58th Annual IEEE Symposium on Foundations of Computer Science

Subdeterminant Maximization via Nonconvex Relaxations and Anti-Concentration

(Extended Abstract)

Javad B. Ebrahimi, Damian Straszak and Nisheeth K. Vishnoi
École Polytechnique Fédérale de Lausanne (EPFL), Switzerland

***Abstract*—Several fundamental problems that arise in optimization and computer science can be cast as follows: Given vectors $v_1, \ldots, v_m \in \mathbb{R}^d$ and a *constraint* family $\mathcal{B} \subseteq 2^{[m]}$, find a set $S \in \mathcal{B}$ that maximizes the squared volume of the simplex spanned by the vectors in S. A motivating example is the ubiquitous data-summarization problem in machine learning and information retrieval where one is given a collection of feature vectors that represent data such as documents or images. The volume of a collection of vectors is used as a measure of their *diversity*, and partition or matroid constraints over $[m]$ are imposed in order to ensure resource or fairness constraints. Even with a simple cardinality constraint ($\mathcal{B} = \binom{[m]}{r}$), the problem becomes NP-hard and has received much attention starting with a result by Khachiyan [1] who gave an $r^{O(r)}$ approximation algorithm for this problem. Recently, Nikolov and Singh [2] presented a convex program and showed how it can be used to *estimate* the value of the most diverse set when there are multiple cardinality constraints (i.e., when $\mathcal{B}$ corresponds to a partition matroid). Their proof of the integrality gap of the convex program relied on an inequality by Gurvits [3], and was recently extended to regular matroids [4], [5]. The question of whether these estimation algorithms can be converted into the more useful *approximation* algorithms – *that also output a set* – remained open.**

The main contribution of this paper is to give the first approximation algorithms for both partition and regular matroids. We present novel formulations for the subdeterminant maximization problem for these matroids; this reduces them to the problem of finding a point that maximizes the absolute value of a nonconvex function over a Cartesian product of probability simplices. The technical core of our results is a new anti-concentration inequality for dependent random variables that arise from these functions which allows us to relate the optimal value of these nonconvex functions to their value at a random point. Unlike prior work on the constrained subdeterminant maximization problem, our proofs do not rely on real-stability or convexity and could be of independent interest both in algorithms and complexity where anti-concentration phenomena has recently been deployed.

***Keywords*-Anti-concentration, Subdeterminant Maximization, Polynomials, Nonconvexity**

I. Introduction

A variety of problems in computer science and optimization can be formulated as the following constrained subdeterminant maximization problem: Given a positive semidefinite (PSD) matrix $L \in \mathbb{R}^{m\times m}$ and a family $\mathcal{B}$ of subsets of $[m] := \{1, 2, \ldots, m\}$, find a set $S \in \mathcal{B}$ that maximizes $\det(L_{S,S})$ where $L_{S,S}$ is the principal sub-matrix of L corresponding to rows and columns from S. Equivalently, if $L = V^\top V$ where $V \in \mathbb{R}^{d\times m}$ is a Cholesky decomposition of L, and $V_1, \ldots, V_m$ correspond to the columns of V, then the problem is to output a set $S \in \mathcal{B}$ that maximizes the squared volume of the parallelepiped spanned by the vectors $\{V_i : i \in S\}$. If the family $\mathcal{B}$ is specified explicitly as a list of its members, this optimization problem, trivially, has an efficient algorithm. The interesting case of the problem is when $|\mathcal{B}|$ is large (possibly exponential in m) and an efficient implicit representation or an appropriate separation oracle is given.

This problem, in its various avatars, has received significant attention in optimization, machine learning and theoretical computer science due to its practical importance and mathematical connections. In geometry and optimization, the vector formulation of the subdeterminant maximization problem for the family $\mathcal{B} = \binom{[m]}{r}$ is related to several volume maximization [6] and matrix low-rank approximation [7] problems. In mathematics, the probability distribution on $2^{[m]}$ in which a set $S \subseteq [m]$ has probability $\mathbf{Pr}(S) \propto \det(L_{S,S})$ is referred to as a determinantal point process (DPP); see [8]. DPPs are important objects of study in combinatorics, probability, physics and, more recently, in computer science as they provide excellent models for diversity in machine learning [9]. Here, the constrained subdeterminant maximization problem corresponds to a constrained MAP-inference problem – that of finding the most probable set from the family $\mathcal{B}$; see [10], [11] for related problems on DPPs. Different constraint families can be employed to ensure various priors, resource, or fairness constraints on the probability distribution.

Algorithmically, even the simplest of constraints make the constrained subdeterminant maximization problem NP-hard; for instance, when $\mathcal{B} = \binom{[m]}{r}$. As the set $\mathcal{B}$ becomes more complicated, algorithms for the constrained subdeterminant maximization problem roughly fall into two classes: 1) *approximation algorithms* that *output a set* $S \in \mathcal{B}$ such that $\det(L_{S,S})$ is within some factor of the optimal value and, (2) *estimation algorithms* that just *output a number* that is within some factor of the optimal value.

Approximation algorithms for the constrained subdetermi-

The full version of the paper is available at *arxiv.org/abs/1707.02757*.

0272-5428/17 $31.00 © 2017 IEEE
DOI 10.1109/FOCS.2017.98

nant maximization problem are rare; Khachiyan [1] proposed the first polynomial time approximation algorithm for the problem when $\mathcal{B} = \binom{[m]}{r}$ which achieved an approximation factor of $r^{O(r)}$ and, importantly, did not depend on the entries of the underlying matrix. This result was improved by Nikolov [12] who presented an approximation algorithm which achieved a factor of e^r. On the other hand, it was shown [13], [14] that there exists a constant $c > 1$ such that approximating the $\mathcal{B} = \binom{[m]}{r}$ case with approximation ratio better than c^r remains NP-hard.

Among estimation algorithms, recently, Nikolov and Singh [2] generalized Nikolov's result to the setting when the family $\mathcal{B}$ corresponds to the bases of a *partition matroid*. They presented an elegant convex program that allowed them to efficiently estimate the value of the maximum determinant set from $\mathcal{B}$ to within a factor of e^r where r is the size of the largest set in the partition matroid $\mathcal{B}$. One of the main ingredients in their proof is an inequality due to Gurvits [3] concerning real stable polynomials. Building on their work, [4], [5] presented estimation algorithms for large classes of families $\mathcal{B}$, such as bases of a regular matroid. While the results of [2], [4], [5] made interesting connections between convex programming, real-stable polynomials and matroids to design estimation algorithms for the constrained subdeterminant maximization problem, the question of whether these estimation algorithms can be converted into approximation algorithms remained open.

Making these approaches constructive is not only crucial for them to be deployed in the practical problems that motivated their study, mathematically, there seem to be barriers in doing so. The main contribution of this paper is to present a new methodology to address the constrained subdeterminant maximization problem that results in approximation algorithms for partition and regular matroids. We obtain our results through a synthesis of novel nonconvex formulations for these constraint families with a new anti-concentration inequality. Together, they allow for a simple polynomial time randomized algorithm that outputs a set $S \in \mathcal{B}$ with high probability. Approximation guarantees of our algorithms are close to prior non-constructive results in several interesting parameter regimes. The simplicity and generality of our results suggests that our techniques, in particular the anti-concentration inequality and its use in understanding nonconvex functions, are likely to find further applications.

A. Overview of Our Contributions

Anti-concentration inequality. We start by describing the common component to both our applications – an anti-concentration inequality. We consider multi-variate functions in which each variable is uniformly and independently distributed over a probability simplex. Roughly, our anti-concentration inequality says that if the restriction of such a function along each variable has a certain *anti-concentration* property then the function is anti-concentrated over the entire domain. Formally, the anti-concentration result applies whenever the multi-variate function satisfies the following property.

Definition I.1 (Anti-concentrated functions). *For $\gamma \geq 1$, a nonnegative measurable*[1] *function $f : \Delta_d \to \mathbb{R}$ is called* γ**-anti-concentrated** *if for every $c \in (0, 1)$*

$$\mathbf{Pr}\left[f(x) \geq c \cdot \mathrm{OPT}\right] \geq 1 - \gamma dc,$$

where x is drawn from the uniform distribution over Δ_d and $\mathrm{OPT} := \max_{z \in \Delta_d} f(z)$ is the maximum value f takes on Δ_d.[2]

Similarly, for any $r \geq 1$ and any $p_1, p_2, \ldots, p_r \geq 0$, a nonnegative function $f : \prod_{i=1}^{r} \Delta_{p_i} \to \mathbb{R}$ is said to be γ-anti-concentrated if for every coordinate $i \in \{1, 2, \ldots, r\}$, and for every choice of $a_j \in \Delta_{p_j}$ for $j \neq i$, the function $x \mapsto f(a_1, \ldots, a_{i-1}, x, a_{i+1}, \ldots, a_r)$ is γ-anti-concentrated.

Perhaps one of the simplest examples of an anti-concentrated function is the univariate map $t \mapsto |at + b|$ over the domain $[0, 1]$. It is not hard to see that it satisfies the condition of Definition I.1 for $\gamma = 2$ (see Fact IV.3). It also follows that for every multi-affine polynomial $p \in \mathbb{R}[x_1, x_2, \ldots, x_r]$ the function $x \mapsto |p(x)|$ is 2-anti-concentrated. Another class of functions that satisfy such an anti-concentration property arise by considering norms and volumes in Euclidean spaces; for instance, functions of the form $t \mapsto \|ut + (1 - t)v\|_2$ for vectors u, v.

Theorem I.1 (Anti-concentration inequality). *Let $\gamma \geq 1$ be a constant. Let $r \geq \gamma$ and $p_1, \ldots, p_r$ be positive integers. For every γ-anti-concentrated function $f : \prod_{i=1}^{r} \Delta_{p_i} \to \mathbb{R}$, if x is sampled from the uniform distribution on $\prod_{i=1}^{r} \Delta_{p_i}$, then*

$$\mathbf{Pr}\left[f(x) \geq (\gamma e^2)^{-r} \cdot \prod_{i=1}^{r} \frac{1}{p_i} \cdot \mathrm{OPT}\right] \geq \frac{1}{e^{\gamma} \log r},$$

where $\mathrm{OPT} := \max\{f(z) : z \in \prod_{i=1}^{r} \Delta_{p_i}\}$ is the maximum value f takes on its domain.

Consequently, the value of a γ-anti-concentrated function at a random point of its domain gives an estimate of its maximum value. As an important special case of Theorem I.1, consider the setting in which $p_i = 2$ for $i = 1, 2, \ldots, r$ (i.e., the domain is the hypercube $[0, 1]^r$) and $f(x) := |p(x)|$ where $p \in \mathbb{R}[x_1, \ldots, x_r]$ is a multi-affine polynomial. Using the previous observation that such an f is 2-anti-concentrated, we conclude from Theorem I.1 that for some

[1] We always assume that the functions we deal with are *regular enough*. Formally, we require measurability with respect to the Lebesgue measure.

[2] Δ_d denotes the standard $(d-1)-$simplex, i.e., $\Delta_d := \left\{x \in \mathbb{R}^d : \sum_{i=1}^{d} x_i = 1, x \geq 0\right\}$.

absolute constant $c > 1$ and a uniformly random choice of $x \in [0,1]^r$ it holds that

$$\mathbf{Pr}\left[|p(x)| \geq c^{-r} \cdot \max_{z \in [0,1]^r} |p(z)|\right] \geq \Omega\left(\frac{1}{\log r}\right). \quad (1)$$

It is also not hard to see that the bound in Theorem I.1 is tight: for $p(x) = \prod_{i=1}^r x_i$, the probability that $|p(x)| \geq (3/4)^r$ over a random choice of $x \in [0,1]^r$ is exponentially small. The bound (1) gives us a way to *estimate* the maximum of $|p(x)|$ over $[0,1]^r$ by just evaluating it on a certain number of random points and outputting the largest one. However, this observation does not directly give us much insight about the problem we typically would like to solve; that of maximizing $|p(b)|$ over binary vectors $b \in \{0,1\}^r$. Towards this, note that for a multi-affine polynomial p,

$$\max_{z \in \{0,1\}^r} |p(z)| = \max_{z \in [0,1]^r} |p(z)|.$$

Moreover, the above has a simple algorithmic proof (see Lemma IV.2) which follows from the convexity of $x \mapsto |p(x)|$ restricted to coordinate-aligned lines. This allows us to use the above algorithm to find a point $b \in \{0,1\}^r$ whose value is at most c^r times worse than optimal given only an evaluation oracle for p. In particular, no assumptions are made on the analytic properties of p, such as concavity or real stability. In fact, in most interesting cases, such functions are highly nonconvex, hence standard convex optimization tools do not apply.

Partition matroids. As a first application of Theorem I.1, we provide an approximation algorithm for the problem of subdeterminant maximization under partition constraints. Let $\mathcal{P} := \{M_1, M_2, \ldots, M_t\}$ be a partition of $[m] := \{1, 2, \ldots, m\}$ into non-empty, pairwise disjoint subsets and let $b = (b_1, b_2, \ldots, b_t)$ be a sequence of positive integers. Then the set $\mathcal{B} := \{S \subseteq [m] : |S \cap M_i| = b_i \text{ for all } i = 1, 2, \ldots, t\}$ is called a partition family induced by $\mathcal{P}$ and b. We first show that the problem of finding the determinant-maximizing set under partition constraints can be reformulated as

$$\max_{x \in \Delta} \det\left(W(x)^\top W(x)\right)^{1/2}$$

where Δ is a certain product of simplices, and $W(x)$ is a matrix whose i-th column is a convex combination of certain vectors derived from $L = V^\top V$ and the variables in x. Subsequently, we show that such functions are 2-anti-concentrated, which allows us to apply Theorem I.1 to obtain the following result.

Theorem I.2 (Subdeterminant maximization under partition constraints)**.** *There exists a polynomial time randomized algorithm such that given a PSD matrix $L \in \mathbb{R}^{m \times m}$, a partition $\mathcal{P} = \{M_1, M_2, \ldots, M_t\}$ of $[m]$ and a sequence of numbers $b = (b_1, b_2, \ldots, b_t) \in \mathbb{N}^t$ with $\sum_{i=1}^t b_i = r$, outputs a set S in the induced partition family $\mathcal{B}$ such that with high probability*

$$\det(L_{S,S}) \geq \mathrm{OPT} \cdot (2e)^{-2r} \cdot \prod_{i=1}^t \left(\frac{1}{p_i}\right)^{b_i},$$

where $\mathrm{OPT} := \max_{S \in \mathcal{B}} \det(L_{S,S})$ *and* $p_i := |M_i|$ *for* $i = 1, 2, \ldots, t$.

Prior work by Nikolov and Singh [2] outputs a random set whose value is at most e^r times worse than OPT in *expectation* and unlike the theorem above, does not yield a polynomial time approximation algorithm, as the probability of success can be exponentially small. Further, in the case when $p_i = O(1)$ for all i and $b_i = 1$ for all i (i.e., when every part has constant size and exactly one vector from every part has to be selected) the approximation ratio of our algorithm is c^r for some constant $c > 1$, which, up to the constant in the base of the exponent, matches their result.

Regular matroids. Our second result for the constrained subdeterminant maximization problem is for the case of regular matroids (i.e., when the constraint family $\mathcal{B}$ arises as a set of bases of a regular matroid; see Section II). To apply Theorem I.1 we consider the polynomial

$$h(x) = \det(VXB^\top),$$

where X is a diagonal matrix with $X_{i,i} := x_i$, $B \in \mathbb{R}^{d \times m}$ is the linear representation of $\mathcal{B}$ and $V \in \mathbb{R}^{d \times m}$ is such that $V^\top V = L$. We remark that this polynomial has also appeared in previous work on *matroid intersection* (see [15], [16], [17]). We observe that $|h(x)|$ is 2-anti-concentrated and has a number of desirable properties, which allows us to prove

Theorem I.3 (Subdeterminant maximization under regular matroid constraints)**.** *There exists a polynomial time randomized algorithm such that given a PSD matrix $L \in \mathbb{R}^{m \times m}$ of rank d, and a totally unimodular matrix B that is a representation of a rank-d regular matroid with bases $\mathcal{B} \subseteq 2^{[m]}$, outputs a set $S \in \mathcal{B}$ such that with high probability*

$$\det(L_{S,S}) \geq \max(2^{-O(m)}, 2^{-O(d \log m)}) \cdot \mathrm{OPT},$$

where $\mathrm{OPT} := \max_{S \in \mathcal{B}} \det(L_{S,S})$.

There are two recent results for this setting ([4] and [5]) that provide e^m- and e^d-estimation algorithms respectively. Similarly as for the case of partition matroids, these results only give an estimate on the value of the optimal solution, and are not constructive. Our algorithm matches the approximation guarantee of the above mentioned results in certain regimes and also outputs an approximately optimal set.

B. Discussion and Future Work

To summarize, motivated by applications in machine learning, we propose and analyze two algorithms for subdeterminant maximization under matroid constraints. Both are based on random sampling and the bounds on their approximation guarantees follow from our anti-concentration result. These algorithms provide both an estimate to the value of the optimal solution as well as a set with the claimed guarantee. The anti-concentration inequality allows us to relate the value of a multi-variate nonconvex function at a random point to its value at the optimal point, and multi-affinity allows us to round this random solution. Furthermore, the anti-concentration result can be applied to *any* multi-affine polynomial and more general functions involving norms and volumes. In particular, it neither relies on real stability nor any other convexity-like property of the polynomial; this should be of independent interest.

An interesting question that arises is whether our algorithms for subdeterminant maximization, or more generally, our anti-concentration results, can be derandomized. In other words, given an anti-concentrated function on the hypercube, can one efficiently and *deterministically* find a point matching the guarantee of Theorem I.1? Another question is whether Theorem I.1 can be extended to more general convex bodies – other than products of simplices. Of interest are, for instance, sets that arise as an intersection of the hypercube $[0,1]^m$ with an affine subspace of dimension $(m-1)$. An anti-concentration inequality for such sets, together with an improved rounding scheme, would imply approximation ratios which depend on the rank of the underlying matroid only – not on the number of elements.

C. Other Related Work

A very general anti-concentration result for polynomial functions over convex domains was obtained by Carbery and Wright [18], however there seem to be two issues in applying their result to our setting: A) it implies a weaker bound of $r^{-O(r)}$ in Equation (1) to obtain a significant probability of success and, B) it does not seem to directly apply to the product of simplices as we need. A more detailed discussion is presented in Section IV. The result by Carbery and Wright and, more generally, the anti-concentration phenomena has found several applications in theoretical computer science, especially for Gaussian measures; see for instance [19], [20], [21], [22]. Finally, our use of rounding using multi-affinity resembles a similar phenomena in algorithms to optimize concave or sub-modular functions; see for instance a survey by Vondrák [23].

D. Technical Overview

We start by describing the approach of Nikolov and Singh for the case of partition matroids. Consider the following simple variant of the constrained subdeterminant maximization problem for partition matroids: Given vectors $v_1, \ldots, v_r, u_1, \ldots, u_r \in \mathbb{R}^r$ the goal is to pick a vector $w_i \in \{v_i, u_i\}$ for each i so as to maximize $|\det(W)|$, where $W \in \mathbb{R}^{r\times r}$ is a matrix that has the w_is as its columns. Denote by OPT the maximum value of the determinant in the above problem.

They start by reformulating the problem as polynomial maximization problem as follows. First, define matrices $A_i(x_i) := x_i v_i v_i^\top + (1-x_i)u_i u_i^\top$ for $i = 1, 2, \ldots, r$. Then, consider the polynomial $p(x, y) := \det\left(\sum_{i=1}^r y_i A_i(x_i)\right)$ and let $g(x)$ be the polynomial that appears as the coefficient of $\prod_{i=1}^r y_i$ in $p(x, y)$.[3] Multi-affinity of g can be used to reduce the task of finding OPT to that of finding $\max_{x\in[0,1]^r} g(x)$. Then, the difficulty that arises is that $g(x)$ is hard to evaluate. To bypass this, a general idea by Gurvits [3] allows them to approximate $g(x)$ by $\inf_{y>0} \frac{p(x,y)}{\prod_{i=1}^r y_i}$, giving rise to the following optimization problem involving two sets of variables

$$\max_{x\in[0,1]^r} \inf_{y>0} \frac{p(x,y)}{\prod_{i=1}^r y_i}. \tag{2}$$

Real stability of $p(x, y)$ for any fixed x implies that this program can be efficiently solved using convex programming. Their main result is that the value of this program is within a factor of e^r of OPT. The key component in the proof of this bound is the above-mentioned result by Gurvits that, in this context where $p(x, y)$ is real-stable with respect to y, implies that, for all $x \in [0,1]^r$

$$g(x) \leq \inf_{y>0} \frac{p(x,y)}{\prod_{i=1}^r y_i} \leq e^r \cdot g(x). \tag{3}$$

While this immediately implies that one can obtain a number that is within an e^r factor of OPT, when trying to obtain an integral solution $x \in \{0,1\}^r$ from the fractional optimal solution $x^\star \in [0,1]^r$ to (2), the intractability of $g(x)$ becomes a bottleneck.[4] Nikolov and Singh present a rounding algorithm which, unfortunately, can require an exponential number of trials to find an e^r-approximate solution; we refer to the full version of the paper for an example.

Overview of the proof of Theorem I.2. Our approach is based on a different formulation of the problem as polynomial maximization, which has the advantage over $g(x)$ that *it is easy to evaluate and does not rely on real-stability*. For every $i = 1, 2, \ldots, r$ and $t \in [0,1]$ define a vector $w_i(t) := (1-t)v_i + tu_i$. Furthermore, for $x \in [0,1]^r$, let $W(x) \in \mathbb{R}^r$ be a matrix with columns $w_1(x_1), w_2(x_2), \ldots, w_r(x_r)$. The polynomial that we consider is

$$\det(W(x))$$

which is easy to evaluate for any x. As before, the multi-affinity of $\det(W(x))$ implies the following:

$$\max_{x\in[0,1]^r} |\det(W(x))| = \max_{x\in\{0,1\}^r} |\det(W(x))| = \text{OPT}. \tag{4}$$

[3] $g(x)$ is also called the mixed-discriminant of the matrices $A_i(x_i)$.

[4] One can use Equation (3) r times to give an approximation algorithm with factor e^{r^2}; we omit the details.

Indeed, if we let $f(x) := |\det(W(x)|$, then the multi-affinity of $\det(W(x))$ implies that whenever we fix all but one of the arguments of f, i.e., $s(t) := f(t, y_2, y_3, \ldots, y_r)$ for some $y_2, y_3, \ldots, y_r \in [0,1]$, then s attains its maximum at either 0 or 1. This means, in particular, that given any point $x \in [0,1]^r$, one can efficiently find a point $\tilde{x} \in \{0,1\}^r$ such that $f(\tilde{x}) \geq f(x)$.

However, the nonconvexity of this formulation is a serious obstacle to solving the optimization problem in Equation (4). This is where a key insight comes in: f shows a remarkable anti-concentration property which, in turn, allows us to get an estimate of OPT by evaluating f at a random point. Formally, the anti-concentration inequality (Theorem I.1) applies to f and allows us to deduce that

$$\mathbf{Pr}\left[f(x) \geq c^{-r} \cdot \mathrm{OPT}\right] \geq \tfrac{1}{e^2 \log r}$$

for some constant $c > 1$. This also results in a simple approximation algorithm to maximize f: Sample a point $x \in [0,1]^r$ uniformly at random, round x to a vertex $\tilde{x} \in \{0,1\}^r$ such that $f(\tilde{x}) \geq f(x)$ as above, and output $\tilde{x}$ as a solution.

We should mention that at this point we could also attempt to invoke the following anti-concentration result (here translated to our setting) proved by Carbery and Wright.

Theorem I.4 (Theorem 2 in [18])**.** *Let $p \in \mathbb{R}[x_1, x_2, \ldots, x_r]$ be a polynomial of degree r. If a point x is sampled uniformly at random from the hypercube $[0,1]^r$, then for every $\beta \in (0,1)$*

$$\mathbf{Pr}\left[|p(x)| \leq \beta^r \cdot \mathrm{OPT}\right] \leq C \cdot \beta \cdot r,$$

where $C > 0$ is an absolute constant.

When applied to our setting, observe that $\det(W(x))$ is indeed a degree-r polynomial in r variables. We have to pick β so as to make $C \cdot \beta \cdot r < 1$, i.e., for $\beta = O(1/Cr)$, we obtain

$$\mathbf{Pr}\left[f(x) \geq r^{-O(r)} \cdot \mathrm{OPT}\right] \geq \tfrac{1}{2}.$$

This implies that the algorithm described above achieves an approximation ratio of (roughly) r^r. Our Theorem I.1 is a certain strengthening of Theorem I.4 which asserts that under the same assumptions

$$\mathbf{Pr}\left[|p(x)| \geq c^{-r} \cdot \mathrm{OPT}\right] \geq \frac{1}{e^2 \log r},$$

for some absolute constant $c > 1$. In fact, Theorem I.1 is a generalization of the above for a larger class of functions (not only polynomials) and for more general domains – this is useful in the case of general partition matroids.

We now show how to extend our algorithm to a general instance of the constrained subdeterminant maximization problem under partition constraints and sketch a proof of Theorem I.2. Recall that in this problem we are given a PSD matrix $L \in \mathbb{R}^{m \times m}$ of rank d and a partition family $\mathcal{B}$ induced by a partition of $[m]$ into disjoint sets $M_1, M_2, \ldots, M_t$ and numbers $b_1, b_2, \ldots, b_t \in \mathbb{N}$ with $\sum_{i=1}^t b_i = r$. The goal is to find a subset $S \in \mathcal{B}(\mathcal{M})$ such that $\det(L_{S,S})$ is maximized. If we consider a decomposition of L into $L = V^\top V$ for $V \in \mathbb{R}^{d \times m}$ then the objective can be rewritten as $\det(L_{S,S}) = \det(V_S^\top V_S)$. For simplicity, we assume that $b_1 = b_2 = \cdots = b_t = 1$, which can be achieved by a simple reduction. To define the relaxation for the general case, for every part M_i for $i = 1, 2, \ldots, t$, introduce a vector $x^i \in \Delta_{p_i}$ where $p_i := |M_i|$ and define a vector $w^i(x^i)$ to be

$$w^i(x^i) := \sum_{j=1}^{p_i} x^i_j v^i_j$$

where $v^i_1, v^i_2, \ldots, v^i_{p_i}$ are the columns of V corresponding to indices in M_i. We denote by x the vector $(x^1, x^2, \ldots, x^r)$ and by $W(x) \in \mathbb{R}^{d \times r}$ the matrix with columns $w^1(x^1), w^2(x^2), \ldots, w^r(x^r)$. Finally we let

$$f(x^1, x^2, \ldots, x^r) := \det(W(x)^\top W(x))^{1/2}.$$

Note that $f(x)$ is no longer a multi-affine polynomial, but as we show in Lemma IV.1 it is 2-anti-concentrated. Having established this property, Theorem I.2 follows. Indeed, as in the illustrative example in the beginning, we can prove that given any fractional point x, we can efficiently find its integral rounding (i.e., round every component x^i to a vertex of the corresponding simplex Δ_{p_i}, for $i = 1, 2, \ldots, t$) which then provides us with a suitable approximate solution.

Overview of the proof of Theorem I.3. In the setting of Theorem I.3 we are given a PSD matrix $L \in \mathbb{R}^{m \times m}$ of rank d and a family of bases $\mathcal{B} \subseteq 2^{[m]}$ of a regular matroid of rank d. The goal is to find a set that attains $\mathrm{OPT} := \max_{S \in \mathcal{B}} \det(L_{S,S})$. The approach of [4] (and similarly [5]) to obtain an estimate on OPT was inspired by that of [2] for the partition matroid case and is as follows: Given the matrix $L = V^\top V$, first, define the following polynomial

$$g(x) := \textstyle\sum_{S \in \mathcal{B}} x^S \det(V_S^\top V_S).$$

This polynomial again turns out to be hard to evaluate. As before, an optimization problem involving two sets of variables, x and y is set up in [4]. The purpose of y variables is to give estimates of values of $g(x)$ and the x variables are constrained to be in the matroid base polytope corresponding to $\mathcal{B}$. On the one hand, real stability along with the fact that $\mathcal{B}$ is a matroid allows them to compute the optimal solution to this bivariate problem, on the other hand, with some additional effort, they are able to push Gurvits' result to obtain roughly an e^m estimate of OPT. However, the main bottleneck is that an iterative rounding approach for finding an approximate integral point does not seem possible as the matroid polytope corresponding to $\mathcal{B}$ may not have a product structure as in the partition matroid case.

We present a new formulation to capture OPT that does not suffer from the intractability of the objective function and allows for rounding via a relaxation that maximizes a certain function h over the hypercube $[0,1]^m$. Start by noting that the objective becomes $\det(L_{S,S}) = \det(V_S^\top V_S) = \det(V_S)^2$, which we can simply think of as maximizing $|\det(V_S)|$ over $S \in \mathcal{B}$. Let $B \in \mathbb{Z}^{m\times d}$ be the linear representation of the matroid $\mathcal{B}$; i.e., for every set $S \subseteq [m]$ of size d, if $S \in \mathcal{B}$ then $|\det(B_S)| = 1$, and $\det(B_S) = 0$ otherwise. Next, consider $h : [0,1]^m \to \mathbb{R}$ given by

$$h(x) := \det(VXB^\top),$$

where $X \in \mathbb{R}^{m\times m}$ is a diagonal matrix with $X_{i,i} := x_i$ for all $i = 1, 2, \ldots, m$. It is not hard to see that $h(x)$ is a polynomial in x and (using the Cauchy-Binet formula) can be written as

$$h(x) = \sum_{S\subseteq[m],|S|=d} x^S \det(V_S)\det(B_S),$$

where x^S denotes $\prod_{i\in S} x_i$. Such a function was studied before in the context of matroid intersection problems [15], [16], [17]. Importantly, the restriction of $h(x)$ to indicator vectors of sets of size d is particularly easy to understand. Indeed, let 1_S be the indicator vector of some set $S \subseteq [m]$ with $|S| = d$. We have

$$h(1_S) = \det(V_S)\det(B_S) = \begin{cases} \pm\det(V_S) & \text{if } S \in \mathcal{B}, \\ 0 & \text{if } S \notin \mathcal{B}. \end{cases}$$

Hence, we are interested in the largest magnitude coefficient of a multi-affine polynomial $h(x)$. The maximum of $|h(x)|$ over $[0,1]^m$ is an upper bound for this quantity. The algorithm then simply selects a point $x \in [0,1]^m$ at random, which by Theorem I.1 can be related to the maximum value of $|h(x)|$, and then performs a rounding.

First, given $x \in [0,1]^m$ it constructs a binary vector $\tilde{x} \in \{0,1\}^m$ such that $|h(\tilde{x})| \geq |h(x)|$; this is possible because the function $|h(x)|$ is convex along any coordinate direction. The vector $\tilde{x}$ is then treated as a set $S_0 \subseteq [m]$, but its cardinality is typically larger than d. We then run another procedure which repeatedly removes elements from S_0 while not loosing too much in terms of the objective. It is based on using $h(1_{S_0})$ as a certain proxy for the sum $\sum_{S\subseteq S_0} |\det(V_S)\det(B_S)|$. This allows us to finally arrive at a set $S \subseteq S_0$ of cardinality d, such that $|h(1_S)| \geq \binom{m}{d}^{-1} |h(1_{S_0})|$. The set S is then the final output.

By applying Theorem I.1 one can conclude that $h(1_{S_0})$ is within a factor of c^m of the maximal value of $|h(x)|$, which results in a $2^{O(m)}$-approximation guarantee. Alternatively, by utilizing the fact that h is a polynomial of degree d, one can apply the result by Carbery-Wright (see Theorem I.4) to obtain a bound of roughly $m^{O(d)}$, which is better whenever m is large compared to d.

Overview of the proof of Theorem I.1. For the sake of clarity, we present only the hypercube case of the anti-concentration inequality, which corresponds to taking $p_1 = p_2 = \cdots = p_r = 2$ in the statement of Theorem I.1. Recall the setting: We are given a function $f : [0,1]^r \to \mathbb{R}_{\geq 0}$ that satisfies a one-dimensional anti-concentration inequality. I.e., for every function of the form $g(t) := f(x_1, x_2, \ldots, x_{i-1}, t, x_{i+1}, \ldots, x_r)$ where $x_j \in [0,1]$ for $j \neq i$ are fixed and $t \in [0,1]$, it holds that

$$\mathbf{Pr}\left[g(t) < c \cdot \max_{s\in[0,1]} g(s)\right] \leq 2c, \tag{5}$$

where the probability is over a random choice of $t \in [0,1]$. The goal is to prove a similar statement for $f(x)$, i.e., $\mathbf{Pr}[f(x) < \alpha \cdot \mathrm{OPT}]$ is small, where OPT is the maximum value f takes on the hypercube and α is a parameter which we want to be as large as possible.

As an initial approach, one can define (for a fixed constant $c > 0$) events of the form $A_i := \{x \in [0,1]^r : f(x_1, \ldots, x_i, x^\star_{i+1}, \ldots, x^\star_r) \geq c \cdot f(x_1, \ldots, x_{i-1}, x^\star_i, \ldots, x^\star_r)\}$, where $x^\star := \mathrm{argmax}_x f(x)$. Note crucially that the events $A_1, A_2, \ldots, A_r$ are not independent. However, we can still write

$$\begin{aligned}\mathbf{Pr}[f(x) \geq c^n \cdot \mathrm{OPT}] &\geq \mathbf{Pr}[A_1 \cap A_2 \cap A_3 \cdots \cap A_r] \\ &= \prod_{i=1}^{r} \mathbf{Pr}[A_i | A_1, A_2, \ldots, A_{i-1}].\end{aligned}$$

From assumption (5) we know that

$$\mathbf{Pr}[A_i | A_1, A_2, \ldots, A_{i-1}] \geq 1 - 2c$$

for all $i = 1, 2, \ldots, r$ and hence

$$\mathbf{Pr}[f(x) \geq c^r \cdot \mathrm{OPT}] \geq (1-2c)^r.$$

To get a probability that is not exponentially small, one has to take c roughly $O(1/r)$, in which case we recover the result of Carbery and Wright [18] in our setting. To go beyond this, a tighter analysis is required.

In what follows, let $k \approx \log r$ and $\delta \approx \frac{1}{k}$. First, using a recursive procedure, we construct a family of k^r sets $\mathcal{S}(i_1, i_2, \ldots, i_r)$ for $i_1, i_2, \ldots, i_r \in \{1, 2, \ldots, k\}$ that are pairwise disjoint and each of them has the same volume (roughly δ^r). In particular, the total volume of all of the sets (which we call *cells*) is $k^r \cdot \delta^r = \Omega(1)$, and hence, form a significant part of the probability space. Additionally, the construction guarantees that for all points x in a given cell $\mathcal{S}(i_1, i_2, \ldots, i_r)$,

$$f(x) \geq \mathrm{OPT} \cdot \left(\frac{r}{m}\right)^r \prod_{j\in[r], i_j\neq k} \frac{k - i_j}{k} \prod_{j\in[r], i_j = k} \frac{1}{r}. \tag{6}$$

Notice that in the above bound, if $i_j = k$ for all $j \in [r]$, then we obtain a very weak bound $f(x) \geq \mathrm{OPT} \cdot \left(\frac{1}{m}\right)^{-r}$ for the corresponding cell.

In the proof we identify a set of cells G (which we call *good*) such that (6) guarantees that $f(x) \geq c^{-r} \cdot \left(\frac{r}{m}\right)^r$ for a constant $c > 0$. Subsequently, we prove that at least $\frac{1}{k}$-fraction of all cells are good. This is achieved by defining an action of the cyclic group of order k on the set of cells, and observing that at least one cell in each orbit is good. The reason is as follows: If we repeatedly apply (*entrywise*) a cyclic shift ($i \mapsto (i+1) \bmod k$) to a tuple $(i_1, \ldots, i_r)$, we obtain k different tuples each of which defines a cell. We prove that at least one of them is good. To this ends let us take the product of all upper bounds following from (6) for k cells in one such orbit. We obtain

$$\mathrm{OPT}^k \cdot \left(\frac{r}{m}\right)^{rk} \cdot \left(\frac{1}{r}\right)^r \cdot \prod_{i=1}^{k-1} \left(\frac{i}{k}\right)^r \approx \mathrm{OPT}^k \cdot \left(\frac{r}{m}\right)^{rk} \cdot e^{-kr} \cdot r^{-r}$$

Hence by taking the k-th root of the above, we can conclude that for at least one of the cells $\mathcal{S}(i'_1, i'_2, \ldots, i'_r)$ in the considered orbit the following bound holds for all $x \in \mathcal{S}(i'_1, i'_2, \ldots, i'_r)$

$$f(x) \geq \mathrm{OPT} \cdot \left(\frac{r}{m}\right)^r \cdot e^{-r} \cdot r^{-r/k} \geq \mathrm{OPT} \cdot \left(\frac{r}{m}\right)^r \cdot c^{-r}$$

for some constant $c > 0$, since $k \approx \log r$. As all the cells are disjoint, have the same volume and the volume of their union is $\Omega(1)$, the inequality $f(x) \geq \mathrm{OPT} \cdot \left(\frac{r}{m}\right)^r c^{-r}$ holds for at least a $\Omega\left(\frac{1}{\log(r)}\right)$ fraction of the space. This completes the sketch of the proof of Theorem I.1.

E. Organization of the Rest of the Paper

We introduce notation and give some background on matroids in Section II. In Section III we present the proof of our anti-concentration result, Theorem I.1. In Section IV we give a proof of Theorem I.2 for partition matroids. The proof of Theorem I.3 (for regular matroids) appears in Section V. Due to space limitations, some proofs are omitted; they appear in the full version of the paper.

II. Preliminaries

Notation. Let $[m]$, $2^{[m]}$ and $\binom{[m]}{d}$ denote the sets $\{1, 2, \ldots, m\}$, the set of all subsets of $[m]$ and the set of all subsets of $[m]$ of size d, respectively. For any subset S of $[m]$, we denote the indicator vector of S by $1_S \in \mathbb{R}^d$. The standard basis vectors for $\mathbb{R}^d$ are denoted by $e_1, e_2, \ldots, e_d$, i.e., e_i stands for the vector having 1 in the i-th coordinate and zeros everywhere else. For a matrix $V \in \mathbb{R}^{d \times m}$, the columns of V are denoted by $V_1, V_2, \ldots, V_m \in \mathbb{R}^d$. The d-dimensional Lebesgue measure (volume) on $\mathbb{R}^d$ is denoted by λ_d. When the dimension is clear from the context, we use λ to denote the volume. Throughout this paper, the probability distributions we consider, are typically uniform over an appropriate domain.

The standard $(d-1)$-simplex, denoted by Δ_d is defined as the convex hull of $e_1, e_2, \ldots, e_d \in \mathbb{R}^d$. Notice that Δ_d is a $(d-1)$-dimensional polytope which is embedded in $\mathbb{R}^d$, and it inherits a $(d-1)$-dimensional Lebesgue measure from the hyperplane it lies on. We use μ_d to denote the induced measure λ_d on the simplex Δ_d, normalized so that $\mu_d(\Delta_d) = 1$. We often deal with Cartesian products of simplices, which we denote by $\Delta = \prod_{i=1}^r \Delta_{p_i}$, for some sequence $p_1, p_2, \ldots, p_r \in \mathbb{N}$. For a point $x \in \Delta$, by x^i we denote i-th component of x belonging to Δ_{p_i} and x^i_j for $j \in [p_i]$ are the components of x^i within Δ_{p_i}. By $V(\Delta)$, we denote the set of points of Δ with integer coordinates. We call $V(\Delta)$ the set of vertices of Δ.

For any vector $x \in \mathbb{R}^m$ by $X \in \mathbb{R}^{m \times m}$ we denote the diagonal matrix, such that $X_{i,i} = x_i$ for all $i \in [m]$. For any two closed subsets $S_1, S_2 \subseteq \mathbb{R}^d$, we denote by $\mathrm{dist}(S_1, S_2)$ the distance between these two sets, formally defined as

$$\mathrm{dist}(S_1, S_2) := \min_{s_1 \in S_1, s_2 \in S_2} \|s_1 - s_2\|_2$$

where $\|\cdot\|_2$ is the standard ℓ_2-norm.

Multi-affine functions. A function $f : \mathbb{R}^m \to \mathbb{R}$ is called affine when f is a polynomial whose total degree is at most one. A function $f : \mathbb{R}^m \to \mathbb{R}$ is called multi-affine if f is a polynomial function where the degree of each variable is at most 1. Suppose that $x_1, \ldots, x_m$ are m variables. We denote the monomial $\prod_{i \in S} x_i$ by x^S for every $S \subseteq [m]$. Every multi-affine function can be written in the form $f(x) = \sum_{S \subseteq [m]} f_S x^S$ where f_S's are real numbers, called the coefficients of f.

Matroids. For a comprehensive treatment of matroid theory we refer the reader to [24]. Below we state the most important definitions and examples of matroids, which are most relevant to our results. A matroid is a pair $\mathcal{M} = (U, \mathcal{I})$ such that U is a finite set and $\mathcal{I} \subseteq 2^U$ satisfies the following three axioms: (1) $\emptyset \in \mathcal{I}$, (2) if $S \in \mathcal{I}$ and $S' \subseteq S$ then $S' \in \mathcal{I}$, (3) if $A, B \in \mathcal{I}$ and $|A| > |B|$, then there exists an element $a \in A \setminus B$ such that $B \cup \{a\} \in \mathcal{I}$. The collection $\mathcal{B} \subseteq \mathcal{I}$ of all inclusion-wise maximal elements of $\mathcal{M}$ is called the set of bases of the matroid. It is known that all the sets in $\mathcal{B}$ have the same cardinality, which is called the rank of the matroid. In this paper we often work with sets of bases $\mathcal{B}$ of matroids instead of independent sets $\mathcal{I}$, for this reason we will also refer to a pair $(U, \mathcal{B})$ as a matroid.

Linear and regular matroids. Let $U = \{w_1, w_2, \ldots, w_m\} \subseteq \mathbb{R}^n$ be a set of vectors. Let $\mathcal{B}$ consist of all subsets of U which form a basis for the linear space generated by all the vectors in U. $\mathcal{M} = (U, \mathcal{B})$ is called a linear matroid. A matrix $A \in \mathbb{R}^{r \times m}$ is called a representation of a matroid $\mathcal{M} = ([m], \mathcal{B})$, if for every set $S \subseteq [m]$, S is independent in $\mathcal{M}$ if and only if the corresponding set of columns $\{A_i : i \in S\}$ is linearly independent. A matroid $\mathcal{M} = (M, \mathcal{B})$ is called a regular matroid if it is representable by a totally unimodular real matrix. A matrix is called totally unimodular if the determinant of any of its square submatrices belongs to the set $\{-1, 0, 1\}$.

Partition matroids. A matroid $\mathcal{M} = (M, \mathcal{B})$ is said to be a partition matroid if there exists a partition $\mathcal{P} = \{M_1, M_2, \ldots, M_t\}$ of the ground set M and a sequence of non-negative integers $b = (b_1, b_2, \ldots, b_t)$ such that $|B \cap M_i| = b_i$ for all $B \in \mathcal{B}$ and $i = 1, 2, \ldots, t$.

III. Anti-Concentration Inequality: Proof of Theorem I.1

Our proof consists of two phases. In the first phase, we divide the space into exponentially many disjoint subsets of equal volume, which we call *cells*, such that, within any cell, the value of f is uniformly lower bounded by a factor that only depends on the cell.

In the second phase, we show that the cells can be partitioned into small size groups in such a way that each group has the same number of cells and within every group, there exists at least one cell where the function f takes relatively large values.

Let us denote $k := \lceil \log(r) \rceil$ and take $x^\star = (x^{\star 1}, \ldots, x^{\star r}) \in \prod_{i=1}^r \Delta_{p_i}$ to be any point at which f attains its optimal value, i.e., $\mathrm{OPT} = f(x^\star)$. For $q = (q^1, \ldots, q^t) \in \prod_{i \in [t]} \Delta_{p_i}$ define

$$\widetilde{f}(q) := f(q^1, \ldots, q^t, x^{\star t+1}, \ldots, x^{\star r}).$$

Notice that for $t = r$, i.e., when $q \in \prod_{i \in [r]} \Delta_{p_i}$, we have $\widetilde{f}(q) = f(q)$.

Phase 1: Cell Construction

In the first phase of the proof, we show that there exists a collection of disjoint sets $\mathcal{S}(i_1, i_2, \ldots, i_r) \subseteq \Delta := \prod_{i \in [r]} \Delta_{p_i}$, called cells, such that the following hold for every $(i_1, \ldots, i_r) \in [k]^r$ and for every point $q \in \mathcal{S}(i_1, i_2, \ldots, i_r)$

$$\mu\left(\mathcal{S}(i_1, i_2, \ldots, i_r)\right) = \left(\frac{1}{k} - \frac{\gamma}{rk}\right)^r,$$
$$f(q) \geq \mathrm{OPT} \prod_{j \in [r]} \frac{1}{\gamma p_j} \left(\frac{k - i_j}{k} + \frac{\gamma i_j}{rk}\right). \tag{7}$$

This fact is a direct consequence of the following, more general lemma, when $t = r$.

Lemma III.1 (Cell construction)**.** *Let $f : \prod_{i \in [r]} \Delta_{p_i} \to \mathbb{R}$ be a γ-anti-concentrated function which attains its maximum value at $x^\star$. If $r \geq \gamma$, then for every $t \leq r$ there exists a family of subsets of $\prod_{i \in [t]} \Delta_{p_i}$,*

$$\left\{\mathcal{S}(i_1, \ldots, i_t) : (i_1, \ldots, i_t) \in [k]^t\right\},$$

such that the following conditions are satisfied

1) *(Equal volume) $\mu\left(\mathcal{S}(i_1, i_2, \ldots, i_t)\right) = \left(\frac{1}{k} - \frac{\gamma}{rk}\right)^t$.*
2) *(Uniform lower bound) for all $q \in \mathcal{S}(i_1, i_2, \ldots, i_t)$, $\widetilde{f}(q) \geq \mathrm{OPT} \cdot \prod_{j \in [t]} \frac{1}{\gamma p_j} \left(\frac{k - i_j}{k} + \frac{\gamma i_j}{rk}\right)$.*
3) *(Disjointness) The sets $\mathcal{S}(i_1, \ldots, i_t)$ for $(i_1, \ldots, i_t) \in [k]^t$ are pairwise disjoint.*

The proof of Lemma III.1 appears in the full version of the paper. For the case when $t = r$, Lemma III.1 says that every cell has volume $\left(\frac{1}{k} - \frac{\gamma}{rk}\right)^t$ and since there are k^r disjoint cells, the volume of the union of these cells is equal to

$$k^r \left(\frac{1}{k} - \frac{\gamma}{rk}\right)^r = \left(1 - \frac{\gamma}{r}\right)^r \approx \frac{1}{e^\gamma}.$$

Let us denote $\zeta_i := \frac{k-i}{k}$ for $i = 1, 2, \ldots, k-1$ and $\zeta_k := \frac{\gamma}{r}$. Then it is easy to see that $\frac{k-i}{k} + \frac{\gamma i}{rk} \geq \zeta_i$ for all $i \in [k]$ and hence from the uniform lower bound property it follows that for every $q \in \mathcal{S}(i_1, \ldots, i_r)$

$$f(q) \geq \mathrm{OPT} \cdot \prod_{j \in [r]} \frac{1}{\gamma p_j} \cdot \prod_{j \in [r]} \zeta_{i_j}. \tag{8}$$

Phase 2: Counting Good Cells

We construct a subset of Δ where f is "large", by taking a union of appropriate cells. Equation (8) gives us a convenient lower-bound on the value of f on each cell $\mathcal{S}(i_1, \ldots, i_r)$. By the equal volume condition in Lemma III.1, all sets $\mathcal{S}(i_1, \ldots, i_r)$ have the same volume. What remains, is to count cells with a large enough lower bound on $f(q)$ following from (8). Let us define a cell $(i_1, i_2, \ldots, i_r)$ to be *good* if

$$f(q) \geq (\gamma e^2)^{-r} \prod_{i \in [r]} \frac{1}{p_i} \cdot \mathrm{OPT} \quad \text{for all } q \in \mathcal{S}(i_1, \ldots, i_r)$$

and denote by G the set of all good cells. We show that at least $\frac{1}{k}$ fraction of cells are good, i.e., that $|G| \geq k^{r-1}$.

To this end, let σ be the cyclic permutation on the set $[k]$, i.e., $\sigma(i) = i + 1$, for $i \in [k-1]$ and $\sigma(k) = 1$. Consider the action of σ on r-tuples $(i_1, \ldots, i_r) \in [k]^r$ defined by

$$\sigma(i_1, \ldots, i_r) := (\sigma(i_1), \ldots, \sigma(i_r)).$$

Let σ^l be the permutation σ composed l times with itself. Now, define the following equivalence relation on cells. Two cells $\mathcal{S}(i_1, \ldots, i_r), \mathcal{S}\left(i'_1, \ldots, i'_r\right)$ are said to be in relation if

$$\exists l \in [k] \qquad \sigma^l\left(i_1, \ldots, i_r\right) = \left(i'_1, \ldots, i'_r\right).$$

Observe that every equivalence class (which we call an orbit) contains exactly k elements.

We show that for any cell $\mathcal{S}(i_1, \ldots, i_r)$, there exists at least one good cell in its orbit, i.e., of the form $\mathcal{S}\left(\sigma^l(i_1 \ldots, i_r)\right)$, for some $l \in [k]$. To this end it is enough (because of (8)) to show that there exists an $l \in [k]$ such that

$$\prod_{j \in [r]} \frac{1}{\gamma p_j} \cdot \prod_{j \in [r]} \zeta_{\sigma^l(i_j)} \geq (\gamma e^2)^{-r} \prod_{i \in [r]} \frac{1}{p_i}. \tag{9}$$

Consider the product of left hand sides of (9) over all $l \in [k]$:

$$\prod_{j\in[r]} \frac{1}{(\gamma p_j)^k} \cdot \prod_{j\in[r]} \prod_{l\in[k]} \zeta_{\sigma^l(i_j)} =$$
$$= \prod_{j\in[r]} \frac{1}{(\gamma p_j)^k} \cdot \prod_{l\in[k]} \prod_{j\in[r]} \zeta_{\sigma^l(i_j)}$$
$$= \prod_{j\in[r]} \frac{1}{(\gamma p_j)^k} \left(\frac{\gamma}{r}\right)^r \prod_{t\in[k-1]} \left(\frac{k-t}{k}\right)^r$$
$$= \prod_{j\in[r]} \frac{1}{(\gamma p_j)^k} \left(\frac{\gamma}{r}\right)^r \left(\frac{(k-1)!}{k^{k-1}}\right)^r$$
$$\geq \prod_{j\in[r]} \frac{1}{(\gamma p_j)^k} \left(\frac{\gamma}{r}\right)^r \frac{1}{e^{kr}}.$$

By taking the k-th root of the right hand side above we obtain a lower bound of

$$\prod_{j\in[r]} \frac{1}{(\gamma p_j)} \cdot \left(\frac{\gamma}{r}\right)^{r/k} \frac{1}{e^r} \geq \gamma^{-r} \cdot \frac{1}{e^{2r}} \cdot \prod_{j\in[r]} \frac{1}{p_j}$$

The last inequality is due to the fact that $k = \lceil \log(r) \rceil$. Hence, concluding, there exists $l \in [k]$ such that for all points $q \in \mathcal{S}(\sigma^l(i_1, i_2, \ldots, i_r))$ we have

$$f(q) \geq (\gamma e^2)^{-r} \cdot \prod_{j\in[r]} \frac{1}{p_j} \cdot \mathrm{OPT}.$$

Thus indeed at least $\frac{1}{k}$-fraction of all cells is good. On the other hand, we proved that the total volume of cells is approximately $\frac{1}{e^\gamma}$. Hence, the volume of the union of good cells is at least $\frac{1}{ke^\gamma}$, which concludes the proof of Theorem I.1.

IV. Partition Matroids: Proof of Theorem I.2

We begin by introducing some useful notation. Let d, r be two positive integers such that $d \geq r$. Let p_i for $i \in [r]$ be r positive integers. Denote $\Delta := \prod_{i=1}^r \Delta_{p_i}$. Fix an arbitrary tuple $\mathcal{V} = \left(v^{(i,j)} : i \in [r], j \in [p_i]\right)$ of vectors in $\mathbb{R}^d$. For every $i \in [r]$ and for every vector $y \in \Delta_{p_i}$, define

$$\vartheta_i(y) := \sum_{j\in[p_i]} y_j v^{(i,j)}. \tag{10}$$

For any vectors $u^1, \ldots, u^r \in \mathbb{R}^d$ define

$$g(u^1, \ldots, u^r) := \det(U^\top U)^{\frac{1}{2}},$$

where U is the $d \times r$ matrix whose i-th column is u^i. Equivalently, g evaluates the r-dimensional volume of the parallelepiped formed by the vectors $u^i, i \in [r]$. Define

$$f_{\mathcal{V}}(x) := g\left(\vartheta_1(x^1), \ldots, \vartheta_r(x^r)\right) \tag{11}$$

For any tuple $y = \left(y^j \in \Delta_{p_j} : j \in \{2, 3, \ldots, r\}\right)$ of $(r-1)$ vectors, define the function $f_y : \Delta_{p_1} \to \mathbb{R}$ by

$$f_y(z) := g\left(\vartheta_1(z), \vartheta_2(y^2), \ldots, \vartheta_r(y^r)\right).$$

For an alternative definition of $f_y(z)$ define by P_y the $(r-1)$-dimensional parallelepiped spanned by $\vartheta_i(y^i)$ for $i \in \{2, 3, \ldots, r\}$. Then

$$f_y(z) = \mathrm{dist}\left(\vartheta_1(z), \mathrm{span}(P_y)\right) \cdot \lambda_{r-1}(P_y). \tag{12}$$

Where $\lambda_{r-1}(P_y)$ denotes the $(r-1)$-dimensional measure of P_y and $\mathrm{span}(P_y)$ is the $(r-1)$-dimensional subspace spanned by P_y.

A. *Proof of Theorem I.2*

We start by observing that it suffices to prove the Theorem for the case when $b_1 = b_2 = \cdots = b_t = 1$. Indeed, when b_i's are not all equal to 1, we can perform a simple reduction to the all-ones case. Namely, we construct a new instance of the problem, where every part M_i is repeated b_i times. After doing so, we obtain a new instance with r parts $M_1', M_2', \ldots, M_r'$ and $b_1' = b_2' = \ldots = b_r' = 1$.

Every feasible solution to the original instance corresponds to a feasible solution to the new instance (with the same value). Conversely, every feasible solution *with non-zero value* corresponds to a feasible solution in the original instance.

Finally, the bound on the approximation ratio follows easily by translating the bound in the simple case $b_1 = b_2 = \ldots = b_r = 1$ to the instance after reduction.

Hence, from now on we assume that $b_1 = b_2 = \cdots = b_t = 1$; in this case $t = r$. Let

$$L = V^\top V$$

be the Cholesky decomposition of the PSD matrix L with $V \in \mathbb{R}^{d\times m}$. One can easily see that

$$L_{S,S} = V_S^\top V_S, \quad \text{for all } S \subseteq [m].$$

The quantity $\det\left(V_S^\top V_S\right)$ is equal to the squared volume of the parallelepiped formed by vectors $\{V_i : i \in S\}$. Therefore, maximizing $\det(L_{S,S})$ subject to $S \in \mathcal{B}(\mathcal{M})$ is equivalent to finding a basis S of the matroid such that the volume of the parallelepiped formed by $\{V_i : i \in S\}$ is maximized. For convenience, identify M_i with $\{(i,j) : j \in [p_i]\}$ and also index the corresponding columns of V by $v^{(i,j)}$ for $j \in [p_i]$. Further, to each pair (i,j) (for $i \in [t]$ and $j \in [p_i]$) assign a variable x_j^i. Let

$$x^i := \left(x_j^i, j \in M_i\right), \quad \text{and} \quad x := (x^i, i \in [r]).$$

Let $f_{\mathcal{V}}$ be the function defined in (11). When each x^i is a vertex of Δ_{p_i}, precisely one x_j^i is equal to 1 and the others are equal to 0. Thus, there exists a natural bijection between the elements of $\mathcal{B}$ (bases of the partition matroid) and the vertices of $\Delta = \prod_{i=1}^t \Delta_{p_i}$. Therefore, the optimization problem can be stated as the problem of maximizing $f_{\mathcal{V}}$ over the vertices of Δ. That is

$$\max \; \{f_{\mathcal{V}}(x) : x \in V(\Delta)\}. \tag{13}$$

The next lemma states that $f_\mathcal{V}$ is 2-anti-concentrated. The proof appears in the next subsection.

Lemma IV.1 (2-Anti-concentration of the volume function)**.** *Let d, r, with $d \geq r$ be two positive integers. Let $p_1, p_2, \ldots, p_r$ be positive integers. For any tuple $\mathcal{V} = \left(v^{(i,j)} \in \mathbb{R}^d : i \in [r], j \in [p_i]\right)$, $f_\mathcal{V}$ is 2-anti-concentrated.*

From the above lemma, and Theorem I.1 we deduce

$$\mathbf{Pr}\left[f_\mathcal{V}(x) > (2e^2)^{-r}\prod_{i=1}^{r}\frac{1}{p_i}\cdot \mathrm{OPT}\right] \geq \frac{1}{e^2 \log r}.$$

By drawing polynomially many independent samples we can ensure that with probability approaching 1, at least one of the samples y satisfies the condition

$$f_\mathcal{V}(y) > (2e^2)^{-r}\prod_{i=1}^{r}\frac{1}{p_i}\cdot \mathrm{OPT}.$$

The next lemma guarantees that in polynomial time, we can round such a y to an integral solution.

Lemma IV.2 (Rounding for the volume function)**.** *Let d, r, with $d \geq r$ be two positive integers. Let $p_1, p_2, \ldots, p_r$ be positive integers. For any tuple $\mathcal{V} = \left(v^{(i,j)} \in \mathbb{R}^d : i \in [r], j \in [p_i]\right)$, $f_\mathcal{V}$ has a polynomial time rounding algorithm, i.e., there exists an algorithm which given a point $x \in \Delta = \prod_{i=1}^{r}\Delta_{p_i}$ outputs (in polynomial time) a vertex $\widetilde{x} \in V(\Delta)$ such that $f_\mathcal{V}(x) \leq f_\mathcal{V}(\widetilde{x})$.*

The proof of the Lemma IV.2 is presented in Section IV-B. As demonstrated above, Theorem I.2 then follows from Lemmas IV.2 and IV.1.

B. Proofs of Lemmas

The proof of Lemma IV.1 relies on the following more general Fact.

Fact IV.3 (2-Anti-concentration of the distance function)**.** *Let t, d be two positive integers. Suppose that $w^1, \ldots, w^t$ are vectors in $\mathbb{R}^d$. The function $f : \Delta_t \to \mathbb{R}$ defined by $f(x) := \|\sum_{i\in[t]} x_i w^i\|_2$ is 2-anti-concentrated.*

The proof of Fact IV.3 appears in the full version of the paper. Here we show how to deduce Lemma IV.1 from it.

Proof of Lemma IV.1: We show that fixing the values of any $(r-1)$ variables results in a 2-anti-concentrated function of the remaining variables. Because of symmetry, we only need to verify this claim for the last $(r-1)$ block-coordinates. Fix an arbitrary tuple $y = (y^2, \ldots, y^r) \in \prod_{i=2}^{r}\Delta_{p_i}$. We show that f_y is 2-anti-concentrated, i.e.,

$$\forall c \in (0,1), \quad \mathop{\mathbf{Pr}}_{z}\left[f_y(z) < c\cdot \mathrm{OPT}\right] < 2cp_1, \tag{14}$$

where OPT is the maximum value of f_y over the simplex Δ_{p_1}. Recall from (12) that

$$f_y(z) = \mathrm{dist}\left(\vartheta_1(z), \mathrm{span}(P)\right)\cdot \lambda_{r-1}(P),$$

where $P = P_y$, as defined in the previous subsection. In particular

$$\mathrm{OPT} = \max_{z\in\Delta_{p_1}} \mathrm{dist}\left(\vartheta_1(z), \mathrm{span}(P)\right)\cdot \lambda_{r-1}(P).$$

Therefore, the event (over a random choice of $z \in \Delta_{p_1}$)

$$f_y(z) < c\cdot \mathrm{OPT}$$

coincides with

$$\mathrm{dist}(\vartheta_1(z), \mathrm{span}(P)) < c\cdot \max_{x^1\in\Delta_{p_1}} \mathrm{dist}\left(\vartheta_1(x^1), \mathrm{span}(P)\right). \tag{15}$$

Define the function $f : \Delta_{p_1} \to \mathbb{R}$ by

$$f(z) := \|\mathrm{dist}(\vartheta_1(z), \mathrm{span}(P))\|_2 = \|\sum_{j\in[p_1]} z_j w^j\|_2,$$

where w^j is the projection of $v^{(1,j)}$ on the space orthogonal to $\mathrm{span}(P)$. Fact IV.3 together with (15) imply then that f is 2-anti-concentrated. Thus, by definition, f_y is 2-anti-concentrated. Consequently, $f_\mathcal{V}$ is 2-anti-concentrated. ■

Proof of Lemma IV.2: We prove that for every setting of $r-1$ block-coordinates of $f_\mathcal{V}$, the induced function (of the remaining coordinate) attains its maximum at one of the vertices of the remaining coordinate. Clearly such a property implies the Lemma, as we can round any point x by rounding one coordinate at a time, without decreasing the value of the function. For each coordinate i, only p_i calls to the evaluation oracle are required, one per each vertex.

Fix any $y = (y^2, \ldots, y^r) \in \prod_{i=2}^{r}\Delta_{p_i}$, we show that the maximum of $f_y(z)$ is attained at a vertex of Δ_{p_1}. Recall that $f_y(z) = g\left(\vartheta_1(z), \vartheta_2(y^2), \ldots, \vartheta_r(y^r)\right)$ is the restriction of $f_\mathcal{V}$ when the last $r-1$ arguments are fixed. By (12), we have

$$f_y(z) = \mathrm{dist}\left(\vartheta_1(z), \mathrm{span}(P)\right)\cdot \lambda_{r-1}(P).$$

and further

$$\mathrm{dist}\left(\vartheta_1(z), \mathrm{span}(P)\right) = \|\sum_{j\in[p_1]} z_j w^j\|_2 \leq \max_{j\in[p_1]}\|w^j\|_2,$$

where w^j is the projection of $v^{(1,j)}$ on the space orthogonal to $\mathrm{span}(P)$. The last inequality follows from the triangle inequality and from the fact that $z_j \in \Delta_{p_j}$. Thus, f_y is maximized at one of the vertices of Δ_{p_1}. ■

V. Regular Matroids: Proof of Theorem I.3

We start by reducing the subdeterminant maximization problem under a regular matroid constraint to a polynomial optimization problem as follows. Let $B_1, B_2, \ldots, B_m \in \mathbb{R}^d$ be the columns of B. Since B is a representation of the matroid $\mathcal{M}$, a set $S \subseteq M$ is a basis of $\mathcal{M}$ if and only the set of the vectors $\{B_i : i \in S\}$ is linearly independent. Let $L = V^\top V$ be a Cholesky decomposition of the PSD matrix L, for $V \in \mathbb{R}^{d\times m}$.

Let us now consider any set $S \in \binom{[m]}{d}$ and define $I_S := \mathrm{Diag}(1_S)$. For any $S \in \binom{[m]}{d}$ we have

$$\det\left(VI_SB^\top\right) = \det\left(\sum_{i\in S} V_iB_i^\top\right) = \det\left(V_S\right)\det\left(B_S^\top\right).$$

Since B is a totally unimodular matrix, $|\det(B_S)| = 1$ if $S \in \mathcal{B}(\mathcal{M})$ and 0 otherwise. Thus for all $S \in \binom{[m]}{d}$

$$\left|\det\left(VI_SB^\top\right)\right| = \begin{cases} |\det(V_S)| & \text{if } S \in \mathcal{B}, \\ 0 & \text{otherwise.}\end{cases}$$

Since for all $S \in \binom{[m]}{d}$, $\det(L_{S,S}) = \det(V_S^\top V_S) = \det(V_S)^2$, maximizing $\det(L_{S,S})$ over $S \in \mathcal{B}$ is equivalent to maximizing $|f(x)|$ for $f(x) := \det(VXB^\top)$ over all the 0-1 vectors $x \in \{0,1\}^m$ subject to $\sum_{i=1}^m x_i = d$. We give an approximation algorithm for this problem which proceeds in two phases.

Phase 1: Finding a Fractional Solution.

In the first phase, we drop the $\sum_{i=1}^m x_i = d$ condition and relax the $0-1$ condition to $x \in [0,1]^m$. Our optimization problem then becomes

$$\begin{aligned} \max_x \quad & |f(x)|, \\ \text{s.t.} \quad & x \in [0,1]^m. \end{aligned} \tag{16}$$

Our algorithm to find an approximate solution to (16) is as follows. We sample a polynomial number of points x from $[0,1]^m$ uniformly and independently at random. Then, we output the point with the largest value of $|f(x)|$. We analyze the performance of this algorithm in two different regimes.

Large d. It follows from the Cauchy-Binet formula that

$$f(x) = \sum_{S\in\mathcal{B}} x^S \det(V_S)\det(B_S). \tag{17}$$

Moreover, $f(x)$ is multi-affine and easy to compute (because it is just a determinant of an $m\times m$ matrix). We show that $|f|$ is 2-anti-concentrated. To this end, we show that for every $i \in [m]$ and every choice of $y_j \in [0,1], j \in [m]\setminus\{i\}$, the univariate function

$$\tau \mapsto |f\left(y_1,\ldots,y_{i-1},\tau,y_{i+1},\ldots,y_m\right)|$$

is 2-anti-concentrated. Such a function is of the form $\tau \mapsto |a\tau + b|$ for some $a, b \in \mathbb{R}$. 2-anti-concentration of such functions follows easily from Fact IV.3. Indeed, by setting $d = 1$ and $t = 2$ in Fact IV.3 we obtain the 2-anti-concentration of $(\tau_1,\tau_2) \mapsto |\tau_1 a_1 + \tau_2 a_2|$, which implies our claim.

Theorem I.1 implies now that if we sample a uniform point x from $[0,1]^m$ then

$$\mathbf{Pr}\left[|f(x)| > 2^{-m}(2e^2)^{-m}\cdot \mathrm{OPT}\right] \geq \frac{1}{e^2\log m}.$$

Where $\mathrm{OPT} := \max_{x\in[0,1]^m}|f(x)|$ is clearly an upper bound on $\max_{S\in\mathcal{B}}|\det(V_S)|$. We can amplify the probability of success by repeating the experiment several times and hence, with high probability obtain a point $\hat{x}$ such that

$$|f(\hat{x})| > (2e)^{-2m}\cdot\mathrm{OPT}. \tag{18}$$

Small d. From (17) it is clear that the function f is a polynomial of degree d in m variables. According to Theorem 2 in [18], if we sample x uniformly from the unit hypercube $[0,1]^m$, then

$$\mathbf{Pr}\left[|f(x)| \leq \beta^d\cdot\mathrm{OPT}\right] \leq C\cdot\beta\cdot m,$$

for any $\beta > 0$ and some absolute constant $C > 0$. By picking $\beta = \frac{1}{2C\cdot m}$, we conclude that with constant probability we obtain a vector $\hat{x}$ such that

$$|f(\hat{x})| > \left(\frac{1}{2mC}\right)^d\cdot\mathrm{OPT}. \tag{19}$$

Phase 2: Rounding the Fractional Solution.

We first round $\hat{x}$ obtained in the previous phase to a $0-1$ vector, and then finally to a set $\hat{S} \in \binom{[m]}{d}$. Since f is multi-affine, the restriction of f to the first coordinate is a 1-dimensional affine function. Therefore, either

$$|f(0,\hat{x}_2,\ldots,\hat{x}_d)| \geq |f(\hat{x})| \quad \text{or} \quad |f(1,\hat{x}_2,\ldots,\hat{x}_d)| \geq |f(\hat{x})|.$$

Hence, we can round the first coordinate without decreasing the value of $|f(\hat{x})|$, using one call to the evaluation oracle. We proceed to the next coordinates and round them one at a time. Let $y \in \{0,1\}^m$ be the outcome of the above rounding algorithm.

Let $S_0 \subseteq [m]$ such that $1_{S_0} = y$. It is likely that $|S_0| > d$, hence we will need to remove several elements from S_0 to obtain a set of cardinality d. Define a function $g : 2^{[m]} \to \mathbb{R}$ to be

$$g(S) := f(1_S) = \det(V_SB_S^\top).$$

Note in particular that g can be computed efficiently. Furthermore, by the Cauchy-Binet formula, we have

$$g(S) = \sum_{T\in\binom{[m]}{d}} g(T) = \sum_{T\in\binom{[m]}{d}} \det(V_T)\det(B_T) \tag{20}$$

for every subset $S \in 2^{[m]}$. We have $|f(y)| = |f(1_{S_0})| = |g(S_0)|$. Further, (20) implies that

$$\sum_{i\in S_0} g(S_0\setminus\{i\}) = (|S_0|-d)\sum_{T\in\binom{S_0}{d}} g(T) = (|S_0|-d)g(S_0).$$

Consequently, there exists an $i \in S_0$ such that:

$$|g(S_0\setminus\{i\})| \geq \frac{|S_0|-d}{|S_0|}|g(S_0)|.$$

In our algorithm we find such an i and consider $S_1 := S_0\setminus\{i\}$. This step of removing one element is repeated until we arrive at a set $\hat{S} \subseteq [m]$ of cardinality d. In this process we can guarantee that

$$|g(\hat{S})| \geq |g(S_0)| \cdot \prod_{j=1}^{|S_0|-d} \frac{j}{j+d} \geq \frac{|g(S_0)|}{\binom{m}{d}}.$$

Finally, since $|g(\hat{S})| = |\det(V_{\hat{S}})|$, we conclude:

$$|\det(V_{\hat{S}})| \geq \frac{|f(y)|}{\binom{m}{d}} > \frac{1}{\binom{m}{d}} \max\left((2e)^{-2m}, (2dC)^{-d}\right) \cdot \mathrm{OPT}$$

hence $|\det(V_{\hat{S}})| > \max\left(2^{-O(m)}, 2^{-O(d \log m)}\right) \cdot \mathrm{OPT}$, and Theorem I.3 follows.

References

[1] L. Khachiyan, "On the complexity of approximating extremal determinants in matrices," *Journal of Complexity*, vol. 11, no. 1, pp. 138–153, 1995.

[2] A. Nikolov and M. Singh, "Maximizing determinants under partition constraints," in *Proceedings of the 48th Annual ACM Symposium on Theory of Computing*, 2016, pp. 192–201.

[3] L. Gurvits, "Hyperbolic polynomials approach to Van der Waerden/Schrijver-Valiant like conjectures: sharper bounds, simpler proofs and algorithmic applications," in *Proceedings of the thirty-eighth annual ACM symposium on Theory of computing*. ACM, 2006, pp. 417–426.

[4] D. Straszak and N. K. Vishnoi, "Real stable polynomials and matroids: Optimization and counting," in *Proceedings of the 49th Annual ACM SIGACT Symposium on Theory of Computing*, ser. STOC 2017, 2017, pp. 370–383.

[5] N. Anari and S. Oveis Gharan, "A generalization of permanent inequalities and applications in counting and optimization," in *Proceedings of the 49th Annual ACM SIGACT Symposium on Theory of Computing*, ser. STOC 2017, 2017, pp. 384–396.

[6] P. Gritzmann, V. Klee, and D. G. Larman, "Largest j-simplices n-polytopes," *Discrete and Computational Geometry*, pp. 477–517, 1995.

[7] S. A. Goreinov and E. E. Tyrtyshnikov, "The maximal-volume concept in approximation by low-rank matrices," *Contemporary Mathematics*, vol. 280, pp. 47–51, 2001.

[8] R. Lyons, "Determinantal probability measures," *ArXiv Mathematics e-prints*, Apr. 2002.

[9] A. Kulesza and B. Taskar, *Determinantal Point Processes for Machine Learning*. Hanover, MA, USA: Now Publishers Inc., 2012.

[10] L. E. Celis, A. Deshpande, T. Kathuria, and N. K. Vishnoi, "How to be fair and diverse?" *Fairness, Accountability, and Transparency in Machine Learning*, 2016.

[11] L. E. Celis, A. Deshpande, T. Kathuria, D. Straszak, and N. K. Vishnoi, "On the Complexity of Constrained Determinantal Point Processes," in *Approximation, Randomization and Combinatorial Optimization. Algorithms and Techniques, 21st International Workshop, RANDOM 2017, Berkeley, CA, USA, August 16-18, 2017.*, 2017.

[12] A. Nikolov, "Randomized Rounding for the Largest Simplex Problem," in *Proceedings of the Forty-seventh Annual ACM Symposium on Theory of Computing*, ser. STOC '15. ACM, 2015, pp. 861–870.

[13] M. D. Summa, F. Eisenbrand, Y. Faenza, and C. Moldenhauer, "On largest volume simplices and sub-determinants," in *Proceedings of the Twenty-Sixth Annual ACM-SIAM Symposium on Discrete Algorithms*. SIAM, 2015, pp. 315–323.

[14] A. Çivril and M. Magdon-Ismail, "On selecting a maximum volume sub-matrix of a matrix and related problems," *Theor. Comput. Sci.*, vol. 410, no. 47-49, pp. 4801–4811, 2009.

[15] L. Lovász, "Singular spaces of matrices and their application in combinatorics," *Boletim da Sociedade Brasileira de Matemática-Bulletin/Brazilian Mathematical Society*, vol. 20, no. 1, pp. 87–99, 1989.

[16] N. Harvey, "Algebraic algorithms for matching and matroid problems," *SIAM Journal on Computing*, vol. 39, no. 2, pp. 679–702, 2009.

[17] R. Gurjar and T. Thierauf, "Linear matroid intersection is in quasi-NC," in *Proceedings of the 49th Annual ACM SIGACT Symposium on Theory of Computing, STOC 2017, Montreal, QC, Canada, June 19-23, 2017*, 2017, pp. 821–830.

[18] A. Carbery and J. Wright, "Distributional and L^q norm inequalities for polynomials over convex bodies in $\mathbb{R}^n$," *Mathematical research letters*, vol. 8, no. 3, pp. 233–248, 2001.

[19] R. O'Donnell, *Analysis of Boolean Functions*. New York, NY, USA: Cambridge University Press, 2014.

[20] A. De, I. Diakonikolas, and R. A. Servedio, "A robust Khintchine inequality, and algorithms for computing optimal constants in fourier analysis and high-dimensional geometry," *SIAM Journal on Discrete Mathematics*, vol. 30, no. 2, pp. 1058–1094, 2016.

[21] K. P. Costello, T. Tao, and V. Vu, "Random symmetric matrices are almost surely nonsingular," *Duke Math. J.*, vol. 135, no. 2, pp. 395–413, 11 2006.

[22] A. Razborov and E. Viola, "Real advantage," *ACM Transactions on Computation Theory (TOCT)*, vol. 5, no. 4, p. 17, 2013.

[23] J. Vondrák, "Submodularity and curvature: the optimal algorithm," *RIMS Kokyuroku Bessatsu*, vol. B 23, pp. 253–266, 2010.

[24] J. G. Oxley, *Matroid theory*. Oxford University Press, USA, 2006, vol. 3.