Integrality Gaps for Random Integer Programs via Discrepancy

Sander Borst*[1], Daniel Dadush*[1], and Dan Mikulincer†[2]

[1]*Centrum Wiskunde & Informatica (CWI), Amsterdam, The Netherlands,*
{sander.borst,dadush}@cwi.nl
[2]*MIT, Cambridge, USA*
danmiku@mit.edu

March 23, 2022

Abstract

We give 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 range of distributions on (A,b,c). Dyer and Frieze (MOR '89) and Borst et al (IPCO '21) respectively, showed that for 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 $O_m(s\log^2 n/n)$ with probability at least $1-1/n-e^{-\Omega_m(s)}$ for $s \geq 1$. In this paper, we extend these results to the case where A is discretely distributed (e.g., entries $\{-1,0,1\}$), and to the case where the columns of A have a logconcave distribution. Second, we improve the success probability from constant, for fixed s and m, to $1-1/\operatorname{poly}(n)$. Using a connection between integrality gaps and Branch-and-Bound due to Dey, Dubey, and Molinaro (SODA '21), our gap results imply that Branch-and-Bound is polynomial for these IPs.

Our main technical contribution and the key for achieving the above results is a new discrepancy theoretic theorem which gives general conditions for when a target t is equal 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. We prove this lemma using a Fourier analytic approach, building on the work of Hoberg and Rothvoss (SODA '19) and Franks and Saks (RSA '20), who studied similar questions for $\{-1,1\}$ combinations.

*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).
†Dan Mikulincer was partially supported by a European Research Council grant no. 803084.

1 Introduction

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

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

In practice, one can observe that a key 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

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

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

Complexity of Branch-and-Bound. Recently, Dey, Dubey, and Molinaro [1] gave theoretical support for this observation, in the context of solving *random packing* IPs via Branch-and-Bound. We state a generalization of their result due to [2] to the case of random logconcave IPs below:

Theorem 1 ([1, 2]). *Let $n \ge \Omega(m)$, $b \in \mathbb{R}^m$, 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 $G \ge 0$, with probability at least $1 - \Pr_{A,c}[\mathsf{IPGAP}(A,b,c) \ge G] - 1/\mathrm{poly}(n)$, the best bound first Branch-and-Bound algorithm applied to* (Primal IP) *produces a tree of size at most $n^{O(m)} e^{2\sqrt{2nG}}$.*

Note that the class of logconcave distributions is quite rich and includes, for example, the uniform distribution on any convex body. From the above, we see that probabilistic bounds for the integrality gap of random logconcave integer programs immediately imply corresponding bounds for the complexity of Branch-and-Bound. This directly motivates the question: which classes of distributions admit small integrality gaps? Furthermore, given that realistic IPs often have discrete coefficients, to what extent do the above results extend to discrete distributions?

Gap Bounds. For models directly captured by the above theorem, suitable integrality have been proven for random packing [3, 4, 5] and Gaussian IPs [6], where the entries of (A,c) are either independent uniform $[0,1]$ or $\mathcal{N}(0,1)$. For random packing IPs, when the entries of $b \in n(0,1/2)^m$, Dyer and Frieze [3] proved that $\mathsf{IPGAP}(A,b,c) \le 2^{O(m)} \log^2(n)/n$ with probability at least $1 - 1/\mathrm{poly}(n) - 2^{-\mathrm{poly}(m)}$. For Gaussian IPs, when $\|b^-\|_2 \le n/10$, Borst et al [6] proved that $\mathsf{IPGAP}(A,b,c) \le \mathrm{poly}(m)\log^2(n)/n$ with probability at least $1 - 1/\mathrm{poly}(n) - 2^{-\mathrm{poly}(m)}$. For these models, the results imply that Branch-and-Bound is polynomial for fixed m with reasonable probability.

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 Branch-and-Bound trees. In particular, Frieze and Sorkin [7] 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. Very recently however, Frieze [8] showed that any Branch-and-Bound tree using the cycle cover relaxation to prune nodes[1] will have expected size $\Omega(2^{n^\alpha})$, for some $0 < \alpha < 1/20$. Interestingly, the number of variables in ATSP is $n(n-1)$, and thus a non-rigorous extrapolation from Theorem 1 may suggest a non-trivial $n^{O(\sqrt{n})}$ upper bound on the tree size of ATSP with good probability. Proving such a sub-exponential upper bound for a Branch-and-Bound algorithm for random ATSP remains an interesting open problem.

To explain the difficulties in extending the gap bounds for random packing and Gaussian IPs, we outline the high level strategy for obtaining these bounds. The basic strategy is to bound the gap by suitably rounding an optimal basic primal solution x^* and dual solution u^* to the LP relaxation (for possibly a slightly perturbed right hand side b) to a nearly optimal solution to the IP. To do this, one first rounds down the at most m fractional components of x^* to get x' (note that m is fixed). Afterwards, one selects a small subset of variables $T \subseteq \{i : x_i^* = 0\}$, $|T| = O_m(\log n)$, with tiny reduced costs, namely $i \in [n]$ with $|c_i - u^* a_i| = O_m(\log n/n)$. This subset is carefully chosen such that flipping the contained variables from 0 to 1 allows us to "fix" the slacks, i.e., such that $Ax^* \approx A(x' + 1_T)$ (in particular, to ensure feasibility). A crucial property in both cases is that x^* has at least $\Omega(n)$ zeros and that the corresponding columns of A are independent subject to having negative reduced cost (see Lemma 22 and Lemma 23). This ensures that one has a large universe of independent columns to choose from when constructing T. To extend these results to more general distributions, at least with respect to the constraint matrix A, a major difficulty was the lack of a general enough condition that ensures the existence of a scheme to properly fix the slacks.

The Discrepancy Problem. Stated abstractly, it is a very natural discrepancy theoretic problem: given a target t (e.g., the difference in the slack vectors) and a random matrix A with "nicely" distributed independent columns, when can we ensure, with high probability, that t is equal or very close to a $\{0,1\}$ combination of the columns of A? The answer given in [6, Lemma 1], improving upon [3, Lemma 3.4], required rather strict conditions on the entries of A, namely that they 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 theoretic lemma. We state it below, restricted to the relevant slightly simplified cases for our applications (see Theorem 8 for the general result).

Theorem 2 (Linear Discrepancy Theorem for Random Matrices)**.**

- *Discrete case: Let $A \in \mathbb{Z}^{m\times\bar{n}}$, $\bar{n} \geq \mathrm{poly}(m)$, where A_{ij} is distributed uniformly and independently on $\{a, a+1, \ldots, a+k\}$ with $k > 1$. Take $p \in [0,1]$ such that $p \leq \frac{s}{100m^5}$, for some small constant $s > 0$, and $p^4 = \omega(\frac{m^3}{\bar{n}})$. Then with probability at least $1 - e^{-\Omega(p\bar{n})}$, $\forall t \in \mathbb{Z}^m$ satisfying $\|t - \bar{n}p(a+k/2)\mathbf{1}_m\|_2 \leq skp\sqrt{m\bar{n}}$, $\exists x \in \{0,1\}^{\bar{n}}$ with $\|x\|_1 = \Theta(p\bar{n})$ such that $Ax = t$.*

[1] The 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). Branch-and-Bound must therefore branch on variables that are integral in the current relaxation, which is somewhat non-standard.

- *Continuous case: Let $A = (a_1, \dots, a_{\bar{n}}) \in \mathbb{R}^{m \times \bar{n}}$, $\bar{n} \geq \operatorname{poly}(m)$, where $a_1, \dots, a_{\bar{n}}$ are independent, logconcave random vectors with $\mathbb{E}[a_i] = \mu$ and $\operatorname{Cov}(a_i) = I_m$, $\forall i \in [\bar{n}]$, and $\|\mu\| = O(e^m)$. Take $p \in [0,1]$ such that $p \leq \frac{s}{100m^5}$, for some small constant $s > 0$, and $p^4 = \omega(\frac{m^3}{\bar{n}})$. Then, with probability at least $1 - e^{-\Omega(p\bar{n})}$, $\forall t \in \mathbb{R}^m$ satisfying $\|t - \bar{n}p\mu\|_2 \leq sp\sqrt{m\bar{n}}$, there $\exists x \in \{0,1\}^{\bar{n}}$, $\|x\|_1 = \Theta(p\bar{n})$ such that $\|t - Ax\|_2 \leq e^{-\Omega(p\bar{n}/m)}$.*

In the intended use case above, $\bar{n}$ will roughly be $O(\operatorname{poly}(m)\log n)$ and $p = 1/\operatorname{poly}(m)$, which will correspond to the very cheap 0-columns that we can pick T from. The corresponding success probability will now be $1 - n^{-\operatorname{poly}(m)}$ as opposed to $\Theta(1)$, which will allow us to get much better tail-bounds for the gap.

Generalized Gap Bounds and Their Application. Using the above, we obtain a substantial generalization of the Gaussian gap result of Borst et al [6], which we state below.

Theorem 3 (Centered IPs). *For $m \geq 1$, $n \geq \operatorname{poly}(m)$, $b \in \mathbb{R}^m$ 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*

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

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 fixed $k \geq 1$.

In this result b^- refers to the vector with $b_i^- = \min(b_i, 0)$. We also obtain a discrete variant of the random packing IP gap bound of [3].

Theorem 4 (Discrete Packing IPs). *For $m \geq 1$, $k \geq 3$, $\beta \in (0, 1/4)$, $n \geq \operatorname{poly}(m)\exp(\Omega(1/\beta))$, $b \in ((kn\beta, kn(1/2 - \beta)) \cap \mathbb{Z})^m$, if c has i.i.d. exponential entries and the entries of A are independent and uniform in $\{1, \dots, k\}$, then*

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

As a corollary of these gap bounds, we derive the following complexity bound for Branch-and-Bound.

Corollary 1. *With probability $1 - 1/\operatorname{poly}(n)$, the best-bound first Branch-and-Bound algorithm applied to* (Primal IP) *produces a tree of size at most $n^{\operatorname{poly}(m)}$ in the Centered IP model of Theorem 3 and of size $n^{\exp(O(1/\beta))\operatorname{poly}(m)}$ in the Discrete Packing IP model from Theorem 4.*

We note that, in the discrete case, the last result does not follow directly from Theorem 1, which requires the constraint matrix to have a logconcave distribution. Using the specific properties of these random instances, namely that the optimal dual solutions have small norm and the logconcavity of the objective, we show how to adapt the counting argument in the proof of Theorem 1 to derive the corresponding result (see Section 5 for the proof).

Compared to prior work, the above results give the first gap bounds and Branch-and-Bound complexity guarantees in the random centered and packing IP models where the constraint matrix A is discrete. We see this as stepping stone towards analyzing more realistic random IPs. We also generalize the gap bounds in the centered case to the setting where the columns of A are independent isotropic logconcave with bounded support, which shows that the gap bound is universal in a limited sense. We recall that a random vector $X \in \mathbb{R}^m$ is isotropic if $\mathbb{E}[X] = 0$ (mean zero) and $\mathbb{E}[XX^\mathsf{T}] = \mathrm{I}_m$ (identity covariance). A similar logconcave extension can also be proved in the packing setting, but we omit it for the sake of concision.

As a second technical improvement, our bounds hold with high probability in n as opposed to a probability depending on m. As these bounds are most interesting in the regime where m is constant (note that $m = 1$ in the packing case is the knapsack problem), the result can be viewed as a significant asymptotic improvement. This is made possible by the substantially improved probability of success in our discrepancy theorem.

On a more conceptual level, our results build a stronger bridge between (linear) discrepancy theorems for random matrices and additive gap bounds for random integer programs, which we hope will lead to further exploration. The use of discrepancy theoretic tools in the related context of approximation algorithms has already proven very fruitful. In particular, the best known $O(\log n)$ additive approximation for bin-packing, due Hoberg and Rothvoss [9], crucially relies on tools from discrepancy theory.

We now comment in more detail on the specific requirements of the gap theorems. Firstly, in the discrete case, we require the right-hand side b to be integral as it would be impossible to cancel a non-integral errors in the slack when rounding. Note that this is essentially without loss of generality, since replacing b by $\lfloor b \rfloor$ does not cut off any integer solutions when A is integral. While A can be discrete, we do not expect that the objective c can be made discrete without greatly increasing the gap bound. In particular, it seems unlikely that the reduced costs could be as small as in the continuous setting. Lastly, in the centered setting, we require the norm bound on the support of the columns of A to ensure that the errors in the slacks we need to repair are not too large. In the pure Gaussian setting, this condition already holds with probability $1 - 1/\mathrm{poly}(n)$, which recovers the result from [6].

In the packing setting, a non-trivial difference with [3] is that we require exponentially distributed objective coefficients, i.e., with density e^{-x}, $x \geq 0$, instead of uniform $[0,1]$. This makes the reduced cost filtering we need much milder, allowing us to control the conditional distribution of columns indexed by the 0 coordinates of x^* more easily. This also makes the integrality gap scale with $\exp(O(1/\beta))$ instead of $O(1/\beta)^m$.

As a final remark, we note that, in the discrete case for both the packing and centered setting, we require the entries to be uniformly distributed on an integer interval of size at least 3. We do not know how to prove the same result for intervals of size 2 (e.g., uniform on $\{-1,1\}$ or $\{0,1\}$), though this may be an artifact of our analysis. In the packing setting, a natural question is whether the gap result still holds when the entries of A are independent Bernoulli's with probability $p \in (0,1)$.

Techniques. Given the discrepancy theorem, the proof of the gap bounds mirrors the proofs in [3, 6], though with non-trivial technical adaptations as well as some simplifications. In particular, earlier proofs required repeated trials on disjoint columns of A to find a suitable subset T. Because of the exponentially small probability of failure in the discrepancy theorem, this is no longer needed and in fact now only hurts the analysis. That is, using all the small reduced cost columns together in the theorem both exponentially decreases the probability of failure and increases the size of the targets one can hit.

To prove the discrepancy theorem, we rely on a Fourier analytic approach, which is completely different from the second-moment counting based proofs in [3, 6]. Let $A \in \mathbb{R}^{n\times m}$ be our random matrix with columns having mean $\mu \in \mathbb{R}^m$, and let $0 < p < 1/\mathrm{poly}(m)$ be a parameter. Our 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. 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 $\|np\mu - t\|$ is a most a $\sqrt{p}$ factors times the *average deviation* of AX around $\mathbb{E}_X[AX]$ (the extra $\sqrt{p}$ term is likely an artifact of our proof and is $1/\mathrm{poly}(m)$ in all our applications), where the average deviation is $\mathbb{E}_X[\|\sum_{i=1}^n (X_i - p)A_i\|^2]^{1/2} = \sqrt{p(1-p)\sum_{i=1}^n \|A_i\|_2^2}$. This is done by applying the Fourier inversion formula, showing that the Fourier coefficients are close enough to Gaussian and integrating (see Section 3 for an overview).

Our approach here is in fact heavily inspired by the works of Hoberg and Rothvoss [10] and Franks and Saks [11] which analyzed $\{-1,1\}$ discrepancy of random matrices. In the above notation, they analyzed the probability mass function $Y' = AX'$, where $X_1', \dots, X_n'$ are uniform $\{-1,1\}$. While sharing many similarities, these models have non-trivial qualitative differences. In the $\{-1,1\}$ model Y' must use every column of A, and hence parity considerations play a significant role in determining which t's have positive probability (this is not issue in our model). In constrast, we must contend with the "drift" of X, namely that $\mathbb{E}[X] = p\mathbf{1}_n$. In particular, it is intuitive that $\|t - np\mu\|$ being small is not sufficient for $\Pr[Y = t] \neq 0$ if $\mathbb{E}_X[AX] = pA\mathbf{1}_n$ is far from $\mathbb{E}_{A,X}[AX] = np\mu$. Indeed, $pA\mathbf{1}_n$ is quite poorly concentrated around its mean, yielding concentration in terms of m (which is constant) instead of n, recalling that we aim for failure probability $e^{-\Omega(pn)}$.

To illustrate the problem with an example, if A has standard Gaussian $\mathcal{N}(0,1)$ entries, then $pA\mathbf{1}_n$ is distributed as $\mathcal{N}(0, p^2 n \mathrm{I}_m)$. Furthermore, the average deviation $\sqrt{p(1-p)\sum_{i=1}^n \|A_i\|_2^2}$ of AX is $O(\sqrt{pnm})$ with probability $1 - e^{-\Omega(n)}$ over A by standard concentration estimates. Since the means are all zero, we wish to be able to hit all targets close enough to the origin. However, by similar Gaussian tail-bounds $\|pA\mathbf{1}_n\|_2 \geq Cp\sqrt{mn}$ with probability $2^{-O(C^2 m)}$. Assuming $p = 1/\mathrm{poly}(m)$ and $C = \Theta(1/\sqrt{p})$, the drift $pA\mathbf{1}_n$ induced by this event – which occurs with probability depending only on m – overwhelms the average deviation $\sqrt{pmn}$ of AX, which makes the probability that $Y = AX$ is close to origin too small for standard Fourier analytic estimates to pick up on.

To deal with this, 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 to work with. Note that this subsampling crucially uses our flexibility to drop columns of A, a distinguishing feature of $\{0,1\}$ combinations versus $\{-1,1\}$ combinations. The subsampling process causes dependencies among the columns, but, fortunately, they are manageable enough to allow the Fourier analytic estimates to go through.

Organization The preliminaries to our results can be found in Section 2. Section 3 contains the proof of the discrepancy result, Theorem 2, and the bounds on the integrality gap, given in Theorems 3 and 4, are proven in Section 4. Section 5 is devoted to the proof of Corollary 1. Finally, in Section 6 we prove several anti-concentration results required by our method.

2 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,

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

We denote the positive part of x as, $x^+ := (\max(x_1,0),\dots,\max(x_m,0))$ and the negative part $x^- := -(\max(-x_1,0),\dots,\max(-x_m,0))$. 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 $\lim\limits_{a\to\infty} \frac{a}{b} = 0$ (resp. $\lim\limits_{a\to\infty} \frac{b}{a} = 0$). The identity matrix in $\mathbb{R}^m$ is denoted by I_m.

2.1 Fourier analysis

Our main tool for proving the discrepancy result is Fourier analysis and we review here the necessary details. Fix $X \sim \mathscr{D}$, 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

$$\mathrm{domain}(\mathscr{D}) := \{v \in \mathbb{R}^m : \langle v,w\rangle \in \mathbb{Z},\ \forall w \in \mathrm{support}(\mathscr{D})\}, \tag{1}$$

which leads to the fundamental domain in Fourier space (where we suppress the dependence on $\mathscr{D}$),

$$V := \{\theta \in \mathbb{R}^m : \|\theta\| \leq \inf_{0\neq w\in\mathrm{domain}(\mathscr{D})} \|\theta - w\|\}. \tag{2}$$

Observe that if $\mathscr{D}$ is absolutely continuous with respect to the Lebesgue measure, then $\mathrm{domain}(\mathscr{D}) = V = \mathbb{R}^m$ and if $\mathrm{support}(\mathscr{D}) \subset \mathbb{Z}^m$, for $\mathrm{domain}(\mathscr{D}) = \mathbb{Z}^m$, and $V = [-\frac{1}{2},\frac{1}{2}]^m$. Those are the two cases on which we focus.

The connection between X and its Fourier transform comes from the Fourier inversion formula [12, Theorem 1.20]:

Theorem 5 (Fourier inversion formula)**.** *For* $\lambda \in \text{support}(\mathscr{D})$*:*

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

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

Another desirable property of the Fourier transform is that it is particularly amenable to convolutions (this is clear from the exponential representation but see [12, Theorem 3.18]).

Theorem 6 (Multiplication-convolution theorem)**.** *Let* X *and* Y *be two independent random vectors. Then,*

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

2.2 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 $\text{Cov}(\nu) := \text{Cov}(X) := \mathbb{E}[XX^\mathsf{T}] - \mathbb{E}[X]\,\mathbb{E}[X]^\mathsf{T} \succeq 0$. 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 $\text{Cov}(X) = \mathrm{I}_m$.

Proposition 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. ☐

Let $X_1, \ldots, X_n$ be independent $\{0,1\}$ random variables with $\mu = \mathbb{E}[\sum_{i=1}^n X_i]$. Then, the Chernoff bound gives [13, Corollary 1.10],

$$\Pr[\sum_{i=1}^n X_i \leq \mu(1-\varepsilon)] \leq e^{-\frac{\varepsilon^2 \mu}{2}}, \varepsilon \in [0,1]. \tag{3}$$

$$\Pr[\sum_{i=1}^n X_i \geq \mu(1+\varepsilon)] \leq e^{-\frac{\varepsilon^2 \mu}{3}}, \varepsilon \in [0,1].$$

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

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

To see this bound, apply [13, Theorem 1.10.30] to the martingale $S_i := \sum\limits_{j=1}^{i} X_j - \mathbb{E}[X_j | X_1, \ldots, X_{j-1}]$.

Lemma 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 $\mathrm{Var}(X) = \int_{-\infty}^{\infty} t^2 f_X(t)dt$ under the conditions $f_X \le M$ and $\mathbb{E}[X] = 0$, then the the unique minimizer is $f_x = M \cdot \mathbf{1}_{[-\frac{1}{2M}, \frac{1}{2M}]}$. We then see,

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

□

2.3 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 $\mathscr{N}(\mu, \Sigma)$, the law of the Gaussian with mean μ and covariance Σ. The probability density function of $\mathscr{N}(\mu, \Sigma)$ is given by $\frac{1}{\sqrt{2\pi}^m \det(\Sigma)^{1/2}} e^{-\frac{1}{2}(x-\mu)^\mathsf{T}\Sigma^{-1}(x-\mu)}$, $\forall x \in \mathbb{R}^n$.

The following is a basic concentration fact for the norm of the standard Gaussian (see [14, Lemma 1], for example).

Lemma 2. *Let $G \sim \mathscr{N}(0, \mathrm{I}_m)$ and let $x \ge 7m$. Then,*

$$\Pr\left(\|G\|^2 \ge x\right) \le e^{-\frac{x}{3}}.$$

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

$$\mathbb{E}[e^{\lambda Y}] \le e^{\sigma^2\lambda^2/2}. \tag{5}$$

A standard normal random variable $X \sim \mathscr{N}(0,1)$ 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 tailbound [15, Proposition 2.5.2]:

$$\max\{\Pr[Y \le -\sigma s], \Pr[Y \ge \sigma s]\} \le e^{-\frac{s^2}{2}}, s \ge 0. \tag{6}$$

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

Lemma 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 $x \in [-1,1]$, $\varphi(x) \le \frac{1-x}{2}\varphi(-1) + \frac{1+x}{2}\varphi(1)$. Therefore,

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

□

We also need the following fact about truncated sub-Gaussian random variables, which is a slight generalization of [2, Lemma 7]:

Lemma 4. *Let $X \in \mathbb{R}$ be 1-sub-Gaussian. Then $\mathbb{E}[X^+] \le 1/2$ 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 $\mathbb{E}[X^2] \le 1$. Therefore, by Proposition 1, we have $\mu := \mathbb{E}[X^+] = \mathbb{E}[|X|]/2 \le \mathbb{E}[X^2]^{1/2}/2 \le 1/2$ by Hölder. $\sqrt{2}$-sub-gaussianity of $X^+ - \mu$ now follows verbatim from the proof of [2, Lemma 5] using that $\mu^2 \le 1/3$ and replacing Gaussian by sub-Gaussian. □

2.3.1 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 7 ([16])**.** *Let $X \in \mathbb{R}^d$ be a logconcave random vector. Then, for any surjective linear transformation $T : \mathbb{R}^d \to \mathbb{R}^k$, TX is a logconcave random vector.*

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

Lemma 5 ([17, Lemma 5.5])**.** *Let $X \in \mathbb{R}$ be a logconcave variable. Then its density function is upper bounded by $\frac{1}{\sqrt{\mathrm{Var}[X]}}$.*

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 $v^\mathsf{T} X$ has maximum density at most $1/\sqrt{\mathrm{Var}[v^\mathsf{T} X]} = 1/\|v\|_2$, where we have used $v^\mathsf{T} X$ 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 6 ([18])**.** *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 $t \in \mathbb{R}$, $\Pr[\theta^\mathsf{T} X \geq \theta^\mathsf{T} \mu - t] \geq 1/e - |t|$.*

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

Lemma 7 ([19])**.** *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 $\varphi : \mathbb{R}_+ \to \mathbb{R}$, $\mathbb{E}[\varphi(X)] \leq \mathbb{E}[\varphi(\mu Z)]$. 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 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 $X_r := X_r \mid X \geq 0$, $p_r = \Pr[X \geq 0]$. Note that X_l, X_r are both non-negative logconcave random variables and that $p_l, p_r \geq 1/e$ by Lemma 6. By Proposition 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

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

□

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

Lemma 9. *Let $X_1, \ldots, X_n \in \mathbb{R}_+$ 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. $\Pr[\sum_{i=1}^n X_i \le (1-\varepsilon)\mu n] \le e^{-n(-\ln(1-\varepsilon)+\varepsilon)} = e^{-n(\sum_{j=2}^{\infty} \varepsilon^j/j)}$, $\varepsilon \in [0,1]$.

Proof. By homogeneity, we assume wlog that $\mu = 1$.
Proof of 1. Let $\lambda := \frac{\varepsilon}{1+\varepsilon}$. Then,

$$\Pr[\sum_{i=1}^n X_i \ge (1+\varepsilon)n] \underbrace{\le}_{\text{Markov}} \mathbb{E}[e^{\lambda \sum_{i=1}^n X_i}]e^{-\lambda(1+\varepsilon)n} \underbrace{\le}_{\textit{Lemma 7}} (\frac{1}{1-\lambda})^n e^{-\lambda(1+\varepsilon)n} = e^{-n(\varepsilon - \ln(1+\varepsilon))}.$$

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

$$\Pr[\sum_{i=1}^n X_i \le (1-\varepsilon)n] \underbrace{\le}_{\text{Markov}} \mathbb{E}[e^{-\lambda \sum_{i=1}^n X_i}]e^{\lambda(1-\varepsilon)n} \underbrace{\le}_{\textit{Lemma 7}} (\frac{1}{1+\lambda})^n e^{-\lambda(1+\varepsilon)n} = e^{-n(-\ln(1-\varepsilon)-\varepsilon)}.$$

□

Finally, we will require concentration of truncated sums.

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

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

Proof. Define $X_1', \dots, X_n' \in \mathbb{R}_+$ to be i.i.d. copies of $X_1 \mid X_1 \ge 0$ and let $p = \Pr[X_1 \ge 0]$, where $1 - 1/e \ge p \ge 1/e$ (Lemma 6). Let $C = \sum_{i=1}^n 1[X_i \ge 0]$, 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 (3) and Lemma 9, we have that

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

and

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

□

2.4 Khinchine Inequality

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

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

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

As a consequence, we have

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

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

2.4.1 Discrete Random Variables

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

Proposition 2 (Discrete Symmetric Moments)**.** *For $k \geq 1$, let U be uniformly distributed on $\{0, \pm 1/k, \ldots, \pm 1\}$. 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 $\mathbb{E}[U^4]/\mathbb{E}[U^2]^2 = \frac{9(3k^2+3k-1)}{15(k+1)k} \leq 2$.*

2.5 Combinatorics

Theorems 3 and 4 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 12 ([20, Theorem 3.1])**.** *For all $\alpha \leq \frac{1}{2}$ and all n,*

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

3 Discrepancy

In this section we describe our main discrepancy results. Let $A \in \mathbb{R}^{n \times m}$ be a random matrix, whose columns are *i.i.d.* and sampled from a distribution $\mathscr{D}$ on $\mathbb{R}^m$. We denote $\mu = \mathrm{Mean}(\mathscr{D})$, $\Sigma = \mathrm{Cov}(\mathscr{D})$, and $\sigma^2 := \|\Sigma\|_{\mathrm{op}}$, where $\|\cdot\|_{\mathrm{op}}$ stands for the operator norm.

Before stating the result, let us introduce some notation. We first describe the type of distributions captured by our result, which are those which have a non-negligible mass on each half-space passing through its mean.

Definition 1 (admissible distributions)**.** *A probability distribution $\mathscr{D}$ on $\mathbb{R}^m$, with mean μ, is called* admissible *if, for any $v \in \mathbb{R}^m$,*

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

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 admissible. Moreover, Grünbaum's inequality in Lemma 6 shows that logconcave measures are admissible.*

Given an admissible distribution, the main idea will be to choose $S \subseteq [n]$ randomly with $\Pr[i \in S] = p$ for $p \leq \frac{1}{\mathrm{poly}(m)}$, independently for all $i \in [n]$ and to show that with high probability over A and positive probability over S, $A\mathbf{1}_S$ is close to a target vector λ. Thus, let us define the random vector $D := A\mathbf{1}_S$.

To understand the distribution of D we will consider its Fourier transform, which we denote by $\hat{D}(\theta) := \mathbb{E}[\exp(2\pi i\langle D, \theta\rangle)]$, and use V and $\mathrm{domain}(\mathscr{D})$ to denote its fundamental domains, as in (1) and (2). With this notation, the result is:

Theorem 8. *Suppose $\mathscr{D}$ is an admissible distribution that satisfies the property* (anti-concentration) *in Definition 2 below, with constant $\kappa > 0$, and let $p \in [0,1]$. Assume,*

- $p \leq \frac{\kappa^4}{1000m^5}$ *and* $p \gg m^{\frac{3}{4}}\left(\frac{\|\mu\|}{\sigma}+1\right)^{-\frac{3}{2}} n^{-\frac{1}{4}}$.
- $\kappa = \Omega(1)$.
- $n = \Omega(\mathrm{poly}(m))$.
- $\sigma\sqrt{\frac{m}{\kappa}} + \|\mu\| \leq e^{\frac{\kappa n}{80}}$ *and* $\left(1+\frac{\|\mu\|^2}{\sigma^2}\right)^{\frac{1}{m}} \leq \frac{\kappa^3 \exp\left(\frac{\kappa^3}{3\cdot 80^2 pm^3}\right)}{50000m^2}$.

Then, when $\mathrm{domain}(\mathscr{D}) = \mathbb{R}^m$*, for all $\lambda \in \mathbb{R}^m$ with $\|\lambda - pn\mu\| \leq \sigma p\sqrt{mn}$, with probability $1 - e^{-\Omega(pn)}$, there is a set S of size $|S| \in [\frac{1}{2}pn, \frac{3}{2}pn]$ such that $\|A\mathbf{1}_S - \lambda\| \leq n^4\exp(-\frac{\kappa^3 pn}{m})$. Moreover, if* $\mathrm{domain}(\mathscr{D}) = \mathbb{Z}^m$*, the same result holds, and we can choose S, such that $A\mathbf{1}_S = \lambda$.*

Remark 2. *The statement of the theorem contains several assumptions, which may seem complicated. In essence, those are technical assumptions that express the relationships between the different parameters and are required in order to make the proof work. Note that most of the assumptions can be satisfied by making n larger or p smaller (or both). Hence, not much generality is lost by making these assumptions.*

Recall Fourier's inversion formula (Theorem 5 in Section 2). In light of the formula, it will be enough to show that the integral in Theorem 5 is positive for appropriate λ, for most choices of A. In this case we will get that $\Pr[D = \lambda] > 0$, 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. This will be achieved by the following sequence of steps:

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 14).
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 $\|\theta\| = O(\frac{\mathrm{poly}(m)}{\sigma\sqrt{n}})$. This will allow to establish that the real part of $\hat{D}(\theta)\exp(-2\pi i\langle\theta, \lambda\rangle)$ is large when $\|\lambda - pn\mu\|\|\theta\|$ is small (Lemma 16).

3. Next, we show that $\int\limits_{\|\theta\|\leq r} |\hat{D}(\theta)|d\theta$ is large for some $r = O(\frac{\text{poly}(m)}{\sigma\sqrt{n}})$. In combination with Lemma 16, we shall conclude that $\int\limits_{\|\theta\|\leq r} \hat{D}(\theta)\exp(-2\pi i\langle\theta,\lambda\rangle)d\theta$ is large (Lemma 18).

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 19).

The first bottleneck is Step 1, which will change the distribution and effective size of the random matrix A. Below we show that admissible 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 $\mathscr{D}$ satisfy the following anti-concentration type property.

Definition 2 (anti-concentration)**.** *We say the measure $\mathscr{D}$ is anti-concentrated if there exists a constant $\kappa > 0$, such that for any $\nu \in \mathbb{R}^m$ and any $\theta \in V$,*

$$\Pr_{X\sim\mathscr{D}}\left[d(\theta^{\mathsf{T}}X,\mathbb{Z}) \geq \kappa\min\left(1,\|\theta\|_\infty\sigma\right) \mid \langle\nu,X\rangle \leq \langle\nu,\mu\rangle\right] \geq \kappa, \quad \text{(anti-concentration)}$$

where $d(\theta^{\mathsf{T}}X,\mathbb{Z}) := \inf\limits_{z\in\mathbb{Z}}|\theta^{\mathsf{T}}X - z|$.

Since we are working in the Fourier domain, 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, which would make $|\hat{D}(\theta)|$ large. There is an extra component in the definition which says that the anti concentration continues to hold after conditioning on an arbitrary half-space, passing through the mean. As will become apparent soon, this is a consequence of the subsampling scheme, from Step 1, and could be an artifact of the proof. From now on, we work under the assumption that $\mathscr{D}$ is anti-concentrated with some absolute constant $\kappa > 0$.

Let us just note that the (anti-concentration) property is not vacuous. In fact Theorem 2 is a direct consequence of the following lemma (see the proof in Section 6) and Theorem 8 (note that by Remark 1, both cases are admissible).

Lemma 13. *Suppose that for $X = (X_1,\ldots,X_m) \sim \mathscr{D}$, one of the following holds,*

1. X is logconcave and isotropic.

2. X_i are i.i.d. *uniformly on an integer interval $\{a,a+1,\ldots,a+k\}$, with $k > 1$.*

Then, $\mathscr{D}$ satisfies (anti-concentration) *with a universal constant $\kappa > 0$.*

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 $\kappa > 0$, for $i = 1,\ldots,n$ we define random variables $Y_i \in \{0,1\}$:

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

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 14. *Suppose that $\mathscr{D}$ satisfies* (anti-concentration) *with $\kappa > 0$ and that it is an admissible distribution, in the sense of Definition 1. Then, if $n \gg \frac{m^2}{\kappa}$:*

1.

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

2. With probability $1 - e^{-\Omega(n)}$, $|S_A| \geq \frac{n}{8}$,

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

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

3. $\Pr\left(d(\theta^\top A_i', \mathbb{Z}) \geq \kappa\min\left(1, \|\theta\|_\infty\sigma\right)|A_1Y_1, ..., A_{i-1}Y_{i-1}\right) \geq \frac{\kappa}{2}$*, for every $\theta \in V$, and $i \in [n']$.*

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 $Z_i \overset{\text{law}}{=} A_i|\left(\|A_i - \mu\| \leq 10\sigma\sqrt{\frac{m}{\kappa}}\right)$ 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 $(Z_i - \mu)(Z_i - \mu)^\top \preceq 10\sigma\sqrt{\frac{m}{\kappa}}\mathrm{I}_m$ almost surely, (matrix concentration) follows from the matrix Bernstein inequality [21, Theorem 1.6.2].

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

$$\left\|\sum_{i\in S_A} A_i - \mu\right\|^2 \leq \sum_{1\in S_A}\|A_i - \mu\|^2 = \mathrm{Tr}\left(\sum_{i\in S_A}^{n}(A_i - \mu)(A_i - \mu)^\top\right) \leq 2n\mathrm{Tr}\left(\sigma^2\mathrm{I}_m\right),$$

where the last inequality is monotonicity of the trace. Also, recalling that $\Sigma \preceq \sigma^2\mathrm{I}_d$, 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,

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

If $\mathscr{D}$ is admissible, then, since A_i is independent from $\{Y_j, A_j\}_{j=1}^{i-1}$, by definition,

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

Taken together, the above displays imply

$$\Pr(Y_i = 1|A_1, \ldots, A_{i-1}) = \Omega(1).$$

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

$$\Pr\left(|S_A| \geq \frac{n}{8}\right) = 1 - e^{-\Omega(n)}.$$

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

$$\begin{aligned}
&\Pr\left(d(\theta^\mathsf{T} A'_i, \mathbb{Z}) \geq \kappa \min\left(1, \|\theta\|_\infty \sigma\right) | A_1 Y_1, \ldots, A_{i-1} Y_{i-1}\right) \\
&\quad = \Pr\left(d(\theta^\mathsf{T} A_i, \mathbb{Z}) \geq \kappa \min\left(1, \|\theta\|_\infty \sigma\right) | \langle \nu, A_i - \mu \rangle \leq 0 \text{ and } \|A_i - \mu\| \leq 10\sigma\sqrt{\frac{m}{\kappa}}\right) \\
&\quad \geq \Pr\left(d(\theta^\mathsf{T} A_i, \mathbb{Z}) \geq \kappa \min\left(1, \|\theta\|_\infty \sigma\right) | \langle \nu, A_i - \mu \rangle \leq 0\right) - \frac{\kappa}{100} \geq \frac{\kappa}{2}.
\end{aligned}$$

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

In light of the lemma, in the sequel, all computations will be made conditioned on the high-probability event defined by Lemma 14 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 $2\pi np\langle\theta, \mu\rangle$. Before proceeding, we introduce an auxiliary parameter $\beta = \frac{1}{80p\sqrt{m}}$, which will bound the region in which we control $\hat{D}(\theta)$. We record here some facts, which will be useful later on:

Lemma 15. *Under the assumptions of Theorem 8, if $\beta = \frac{1}{80p\sqrt{m}}$, then for large enough n,*

$$\frac{1}{\sqrt{24\pi^2 p}} \leq \beta \leq \frac{1}{80p\sqrt{m}}$$

$$\beta^3 \leq \frac{\sqrt{n}}{50000p} \min\left(\left(\frac{\|\mu\|}{\sigma} + 1\right)^{-3}, \left(\frac{\kappa}{m}\right)^{\frac{3}{2}}\right)$$

Proof. Since $p \leq \frac{\kappa^4}{1000m^5}$, we immediately get the first set of inequalities. For the second inequality, the assumptions in Theorem 8 ensure,

$$\frac{\sqrt{n}}{50000p}\left(\frac{\kappa}{m}\right)^{\frac{3}{2}} \geq c\frac{\sqrt{n}}{pm^{\frac{3}{2}}} = 80cp^2\sqrt{n}\beta^3 \geq \beta^3,$$

$$\frac{\sqrt{n}}{50000p}\left(\frac{\|\mu\|}{\sigma} + 1\right)^{-3} = cp^2 m^{\frac{3}{2}}\sqrt{n}\left(\frac{\|\mu\|}{\sigma} + 1\right)^{-3}\beta^3 \geq \beta^3$$

for some small constant $c > 0$, and the inequality follows, for large n, since $p \gg m^{\frac{3}{4}}\left(\frac{\|\mu\|}{\sigma} + 1\right)^{-\frac{3}{2}} n^{-\frac{1}{4}}$. □

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

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

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 16. *With probability $1-e^{-\Omega(n)}$ over A, for all $\|\theta\|\le\frac{\beta}{\sigma\sqrt{n}}$,*

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

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 $g(x)=f(x)-2\pi px$ and, with a second-order Taylor approximation of $f(x)$ around $x=0$, we see that for every $x\ge 0$ there is some $x'\in[0,x]$ such that

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

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

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

Now,

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

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

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

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

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

Proof. Lemma 15 implies $p\beta\le\frac{1}{80\sqrt{m}}$, $\beta^3 p\left(\frac{\|\mu\|}{\sigma}+1\right)^3\le\frac{\sqrt{n}}{50000}$ and, $\beta^3 p(m/\kappa)^{\frac{3}{2}}\le\frac{\sqrt{n}}{50000}$. Thus,

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

and the Corollary follows from Lemma 16. $\square$

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 17. *The following holds,*

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

Proof. We have:

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

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

$$\cos(x) \ge 1 - x^2$$
$$\ln(1-x) \ge -\frac{3}{2}x \quad \text{when } |x| \le \frac{1}{2}.$$

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

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

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

$$|\hat{D}(\theta)| \ge \exp\left(-12\pi^2 p\left(n\langle \theta, \mu\rangle^2 + \sum_{i=1}^{n}\langle \theta, A_i - \mu\rangle^2\right)\right) \ge \exp\left(-24\pi^2 np\left(\langle \theta, \mu\rangle^2 + \sigma^2\theta\theta^{\mathsf{T}}\right)\right)$$
$$\ge \exp\left(-24\pi^2\theta\left(\mu^{\mathsf{T}}\mu + \sigma^2 \mathrm{I}_m\right)\theta^{\mathsf{T}} np\right).$$

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

$$\begin{aligned}\int_{\|\theta\| \le \frac{\beta}{\sigma\sqrt{n}}} |\hat{D}(\theta)| d\theta &\ge \int_{\|\theta\| \le \frac{\beta}{\sigma\sqrt{n}}} \exp\left(-24\pi^2\theta\left(\mu^{\mathsf{T}}\mu + \sigma^2\mathrm{I}_m\right)\theta^{\mathsf{T}} np\right) d\theta \\ &= \frac{1}{\sqrt{\det\left(96\pi^3 np\left(\mu^{\mathsf{T}}\mu + \sigma^2\mathrm{I}_m\right)\right)}} \Pr\left(\|Y\| \le \frac{\beta}{\sigma\sqrt{n}}\right) \\ &= \frac{1}{(96\pi^3 np)^{\frac{m}{2}}\left(\|\mu\|^2 + \sigma^2\right)^{\frac{1}{2}}\sigma^{m-1}} \Pr\left(\|Y\| \le \frac{\beta}{\sigma\sqrt{n}}\right). \end{aligned} \tag{7}$$

By Chebyshev's inequality,

$$\Pr\left(\|Y\| \ge \frac{\beta}{\sigma\sqrt{n}}\right) \le \frac{\sigma^2 n}{\beta^2}\operatorname{Tr}\left(\frac{1}{48\pi^2 np}(\mu^{\mathsf{T}}\mu + \sigma^2\mathrm{I}_m)^{-1}\right) \le \frac{m}{48\pi^2 p\beta^2} \le \frac{1}{2},$$

where in the last inequality, we have used $\beta = \frac{1}{80p\sqrt{m}}$ and $p \le \frac{1}{m^5}$. The proof concludes by plugging this estimate into (7). □

To handle the integral in Theorem 5, we note that for small enough λ, as dictated by Theorem 8, by Corollary 2,

$$|\arg(\hat{D}(\theta)\exp(2\pi i\langle \theta, \lambda\rangle))| \le |\arg(\hat{D}(\theta)) - 2\pi np\langle \theta, \mu\rangle| + |\arg(\exp(2\pi i\langle \theta, \lambda - pn\mu\rangle))| \le \frac{1}{4}\pi,$$

whenever $\|\theta\| \le \frac{\beta}{\sigma\sqrt{n}}$ (recall $\beta = \frac{1}{80p\sqrt{m}}$). For a complex number z, we have $\Re(z) = \cos(\arg(z))|z|$. Thus, the previous lemma implies,

Lemma 18. *Let $\lambda \in \mathbb{R}^m$ with $\|\lambda - pn\mu\| \le \sigma p\sqrt{mn}$. Then, for $\|\theta\| \le \frac{\beta}{\sigma\sqrt{n}}$,*

$$\Re\left[\hat{D}(\theta)\exp(-2\pi i\langle \theta, \lambda\rangle)\right] \ge 0,$$

and

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

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. As mentioned above, the main difficulty comes from the fact that, a-priori, $\langle \theta, D \rangle$ can be close to an integer, irregardless of of the value $\|\theta\|$. Our anti-concentration assumption allows us to avoid this. Specifically Item 3 in Lemma 14 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 19. *Assume that* $\sigma\sqrt{\frac{m}{\kappa}} + \|\mu\| \leq e^{\frac{\kappa n}{80}}$. *Then, with probability* $1 - e^{-\Omega(n)}$, *as long as n is large enough, we have*

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

for

$$\theta \in \tilde{V},$$

where V is the fundamental domain, as in (2), *and*

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

Proof. Let N be an ε-net of $\tilde{V}$ for $\varepsilon = \frac{1}{2n^2\left(\sigma\sqrt{m/\kappa} + \|\mu\|\right)}$. Standard arguments show that, under the assumption $\sigma\sqrt{\frac{m}{\kappa}} + \|\mu\| \leq e^{\frac{\kappa n}{80}}$, one can take $|N| \leq \left(\sigma\sqrt{\frac{m}{\kappa}} + \|\mu\|\right) e^{\frac{\kappa}{80}n} \leq e^{\frac{\kappa}{40}n}$, when n is large enough. For $\theta \in N$, define

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

If we fix θ and set $X_i := \mathbf{1}_{i \in E(\theta)}$ then $|E(\theta)| = \sum\limits_{i=1}^{n} X_i$. The statement of Item 3 in Lemma 14 is $\mathbb{E}\left[X_i | X_1, \dots, X_{i-1}\right] \geq \frac{\kappa}{2}$. Applying Azuma's inequality (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:

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

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

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

Thus:

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

where the last inequality holds for large enough n. □

By properly integrating the inequality, we have thus obtained:

Lemma 20. *The following inequality holds:*

$$\int_B |\hat{D}(\theta)|d\theta \le \left(\frac{1}{p\sigma^2 nm^{20}}\right)^{\frac{m}{2}} + \exp\left(-\frac{\kappa^3 pn}{80}\right),$$

where

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

Proof. From Lemma 19, when $p < 0.1$, and $\sigma\|\theta\|_\infty \le \sigma\|\theta\| \le \sqrt{m}$, we have

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

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

$$
\begin{aligned}
\int_{\sqrt{m} \geq \|\theta\| \geq \frac{\beta}{\sigma\sqrt{n}}} |D(\theta)| d\theta &\leq \int_{\|\theta\| \geq \frac{\beta}{\sigma\sqrt{n}}} \exp\left(-\frac{\kappa^3 pn\sigma^2\|\theta\|^2}{10m}\right) d\theta \\
&\leq \left(\frac{5m}{\kappa^3 pn\sigma^2}\right)^{\frac{m}{2}} \mathbb{P}\left(\|Y\| \geq \frac{\beta}{\sigma\sqrt{n}}\sqrt{\frac{\kappa^3 pn\sigma^2}{5m}}\right) \\
&= \left(\frac{5m}{\kappa^3 pn\sigma^2}\right)^{\frac{m}{2}} \mathbb{P}\left(\|Y\|^2 \geq \beta^2\frac{\kappa^3 p}{5m}\right) \\
&= \left(\frac{5m}{\kappa^3 pn\sigma^2}\right)^{\frac{m}{2}} \mathbb{P}\left(\|Y\|^2 \geq \frac{\kappa^3}{80^2 pm^2}\right)
\end{aligned}
$$

where the last equality holds since $\beta = \frac{1}{80p\sqrt{m}}$. By Lemma 2,

$$
\mathbb{P}\left(\|Y\|^2 \geq \frac{\kappa^3}{80^2 pm^2}\right) \leq \exp\left(-\frac{\kappa^3}{3 \cdot 80^2 pm^2}\right).
$$

Hence,

$$
\int_{\frac{\sqrt{m}}{4} \geq \|\theta\| \geq \beta/\sqrt{n}} |D(\theta)| d\theta \leq \left(\frac{5m \exp\left(-\frac{\kappa^3}{3\cdot 80^2 pm^3}\right)}{\kappa^3 p\sigma^2 n}\right)^{\frac{m}{2}}.
$$

Furthermore, for all $\theta \in \left[-\frac{e^{\frac{\kappa^3 pn}{80m}}}{4n^2}, \frac{e^{\frac{\kappa^3 pn}{80m}}}{4n^2}\right]^m$ with $\|\theta\|_\infty \sigma \geq 1$, Lemma 19 also implies,

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

So:

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

Summing the previous two inequalities yields the result. $\square$

Proving Theorem 8: If the support of $\mathscr{D}$ is contained in $\mathbb{Z}^m$, then Lemma 18 and Lemma 20 are enough to prove the discrete case in Theorem 8.

Proof of Theorem 8 when $\text{domain}(\mathscr{D}) = \mathbb{Z}^m$. We begin by noting that the assumption on $\left(1+\frac{\|\mu\|^2}{\sigma^2}\right)^{\frac{1}{m}}$ ensures,

$$\frac{\kappa^3 \exp\left(\frac{\kappa^3}{3\cdot 80^2 pm^3}\right)}{5m} \geq 500\pi^3 m \left(1+\frac{\|\mu\|^2}{\sigma^2}\right)^{\frac{1}{m}} \tag{8}$$

Let B be defined as in Lemma 20. Then, for n large enough, by plugging (8) into Lemma 20 we get,

$$\int\limits_B |D(\theta)|d\theta \leq 0.2 \left(\frac{1}{200\pi^3 nmp\sigma^2 \left(1+\frac{\|\mu\|^2}{\sigma^2}\right)^{\frac{1}{m}}}\right)^{\frac{m}{2}}. \tag{9}$$

Moreover, by Lemma 18,

$$\begin{aligned}
\Re\left[\int_{\|\theta\|\leq\frac{\beta}{\sigma\sqrt{n}}} \hat{D}(\theta)\exp(-2\pi i\langle\theta,\lambda\rangle)d\theta\right] &\geq \cos\left(\frac{\pi}{4}\right)\left(\frac{1}{200\pi^3 nmp}\right)^{\frac{m}{2}} \frac{1}{\sigma^{m-1}\left(\sigma^2+\|\mu\|^2\right)^{\frac{1}{2}}} \\
&= \cos\left(\frac{\pi}{4}\right)\left(\frac{1}{200\pi^3 nmp\sigma^2 \left(1+\frac{\|\mu\|^2}{\sigma^2}\right)^{\frac{1}{m}}}\right)^{\frac{m}{2}}
\end{aligned} \tag{10}$$

So,

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

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

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

Since $n \geq \text{poly}(m)$, for large enough n we have $\Pr[D=\lambda] > 2\exp(-pn/12)$, which implies the existence of a suitable S with $|S| \in [0.5pn, 1.5pn]$ and $A\mathbf{1}_S = \lambda$. $\square$

The proof of the continuous case requires an extra step to control the Fourier transform outside the domain of Lemma 20. To deal with continuous distributions, we define $R \sim \mathscr{N}(0,\gamma \mathrm{I}_m)$ for γ to be determined later and $H = D+R$. If f_H is the density of H, we will show that, for appropriate λ, $f_H(\lambda)$ is positive, from which it will follow that with high probability there exists a suitable set S such that $\|A\mathbf{1}_S - \lambda\|$ is small.

Proof of Theorem 8 when $\operatorname{domain}(\mathscr{D}) = \mathbb{R}^m$. The main difference from the discrete case is that now, by the multiplication-convolution theorem (Theorem 6), we have $\hat{H}(\theta) = \hat{D}(\theta) \cdot \hat{R}(\theta) = e^{-\frac{\|\theta\|^2\gamma}{2}}\hat{D}(\theta)$. So, if $\theta_0 := \exp\left(\frac{\kappa^3 pn}{80m}\right)\frac{1}{n^2}$, the estimates in (9) and (10) imply,

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

where the last inequality holds as long as n is large enough and supposing $\gamma < 1$. Thus, to complete the proof it is enough to choose γ such that,

$$
\int\limits_{\|\theta\|_\infty\geq\theta_0} |\hat{H}(\theta)|d\theta \leq \frac{1}{8}\left(\frac{1}{200\pi^3 nmp\sigma^2\left(1+\frac{\|\mu\|^2}{\sigma^2}\right)^{\frac{1}{m}}}\right)^{\frac{m}{2}}. \tag{11}
$$

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

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

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

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

for which (11) holds, as long as n is large enough. Also, $\gamma < 1$ as required earlier and $\gamma\theta_0^2 \geq 7m$, as required by Lemma 2. If $f_H(\lambda)$ is the density of H at λ, Theorem 5 along with (11) give,

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

Now, recall that $H = D + R$, where $R \sim \mathcal{N}(0, \gamma \mathrm{I}_m)$. So, when n is large, by applying Lemma 2 again,

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

We conclude that with probability $1 - e^{-\Omega(n)}$ over A, there exists some $T \subset [n]$ and some $v \in \mathbb{R}^m$, with $\|v\| \leq \exp\left(-\frac{\kappa^3 pn}{40m}\right) n^4$, such that

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

Or, in other words, $\|A\mathbf{1}_T - \lambda\| = \|v\| \leq \exp\left(-\frac{\kappa^3 pn}{40m^2}\right) n^4$. Finally, we finish, as in the proof of the discrete case, with the multiplicative Chernoff bound:

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

Assume $m^6 = O(n)$. So, for large enough n we have $\Pr[D = \lambda] > 2\exp(-pn/12)$, which implies the existence of a suitable S with $|S| \in [0.5pn, 1.5pn]$ and $A\mathbf{1}_S = \lambda$. $\square$

4 Integrality Gap Bounds

4.1 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:

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

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

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

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

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:

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

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 21 ([6, 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,

$$N_0 := \{i \in [n] | x^*_i = 0\}, \text{and } N_1 := \{i \in [n] | x^*_i = 1\}. \tag{12}$$

Let W be the matrix with columns $W_{\cdot,i} = \begin{bmatrix} c_i & A_{\cdot,i} \end{bmatrix}^\mathsf{T}$. The distribution of the columns of W with indices in N_0 plays an important role in the proofs of Theorems 3 and 4. 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 [6, Lemma 5] for the proof.

Lemma 22. *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 $W_{\cdot,i} \mid i \in N$ is the same as $W_{\cdot,i} \mid u^{*\mathsf{T}} A_{\cdot,i} - c_i > 0$.*

4.2 The Gap Bound for Centered IPs

In this subsection we will prove Theorem 3. In the setting of Theorem 3, 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 $\{0, \pm 1/k, \ldots, \pm 1\}$ 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 $\mathrm{I}_m/3$, and thus essentially isotropic).

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

4.2.1 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 [6, Lemma 4].

Lemma 23. *For $A \in \mathbb{R}^{m\times n}$, $n \geq 10^5 m$, distributed as (LI) or (DSU), $c \sim \mathcal{N}(0, \mathrm{I}_m)$, $\|b^-\| \leq \frac{n}{12\sqrt{2}}$ 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 24. *Let $W^\mathsf{T} := (c, A^\mathsf{T})$ where $c \sim \mathcal{N}(0, \mathrm{I}_n)$ and $A \in \mathbb{R}^{m\times n}$ is distributed as (LI) or (DSU). Then, for $n = \Omega(m)$, we have that*

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

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

$$\frac{1}{6} \leq \mathbb{E}\left[(v^\mathsf{T} W)_i^+\right] \leq \frac{1}{2}. \tag{13}$$

To see the right inequality, by Proposition 1, $\mathbb{E}\left[(v^{\mathsf{T}}W)_i^+\right] = \frac{1}{2}\mathbb{E}\left[|v^{\mathsf{T}}W|_i^+\right]$. Now observe that every entry has variance at most 1, so with Jensen's inequality

$$\mathbb{E}\left[(v^{\mathsf{T}}W)_i^+\right] = \frac{1}{2}\mathbb{E}\left[|v^{\mathsf{T}}W|_i^+\right] \le \frac{1}{2}\sqrt{\mathrm{Var}(|v^{\mathsf{T}}W|_i^+)} \le \frac{1}{2}.$$

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

$$\frac{1}{6} \le \frac{1}{2\sqrt{e}} \le \frac{1}{2}\mathbb{E}\left[|v^{\mathsf{T}}W|_i^+\right] = \mathbb{E}\left[(v^{\mathsf{T}}W)_i^+\right].$$

If the columns of W are discrete, by Proposition 2 every entry satisfies, $\mathrm{Var}(W_{ij}) \ge \frac{1}{3}$. Hence, $\mathrm{Var}\left((v^{\mathsf{T}}W)_i\right) \ge \frac{1}{3}$. Moreover, Proposition 2 also implies, $\mathbb{E}[W_{ji}^4] \le 3\mathbb{E}[W_{ji}^2]^2$. So, by Khinchine's inequality in Lemma 11,

$$\frac{1}{6} \le \frac{\sqrt{\mathrm{Var}\left((v^{\mathsf{T}}W)_i\right)}}{2\sqrt{3}} \le \frac{1}{2}\mathbb{E}\left[|v^{\mathsf{T}}W|_i^+\right] = \mathbb{E}\left[(v^{\mathsf{T}}W)_i^+\right].$$

Now, having established (13), we can bound $\Pr\left[\|(v^{\mathsf{T}}W)^+\|_1 \notin [\frac{n}{8}, \frac{5n}{8}]\right]$. In the log-concave case, Lemma 10 immediately gives,

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

In the discrete case, by Lemma 3, every entry of W is 1-sub-Gaussian, and Lemma 4 shows that $(v^{\mathsf{T}}W)_i^+ - \mathbb{E}\left[(v^{\mathsf{T}}W)_i^+\right]$ -is $\sqrt{2}$-sub-Gaussian. After summing the coordinates we get that $\|(v^{\mathsf{T}}W)^+\|_1 - \mathbb{E}\left[\|(v^{\mathsf{T}}W)^+\|_1\right]$ -is $\sqrt{2n}$-sub-Gaussian. Applying (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 $|N_\varepsilon| \le \left(\frac{3}{\varepsilon}\right)^m$. Hence, by applying a union bound,

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

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 $\|u - \tilde{u}\|_2 \le \varepsilon$. Under E, we have,

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

which is equivalent to,

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

On the other hand,

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

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

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

□

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

Lemma 25. *Let $c \sim \mathcal{N}(0, \mathrm{I}_n)$. Then, for every $\alpha \in [0, 2\sqrt{\log(2)}]$,*

$$\Pr\left[\max_{x \in \{0,1\}^n,\ \|x\|_1 \ge \beta n} c^\mathsf{T} x \ge \alpha n\right] \le e^{\frac{-\alpha^2 n}{2}},$$

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

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

$$\Pr\left(c^\mathsf{T} x \ge \alpha n\right) \le e^{-\frac{\alpha^2 n^2}{2\|x\|_2^2}} \le e^{-\frac{\alpha^2 n}{2}}.$$

We now apply a union bound,

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

□

We now have the ingredients to prove the main lemma.

Proof of Lemma 23. For the proof, we will consider the extended matrix $W^\mathsf{T} := (c, A^\mathsf{T})$. We begin by showing that, for the optimal solution, $c^\mathsf{T} x^*$ is large. Let $u \ge 0$ be any dual solution. Then, under the complement of the event defined in Lemma 24 for W, using (Dual LP),

$$\begin{aligned} \mathrm{val}_b^*(u) &= b^\mathsf{T} u + \|(c - A^\mathsf{T} u)^+\|_1 \ge -\|b^-\|\|u\| + \|((1,-u)^\mathsf{T} W)^+\|_1 \\ &\ge -\|b^-\|\|u\| + \sqrt{1+\|u\|^2}\,\frac{n}{12} \ge \frac{n}{12}\left(-\frac{\|u\|}{\sqrt{2}} + \sqrt{1+\|u\|^2}\right) \ge \frac{n}{12\sqrt{2}}. \end{aligned} \quad (14)$$

The second inequality is the lower bound in Lemma 24 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 $c^\mathsf{T} x^*$ follows by noting,

$$c^\mathsf{T} x^* = \mathrm{val}_c(x^*) = \mathrm{val}^*{}_b(u^*).$$

We now prove that $\|u^*\|$ cannot be too large. Again, under the complement of the event in Lemma 24, but using the upper bound this time,

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

where in the third inequality we have applied (14) 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 (14), we have

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

and by conditioning on the complement of the event in Lemma 25 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 25 and 24. □

4.2.2 Conditional Distribution of 0-columns of IP

Let B be a random variable with the same distribution as the columns of A. By Lemma 13, 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 26. *Let B' have the law of B, conditioned on $|u^{*\mathsf{T}}B| \leq C$. Then,*

1. $\Pr[|u^{*\mathsf{T}}B| \leq C] \geq 1 - \frac{\kappa}{150}$.
2. *We have* $\frac{1}{10}I_m \preccurlyeq \mathrm{Cov}(B') \preccurlyeq 2I_m$.
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. Let $E = \{a \in \mathbb{R}^m : |u^{*\mathsf{T}}a| \leq C\}$. From Chebyshev's inequality, and since distributions we consider satisfy $\mathrm{Cov}(B) \preceq \mathrm{I}_m$,

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

which is the first claim.

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

$$\mathbb{E}\left[|w^\mathsf{T}B'|^2\right] = \frac{\mathbb{E}\left[|w^\mathsf{T}B|^2\mathbf{1}_E\right]}{\Pr(B \in E)} \leq 2\mathbb{E}\left[|w^\mathsf{T}B'|^2\right] \leq 2\|w\|^2.$$

In the (DSU) case, to lower bound $\mathrm{Cov}(B')$, we note that by Proposition 2, $\mathbb{E}[W_{ji}^4] \leq 3\,\mathbb{E}[W_{ji}^2]^2$. As a consequence, Lemma 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 2 we have $\mathbb{E}[|w^\top B|^2] \geq \frac{1}{3}\|w\|^2$. So:

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

proving the second claim for the (DSU) case.

In the (LI) case, $\langle w, B\rangle$ is logconcave. By Lemma 5, $f_{\langle B,v\rangle} \leq \frac{1}{\sqrt{\mathrm{Var}\langle B,v\rangle}} = 1$. So, $f_{\langle v,B'\rangle} \leq \frac{1}{1-1/150} f_{\langle B,v\rangle} = \frac{150}{149}$. Now, Lemma 1 implies that $\mathrm{Var}(\langle B'', v\rangle) \geq \frac{1}{13}$.

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 $\sigma' := \sqrt{\|\mathrm{Cov}(B')\|_{\mathrm{op}}} \leq \sqrt{2}$. 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:

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

The last inequality follows from (anti-concentration),

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

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 3, 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 27. *We can apply rejection sampling to $(c_i, A_{\cdot,i})^\top | u^* A_{\cdot,i} - c_i \geq 0$, such that we have $\mathrm{Law}(A_{\cdot,i} | acceptance) = \mathrm{Law}(B')$ and for all accepted entries $u^* A_{\cdot,i} - c_i \in [0, \delta]$ with the probability of acceptance at least $\Omega(\delta)$, when $\delta = O(1)$ and $\kappa = \Omega(1)$.*

Proof. Let $M = \delta \exp(-\frac{1}{2}(C+\delta)^2)$. Let $N \sim \mathcal{N}(0,1)$ be an independent variable. We sample for $B = A_{\cdot,i}$ by accepting with probability:

$$\Pr[\text{acceptance}|B, c_i] = \begin{cases} \frac{M}{\Pr[u^{*\top}B - N \in [0,\delta] | u^{*\top}B - N \geq 0]} & \text{if } |u^{*\top}B| \leq C \wedge u^{*\top}B - c_i \in [0,\delta] \\ 0 & \text{else} \end{cases}$$

If $|u^{*\top}B| \leq C$, then,

$$\Pr[u^{*\top}B - N \in [0,\delta] \mid u^{*\top}B - N \geq 0] \geq \Pr[u^{*\top}B - N \in [0,\delta]] \geq \delta \exp(-\frac{1}{2}(C+\delta)^2).$$

Thus, the acceptance probability is at most 1, which makes the sampling procedure well-defined.

Now we will prove that $\Pr[\text{acceptance}] = \Omega(1)$:

$$\begin{aligned}
\Pr[\text{acceptance}] &= \int_{y\in\mathbb{R}^m} \Pr[u^{*\top}B - c_i \in [0,\delta]] \Pr[\text{acceptance}|B=y, u^{*\top}y - c_i \in [0,\delta]] dy \\
&= \int_{y:|u^{*\top}y|\le C} f_B(y) M \frac{\Pr[u^{*\top}y - c_i \in [0,\delta]]}{\Pr[u^{*\top}y - N \in [0,\delta]]} dy \\
&= M \Pr[|u^{*\top}B| \le C] \ge \frac{1}{2}M \ge \delta \exp(-2C^2).
\end{aligned}$$

For the last inequality we assumed that $\delta \le C$. Since $C = \frac{\sqrt{2}\|u^*\|}{\sqrt{\kappa}}$, by Lemma 23 and the fact that $\kappa = \Omega(1)$, we get, $\Pr[\text{acceptance}] \ge \delta \exp\left(-\frac{2^{12}}{\kappa}\right) = \Omega(\delta)$.

Let $\bar{B} := B|\text{acceptance}$. Now, for all y with $|u^{*\top}B| \le C$:

$$\begin{aligned}
f_{\bar{B}}(y) &= \frac{\Pr[\text{acceptance}|B=y, u^{*\top}y - c_i \in [0,\delta]] \cdot f_B(y) \cdot \Pr[u^{*\top}y - c_i \in [0,\delta] | u^{*\top}y - c_i \ge 0]}{\Pr[\text{acceptance}]} \\
&= \frac{f_B(y)M}{\Pr[\text{acceptance}]}
\end{aligned}$$

For all other y, we have $f_{\bar{B}}(y) = 0$. Hence, $f_{\bar{B}} \propto f_{B'}$ and therefore $\mathrm{Law}(\bar{B}) = \mathrm{Law}(B')$. Note that while our notation here assumes that B, B' and $\bar{B}$ have continuous distributions, the same argument holds for discrete distributions.

□

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 28. *If B' is logconcave, then using rejection sampling on B' with acceptance probability at least $\Omega(1)$, we can get a distribution B'' such that:*

1. $\mathbb{E}[B''] = 0$.
2. $\mathrm{Cov}(B'') \succcurlyeq \frac{1}{768} I_m$.
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. Let $\mu' := \mathbb{E}[B']$. By Hölder's inequality, we have any unit vector $v \in \mathbb{R}^m$,

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

where the second equality follows since B is isotropic and from

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

Recall $C \geq \sqrt{150}\|u^*\|$. Hence, the concentration bound in [17, Lemma 5.7], coupled with the fact that B is isotropic, gives,

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

So,

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

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 6 we have,

$$\Pr[B'\in H] \geq \Pr[B\in H] - \Pr[|u^{*\mathsf{T}}B|>C] \geq \frac{1}{e} - \frac{1}{12} - \frac{1}{150} = \frac{1}{4}.$$

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 $f:\mathbb{R}^m\to\mathbb{R}$, with $\|f\|_\infty\leq 1$, such that $\frac{\mathbb{E}[f(B')B']}{1/4} = -\mu'$. Let $g(x) = \frac{f(x)+4}{5}$.

We will perform rejection sampling with $\Pr[\text{acceptance}|B'] = g(B')$. Note that, since $0\leq f(B')\leq 1$, this probability is well defined. Call the resulting distribution B''. The probability of acceptance is $\mathbb{E}[g(B')] = \frac{\frac{1}{4}+4}{5} \geq \frac{1}{2}$. 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.

For every halfspace H containing the origin, by the Grünbaum inequality, $\Pr[B\in H] \geq \frac{1}{e}$. So $\Pr[B'\in H] \geq \Pr[B\in H] - \Pr[|u^{*\mathsf{T}}B|>C] \geq \frac{1}{e} - \frac{1}{150} \geq \frac{1}{4}$. Because $f_{B''} = \frac{g\cdot f_{B'}}{\Pr[\text{acceptance}]} \geq g\cdot f_{B'} \geq \frac{4}{5}f_{B'}$, 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:

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

This implies $\mathrm{Var}(\langle v,B''\rangle) \preccurlyeq 2\,\mathrm{Var}(\langle v,B'\rangle) \preccurlyeq 2I_m$. Because $\mathrm{Cov}(B') \geq \frac{1}{10}I_m$ and the fact that $\langle B',v\rangle$ is logconcave, by Lemma 5, $f_{\langle B',v\rangle} \leq \frac{1}{\sqrt{\mathrm{Var}\langle B',v\rangle}} \leq 4$. Hence, $f_{\langle v,B''\rangle} \leq 2\cdot f_{\langle v,B'\rangle} \leq 768$. Now, Lemma 1 implies that $\mathrm{Var}(\langle B'',v\rangle) \geq \frac{1}{768}$. Hence $\frac{1}{768}I_m \preccurlyeq \mathrm{Cov}(B') \preccurlyeq 2I_m$. Now by Lemma 34 anti-concentration holds with a constant parameter. $\square$

4.2.3 Proof of Theorem 3

Proof of Theorem 3. 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 23, 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 is gap is small with high probability over the randomness of A_{N_0}.

Set $\delta := \frac{\mathrm{poly}(m)\log n}{n}$, where the polynomial factor is the same one as dictated by Theorem 8. 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 8 to round x^* to a near optimal solution. By Lemma 22, first note that $(c_i, A_{;i}), i \in N_0$ are independent and distributed according to $u^{*\mathsf{T}} A_{\cdot,i} - c_i > 0$.

By Lemma 27, using rejection sampling we can sample a set $Z \subseteq N_0$, such that $\mathrm{Law}(A_{\cdot,i} | i \in Z) = \mathrm{Law}(B')$, $\Pr[i \in Z | i \in N_0] = \Omega(\delta)$ and such that $u^{*\mathsf{T}} A_{\cdot,i} - c_i \in [0, \delta]$ for all $i \in Z$. If columns of A have a discrete symmetric distribution (DSU), then $\mathbb{E}[B'] = 0$. Moreover, by Lemma 26, the distribution of B' 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 28, which achieves that the law of $A_{\cdot,i}, i \in Z$, is that of B'', which also admissible, anti-concentrated with parameter $\kappa = \Omega(1)$, and satisfies $\mathbb{E}[B''] = 0$. 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 $\mathbb{E}[Z] \geq \Omega(\delta |N_0|) = \Omega(\mathrm{poly}(m) \log n)$. Thus, by the Chernoff bound (3), $|Z| \geq \Omega(\mathrm{poly}(m) \log n)$ with probability at $1 - n^{-\mathrm{poly}(m)}$. 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 $p = \frac{\varepsilon \cdot \kappa^4}{1000 m^5}$, where $\varepsilon > 0$ is chosen small enough to have $\kappa^3 \exp\left(\frac{\kappa^3}{3 \cdot 80^2 p m^3}\right) \geq 50000 m^2$. We consider the rounded vector x', from Lemma 21, 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 8 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 21, which coupled with $|Z| p = \Omega(\mathrm{poly}(m) \log(n))$ and the fact $\max\limits_{i \in [n]} \|A_{\cdot,i}\| = O(\sqrt{\log(n)} + \sqrt{m})$, shows that, as long as the polynomial in the degree of δ is large enough,

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

Thus, for large n, Theorem 8 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 $1 - e^{-\Omega(p|Z|)} = 1 - n^{-\mathrm{poly}(m)}$ there exists a set $T \subseteq Z$ such that $|T| \leq \frac{3}{2} p |Z|$, and $\| \sum_{i \in T} A_i - t \| \leq 32 n^4 \exp(-\kappa^3 p |Z|/m)$ 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

$$
Ax'' \leq b \quad \text{and} \quad u^{*\mathsf{T}}(b - Ax'') \leq 1/\mathrm{poly}(n). \tag{15}
$$

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,

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

We conclude that $u^{*\mathsf{T}}(b - Ax'') = 0$ as needed.

In the (LI) setting, we first note that

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

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

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

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

$$\begin{aligned} \mathsf{IPGAP}(A,b,c) &= u^{*\mathsf{T}}(b - Ax'') + \left(\sum_{i=1}^n x_i''(A^\mathsf{T} u^* - c)_i^+ + (1 - x_i'')(c - A^\mathsf{T} u^*)_i^+\right) \\ &= u^{*\mathsf{T}}(b - Ax'') + \sum_{i\in T}(A^\mathsf{T} u^* - c)_i \quad \text{(by complementary slackness)} \\ &\le 1/\operatorname{poly}(n) + |T|\cdot\delta \qquad \text{(by (15) and } T \subseteq Z) \\ &\le O\left(\frac{\operatorname{poly}(m)\log(n)^2}{n}\right). \end{aligned}$$

□

4.3 The Gap Bound for Packing IPs

In this section we will prove Theorem 4. 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 $\{1,\dots,k\}$, $k \ge 3$.

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 [3].

We want to show $\mathsf{IPGAP}(A,b,c) \le \frac{\exp(O(1/\beta))\operatorname{poly}(m)(\log n)^2}{n}$ with probability at least $1 - n^{-\operatorname{poly}(m)}$. 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 $\|A(x^* - x')\| \le \sum_{i:x_i^*\in(0,1)} \|A_{\cdot,i}\| \le m\sqrt{m}$.

Similar to the proof of Theorem 3, 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 $A_{\cdot,i} - c_i \ge 0$. To be able to apply Theorem 2, we convert the conditional distribution of the columns of A back into their original distribution using rejection sampling:

Lemma 29. *Using rejection sampling on $(c_i, A_{\cdot,i}) \mid u^{*\top}A_{\cdot,i} - c_i \geq 0$, with an acceptance probability of at least $\frac{2}{3}\delta \exp(-\|u^*\|_1)$ as long as $\delta \leq \max(\frac{1}{k}, \frac{1}{m})\|u^*\|_1$, we can make sure that the accepted columns satisfy $u^{*\top}A_{\cdot,i} - c_i \in [0,\delta]$ and that $A_{\cdot,i}$, conditioned on acceptance, is distributed uniformly on $\{\frac{1}{k}, \dots, 1\} \cap [\frac{1}{3m}, 1]$.*

Proof. Let $B = A_{\cdot,i}$, set $M = \delta \exp(-\|u^*\|_1)$, and define the sampling procedure which accepts with probability:

$$\Pr[\text{acceptance} \mid B = y, c_i = x] = \begin{cases} \frac{M}{\Pr_{c_i}[u^{*\top}y - c_i \in [0,\delta] \mid u^{*\top}y - c_i \geq 0]} & \text{if } u^{*\top}y - x \in [0,\delta] \wedge y \in [\frac{1}{3m}, 1]^m \\ 0 & \text{else} \end{cases}$$

For $y \in \{\frac{1}{k}, \dots, 1\}^m \wedge [\frac{1}{3m}, 1]$ and $\delta \leq \max(\frac{1}{k}, \frac{1}{3m})\|u^*\|_1$, then $u^{*\top}y - \delta \geq 0$. Because c_i is exponentially distributed with parameter $\lambda = 1$, we have:

$$\begin{aligned} \Pr_{c_i}[u^{*\top}y - c_i \in [0,\delta] \mid u^{*\top}y - c_i \geq 0] &= \Pr_{c_i}[c_i \in [u^{*\top}y - \delta, u^{*\top}y] \mid c_i \leq u^{*\top}y] \\ &\geq \Pr_{c_i}[c_i \in [u^{*\top}y - \delta, u^{*\top}y]] \geq \delta \exp(-u^{*\top}y) \\ &\geq \delta \exp(-\|u^*\|_1) = M. \end{aligned}$$

Hence, the conditional acceptance probability always lies in $[0,1]$, so the sampling procedure is well defined.

Because, $\Pr[y \in [\frac{1}{3m}, 1]^m] \geq (1 - \frac{1}{3m})^m \geq \frac{2}{3}$, we have:

$$\begin{aligned} \Pr[\text{acceptance}] &= \mathbb{E}_{B,c_i}[\Pr[\text{acceptance} \mid B, c_i] \mid u^{*\top}B - c_i \geq 0] \\ &= \mathbb{E}_B\left[\frac{M \Pr_{c_i}\left[u^{*\top}B - c_i \in [0,\delta] \mid u^{*\top}y - c_i \geq 0\right] \mathbf{1}_{y \in [\frac{1}{3m},1]^m}}{\Pr_{c_i}\left[u^{*\top}B - c_i \in [0,\delta] \mid u^{*\top}y - c_i \geq 0\right]}\right] \\ &= M \cdot \Pr[y \in [\frac{1}{3m}, 1]^m] \geq \frac{2}{3}M. \end{aligned}$$

Call the law of B conditioned on acceptance B'. For $y \in \{\frac{1}{k}, \dots, 1\}^m \cap [\frac{1}{3m}, 1]^m$ with $u^{*\top}y - c_i \in [0,\delta]$ we have:

$$\begin{aligned} \Pr[B' = y] &= \frac{\Pr[B = y]\,\mathbb{E}_{c_i}[\Pr[\text{acceptance} \mid B = y, c_i] \mid u^{*\top} - c_i \geq 0]}{\Pr[\text{acceptance}]} \\ &= \frac{\Pr[B = y] M \Pr[u^{*\top}y - c_i \in [0,\delta] \mid u^{*\top}y - c_i \geq 0]}{\Pr[\text{acceptance}]\Pr[u^{*\top}y - c_i \in [0,\delta] \mid u^{*\top}y - c_i \geq 0]} = \Pr[B = y \mid y \in [\frac{1}{3m}, 1]^m]. \end{aligned}$$

So the distribution of B' is equal to that of B. □

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

$$3n \geq \sum_{i=1}^{n} c_i = c^{\mathsf{T}} \mathbf{1}_n \geq \mathrm{val}_{\mathsf{LP}}(x^*) = \mathrm{val}^*(u^*) \geq \sum_{i=1}^{m} b_i' u_i^* \geq \frac{n \cdot \beta}{2} \|u^*\|_1.$$

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

For the second claim, let $H : (0, \frac{1}{2}] \to (0, -\log(\frac{1}{2})]$ 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,

$$\begin{aligned}
\Pr\left[(Ax)_1 \leq b_1'\right] &\leq \Pr\left[(Ax)_1 \leq \frac{1-\beta}{2} n\right] = \Pr\left[(Ax)_1 - \frac{1-\alpha}{2} n \leq -\frac{\beta - \alpha}{2} n\right] \\
&\leq \Pr\left[(Ax)_1 - \frac{K}{2} \leq -\frac{\beta-\alpha}{2} n\right] = \Pr\left[(Ax)_1 - \mathbb{E}[(Ax)_1] \leq -\frac{\beta - \alpha}{2} n\right] \\
&\leq \exp\left(-(\beta-\alpha)^2 n\right) \leq \exp\left(-\frac{1}{4}\beta^2 n\right).
\end{aligned}$$

Let $S = \{x \in \{0,1\}^n : |\{i : x_i = 1\}| \geq (1-\alpha)n\}$. Note that by Lemma 12, $|S| \leq \sum_{i=0}^{\lfloor \alpha n \rfloor} \binom{n}{i} \leq \exp(H(\alpha)n)$. 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 $c_i - u^{*\mathsf{T}} A_{\cdot,i} \leq 0$, 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 (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 of Theorem 4. Let $r = \lceil \frac{10^6 m^{12} \log(n)}{s^2} \rceil$, where s is the constant given in Theorem 2. Let $\mu = \frac{k + \lceil \frac{k}{3m} \rceil}{2k} = \mathbb{E}[U]$, where $U \sim \mathrm{Uniform}(\{\frac{1}{k}, \dots, 1\} \cap [\frac{1}{3m}, 1])$. Now we define

$$\gamma = \frac{sr}{1000 m^5} \mu \text{ and } b' = b - \gamma \mathbf{1}. \tag{16}$$

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 30 this happens probability $1 - e^{-\Omega(\beta^2 n)}$. Subject to this, we condition on the exact values of x^*, u^*.

Let $\delta := \frac{3\exp(C_1/\beta)r}{C_2\beta^4 n}$. Note that by our assumption that $n \geq \mathrm{poly}(m)\exp(\Omega(1/\beta))$, we may assume that $\delta \leq \beta^4/(C_1 m) \leq \frac{\|u^*\|_1}{3m}$. Thus, by Lemma 29 we can sample a set $Z \subseteq N_0$ such that for $\mathrm{Law}(A_{\cdot,i}|i \in Z) = \mathrm{Uniform}((\{\frac{1}{k},\dots,1\} \cap [\frac{1}{3m},1])^m)$ and that $\Pr[i \in Z | i \in N_0] = \exp(-\|u^*\|_1)$. Noting that $\mathbb{E}[|Z|] = 2r$, by Chernoff's inequality (3), with probability at least $1 - n^{-\mathrm{poly}(m)}$, 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 2 on $kA_{\cdot,Z} \in \mathbb{Z}^{m \times r}$ and $kt \in \mathbb{Z}^m$ with parameter $p = \frac{\gamma}{\mu r} = \frac{s}{100m^5}$, 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 2, we first note that by Lemma 22 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 $|Z|p\mu$:

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

As a result, with probability $1 - \exp(-p|Z|) \geq 1 - n^{-\mathrm{poly}(m)}$ 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:

$$\begin{aligned} \mathsf{IPGAP}(A,b,c) &= \mathrm{val}_{\mathsf{LP}}(A,b,c) - \mathrm{val}_{\mathsf{IP}}(A,b,c) \leq \mathrm{val}_b^*(u^*) - \mathrm{val}_c(x'') \\ &= b^\top u^* + \sum_{i=1}^n (c - A^\top u^*)_i^+ - c^\top x'' \\ &= (b - Ax'')^\top u^* + \left(\sum_{i=1}^n x_i''(A^\top u^* - c)_i^+ + (1 - x_i'')(c - A^\top u^*)_i^+\right) \\ &= (b - Ax' - t)^\top u^* + \sum_{i \in T} (A^\top u^* - c)_i \qquad \text{(by complementary slackness)} \\ &= \sum_{i \in T} (A^\top u^* - c)_i \qquad \text{(since } (b - Ax' - t)^\top u^* = 0) \\ &\leq \delta|T| \leq \frac{(\exp(C_1/\beta)\,\mathrm{poly}(m)\log n)^2}{n}. \qquad \text{(by Lemma 29)} \end{aligned}$$

$\square$

5 The Size of the Branch-and-Bound Tree

In this section, we prove Corollary 1 in the discrete setting. This will follow by adapting the proof of Theorem 1 from [2].

The first ingredient is the key theorem from [22, Theorem 3], which relates the Branch-and-Bound tree size to the size of a certain knapsack.

Theorem 9. *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*

$$2n \cdot |\{x \in \{0,1\}^n : \sum_{i=1}^{n} x_i |(A^\mathsf{T} u^* - c)_i| \le \mathsf{IPGAP}(A,b,c)\}| + 1, \tag{17}$$

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 (17) in the setting of Corollary 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 31 ([2, Lemma 11])**.** *Let $\omega_1, \ldots, \omega_n \in \mathbb{R}$ be independent continuous random variables with maximum density at most* 1*. Then, for any $G \ge 0$, we have*

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

To get a bound on (17), we will use the above lemma for G being an upper bound on $\mathsf{IPGAP}(A,b,c)$, 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 in [2] 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 $W^\mathsf{T} := (c, A^\mathsf{T})$. 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 $(A^\mathsf{T} u - c)$ are also independent and have maximum density 1. By applying an appropriate union bound (e.g., see the proof of [2, Lemma 16]), one can derive the following bound over a family of knapsacks:

Lemma 32. *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 $n \ge m, R \ge 2, G \ge 1/n$, we have that*

$$\Pr[\max_{\|u\|_2 \le R} |\{x \in \{0,1\}^n : \sum_{i=1}^{n} x_i |(A^\mathsf{T} u - c)_i| \le G\}| \ge (nR)^{\Theta(m)} e^{2\sqrt{2nG}}] \le n^{-\Omega(m)}.$$

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

Lemma 31 implies that for any $u \in N'$, $\mathbb{E}[|K(u,G)|] \le e^{2\sqrt{2nG}}$. So, by the Markov bound, we see:

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

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

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

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

$$\sum_{i=1}^{n} x_i |(A^\top u' - c)_i| \le \sum_{i=1}^{n} \left(x_i |(A^\top u - c)_i| + x_i |(A^\top (u' - u))_i| \right) \le \sum_{i=1}^{n} x_i |(A^\top u - c)_i| + \|A^\top (u' - u)\|_1$$
$$\le \sum_{i=1}^{n} x_i |(A^\top u - c)_i| + \sqrt{m} n \varepsilon \le G + \frac{1}{n}.$$

So:

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

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

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

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

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

Proof of Corollary 1. As in section 4, 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 $\{0, \pm 1/k, \pm 2/k, \ldots, 1\}$ in the centered case or in $\{1/k, 2/k, \ldots, 1\}$ in the packing case, and hence all contained in $[-1, 1]$.

Let $K = \{x \in \{0,1\}^n : \sum_{i=1}^n x_i |(A^\top u^* - c)_i| \le \mathsf{IPGAP}(A,b,c)\}$. By Theorem 9, 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 3 we have that $\mathsf{IPGAP}(A,b,c)$ is at most $G := \mathrm{poly}(m)(\log n)^2/n$ with probability $1 - n^{-\mathrm{poly}(m)}$, by Lemma 23 that $\|u^*\| \le R := 32$ with probability $1 - e^{-\Omega(n)}$. Applying Lemma 32 with G, R together with the union bound, we conclude that $|K| \le n^{\mathrm{poly}(m)}$ with probability $1 - n^{-\Omega(m)}$, as needed.

In the packing case, by Theorem 4 the integrality gap $\mathsf{IPGAP}(A,b,c)$ is upper bound by $G := \exp(O(1/\beta))\,\mathrm{poly}(m)(\log n)^2/n$ with probability $1 - n^{-\mathrm{poly}(m)}$, and by Lemma 30 we have that $\|u^*\|_2 \le \|u^*\|_1 \le R := O(1/\beta)$ with probability $1 - e^{-\Omega(n)}$. Applying Lemma 32 with G, R together with the union bound, we conclude that $|K| \le n^{\exp(O(1/\beta))\,\mathrm{poly}(m)}$ with probability $1 - n^{-\Omega(m)}$, as needed. □

6 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 essentially prove Lemma 13.

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

Lemma 33. *Let X be a random variable with $\mathbb{E}[X] = \mu$ and $\mathrm{Var}(X) = \sigma^2$. 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 $\delta = \frac{\varepsilon^2}{12C}$,

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

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

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

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

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

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

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

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

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

We then arrive at the same conclusion. $\square$

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

Lemma 34. *Let X be a random vector in $\mathbb{R}^m$ with $\Sigma := \mathrm{Cov}(X)$. 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 $\theta \in \mathbb{R}^m$, $x \in \mathbb{R}$, $\rho_\theta(x) \leq \frac{C}{\sqrt{\mathrm{Var}(\langle X,\theta\rangle)}}$.*
- *For every $\nu \in \mathbb{R}^m$, $\Pr(\langle \nu, X\rangle \leq \langle \nu, \mu\rangle) \geq \frac{1}{C}$.*
- *$\|\Sigma\|_{\mathrm{op}}\|\Sigma^{-1}\|_{\mathrm{op}} \leq C$.*

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

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

where $\sigma := \|\Sigma\|_{\mathrm{op}}$. In other words, X satisfies (anti-concentration) *with constant $\frac{1}{48C^{3.5}}$.*

Before proving the result, we just note that by Lemmas 6 and 5, Lemma 34 applies to isotropic logconcave distributions and hence proves the first half of Lemma 13.

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

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

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

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

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 33, then

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

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

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

Thus, with a union bound

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

$\square$

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

Lemma 35. *Let $X = (X_1, \ldots, X_m)$ be a random vector in $\mathbb{R}^m$, such that $\{X_i\}_{i=1}^m$ are* i.i.d. *uniformly on $\{a, a+1, \ldots, a+k\}$, for some $a, k \in \mathbb{N}$, with $k > 1$. Set $\mu = \mathbb{E}[X_1]$ and $\sigma = \sqrt{\mathrm{Var}(X_1)}$. Then, for every $\theta \in [-\frac{1}{2}, \frac{1}{2}]^m$, and every $\nu \in \mathbb{R}^m$,*

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

In other words, X satisfies (anti-concentration) *with constant $\frac{1}{40}$.*

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

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

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

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

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

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

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

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

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

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

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

$$\frac{|F|}{|\mathrm{Support}(X_m|E)|} \geq \frac{1}{10}. \tag{18}$$

This can be done by inspecting the trajectory $a\theta_m + \{0, \theta_m, 2\theta_m, ..., \lceil \frac{k+1}{2} \rceil \theta_m\} \mod 1$ and noting that at most a $\frac{9}{10}$ fraction of the set $\{0, \theta_m, 2\theta_m, ..., \lceil \frac{k+1}{2} \rceil \theta_m\} \mod 1$ 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 (18) 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] 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.

[2] S Borst, D Dadush, S Huiberts, and S Tiwari. "On the Integrality Gap of Binary Integer Programs with Gaussian Data". In: *arXiv preprint arXiv:2012.08346 (v2)* (2021).

[3] 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.

[4] 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.

[5] 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.

[6] S Borst, D Dadush, S Huiberts, and S Tiwari. "On the Integrality Gap of Binary Integer Programs with Gaussian Data". In: *Integer Programming and Combinatorial Optimization*. Ed. by M Singh and DP Williamson. Cham: Springer International Publishing, 2021, pp. 427–442.

[7] 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.

[8] AM Frieze. *On the expected efficiency of branch and bound for the asymmetric TSP*. Preprint available at . 2020.

[9] 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.

[10] 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.

[11] G Kuperberg, S Lovett, and R Peled. "Probabilistic existence of regular combinatorial structures". In: *Geometric and Functional Analysis* 27.4 (2017), pp. 919–972.

[12] 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.

[13] 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.

[14] 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.

[15] R Vershynin. *High-dimensional probability: An introduction with applications in data science*. Vol. 47. Cambridge university press, 2018.

[16] A Prékopa. “Logarithmic concave measures with application to stochastic programming”. In: *Acta Scientiarum Mathematicarum* 32 (1971), pp. 301–316.

[17] L Lovász and S Vempala. “The geometry of logconcave functions and sampling algorithms”. In: *Random Structures & Algorithms* 30.3 (2007), pp. 307–358.

[18] 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.

[19] M Fradelizi. “Hyperplane sections of convex bodies in isotropic position”. In: *Beiträge Algebra Geom* 40.1 (1999), pp. 163–183.

[20] D Galvin. “Three Tutorial Lectures on Entropy and Counting”. en. In: *arXiv:1406.7872 [math]* (June 2014). arXiv: 1406.7872 [math].

[21] JA Tropp. “An introduction to matrix concentration inequalities”. In: *arXiv preprint arXiv:1501.01571* (2015).

[22] SS Dey, Y Dubey, and M Molinaro. *Branch-and-Bound Solves Random Binary IPs in Polytime*. arXiv preprint arXiv:2007.15192 (v4). 2021.