| \geq \lceil \frac{k+1}{2} \rceil$, it is not hard to see that, as long as $k \geq 2$,

$$(5.12) \quad \frac{|F|}{|\text{Support}(X_m|E)|} \geq \frac{1}{10}.$$

This can be done by inspecting the trajectory $a\theta_m + \{0, \theta_m, 2\theta_m, \dots, \lceil \frac{k+1}{2} \rceil \theta_m\} \bmod 1$ and noting that at most a $\frac{9}{10}$ fraction of the set $\{0, \theta_m, 2\theta_m, \dots, \lceil \frac{k+1}{2} \rceil \theta_m\} \bmod 1$ can occupy any interval of length $\frac{1}{5} \min(\theta_m \sigma, 1)$.

Indeed, let I be such an interval. If $\sigma_m > \frac{1}{5} \min(\theta_m \sigma, 1)$, then it cannot be the case that for some j , both $j\theta_m, (j+1)\theta_m \in I + \mathbb{Z}$. On the other hand, if $\theta_m \leq \frac{1}{5} \min(\theta_m \sigma, 1)$, then if, for some j , $j\theta_m \in I$, necessarily, $(j + \min(\frac{\sigma}{5}, \frac{1}{5\theta_m}))\theta_m \notin I$ and (5.12) follows since $\min(\frac{\sigma}{5}, \frac{1}{5\theta_m}) \leq \frac{4}{5}k$. Thus, by invoking the law of total probability on all possible values of r ,

$$\Pr\left(d(\theta^\top X, \mathbb{Z}) \geq \frac{1}{10} \min(\theta_m \sigma, 1) \mid X \in E\right) \geq \frac{|F|}{|\text{Support}(X_m|E)|} \geq \frac{1}{10}.$$

□

References

- [1] R Bixby and E Rothberg. “Progress in computational mixed integer programming—a look back from the other side of the tipping point”. In: *Annals of Operations Research* 149.1 (2007), p. 37.
- [2] T Achterberg and R Wunderling. “Mixed integer programming: Analyzing 12 years of progress”. In: *Facets of combinatorial optimization*. Springer, 2013, pp. 449–481.
- [3] MW Savelsbergh. “Preprocessing and probing techniques for mixed integer programming problems”. In: *ORSA Journal on Computing* 6.4 (1994), pp. 445–454.
- [4] M Conforti, G Cornuéjols, and G Zambelli. “Polyhedral approaches to mixed integer linear programming”. In: *50 years of integer programming 1958-2008*. Springer, 2010, pp. 343–385.
- [5] E Balas, S Ceria, and G Cornuéjols. “Mixed 0-1 programming by lift-and-project in a branch-and-cut framework”. In: *Management Science* 42.9 (1996), pp. 1229–1246.

- [6] M Fischetti and D Salvagnin. “Approximating the split closure”. In: *INFORMS Journal on Computing* 25.4 (2013), pp. 808–819.
- [7] M Fischetti, A Lodi, M Monaci, D Salvagnin, and A Tramontani. “Improving branch-and-cut performance by random sampling”. In: *Mathematical Programming Computation* 8.1 (2016), pp. 113–132.
- [8] SS Dey, Y Dubey, and M Molinaro. “Branch-and-Bound Solves Random Binary IPs in Polytime”. In: *Proceedings of the 2021 ACM-SIAM Symposium on Discrete Algorithms (SODA)*. Society for Industrial and Applied Mathematics, Jan. 2021, pp. 579–591. DOI: 10.1137/1.9781611976465.35.
- [9] S Borst, D Dadush, S Huiberts, and S Tiwari. “On the Integrality Gap of Binary Integer Programs with Gaussian Data”. In: *Mathematical Programming* (May 2022). DOI: 10.1007/s10107-022-01828-1.
- [10] LA Wolsey and GL Nemhauser. *Integer and combinatorial optimization*. Vol. 55. John Wiley & Sons, 1999.
- [11] M Conforti, G Cornuéjols, G Zambelli, et al. *Integer programming*. Vol. 271. Springer, 2014.
- [12] M Dyer and A Frieze. “Probabilistic Analysis of the Multidimensional Knapsack Problem”. en. In: *Mathematics of OR* 14.1 (Feb. 1989), pp. 162–176. DOI: 10.1287/moor.14.1.162.
- [13] A Goldberg and A Marchetti-Spaccamela. “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] GS Lueker. “On the Average Difference between the Solutions to Linear and Integer Knapsack Problems”. In: *Applied Probability-Computer Science: The Interface Volume 1*. Birkhäuser Boston, 1982, pp. 489–504. DOI: 10.1007/978-1-4612-5791-2.
- [15] M Dyer and A Frieze. “Probabilistic analysis of the generalised assignment problem”. In: *Math. Program.* 55.1-3 (Apr. 1992), pp. 169–181. DOI: 10.1007/bf01581197.
- [16] R Beier and B Vöcking. “Probabilistic analysis of knapsack core algorithms”. In: *Proceedings of the Fifteenth Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2004, New Orleans, Louisiana, USA, January 11-14, 2004*. Ed. by JI Munro. SIAM, 2004, pp. 468–477.
- [17] H Röglin and B Vöcking. “Smoothed analysis of integer programming”. In: *Math. Program.* 110.1 (Jan. 2007), pp. 21–56. DOI: 10.1007/s10107-006-0055-7.
- [18] A Frieze and GB Sorkin. “The probabilistic relationship between the assignment and asymmetric traveling salesman problems”. In: *SIAM Journal on Computing* 36.5 (2007), pp. 1435–1452.
- [19] AM Frieze. *On the expected efficiency of branch and bound for the asymmetric TSP*. Preprint available at . 2020.
- [20] L Lovász, J Spencer, and K Vesztegombi. “Discrepancy of set-systems and matrices”. In: *European Journal of Combinatorics* 7.2 (1986), pp. 151–160.
- [21] J Matousek. *Geometric discrepancy: An illustrated guide*. Vol. 18. Springer Science & Business Media, 1999.
- [22] B Chazelle. *The discrepancy method: randomness and complexity*. Cambridge University Press, 2000.
- [23] N Bansal. “Constructive algorithms for discrepancy minimization”. In: *2010 IEEE 51st Annual Symposium on Foundations of Computer Science*. IEEE, 2010, pp. 3–10.
- [24] S Lovett and R Meka. “Constructive discrepancy minimization by walking on the edges”. In: *SIAM Journal on Computing* 44.5 (2015), pp. 1573–1582.
- [25] T Rothvoss. “Constructive discrepancy minimization for convex sets”. In: *SIAM Journal on Computing* 46.1 (2017), pp. 224–234.
- [26] R Eldan and M Singh. “Efficient algorithms for discrepancy minimization in convex sets”. In: *arXiv preprint arXiv:1409.2913* (2014).
- [27] N Bansal, D Dadush, S Garg, and S Lovett. “The Gram-Schmidt walk: a cure for the Banaszczyk blues”. In: *Proceedings of the 50th Annual ACM SIGACT Symposium on Theory of Computing*. 2018, pp. 587–597.
- [28] LC Lau, R Ravi, and M Singh. *Iterative methods in combinatorial optimization*. Vol. 46. Cambridge University Press, 2011.

- [29] R Hoberg and T Rothvoss. “A logarithmic additive integrality gap for bin packing”. In: *Proceedings of the Twenty-Eighth Annual ACM-SIAM Symposium on Discrete Algorithms*. SIAM. 2017, pp. 2616–2625.
- [30] N Bansal. “On a generalization of iterated and randomized rounding”. In: *Proceedings of the 51st Annual ACM SIGACT Symposium on Theory of Computing*. 2019, pp. 1125–1135.
- [31] N Bansal, L Rohwedder, and O Svensson. “Flow Time Scheduling and Prefix Beck-Fiala”. In: *Proceedings of the 54th Annual ACM SIGACT Symposium on Theory of Computing*. STOC 2022. Rome, Italy, 2022, pp. 331–342.
- [32] KP Costello. “Balancing Gaussian vectors”. In: *Israel Journal of Mathematics* 172.1 (2009), pp. 145–156.
- [33] G Kuperberg, S Lovett, and R Peled. “Probabilistic existence of regular combinatorial structures”. In: *Geometric and Functional Analysis* 27.4 (2017), pp. 919–972.
- [34] E Ezra and S Lovett. “On the Beck-Fiala conjecture for random set systems”. In: *Random Structures & Algorithms* 54.4 (2019), pp. 665–675.
- [35] R Hoberg and T Rothvoss. “A fourier-analytic approach for the discrepancy of random set systems”. In: *Proceedings of the Thirtieth Annual ACM-SIAM Symposium on Discrete Algorithms*. SIAM. 2019, pp. 2547–2556.
- [36] C Franks and M Saks. “On the discrepancy of random matrices with many columns”. In: *Random Structures & Algorithms* 57.1 (2020), pp. 64–96.
- [37] A Potukuchi. “Discrepancy in random hypergraph models”. In: *arXiv preprint arXiv:1811.01491* (2018).
- [38] N Bansal and R Meka. “On the discrepancy of random low degree set systems”. In: *Random Structures & Algorithms* 57.3 (2020), pp. 695–705.
- [39] C MacRury, T Masařík, L Pai, and X Pérez-Giménez. “The Phase Transition of Discrepancy in Random Hypergraphs”. In: *arXiv preprint arXiv:2102.07342* (2021).
- [40] G Kuperberg, S Lovett, and R Peled. “Probabilistic existence of regular combinatorial structures”. In: *Geometric and Functional Analysis* 27.4 (2017), pp. 919–972.
- [41] D Lokshantov, N Narayanaswamy, V Raman, M Ramanujan, and S Saurabh. “Faster parameterized algorithms using linear programming”. In: *ACM Transactions on Algorithms (TALG)* 11.2 (2014), pp. 1–31.
- [42] Y Iwata, M Wahlstrom, and Y Yoshida. “Half-integrality, LP-branching, and FPT algorithms”. In: *SIAM Journal on Computing* 45.4 (2016), pp. 1377–1411.
- [43] M Wahlström. “LP-branching algorithms based on biased graphs”. In: *Proceedings of the Twenty-Eighth Annual ACM-SIAM Symposium on Discrete Algorithms*. SIAM. 2017, pp. 1559–1570.
- [44] AK Lenstra, HW Lenstra, and L Lovász. “Factoring polynomials with rational coefficients”. In: *Mathematische annalen* 261.ARTICLE (1982), pp. 515–534.
- [45] H Lenstra. “Integer Programming with a Fixed Number of Variables”. In: *Mathematics of OR* 8.4 (Nov. 1983), pp. 538–548. DOI: 10.1287/moor.8.4.538. eprint: <https://doi.org/10.1287/moor.8.4.538>.
- [46] ML Furst and R Kannan. “Succinct certificates for almost all subset sum problems”. In: *SIAM Journal on Computing* 18.3 (1989), pp. 550–558.
- [47] G Pataki, M Tural, and EB Wong. “Basis Reduction and the Complexity of Branch-and-Bound”. en. In: *Proceedings of the Twenty-First Annual ACM-SIAM Symposium on Discrete Algorithms*. Society for Industrial and Applied Mathematics, Jan. 2010, pp. 1254–1261. DOI: 10.1137/1.9781611973075.100.
- [48] JA Tropp. “An introduction to matrix concentration inequalities”. In: *arXiv preprint arXiv:1501.01571* (2015).
- [49] L Lovász and S Vempala. “The geometry of logconcave functions and sampling algorithms”. In: *Random Structures & Algorithms* 30.3 (2007), pp. 307–358.
- [50] SS Dey, Y Dubey, and M Molinaro. *Branch-and-Bound Solves Random Binary IPs in Polytime*. arXiv preprint arXiv:2007.15192 (v4). 2021.
- [51] EM Stein and G Weiss. *Introduction to Fourier analysis on Euclidean spaces*. Princeton Mathematical Series, No. 32. Princeton University Press, Princeton, N.J., 1971, pp. x+297.

- [52] B Doerr. “Analyzing Randomized Search Heuristics: Tools from Probability Theory”. In: *Series on Theoretical Computer Science*. WORLD SCIENTIFIC, Feb. 2011, pp. 1–20. DOI: 10.1142/9789814282673_0001. URL: https://doi.org/10.1142/9789814282673_0001.
- [53] B Laurent and P Massart. “Adaptive estimation of a quadratic functional by model selection”. In: *Ann. Statist.* 28.5 (2000), pp. 1302–1338. DOI: 10.1214/aos/1015957395. URL: <https://doi.org/10.1214/aos/1015957395>.
- [54] R Vershynin. *High-dimensional probability: An introduction with applications in data science*. Vol. 47. Cambridge university press, 2018.
- [55] A Prékopa. “Logarithmic concave measures with application to stochastic programming”. In: *Acta Scientiarum Mathematicarum* 32 (1971), pp. 301–316.
- [56] D Bertsimas and S Vempala. “Solving Convex Programs by Random Walks”. In: *Journal of the ACM* 51.4 (July 2004), pp. 540–556. DOI: 10.1145/1008731.1008733.
- [57] M Fradelizi. “Hyperplane sections of convex bodies in isotropic position”. In: *Beiträge Algebra Geom* 40.1 (1999), pp. 163–183.
- [58] D Galvin. “Three Tutorial Lectures on Entropy and Counting”. en. In: *arXiv:1406.7872 [math]* (June 2014). arXiv: 1406.7872 [math].

A Preliminaries

We begin this section by introducing some notation, to be used throughout the paper. If $(x_1, \dots, x_m) \in \mathbb{R}^m$ and $p \geq 1$, the p -norm is defined by,

$$\|x\|_p := \left(\sum_{i=1}^m |x_i|^p \right)^{\frac{1}{p}}.$$

When $p = 2$, i.e. the Euclidean norm, we will sometimes omit the subscript, so $\|x\| = \|x\|_2$. We interpret $p = \infty$ in the limiting sense,

$$\|x\|_\infty := \max_{1 \leq i \leq m} |x_i|.$$

We denote the positive part of x as, $x^+ := (\max(x_1, 0), \dots, \max(x_m, 0))$ and the negative part $x^- := (\max(-x_1, 0), \dots, \max(-x_m, 0))$. In this way, $x = x^+ - x^-$. The all-ones vector is denoted $\mathbf{1}_m := (1, \dots, 1)$. We will sometimes omit the subscript, when the dimension is clear from the context. If S is a subset of indices, we will write $\mathbf{1}_S$ for a vector such that $(\mathbf{1}_S)_i = 1$ if $i \in S$ and 0 otherwise. If a and b are quantities that depend on the problem’s parameters we will write $a = O(b)$ (resp. $a = \Omega(b)$) to mean $a \leq Cb$, (resp. $a \geq Cb$) for some numerical constant $C > 0$. We also write $a \ll b$ or $a = o(b)$ (resp. $a \gg b$ or $a = \omega(b)$) to mean $\lim_{a \rightarrow \infty} \frac{a}{b} = 0$ (resp. $\lim_{a \rightarrow \infty} \frac{b}{a} = 0$). The identity matrix in \mathbb{R}^m is denoted by I_m .

A.1 Fourier analysis Our main tool for proving the discrepancy result is Fourier analysis and we review here the necessary details. Fix $X \sim \mathcal{D}$, a random vector in \mathbb{R}^m . The Fourier transform of X (sometimes also called the characteristic function) is the complex-valued function defined by $\hat{X}(\theta) := \mathbb{E}[\exp(2\pi i \langle X, \theta \rangle)]$.

To understand the natural domain for θ , we first define the (dual) period of \mathcal{D} . For this, choose an arbitrary $a \in \text{support}(\mathcal{D})$ and denote,

$$(A.1) \quad \text{period}(\mathcal{D}) := \{v \in \mathbb{R}^m : \langle v, w - a \rangle \in \mathbb{Z}, \forall w \in \text{support}(\mathcal{D})\}.$$

The definition of $\text{period}(\mathcal{D})$ does not depend on the choice of a . It is readily seen that when \mathcal{D} is absolutely continuous with respect to the Lebesgue measure $\text{period}(\mathcal{D}) = \{0\}$, while $\text{period}(\mathcal{D}) = \mathbb{Z}^m$, when $\text{support}(\mathcal{D}) \subset \mathbb{Z}^m$. These are the cases on which we focus. Using the period, we define the fundamental domain of \mathcal{D} , in Fourier space (where we suppress the dependence on \mathcal{D}),

$$(A.2) \quad V := \{\theta \in \mathbb{R}^m : \|\theta\| \leq \inf_{0 \neq w \in \text{period}(\mathcal{D})} \|\theta - w\|\}.$$

Observe that if \mathcal{D} is absolutely continuous with respect to the Lebesgue measure, then $V = \mathbb{R}^m$ and when $\text{support}(\mathcal{D}) \subset \mathbb{Z}^m$, $V = [-\frac{1}{2}, \frac{1}{2}]^m$.

The connection between X and its Fourier transform comes from the Fourier inversion formula [51, Theorem 1.20]:

THEOREM A.1. (FOURIER INVERSION FORMULA) *Suppose that either \mathcal{D} is absolutely continuous, or $\text{support}(\mathcal{D}) \subset \mathbb{Z}^m$. Then, for $t \in \text{support}(\mathcal{D})$:*

$$\Pr[X = t] = \int_{\theta \in V} \hat{X}(\theta) \exp(-2\pi i \langle \theta, t \rangle) d\theta$$

If X is absolutely continuous, we interpret $\Pr[X = t]$ as the density of X at t .

Another desirable property of the Fourier transform is that it is particularly amenable to convolutions (this is clear from the exponential representation but see [51, Theorem 3.18]).

THEOREM A.2. (MULTIPLICATION-CONVOLUTION THEOREM) *Let X and Y be two independent random vectors. Then,*

$$\widehat{(X + Y)}(\theta) = \hat{X}(\theta) \hat{Y}(\theta).$$

A.2 Probability Distributions Let $X \in \mathbb{R}^m$ be a random vector distributed according to a probability measure ν on \mathbb{R}^m . We will use f_X to refer to the probability density function of X . We define the mean $\text{Mean}(\nu) := \mathbb{E}[X] \in \mathbb{R}^m$ and covariance matrix by $\text{Cov}(\nu) := \text{Cov}(X) := \mathbb{E}[XX^T] - \mathbb{E}[X] \mathbb{E}[X]^T \succeq 0$. If $X \in \mathbb{R}$ is a real random variable, we use the notation $\text{Var}[X]$ to write the variance instead of $\text{Cov}(X)$. We say that X , or its law ν , is isotropic if and $\mathbb{E}[X] = 0$ and $\text{Cov}(X) = I_m$.

PROPOSITION A.1. *Let $X \in \mathbb{R}^n$ satisfy $\mathbb{E}[X] = 0$. Then $\mathbb{E}[X^+] = \mathbb{E}[|X|]/2$.*

Proof. Note that $0 = \mathbb{E}[X] = \mathbb{E}[X^+ - X^-] \Rightarrow \mathbb{E}[X^+] = \mathbb{E}[X^-]$. Thus, $\mathbb{E}[|X|] = \mathbb{E}[X^+] + \mathbb{E}[X^-] = 2\mathbb{E}[X^+]$, as needed. \square

Let X_1, \dots, X_n be independent $\{0, 1\}$ random variables with $\mu = \mathbb{E}[\sum_{i=1}^n X_i]$. Then, the Chernoff bound gives [52, Corollary 1.10],

$$(A.3) \quad \begin{aligned} \Pr\left[\sum_{i=1}^n X_i \leq \mu(1 - \varepsilon)\right] &\leq e^{-\frac{\varepsilon^2 \mu}{2}}, \varepsilon \in [0, 1], \\ \Pr\left[\sum_{i=1}^n X_i \geq \mu(1 + \varepsilon)\right] &\leq e^{-\frac{\varepsilon^2 \mu}{3}}, \varepsilon \in [0, 1]. \end{aligned}$$

A more refined version is given by Azuma's inequality which allows the random variables to admit some mild dependencies. Let X_1, \dots, X_n be $\{0, 1\}$ random variables with $\mu = \sum_{i=1}^n \mathbb{E}[X_i | X_1, \dots, X_{i-1}]$. Then,

$$(A.4) \quad \Pr\left[\sum_{i=1}^n X_i \geq (1 + \varepsilon)\mu\right] \leq e^{-\frac{\varepsilon^2 \mu^2}{2n}}, \varepsilon \in [0, 1].$$

To see this bound, apply [52, Theorem 1.10.30] to the martingale $S_i := \sum_{j=1}^i X_j - \mathbb{E}[X_j | X_1, \dots, X_{j-1}]$.

LEMMA A.1. *If the density function f_X of X is bounded from above by M , then:*

$$\text{Var}(X) \geq \frac{1}{12M^2}.$$

Proof. If we want to minimize $\text{Var}(X) = \int_{-\infty}^{\infty} t^2 f_X(t) dt$ under the conditions $f_X \leq M$ and $\mathbb{E}[X] = 0$, then the unique minimizer is $f_X = M \cdot \mathbf{1}_{[-\frac{1}{2M}, \frac{1}{2M}]}$. We then see,

$$\text{Var}(X) = \int_{-\frac{1}{2M}}^{\frac{1}{2M}} t^2 \cdot M dt = \left[\frac{1}{3} M \cdot t^3 \right]_{-\frac{1}{2M}}^{\frac{1}{2M}} = \frac{1}{12M^2}$$

\square

A.3 Gaussian and Sub-Gaussian Random Variables If $\mu \in \mathbb{R}^m$ and $\Sigma \succ 0$ is an $m \times m$ positive-definite matrix, we denote by $\mathcal{N}(\mu, \Sigma)$, the law of the Gaussian with mean μ and covariance Σ . The probability density function of $\mathcal{N}(\mu, \Sigma)$ is given by $\frac{1}{\sqrt{2\pi}^m \det(\Sigma)^{1/2}} e^{-\frac{1}{2}(x-\mu)^\top \Sigma^{-1}(x-\mu)}$, $\forall x \in \mathbb{R}^m$.

The following is a basic concentration fact for the norm of the standard Gaussian (see [53, Lemma 1], for example).

LEMMA A.2. *Let $G \sim \mathcal{N}(0, I_m)$ and let $x \geq 7m$. Then,*

$$\Pr(\|G\|^2 \geq x) \leq e^{-\frac{x}{3}}.$$

A random variable $Y \in \mathbb{R}$ is σ -sub-Gaussian if for all $\lambda \in \mathbb{R}$, we have

$$(A.5) \quad \mathbb{E}[e^{\lambda Y}] \leq e^{\sigma^2 \lambda^2 / 2}.$$

A standard normal random variable $X \sim \mathcal{N}(0, 1)$ is 1-sub-Gaussian. If variables $Y_1, \dots, Y_k \in \mathbb{R}$ 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 [54, Proposition 2.5.2]:

$$(A.6) \quad \max\{\Pr[Y \leq -\sigma s], \Pr[Y \geq \sigma s]\} \leq e^{-\frac{s^2}{2}}, s \geq 0.$$

The following standard lemma shows that bounded random variables are sub-Gaussian.

LEMMA A.3. *Let $X \in [-1, 1]$ be a mean-zero random variable. Then X is 1-sub-Gaussian.*

Proof. Let $\varphi(x) := e^{\lambda x}$ for $\lambda \in \mathbb{R}$. By convexity of φ , note that for $x \in [-1, 1]$, $\varphi(x) \leq \frac{1-x}{2}\varphi(-1) + \frac{1+x}{2}\varphi(1)$. Therefore,

$$\begin{aligned} \mathbb{E}[\varphi(X)] &\leq \mathbb{E}\left[\frac{1-X}{2}\varphi(-1) + \frac{1+X}{2}\varphi(1)\right] = \frac{1}{2}(\varphi(-1) + \varphi(1)) = \frac{1}{2}(e^{-\lambda} + e^{\lambda}) \\ &= \sum_{i=0}^{\infty} \frac{\lambda^{2i}}{(2i)!} \leq \sum_{i=0}^{\infty} \frac{(\lambda^2/2)^i}{i!} = e^{\lambda^2/2}, \text{ as needed.} \end{aligned}$$

□

We also need the following fact about truncated sub-Gaussian random variables, which is a slight generalization of [9, Lemma 7]:

LEMMA A.4. *Let $X \in \mathbb{R}$ be 1-sub-Gaussian. Then $\mathbb{E}[X^+] \leq 1/2$ and $X^+ - \mathbb{E}[X^+]$ is $\sqrt{2}$ -sub-Gaussian.*

Proof. Since X is 1-sub-Gaussian, note that $\mathbb{E}[X] = 0$ and that $\mathbb{E}[X^2] \leq 1$. Therefore, by Proposition A.1, we have $\mu := \mathbb{E}[X^+] = \mathbb{E}[|X|]/2 \leq \mathbb{E}[X^2]^{1/2}/2 \leq 1/2$ by Hölder. $\sqrt{2}$ -sub-Gaussianity of $X^+ - \mu$ now follows verbatim from the proof of [9, Lemma 5] using that $\mu^2 \leq 1/3$ and replacing Gaussian by sub-Gaussian. □

A.3.1 Logconcave Measures If a measure ν has a density that is a logconcave function, we call ν logconcave. Logconcave distributions have many useful analytical properties. In particular, the marginals of logconcave random vectors are also logconcave.

THEOREM A.3. ([55]) *Let $X \in \mathbb{R}^d$ be a logconcave random vector. Then, for any surjective linear transformation $T : \mathbb{R}^d \rightarrow \mathbb{R}^k$, TX is a logconcave random vector.*

The following gives a (essentially tight) bound on the maximum density of any one dimensional logconcave \mathbb{R} in terms of the variance.

LEMMA A.5. ([49, LEMMA 5.5]) *Let $X \in \mathbb{R}$ be a logconcave variable. Then its density function is upper bounded by $\frac{1}{\sqrt{\text{Var}[X]}}$.*

The above has an important consequence. If $X \in \mathbb{R}^n$ is logconcave and isotropic, then for any vector $v \in \mathbb{R}^n \setminus \{0\}$, the random variable $v^\top X$ has maximum density at most $1/\sqrt{\text{Var}[v^\top X]} = 1/\|v\|_2$, where we have used $v^\top X$ is logconcave.

By a result of Grünbaum, the mean of logconcave measure is also an approximate median. In particular, any halfspace containing the mean has measure at least $1/e$. We will use the following generalization of this result.

LEMMA A.6. ([56]) Let $X \in \mathbb{R}^n$ be a logconcave measure with mean $\mathbb{E}[X] = \mu$. Then for any $\theta \in \mathbb{S}^{n-1}$ and $t \in \mathbb{R}$, $\Pr[\theta^\top X \geq \theta^\top \mu - t] \geq 1/e - |t|$.

We shall require the fact that logconcave random variables satisfy the following comparison inequality.

LEMMA A.7. ([57]) Let $X \in \mathbb{R}_+$ logconcave with $\mathbb{E}[X] = \mu$ and let Z have density e^{-x} (exponential distribution), $x \geq 0$. Then, for any convex function $\varphi : \mathbb{R}_+ \rightarrow \mathbb{R}$, $\mathbb{E}[\varphi(X)] \leq \mathbb{E}[\varphi(\mu Z)]$. In particular,

1. $\mathbb{E}[X^2] \leq 2\mu^2$.
2. $\mathbb{E}[e^{\lambda X}] \leq \frac{1}{1-\lambda\mu}$, $\lambda < 1/\mu$.

LEMMA A.8. For $X \in \mathbb{R}$ mean-zero and logconcave, we have $\mathbb{E}[X^2] \leq e\mathbb{E}[|X|]^2$.

Proof. Let $X_l := -X \mid X \leq 0$, $p_l = \Pr[X \leq 0]$ and $X_r := X \mid X \geq 0$, $p_r = \Pr[X \geq 0]$. Note that X_l, X_r are both non-negative logconcave random variables and that $p_l, p_r \geq 1/e$ by Lemma A.6. By Proposition A.1, $p_l \mathbb{E}[X_l] = \mathbb{E}[X^-] = \mathbb{E}[|X|]/2 = \mathbb{E}[X^+] = p_r \mathbb{E}[X_r]$. We now see that

$$\mathbb{E}[X^2] = p_l \mathbb{E}[X_l^2] + p_r \mathbb{E}[X_r^2] \stackrel{\text{Lemma A.7}}{\leq} 2(p_l \mathbb{E}[X_l]^2 + p_r \mathbb{E}[X_r]^2) = \mathbb{E}[|X|](\mathbb{E}[X_l] + \mathbb{E}[X_r]) \leq e\mathbb{E}[|X|]^2.$$

□

We will also require the following concentration inequality for sums of non-negative logconcave random variables.

LEMMA A.9. Let $X_1, \dots, X_n \in \mathbb{R}_+$ be independent non-negative logconcave random variables with mean μ . Then, the following holds:

1. $\Pr[\sum_{i=1}^n X_i \geq (1 + \varepsilon)\mu n] \leq e^{-n(\varepsilon - \ln(1 + \varepsilon))} \leq e^{-n(\frac{\varepsilon^2}{2(1 + \varepsilon)^2})}$, $\varepsilon > 0$.
2. $\Pr[\sum_{i=1}^n X_i \leq (1 - \varepsilon)\mu n] \leq e^{-n(-\ln(1 - \varepsilon) + \varepsilon)} = e^{-n(\sum_{j=2}^\infty \varepsilon^j / j)}$, $\varepsilon \in [0, 1]$.

Proof. By homogeneity, we assume wlog that $\mu = 1$.

Proof of 1. Let $\lambda := \frac{\varepsilon}{1 + \varepsilon}$. Then,

$$\Pr[\sum_{i=1}^n X_i \geq (1 + \varepsilon)n] \stackrel{\text{Markov}}{\leq} \mathbb{E}[e^{\lambda \sum_{i=1}^n X_i}] e^{-\lambda(1 + \varepsilon)n} \stackrel{\text{Lemma A.7}}{\leq} \left(\frac{1}{1 - \lambda}\right)^n e^{-\lambda(1 + \varepsilon)n} = e^{-n(\varepsilon - \ln(1 + \varepsilon))}.$$

Proof of 2. Let $\lambda := \frac{\varepsilon}{1 - \varepsilon}$. Then,

$$\Pr[\sum_{i=1}^n X_i \leq (1 - \varepsilon)n] \stackrel{\text{Markov}}{\leq} \mathbb{E}[e^{-\lambda \sum_{i=1}^n X_i}] e^{\lambda(1 - \varepsilon)n} \stackrel{\text{Lemma A.7}}{\leq} \left(\frac{1}{1 + \lambda}\right)^n e^{-\lambda(1 - \varepsilon)n} = e^{-n(-\ln(1 - \varepsilon) - \varepsilon)}.$$

□

Finally, we will require concentration of truncated sums.

LEMMA A.10. Let $X_1, \dots, X_n \in \mathbb{R}$ be i.i.d. mean zero logconcave random variables with $\mathbb{E}[X_1^+] = \alpha$. Then, for $\varepsilon \in [0, 1/2]$, we have that

1. $\Pr[\sum_{i=1}^n X_i^+ \geq (1 + \varepsilon)^2 n \alpha] \leq e^{-\frac{n\varepsilon^2}{3\varepsilon}} + e^{-\frac{n\varepsilon^2}{2\varepsilon(1 + \varepsilon)}}$.
2. $\Pr[\sum_{i=1}^n X_i^+ \leq (1 - \varepsilon)^2 n \alpha] \leq e^{-\frac{n\varepsilon^2}{2\varepsilon}} + e^{-\frac{n(1 - \varepsilon)\varepsilon^2}{2\varepsilon}}$.

Proof. Define $X'_1, \dots, X'_n \in \mathbb{R}_+$ to be i.i.d. copies of $X_1 \mid X_1 \geq 0$ and let $p = \Pr[X_1 \geq 0]$, where $1 - 1/e \geq p \geq 1/e$ (Lemma A.6). Let $C = \sum_{i=1}^n 1[X_i \geq 0]$, which a binomial distribution with parameters n and p . Since X_1, \dots, X_n are i.i.d., $\sum_{i=1}^n X_i^+$ has the same law as $\sum_{i=1}^C X'_i$. Therefore, by (A.3) and Lemma A.9, we have that

$$\begin{aligned} \Pr\left[\sum_{i=1}^C X'_i \geq (1 + \varepsilon)^2 n \alpha\right] &\leq \Pr[C \geq \lceil (1 + \varepsilon)pn \rceil] + \Pr\left[\sum_{i=1}^{\lfloor (1+\varepsilon)pn \rfloor} X'_i \geq (1 + \varepsilon)^2 n \alpha\right] \\ &\leq e^{-\frac{np\varepsilon^2}{3}} + e^{-\frac{(1+\varepsilon)np\varepsilon^2}{2(1+\varepsilon)^2}} \leq e^{-\frac{n\varepsilon^2}{3e}} + e^{-\frac{n\varepsilon^2}{e(1+\varepsilon)}}, \end{aligned}$$

and

$$\begin{aligned} \Pr\left[\sum_{i=1}^C X'_i \leq (1 - \varepsilon)^2 n \alpha\right] &\leq \Pr[C \leq \lfloor (1 - \varepsilon)pn \rfloor] + \Pr\left[\sum_{i=1}^{\lceil (1-\varepsilon)pn \rceil} X'_i \leq (1 - \varepsilon)^2 n \alpha\right] \\ &\leq e^{-\frac{np\varepsilon^2}{2}} + e^{-\frac{np(1-\varepsilon)\varepsilon^2}{2}} \leq e^{-\frac{n\varepsilon^2}{2e}} + e^{-\frac{n(1-\varepsilon)\varepsilon^2}{2e}}. \end{aligned}$$

□

A.4 Khinchine Inequality

LEMMA A.11. *Let $X_1, \dots, X_n \in \mathbb{R}$ be independent mean zero random variables satisfying $\mathbb{E}[X_i^4] \leq 3\mathbb{E}[X_i^2]^2 < \infty, \forall i$. Then, for any scalars $a_1, \dots, a_n \in \mathbb{R}$, we have that $\sqrt{\frac{1}{3}\mathbb{E}[|\sum_i a_i X_i|^4]} \leq \mathbb{E}[|\sum_i a_i X_i|^2] \leq 3\mathbb{E}[|\sum_i a_i X_i|^2]^2$.*

Proof. Letting $Z := |\sum_i a_i X_i| \geq 0$, we wish to show $\sqrt{\mathbb{E}[Z^4]} \leq \mathbb{E}[Z^2] \leq 3\mathbb{E}[Z^2]^2$. We first show that $\mathbb{E}[Z^4] \leq 3\mathbb{E}[Z^2]^2$, proving the first part of our claim:

$$\begin{aligned} \mathbb{E}[Z^4] &= \sum_{i,j,k,l} a_i a_j a_k a_l \mathbb{E}[X_i X_j X_k X_l] = \sum_i a_i^4 \mathbb{E}[X_i^4] + \sum_{i \neq j} 3a_i^2 a_j^2 \mathbb{E}[X_i^2] \mathbb{E}[X_j^2] \\ &= \sum_i a_i^4 \underbrace{(\mathbb{E}[X_i^4] - 3\mathbb{E}[X_i^2]^2)}_{\leq 0} + \sum_{i,j} 3a_i^2 a_j^2 \mathbb{E}[X_i^2] \mathbb{E}[X_j^2] \leq 3\left(\sum_i a_i^2 \mathbb{E}[X_i^2]\right)^2 = 3\mathbb{E}[Z^2]^2. \end{aligned}$$

As a consequence, we have

$$\mathbb{E}[Z^2] = \mathbb{E}[Z^{2/3}(Z^4)^{1/3}] \underbrace{\leq}_{\text{H\"older}} \mathbb{E}[Z^{2/3}] \mathbb{E}[Z^4]^{1/3} \underbrace{\leq}_{\mathbb{E}[Z^4] \leq 3\mathbb{E}[Z^2]^2} 3^{1/3} \mathbb{E}[Z^{2/3}] \mathbb{E}[Z^2]^{2/3},$$

which, after rearranging, gives $\mathbb{E}[|\sum_i a_i X_i|^2] \leq 3\mathbb{E}[|\sum_i a_i X_i|^2]^2$. □

A.4.1 Discrete Random Variables

Here we list some of the moments of (DSU) for reference.

PROPOSITION A.2. (DISCRETE SYMMETRIC MOMENTS) *For $k \geq 1$, let U be uniformly distributed on $\{0, \pm 1/k, \dots, \pm 1\}$. Then, $\mathbb{E}[U^2] = \frac{k+1}{3k} \geq 1/3$, $\mathbb{E}[U^4] = \frac{(k+1)(3k^2+3k-1)}{15k^3}$ and $\mathbb{E}[U^4]/\mathbb{E}[U^2]^2 = \frac{9(3k^2+3k-1)}{15(k+1)k} \leq 2$.*

A.5 Rejection sampling In the proofs of Theorems 1.2 and 1.3 we make use of a tool called rejection sampling. This tool allows us to change the distribution of a random variable $X \sim \mathcal{D}$. It works by ‘accepting’ an evaluation of the variable with a probability that depends on its value. Let us formally define a rejection sampling procedure first.

DEFINITION 3. *A rejection sampling procedure on a random variable X is a randomized algorithm ψ that, given a realization of X , outputs either reject or accept.*

Now the variable X conditioned on $\psi(X) = \text{accept}$, will be distributed according to a probability \mathcal{D}' . We can choose the distribution \mathcal{D}' by appropriately setting $\Pr[\psi(X) = \text{accept} \mid X]$. This is formalized in the following lemma.

LEMMA A.12. *Let X, Y be a random variable on some set S , with probability density functions f_X and f_Y , and $\text{Support}(Y) \subseteq \text{Support}(X)$. Suppose that $\frac{f_Y(s)}{f_X(s)} \leq K$ for all $s \in S$. Then there exists a rejection sampling procedure ψ with $\text{Law}(X \mid \psi(X) = \text{accept}) = \text{Law}(Y)$ and $\Pr[\psi(X) = \text{accept}] = \frac{1}{K}$.*

Proof. Define ψ such that:

$$\Pr[\psi(X) = \text{accept} | X = x] = \begin{cases} \frac{f_Y(x)}{K f_X(x)} & x \in \text{Support}(X) \\ 0 & \text{else} \end{cases}.$$

This probability is well defined as $\frac{f_Y(x)}{K f_X(x)} \leq \frac{K}{K} = 1$ for all $x \in \text{Support}(X)$. Now $\Pr[\psi(X) = \text{accept}] = \int_{x \in \text{Support}(X)} \frac{f_Y(x)}{K f_X(x)} f_X(x) dx = \frac{1}{K}$. Call the variable X conditioned ($\psi(X) = \text{accept}$), Z . Now:

$$f_Z(z) = \frac{\frac{f_Y(x)}{K f_X(x)} f_X(x)}{\Pr[\psi(X) = \text{accept}]} = \frac{f_Y(x)}{K \cdot 1/K} = f_Y(x).$$

This proves the statement. \square

A.6 Combinatorics Theorems 1.2 and 1.3 rely on some properties of the optimal dual solution, that hold with high probability. We prove these by taking the union bound over all vectors in $\{0, 1\}^n$ that contain at most αn zeroes. By applying the following lemma, we are able to upper bound the number of these vectors by $\exp(H(\alpha)n)$ where H is the entropy function defined as $H(x) = -x \log(x) - (1-x) \log(1-x)$.

LEMMA A.13. ([58, THEOREM 3.1]) *For all $\alpha \leq \frac{1}{2}$ and all n ,*

$$\sum_{i=0}^{\lfloor \alpha n \rfloor} \binom{n}{i} \leq \exp(H(\alpha)n).$$