Subdeterminant Maximization via Nonconvex Relaxations and Anti-Concentration

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 [Kha95] who gave an $\mathcal{O}(r)$ approximation algorithm for this problem. Recently, Nikolov and Singh [NS16] 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 [Gur06], and was recently extended to regular matroids [SV17, AO17]. 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.
1 Introduction
   1.1 Overview of our contributions ......................................................... 3
   1.2 Discussion and future work .............................................................. 4
   1.3 Other related work ............................................................................. 6
   1.4 Technical overview ............................................................................. 7
   1.5 Organization of the rest of the paper .................................................. 7

2 Preliminaries ............................................................................................. 12

3 Anti-Concentration Inequality: Proof of Theorem 1.1 ................................. 13
   3.1 Proof of Lemma 3.1 ............................................................................ 15

4 Partition Matroids: Proof of Theorem 1.2 .................................................. 18
   4.1 Proof of Theorem 1.2 ....................................................................... 19
   4.2 Proofs of Lemmas ............................................................................. 20

5 Regular Matroids: Proof of Theorem 1.3 ..................................................... 21

A Hard Example for the Nikolov-Singh Algorithm ....................................... 25

B Omitted Proofs
   B.1 Proofs for Section 2 ........................................................................ 26
   B.2 Proofs for Section 1 ........................................................................ 27
1 Introduction

A variety of problems in computer science and optimization can be formulated as the following constrained subdeterminant maximization problem: Given a positive semi-definite (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 $L = V^T V$ 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 $[GKL95]$ and matrix low-rank approximation $[GT01]$ problems. In mathematics, the probability distribution on $2^{|m|}$ in which a set $S \subseteq [m]$ has probability $\Pr(S) \propto \det(L_{S,S})$ is referred to as a determinantal point process (DPP); see $[Lyo02]$. 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 $[KT12]$. 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 $[CDKV16, CDKV17]$ 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 subdeterminant maximization problem are rare; Khachiyan $[Kha95]$ 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 $[Nik15]$ who presented an approximation algorithm which achieved a factor of $e^r$. On the other hand, it was shown $[SEP15, CM09]$ 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 $[NS16]$ 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 $[Gur06]$ concerning real stable polynomials. Building on their work, $[SV17, AO17]$ presented estimation algorithms for large classes of families $\mathcal{B}$, such as bases of a regular matroid. While the results of $[NS16, SV17, AO17]$ 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.

1.1 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 1.1 (Anti-concentrated functions)** For $\gamma \geq 1$, a nonnegative measurable\(^1\) function $f : \Delta_d \to \mathbb{R}$ is called $\gamma$-anti-concentrated if for every $c \in (0, 1)$

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

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

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 $\gamma$-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 $\gamma$-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 1.1 for $\gamma = 2$ (see Fact 1.3). 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 1.1 (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 $\gamma$-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

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

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

---

\(^1\)We always assume that the functions we deal with are regular enough. Formally, we require measurability with respect to the Lebesgue measure.

\(^2\)\(\Delta_d\) denotes the standard $(d-1)$-simplex, i.e., $\Delta_d := \{x \in \mathbb{R}^d : \sum_{i=1}^d x_i = 1, x \geq 0\}$. 
Consequently, the value of a $\gamma$-anti-concentrated function at a random point in its domain gives an estimate of its maximum value. In the simplest non-trivial case, it applies to multi-affine functions over the hypercube $[0,1]^r$ and says that the value of the function at a random point is at least $c^{-r}$ times its optimal value, with significant probability (where $c > 1$ is an absolute constant). It is also easy to see that the bound in Theorem 1.1 is tight: For $p(x) = \prod_{i=1}^r x_i$, one can show that the probability that $|p(x)| \geq (3/4)^r$ over a random choice of $x \in [0,1]^r$ is exponentially small.

As an important special case of Theorem 1.1, 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 fact noted earlier that such an $f$ is 2-anti-concentrated, we conclude from Theorem 1.1 that for some absolute constant $c > 1$ and a uniformly random choice of $x \in [0,1]^r$,

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

(1)

This 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 4.2) 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 1.1 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

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

where $\Delta$ 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 $L = V^T V$ and the variables in $x$. Subsequently, we show that such functions are 2-anti-concentrated, which allows us to apply Theorem 1.1 to obtain the following result.

**Theorem 1.2 (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

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

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

5
Prior work by Nikolov and Singh [NS16] 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 (see Appendix A).

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 $B$ arises as a set of bases of a regular matroid; see Section 2). To apply Theorem 1.1 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 $B$ 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, e.g., by Harvey [Har09]). We observe that $|h(x)|$ is 2-anti-concentrated and has a number of desirable properties, which allows us to prove

**Theorem 1.3 (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 $B \subseteq 2^{[m]}$, outputs a set $S \in B$ such that with high probability

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

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

There are two recent results for this setting ([SV17] and [AO17]) that provide $e^m$- and $e^d$-estimation algorithms respectively. As in the case of the algorithm for 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.

### 1.2 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-linearity allows us to round this random solution. Furthermore, the anti-concentration result can be applied to any multi-linear polynomial and beyond. In particular, it neither relies on real stability nor any other convexity-like property of the polynomial; this should be of independent interest. We leave open the problem of extending the anti-concentration inequality from hypercubes and products of simplices to more general bodies; this might allow us to improve the approximation ratios.
1.3 Other related work

A very general anti-concentration result for polynomial functions over convex domains was obtained by Carbery and Wright [CW01], 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 4. 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 [O'D14, DDS16, CTV06, RV13]. Finally, our use of rounding using multi-linearity resembles a similar phenomena in algorithms to optimize concave or sub-modular functions; see for instance a survey by Vondrák [Von10].

1.4 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 $w_i$s 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 $p(x, y) := \det(\sum_{i=1}^r y_i A_i(x_i))$ and let $g(x)$ be the polynomial that appears as the coefficient of $\prod_{i=1}^r y_i$ in $p(x, y)$. Multi-linearity 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 [Gur06] 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

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

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 $x \in [0,1]^r$

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

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 $x^* \in [0,1]^r$ to (2), the intractability of $g(x)$ becomes a bottleneck. Nikolov and Singh present a rounding algorithm which, unfortunately, can require an exponential number of trials to find an $e^r$-approximate solution; see Appendix A

**Overview of the proof of Theorem 1.2.** 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

---

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.
$w_i(t) := (1 - t)w_i + tw_i$. 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-linearity of $\det(W(x))$ implies the following:

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

(4)

Indeed, if we let $f(x) := |\det(W(x))|$, then the multi-linearity 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 $\hat{x} \in \{0, 1\}^r$ such that $f(\hat{x}) \geq f(x)$.

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 $\text{OPT}$ by evaluating $f$ at a random point. Formally, the anti-concentration inequality (Theorem 2 in [CW01]) applies to $f$ and allows us to deduce that

$$\Pr[f(x) \geq c^{-r} \cdot \text{OPT}] \geq \frac{1}{e^2 \log r}$$

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 1.4 (Theorem 2 in [CW01])** 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 $\beta \in (0, 1)$

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

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 $\beta$ so as to make $C \cdot \beta \cdot r < 1$, i.e., for $\beta = O(1/cr)$, we obtain

$$\Pr[f(x) \geq r^{-O(r)} \cdot \text{OPT}] \geq \frac{1}{r}.$$ 

This implies that the algorithm described above achieves an approximation ratio of (roughly) $r^r$. Our Theorem 1.1 is a certain strengthening of Theorem 1.4 which asserts that under the same assumptions

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

for some absolute constant $c > 1$. In fact, Theorem 1.1 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 1.2. 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 $B$ 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
∑_{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-linear polynomial, but as we show in Lemma \[1.1\] it is 2-anti-concentrated. Having established this property, Theorem \[1.2\] 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 \( \Delta_{p_i} \), for \( i = 1, 2, \ldots, t \)) which then provides us with a suitable approximate solution.

**Overview of the proof of Theorem \[1.3\].** In the setting of Theorem \[1.3\] 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 \( \text{OPT} := \max_{S \in \mathcal{B}} \det(L_{S,S}) \). The approach of \[SV17\] to obtain an estimate on \( \text{OPT} \) was inspired by that of \[NS16\] for the partition matroid case and is as follows:\footnote{The approach of \[AO17\] is also similar.}

Given the matrix \( L = V^\top V \), first, define the following polynomial

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

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. 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 \( \text{OPT} \). 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 \( \text{OPT} \) 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

\[
 h(x) := \det(VX^\top B^\top),
\]

where \( X \) is the vector with components \( x^i \).
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 [Lov89, Har09, GT17]. 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-linear 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 1.1 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 losing 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 1.1 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 for the algorithm. Alternatively, by utilizing the fact that $h$ is a polynomial of degree $d$, one can apply the result by Carbery-Wright (see Theorem 1.4) 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 1.1.** 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 1.1. Recall the setting: We are given a function $f : [0,1]^r \rightarrow \mathbb{R}_{\geq 0}$ 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

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

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., $\Pr[f(x) < \alpha \cdot \text{OPT}]$ is small, where OPT is the maximum value $f$ takes on the hypercube and $\alpha$ is a parameter which we want to be as large as possible.

As an initial approach, one can define (for some $c > 0$) events of the form

$$A_i := \{ x \in [0,1]^r : f(x_1, x_2, \ldots, x_i, x^*_{i+1}, \ldots, x^*_r) \geq c \cdot f(x_1, x_2, \ldots, x_{i-1}, x^*_i, \ldots, x^*_r) \},$$

where $x^* := \arg\max_x f(x)$. Note crucially that the events $A_1, A_2, \ldots, A_r$ are not independent. However, we can still write

$$\Pr[f(x) \geq c^n \cdot \text{OPT}] \geq \Pr[A_1 \cap A_2 \cap A_3 \cdots \cap A_r]$$

$$= \Pr[A_1] \cdot \Pr[A_2 | A_1] \cdot \Pr[A_3 | A_1, A_2] \cdots \Pr[A_r | A_1, A_2, \ldots, A_{r-1}].$$
To get a probability that is not exponentially small, one has to take $c$ of size roughly $O(1/r)$, in which case we recover the result by Carbery and Wright \cite{CW01} 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 $S(i_1, i_2, \ldots, i_r)$ 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 $\delta^r$). 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 $S(i_1, i_2, \ldots, i_r)$,\footnote{From assumption \ref{assumption:cell_volume} we know that}

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

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

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

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

In the proof we identify a set of cells $G$ (which we call *good*) such that the above bound guarantees that $f(x) \geq c^{-r} \cdot \left(\frac{\delta}{m}\right)^r$ for a constant $c > 0$. Subsequently, we prove that at least $\frac{1}{r}$ 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 $(i \mapsto (i + 1) \mod k)$ 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 \ref{assumption:cell_volume} for $k$ cells in one such orbit. We obtain

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

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

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

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 $f(x) \geq \text{OPT} \cdot \left(\frac{r}{m}\right)^r c^{-r}$ holds for at least $\Omega \left(\frac{1}{\log(r)}\right)$ fraction of the space. This completes the sketch of the proof of Theorem \ref{theorem:anti-concentration}.

### 1.5 Organization of the rest of the paper

We present some notation and preliminaries about matroids and measures in Section \ref{section:notation}. In Section \ref{section:anti-concentration-matroids} we present the proof of our anti-concentration result, Theorem \ref{theorem:anti-concentration}. In Section \ref{section:approximation-algorithm} we present a proof of Theorem \ref{theorem:approximation-algorithm} for partition matroids. In Section \ref{section:partitioning} we present a proof of Theorem \ref{theorem:partitioning} for regular matroids. A few technical facts arising in these three sections are delegated to the appendix. In Section \ref{section:example} we present an example to show that the Nikolov-Singh algorithm may not yield an approximation algorithm.
2 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 $\lambda_d$. When the dimension is clear from the context, we use $\lambda$ 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 $\Delta_d$ is defined as the convex hull of $e_1, e_2, \ldots, e_d \in \mathbb{R}^d$. Notice that $\Delta_d$ 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 $\mu_d$ to denote the induced measure $\lambda_d$ on the simplex $\Delta_d$, 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 $\Delta_{p_i}$ and $x^j$ for $j \in [p_i]$ are the components of $x^i$ within $\Delta_{p_i}$. By $V(\Delta)$ we denote the set of points of $\Delta$ with integer coordinates. We call $V(\Delta)$ the set of vertices of $\Delta$.

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 set $S \subseteq \mathbb{R}^d$, we denote by span($S$) linear space spanned by $S$. For any two closed subsets $S_1, S_2 \subseteq \mathbb{R}^d$, we denote by dist($S_1, S_2$) the distance between these two sets, formally defined as

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

where $\|\cdot\|_2$ is the standard $\ell_2$-norm. For a vector $u$ and a set $S$ of the vectors, we denote by $u^\perp_S$ the orthogonal component of the vector $u$ with respect to span($S$).

Multi-linear functions. A function $f : \mathbb{R}^m \to \mathbb{R}$ is called multi-linear 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-linear 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 $f$. 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 $f : \mathbb{R}^m \times \cdots \times \mathbb{R}^r \to \mathbb{R}$ is called block-multi-linear if for every index $i \in [r]$ and for every choice of $y^i_j \in \mathbb{R}^{p_i}, j \in [r] \setminus \{i\}$ the function $f(y^1, \ldots, x^i, \ldots, y^r)$ is an affine function over $\mathbb{R}^{p_i}$.

Matroids. For a comprehensive treatment of matroid theory we refer the reader to [Ox10]. Below we state the most important definitions and examples of matroids, which are most relevant to our results. A matroid is a pair $M = (U, I)$ such that $U$ is a finite set and $I \subseteq 2^U$ satisfies the following three axioms: (1) $\emptyset \in I$, (2) if $S \in I$ and $S' \subseteq S$ then $S' \in I$, (3) if $A, B \in I$ and $|A| > |B|$, then there exists an element $a \in A \setminus B$ such that $B \cup \{a\} \in I$. The collection $B \subseteq I$ 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 $I$, 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 $\mathcal{B}$ consists 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 $\mathcal{M} = (M, B)$ 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 $B \in B$ and $i = 1, 2, \ldots, t$.

**Fact about the Lebesgue measure.** The following simple consequence of Fubini’s theorem is
used several times in the paper. We present a proof in the appendix.

**Lemma 2.1** Let $T$ be a Lebesgue measurable subset of $\mathbb{R}^d$. Let

$$T_1 := \{x \in \mathbb{R} : \exists x_2, \ldots, x_d \in \mathbb{R} \ (x, x_2, \ldots, x_d) \in T\}.$$  

For every $x \in T_1$, let

$$T(x) := \{(x_2, \ldots, x_d) \in \mathbb{R}^{d-1} : (x, x_2, \ldots, x_d) \in T\}.$$  

If $\lambda_{\text{min}}$ and $\lambda_{\text{max}}$ are, respectively, the minimum and the maximum of $\lambda_{d-1}(T(x))$ over $x \in T_1$, then

$$\lambda_1(T_1)\lambda_{\text{min}} \leq \lambda(T) \leq \lambda_1(T_1)\lambda_{\text{max}}.$$  

3 Anti-Concentration Inequality: Proof of Theorem 1.1

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

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

$$\tilde{f}(q) := f(q^1, \ldots, q^t, x^{*t+1}, \ldots, x^*).$$

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

**Phase 1: Cell construction**

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

$$\mu(S(i_1, i_2, \ldots, i_r)) = \left( \frac{1 - \gamma}{k} \right)^r, \quad \text{for all } (i_1, i_2, \ldots, i_r) \in [k]^r,$$

$$f(q) \geq f(x^*) \prod_{j \in [r]} \frac{1}{\gamma p_j} \left( \frac{k - i_j}{k} + \frac{\gamma i_j}{rk} \right), \quad \text{for all } q \in S(i_1, i_2, \ldots, i_r). \quad (7)$$

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

Lemma 3.1 (Cell construction) Let \( f : \prod_{i \in [r]} \Delta_{p_i} \to \mathbb{R} \) be a \( \gamma \)-anti-concentrated function which attains its maximum value at \( x^* \). If \( r \geq \gamma \), then for every \( t \leq r \) there exists a family of subsets of \( \prod_{i \in [r]} \Delta_{p_i} \),
\[
\{ S(i_1, \ldots, i_t) : (i_1, \ldots, i_t) \in [k]^t \},
\]
such that the following conditions are satisfied

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

We prove this lemma in the next subsection. For the case when \( t = r \), Lemma 3.1 says that every cell has volume \( \left( \frac{1}{k} - \frac{\gamma}{rk} \right)^r \) and since there are \( k^r \) disjoint cells, the volume of the union of these cells is equal to
\[
k^r \left( \frac{1}{k} - \frac{\gamma}{rk} \right)^r \approx \frac{1}{e^\gamma}.
\]
Let us denote
\[
\varsigma_i := \frac{k-i}{k}, \quad \text{for } i \in [k-1],
\]
\[
\varsigma_k := \frac{\gamma}{r},
\]
then it is easy to see that \( \frac{k-i}{k} + \frac{\gamma i}{rk} \geq \varsigma_i \) for all \( i \in [k] \) and hence from the uniform lower bound property it follows that for every \( q \in S(i_1, \ldots, i_r) \)
\[
f(q) \geq \text{OPT} \cdot \prod_{j \in [r]} \frac{1}{\gamma p_j} \prod_{j \in [r]} \varsigma_i^j, \tag{8}
\]

Phase 2: Counting Good Cells

We construct a subset of \( \Delta \) 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 \( S(i_1, \ldots, i_r) \). By equal volume condition in Lemma 3.1 all the sets \( S(i_1, \ldots, i_r) \) have the same volume. What remains to do is to count cells with a large enough lower bound on \( f(p) \) following from (8). Let us define the set of good cells to be
\[
G := \left\{ (i_1, i_2, \ldots, i_r) \in [k]^r : f(q) \geq (\gamma e^2)^{-r} \prod_{i \in [r]} \frac{1}{p_i} \cdot \text{OPT} \text{ for all } q \in S(i_1, i_2, \ldots, i_r) \right\}.
\]
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 \( \sigma \) 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 \( \sigma \) 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 \( \sigma^l \) be the permutation \( \sigma \) composed \( l \) times with itself. Now, define the following equivalence relation on cells. Two cells \( S(i_1, \ldots, i_r), S(i'_1, \ldots, i'_r) \) are said to be in relation if
\[
\exists l \in [k] \quad \sigma^l(i_1, \ldots, i_r) = (i'_1, \ldots, i'_r).
\]
Observe that every equivalence class (which we will call an orbit) contains exactly $k$ elements.

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

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

(9)

Let us now consider the product of LHS's of (9) over all $l \in [k]$. We obtain

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

Note now that by taking the $k$-th root of the RHS above we obtain a lower bound of

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

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 $q \in S(\sigma^l(i_1, i_2, \ldots, i_r))$ we have

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

Thus indeed at least $\frac{1}{k^r}$-fraction of all cells is good. On the other hand, we proved that the total volume of cells is approximately $\frac{1}{e^2}$. Hence, the volume of the union of good cells is at least $\frac{1}{ke^2}$, which concludes the proof of Theorem 1.1.

### 3.1 Proof of Lemma 3.1

We prove the lemma by induction on $t$.

**Base case**

For $t = 1$, consider the set

$$
W_1 := \Delta_{p_1}.
$$

Define

$$
T(j) := \left\{ q \in W_1 : f(q) \geq \left(\frac{k-j}{k} + \frac{\gamma j}{rk}\right) \frac{1}{\gamma p_1} f(x^*) \right\}.
$$
Since for $r \geq \gamma$, $\frac{k-j}{k} + \frac{\gamma j}{rk}$ is a non-increasing function of $j$, $T(1) \subseteq T(2) \subseteq \cdots \subseteq T(k)$. Since $f$ is $\gamma$-anti-concentrated, the measure of the set $T(j)$ is at least

$$1 - \gamma p_1 \left( \frac{k-j}{k} \gamma j \frac{1}{rk} \right) = j \left( \frac{1}{k} - \frac{\gamma}{rk} \right).$$

Thus, $T(j)$'s form a nested sequence of sets with the specified lower bounds on their measures.

Next we show that we can restrict each $T(j)$ to its subset $T'(j)$ such that they are still nested and their measure is equal to

$$j \left( \frac{1}{k} - \frac{\gamma}{rk} \right).$$

For this, define the sets $T'(j)$ inductively as follows. Let $T'(1)$ be any subset of $T(1)$ of measure equal to $\frac{1}{k} - \frac{\gamma}{rk}$. Since $\mu(T(1)) \geq \frac{1}{k} - \frac{\gamma}{rk}$, such a subset $T'(1)$ exists. Suppose that for all $j \in [i-1]$, the sets $T'(j)$ are defined such that

$$T'(j - 1) \subseteq T'(j) \subseteq T(j)$$

and $\mu(T'(j)) = j(\frac{1}{k} - \frac{2}{rk})$. To define $T'(i)$, notice that $T'(i - 1) \subseteq T(i - 1) \subseteq T(i)$ and also $\mu(T(i)) \geq i(\frac{1}{k} - \frac{\gamma}{rk})$. Now define $T'(i)$ to be any subset of $T(i)$ of measure $i(\frac{1}{k} - \frac{\gamma}{rk})$ which contains the set $T'(i - 1)$. Let $S(1) := T'(1)$ and define

$$S(j) := T'(j) \setminus T'(j - 1) \quad \text{for} \quad j = 2, \ldots, k.$$ 

It follows that the sets $S(j)$ for $j = 1, 2, \ldots, k$ are disjoint and thus

$$\mu(S(j)) = \frac{1}{k} - \frac{\gamma}{rk}.$$ 

Furthermore, $S(j) \subseteq T'(j) \subseteq T(j)$. Therefore, for every $q \in S(j)$ we have $q \in T(j)$ and by the choice of $T(j)$

$$\tilde{f}(q) \geq \left( \frac{k-j}{k} \gamma j \frac{1}{rk} \right) \frac{1}{\gamma p_1} f(x^*).$$

Equation (10) implies that for every $q \in S(j)$

$$\begin{cases} 
\tilde{f}(q) \geq \frac{k-j}{k} \frac{1}{\gamma p_1} f(x^*) & \text{if} \quad j < k, \\
\tilde{f}(q) \geq \frac{1}{\gamma p_1} f(x^*) & \text{if} \quad j = k.
\end{cases}$$

Induction step

Now, assume that for $t - 1$, the claim is valid. For any $(t-1)$-tuple $(i_1, \ldots, i_{t-1}) \in [k]^{t-1}$, define

$$W_t(i_1, \ldots, i_{t-1}) := S(i_1, \ldots, i_{t-1}) \times \Delta_{p_t}.$$ 

Since the sets $S(i_1, \ldots, i_{t-1})$ are pairwise disjoint by induction hypothesis, the sets $W_t(i_1, \ldots, i_{t-1})$ are pairwise disjoint as well. Let

$$T(i_1, \ldots, i_t) := \left\{ q \in W_t(i_1, \ldots, i_{t-1}) : \tilde{f}(q) \geq \left( \frac{k-i_t}{k} \gamma i_t \frac{1}{rk} \right) \frac{1}{\gamma p_t} \tilde{f}(q^1, \ldots, q^{t-1}) \right\}.$$ 

(12)
The sets $\mathcal{T}(i_1, \ldots, i_t)$ and $\mathcal{T}(i'_1, \ldots, i'_t)$ are disjoint whenever $(i_1, \ldots, i_{t-1}) \neq (i'_1, \ldots, i'_{t-1})$. We claim that

$$\mu(\mathcal{T}(i_1, \ldots, i_t)) \geq t \left( \frac{1}{k} - \frac{\gamma}{rk} \right)^t.$$  \hfill (13)

This is because for any point $q' \in S(i_1, \ldots, i_{t-1})$, the measure of the set

$$\{q' \in \Delta_{p_t} : (q', q'') \in \mathcal{T}(i_1, \ldots, i_t)\}$$

is at least $i_t(\frac{1}{k} - \frac{\gamma}{rk})$ by $\gamma$-anti-concentration of $f$. Since this holds for any $q' \in S(i_1, \ldots, i_{t-1})$, and because of the induction hypothesis $\mu(S(i_1, i_2, \ldots, i_{t-1})) = (\frac{1}{k} - \frac{\gamma}{rk})^{t-1}$, now the claim follows from Lemma 2.1 (after a suitable reparametrization).

Since for $r \geq \gamma$, $\frac{k-i}{k} + \frac{\gamma i}{rk}$ is a non-increasing function of $i_t$, for any fixed tuple $(i_1, \ldots, i_{t-1}) \in [k]^{t-1}$, (12) implies that

$$\mathcal{T}(i_1, \ldots, i_{t-1}, 1) \subseteq \mathcal{T}(i_1, \ldots, i_{t-1}, 2) \subseteq \cdots \subseteq \mathcal{T}(i_1, \ldots, i_{t-1}, k).$$

Similar to the 1-dimension case, we choose $\mathcal{T}'(i_1, \ldots, i_t) \subseteq \mathcal{T}(i_1, \ldots, i_t)$ in such a way that

$$\mathcal{T}'(i_1, \ldots, i_{t-1}, 1) \subseteq \mathcal{T}'(i_1, \ldots, i_{t-1}, 2) \subseteq \cdots \subseteq \mathcal{T}'(i_1, \ldots, i_{t-1}, k),$$

and also, the inequalities in (13) become equalities for $\mathcal{T}'$. That is, we have

$$\mu(\mathcal{T}'(i_1, \ldots, i_t)) = t \left( \frac{1}{k} - \frac{\gamma}{rk} \right)^t, \text{ for all } (i_1, \ldots, i_t) \in [k]^t.$$  \hfill (14)

This equality together with the assumption that $\mathcal{T}'(i_1, \ldots, i_t)$ are nested imply that

$$\mu(S(i_1, \ldots, i_t)) = \left( \frac{1}{k} - \frac{\gamma}{rk} \right)^t$$

in which

$$S(i_1, \ldots, i_{t-1}, 1) := \mathcal{T}'(i_1, \ldots, i_{t-1}, 1),$$

$$S(i_1, \ldots, i_{t-1}, j) := \mathcal{T}'(i_1, \ldots, i_{t-1}, j) \setminus \mathcal{T}'(i_1, \ldots, i_{t-1}, j-1), \text{ for } j = 2, 3, \ldots, k.$$  \hfill (15)

Now, we show that the sets $S(i_1, \ldots, i_t)$ and $S(i'_1, \ldots, i'_t)$ are disjoint for $(i_1, \ldots, i_t) \neq (i'_1, \ldots, i'_t)$. For this, we consider two cases.

**Case 1:** If $(i_1, \ldots, i_{t-1}) \neq (i'_1, \ldots, i'_{t-1})$ then we have

$$S(i_1, \ldots, i_t) \subseteq \mathcal{T}'(i_1, \ldots, i_t) \subseteq \mathcal{T}(i_1, \ldots, i_t) \text{ and } S(i'_1, \ldots, i'_t) \subseteq \mathcal{T}'(i'_1, \ldots, i'_t) \subseteq \mathcal{T}(i'_1, \ldots, i'_t).$$

Recall that $\mathcal{T}(i_1, \ldots, i_t)$ and $\mathcal{T}(i'_1, \ldots, i'_t)$ are disjoint when $(i_1, \ldots, i_{t-1}) \neq (i'_1, \ldots, i'_{t-1})$. Thus, $S(i_1, \ldots, i_t)$ and $S(i'_1, \ldots, i'_t)$ are disjoint, too.

**Case 2:** If $(i_1, \ldots, i_{t-1}) = (i'_1, \ldots, i'_{t-1})$, then clearly $i_t \neq i'_t$. Without loss of generality, we may assume that $i_t \leq i'_t - 1$. Hence,

$$S(i_1, \ldots, i_t) \subseteq \mathcal{T}'(i_1, \ldots, i_t) \subseteq \mathcal{T}'(i_1, \ldots, i'_t - 1) = \mathcal{T}'(i'_1, \ldots, i'_t - 1)$$

and

$$S(i'_1, \ldots, i'_t) \subseteq \mathcal{T}'(i'_1, \ldots, i'_t) \setminus \mathcal{T}'(i'_1, \ldots, i'_t - 1).$$

17
Thus, \( S(i_1, \ldots, i_t) \) and \( S(i'_1, \ldots, i'_t) \) are disjoint.

Finally, we show that the uniform lower bound condition is also satisfied. Since \( S(i_1, \ldots, i_t) \subseteq T(i_1, \ldots, i_t) \) for any \( q \in S(i_1, \ldots, i_t) \), by (12) we have

\[
\tilde{f}(q) \geq \left( k - \frac{i_t}{k} + \frac{\gamma i_t}{r k} \right) \frac{1}{\gamma p_t} \tilde{f}(q^1, \ldots, q^{t-1}).
\]  

(15)

Furthermore, (12) implies that

\[
T(i_1, \ldots, i_t) \subseteq W_t(i_1, \ldots, i_{t-1}).
\]

By the choice of \( W_t(i_1, \ldots, i_{t-1}) \), \((q^1, \ldots, q^{t-1}) \in S(i_1, \ldots, i_{t-1})\). Therefore, by induction hypothesis we conclude that

\[
\tilde{f}(q^1, \ldots, q^{t-1}) \geq f(x^*) \cdot \prod_{j \in [t-1]} \frac{1}{\gamma p_j} \left( \frac{k - i_j}{k} + \frac{\gamma i_j}{r k} \right)
\]

(16)

Now we can simply combine the bounds in (15) and (16) to obtain

\[
\tilde{f}(q) \geq f(x^*) \prod_{j \in [t]} \frac{1}{\gamma p_j} \left( \frac{k - i_j}{k} + \frac{\gamma i_j}{r k} \right), \quad \text{for all } q \in S(i_1, i_2, \ldots, i_t).
\]

### 4 Partition Matroids: Proof of Theorem [1.2]

Let us introduce some terminology, which we use in the remaining part of this section. Let also \( \Delta := \prod_{i=1}^r \Delta_{p_i} \).

**Definition 4.1 (Roundable function)** A function \( f : \Delta \to \mathbb{R} \) is called roundable if there exists a polynomial time algorithm which for every input point \( x \in \Delta \) outputs a vertex \( x' \in V(\Delta) \) such that \( f(x') \geq f(x) \).

Note that in particular it follows that the maximum value of a roundable function is attained at a vertex.

Let \( d, r \) be two positive integers such that \( d \geq r \). Let \( p_i \) for \( i \in [r] \) be \( r \) positive integers. Fix an arbitrary tuple \( V = (v(i,j) : i \in [r], j \in [p_i]) \) of vectors in \( \mathbb{R}^d \). For every \( i \in [r] \) and for every vector \( y \in \Delta_{p_i} \), define

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

(17)

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

\[
f_V(x) := g \left( \sum_{j \in [p_1]} x_j^1 v(1,j), \ldots, \sum_{j \in [p_r]} x_j^r v(r,j) \right) = g \left( \vartheta_1(x^1), \ldots, \vartheta_r(x^r) \right).
\]

(18)

For every tuple \( y = (y^j \in \Delta_{p_j} : j \in [r] \setminus \{1\}) \), define the function \( f_y : \Delta_{p_1} \to \mathbb{R} \) by

\[
f_y(z) := g \left( \vartheta_1(z), \vartheta_2(y^2), \ldots, \vartheta_r(y^r) \right).
\]
Let $P$ be an $(r-1)$-dimensional parallelepiped formed by the vectors $\vartheta_i(y^i)$ for $i \in [r] \setminus \{1\}$. Thus, the function $f_y(z)$ can be viewed as the volume $(r-1)$-dimensional volume of $P$, multiplied by the distance of $\vartheta_1(z)$ from the hyperplane spanned by $P$; that is

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

(19)

For every $j \in p_1$ define

$$w^j := (v^{(1,j)})^\perp_P.$$  

(20)

Since the orthogonal component with respect to a fixed space is a linear transformation, we have

$$\vartheta_1(z)^\perp_P = \left(\sum_{j \in [p_1]} z_j v^{(1,j)}\right)^\perp_P = \sum_{j \in [p_1]} z_j w^j.$$  

(21)

4.1 Proof of Theorem 1.2

We consider two cases.

Case 1: $b_1 = b_2 = \cdots = b_t = 1$

In this case, $t = r$. Let

$$L = V^\top V$$

be the Cholesky decomposition of the PSD matrix $L$ in which $V$ is a $d \times m$ real matrix. One can easily see that

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

The quantity $\det(V_S^\top V_S)$ 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 $x^i_j$. Let

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

Let $f_V$ be the function defined in (18). When each $x^i$ is a vertex of $\Delta_{p_i}$, precisely one $x^i_j$ 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 $f_V$ over the vertices of $\Delta$. That is

$$\max_x f_V(x), \quad \text{s.t. } x \in V(\Delta).$$

(22)

In the next lemma we prove that $f_V$ is 2-anti-concentrated. We postpone the proof of it to the next subsection.

**Lemma 4.1 (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 $V = (v^{(i,j)} \in \mathbb{R}^d : i \in [r], j \in [p_i])$, $f_V$ is 2-anti-concentrated.
From the above lemma, and Theorem 1.1 we deduce

$$
\text{Pr}
\left[
\frac{1}{r} \prod_{i=1}^{r} \frac{1}{p_i} \cdot \text{OPT}
\right]
\geq \frac{1}{e^2 \log r}.
$$

By the standard trick of re-sampling, we can take polynomially many independent samples to ensure that with probability approaching to 1, at least one of the samples satisfies the condition

$$
f_V(y) > (2e^2)^{-r} \prod_{i=1}^{r} \frac{1}{p_i} \cdot \text{OPT}.
$$

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

**Lemma 4.2 (Roundability of the volume function)** Let $d, r$ be two positive integers. Let $p_1, p_2, \ldots, p_r$ be positive integers. For any tuple $V = \{v(i,j) : i \in [r], j \in [p_i]\}$, $f_V$ is roundable.

Thus, we conclude Theorem 1.2 under the assumption that $b_i = 1$ for all $i \in [t]$. The proof of the Lemma 4.2 is presented at the end of this section.

**Case 2: Arbitrary $b_i$’s**

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 $b_1 = b_2 = \ldots = b_r = 1$ for the instance after reduction.

## 4.2 Proofs of Lemmas

**Proof of Lemma 4.1** 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.,

$$
\forall c \in (0, 1), \quad \text{Pr}_z[f_y(z) < c \cdot \text{OPT}] < 2cp_1,
$$

where $\text{OPT}$ is the maximum value of $f_y$ over the simplex $\Delta_{p_1}$. Recall from (19) that

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

In particular

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

Therefore, the event

$$
f_y(z) < c \cdot \text{OPT}
$$
coincides with
\[
\text{dist}(\vartheta_1(z), \text{span}(P)) < c \cdot \max_{x^1 \in \Delta_{p_1}} \text{dist}(\vartheta_1(x^1), \text{span}(P)). \tag{24}
\]

Define the function \( f : \Delta_{p_1} \to \mathbb{R} \) by
\[
f(x^1) := \|\text{dist}(\vartheta_1(x^1), \text{span}P)\|_2 = \| \sum_{j \in [p_1]} x^1_j w^j \|_2.
\]

The second equality is due to the definition of the distance between a vector and a subspace. The next fact implies that \( f \) is 2-anti-concentrated. We present a proof in the appendix.

**Fact 4.3 (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.

This implies that
\[
\Pr_{z}[f_y(z) < c \cdot \text{OPT}] = \Pr_{z}[\text{dist}(\vartheta_1(z), \text{span}(P)) < c \cdot \max_{x^1 \in \Delta_{p_1}} \text{dist}(\vartheta_1(x^1), \text{span}(P))] < 2p_1 c
\]
where the equality is due to the equivalence of the two events and the inequality is a consequence of the 2-anti-concentration of \( f \), guaranteed by Fact 4.3. Thus, by definition, \( f_y \) is 2-anti-concentrated. Consequently, \( f_V \) is 2-anti-concentrated.

**Proof of Lemma 4.2:** We prove \( f_V \) is roundable by showing that for every set of \( r - 1 \) block-coordinates and for every assignment of them, the induced function attains its maximum at one of the vertices of the remaining block-coordinate. Clearly such a property implies that \( f_V \) is roundable because we can round any point \( x \) by rounding one coordinate at a time, without decreasing the value of the function. For each coordinate \( i \), we require \( p_i \) calls to the evaluation oracle, one per each vertex, and then we pick the one with the maximum value of \( f_V \).

We show that for all \( y = (y^2, \ldots, y^r) \in \prod_{i=2}^r \Delta_{p_i} \), the maximum of \( f_y(z) \) is attained at a vertex of \( \Delta_{p_1} \). Recall that \( f_y(z) = g(\vartheta_1(z), \vartheta_2(y^2), \ldots, \vartheta_r(y^r)) \) is the restriction of \( f_V \) when the last \( r - 1 \) arguments are fixed. Due to the symmetry in the problem, such a statement holds for any coordinate. By (19), we have
\[
f_y(z) = \text{dist}(\vartheta_1(z), \text{span}(P)) \cdot \lambda_{r-1}(P).
\]
Thus, the maximum volume of \( f_y \) is achieved when \( \text{dist}(\vartheta_1(z), \text{span}(P)) \) is maximized. By applying triangle inequality to (21) we obtain
\[
\|\vartheta_1(z)^{-r}\|_2 = \| \sum_{j \in [p_1]} z_j w^j \|_2 \leq \sum_{j \in [p_1]} z_j \|w^j\|_2 \leq \max_{j \in [p_1]} \|w^j\|_2.
\]
The last inequality in this chain is because \( z_j \)'s are non-negative numbers whose sum equals 1. Thus, \( f_y \) is maximized at one of the vertices of \( \Delta_{p_1} \).

## 5 Regular Matroids: Proof of Theorem 1.3

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 $M$, a set $S \subseteq M$ is a basis of $M$ 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 $I_S := \text{Diag}(1_S)$. Notice that the matrix $VI_S$ is a $d \times m$ matrix which is obtained from $V$ by replacing the columns with labels not in $S$, by 0. Thus, for any $S \in \binom{[m]}{d}$ we have

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

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

$$\forall S \in \binom{[m]}{d}, \quad |\det(VI_SB^\top)| = \begin{cases} |\det(V_S)| & \text{if } S \in \mathcal{B}, \\ 0 & \text{otherwise}. \end{cases}$$

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 $f(x) := |\det(VXB^\top)|$ 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 integrality condition to $x \in [0,1]^m$. Our optimization problem then becomes

$$\max_{x} |f(x)|, \quad \text{s.t. } x \in [0,1]^m. \quad (25)$$

Our algorithm to find an approximate solution to (25) 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 is not hard to derive the following formula for $f(x)$

$$f(x) = \sum_{S \in \mathcal{B}} x^S \det(V_S) \det(B_S). \quad (26)$$

Hence it is multi-linear 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

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

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 4.3. Indeed, by setting $d = 1$ and $t = 2$ in Fact 4.3 we obtain the 2-anti-concentration of $(\tau_1, \tau_2) \mapsto |\tau_1a_1 + \tau_2a_2|$, which implies our claim.

Theorem 4.1 implies now that if we sample a uniform point $x$ from $[0,1]^m$ then

$$\Pr \left[ |f(x)| > 2^{-m}(2e^2)^{-m} \cdot \text{OPT} \right] \geq \frac{1}{e^2 \log m}.$$
Where \( \text{OPT} = \max_{x \in [0,1]^m} |f(x)| \). At this point, note also that \( \text{OPT} \geq \max_{S \subseteq B} |\text{det}(V_S)| \). 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

\[
|f(\hat{x})| > (2e)^{-2m} \cdot \text{OPT}. \tag{27}
\]

**Small \( d \):** From (26) it is clear that the function \( f \) is a polynomial of degree \( d \) in \( m \) variables. According to Theorem 2 in [CAW01], if we sample a uniform \( x \) from the unit hypercube \([0,1]^m\), then

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

for any \( \beta \). By picking \( \beta = \frac{1}{2eCm} \), we conclude that with constant probability we obtain a vector \( \hat{x} \) such that

\[
|f(\hat{x})| > \left( \frac{1}{2mC} \right)^d \cdot \text{OPT}. \tag{28}
\]

**Phase 2: Rounding the fractional solution.**

Let us now try to round \( \hat{x} \) obtained in the previous phase to a \( 0 - 1 \) vector, and then finally to a set \( \hat{S} \subseteq \binom{[m]}{d} \). Since \( f \) is multi-linear, 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 \( 1_{S_0} = y \). Most likely \( |S_0| > d \), hence we will need to remove several elements from \( S_0 \) to obtain a set of cardinality \( d \). Define a function \( g : 2^{|m|} \rightarrow \mathbb{R} \) to be

\[
g(S) := f(1_S) = \text{det}(V_S B^\top_S)
\]

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

\[
g(S) = \sum_{T \in \binom{[m]}{d}} g(T) = \sum_{T \in \binom{[m]}{d}} \text{det}(V_T) \text{det}(B_T) \tag{29}
\]

for every subset \( S \in \binom{[m]}{d} \).

We have \( |f(y)| = |f(1_{S_0})| = |g(S_0)| \). Further, (29) implies that

\[
\sum_{i \in S_0} g(S_0 \setminus \{i\}) = (|S_0| - d) \sum_{T \in \binom{[S_0 \setminus \{i\}]}{d-1}} g(T) = (|S_0| - d) g(S_0).
\]

For this reason, 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 \( S_1 := S_0 \setminus \{i\} \). We repeat the step of removing one element until we arrive at a set \( \hat{S} \subseteq [m] \) of cardinality \( d \). In this process we can guarantee that

\[
|\text{det}(V_{\hat{S}})| = |g(\hat{S})| \geq |g(S_0)| \cdot \frac{|S_0| - d}{|S_0|} \cdot \frac{|S_0| - 1 - d}{|S_0| - 1} \cdot \ldots \cdot \frac{1}{d+1} \geq \frac{|g(S_0)|}{\binom{m}{d}}.
\]
Finally we conclude:

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

and Theorem 1.3 follows.

References

[AO17] 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, STOC 2017, pages 384–396, 2017.

[CDK+17] 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, 20th International Workshop, APPROX 2017, and 21st International Workshop, RANDOM 2017, Berkeley, CA, USA, August 16-18, 2017., 2017.

[CDKV16] 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.

[ÇM09] A. Çivril and M. Magdon-Ismail. On selecting a maximum volume sub-matrix of a matrix and related problems. Theor. Comput. Sci., 410(47-49):4801–4811, 2009.

[CTV06] K. P. Costello, T. Tao, and V. Vu. Random symmetric matrices are almost surely nonsingular. Duke Math. J., 135(2):395–413, 11 2006.

[CW01] A. Carbery and J. Wright. Distributional and $L^q$ norm inequalities for polynomials over convex bodies in $\mathbb{R}^n$. Mathematical research letters, 8(3):233–248, 2001.

[DDS16] 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, 30(2):1058–1094, 2016.

[GKL95] P. Gritzmann, V. Klee, and D. G. Larman. Largest $j$-simplices $n$-polytopes. Discrete and Computational Geometry, pages 477–517, 1995.

[GT01] S. A. Goreinov and E. E. Tyrtyshnikov. The maximal-volume concept in approximation by low-rank matrices. Contemporary Mathematics, 280:47–51, 2001.

[GT17] 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, pages 821–830, 2017.

[Gur06] 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, pages 417–426. ACM, 2006.

[Har09] N. Harvey. Algebraic algorithms for matching and matroid problems. SIAM Journal on Computing, 39(2):679–702, 2009.
A Hard Example for the Nikolov-Singh Algorithm

In this section we give an example that the algorithm proposed by Nikolov and Singh [NS16] for the subdeterminant maximization under partition constraints might fail to output a set with non-zero subdeterminant with high probability, even though the expected value of the returned solution is high.

Lemma A.1 There exists an instance of the subdeterminant maximization problem under partition constraints, for which the optimal value is equal to 1 and the Nikolov-Singh Algorithm [NS16] outputs a non-zero solution with exponentially small probability.
Proof: Let $L = V^TV$, where $V \in \mathbb{R}^{r \times m}$ is a matrix with $m = r^2$. The columns of $V$ are standard unit vectors $e_1, e_2, \ldots, e_r \in \mathbb{R}^r$ each one repeated $r$ times. We consider the problem of maximizing $\det(V_S^TV_S)$ over sets $S \subseteq [m]$ of cardinality $r$. This is an instance of the subdeterminant maximization problem under partition constraints, when there is only one partition of size $m$ and $b_1 = r$. For such instances the algorithm of [NS16] specializes to that of [Nik15]. It first solves the convex program

$$\max_{x \in P} \log \det \left( \sum_{i=1}^m x_i v_i v_i^\top \right)$$

where $P = \{x \in \mathbb{R}^m : \sum_{i=1}^m x_i = r, 0 \leq x \leq 1\}$. It is not hard to see that the point

$$z = (1/r, 1/r, \ldots, 1/r) \in \mathbb{R}^m$$

is an optimal solution to the above optimization problem.

The output of the Nikolov-Singh algorithm is a random set $S$ sampled according to a distribution $\rho$ given by $\rho(S) \propto z^S$ for $|S| = r$. It can be simply seen to be the uniform distribution over all subsets of $[m]$ of size $r$.

Suppose that $S$ is distributed according to $\rho$. It is immediate to see that $\det(V_S^TV_S) \in \{0, 1\}$. Moreover, the determinant is 1 if and only if exactly one vector is picked from every group of $r$ copies of standard unit vectors.

$$\Pr \left[ \det(V_S^TV_S) = 1 \right] = \frac{r^r}{r^r} \approx \frac{r^r r!}{(r^r)^2} \approx \frac{r^r \cdot r^r}{(r^r)^2 e^r} = e^{-r}.$$ 

In the above estimate we used Stirling approximation and ignored small polynomial factors in $r$.

The above calculation implies that with probability exponentially close to one, the randomized algorithm of [NS16] returns a trivial solution 0. To obtain a solution of value at least the expectation (which is also roughly $e^{-r}$) one needs to run this algorithm about $e^r$ times.

\[\square\]

B Omitted Proofs

B.1 Proofs for Section 2

Proof of Lemma 2.1. Denote by $1_T, 1_{T_1}, 1_{T(x)}$ the indicator function of $T$, $T_1$ and $T(x)$, respectively. Since $T, T_1, T(x)$ are measurable sets, $1_T, 1_{T_1}, 1_{T(x)}$ are measurable functions. Also, observe that

$$1_T(x_1, \ldots, x_d) = 1_{T_1}(x_1)1_{T(x)}(x_2, \ldots, x_d).$$

Therefore

$$\lambda_d(T) = \int_R 1_T d\lambda_d = \int_R 1_{T_1} \int_{R^{d-1}} 1_{T(x)} d\lambda_{d-1} d\lambda_1.$$ 

Since for each $x \in T_1$ we have $\lambda_{\min} \leq \lambda_{d-1}(T(x)) = \int_{R^{d-1}} 1_{T(x)} d\lambda_{d-1} \leq \lambda_{\max}$ and $\int_{R} 1_{T_1} d\lambda_1 = \lambda_1(T)$, we conclude that

$$\lambda_{\min} \lambda_1(T_1) \leq \int_{R} 1_{T_1} \int_{R^{d-1}} 1_{T(x)} d\lambda_{d-1} d\lambda_1 \leq \lambda_{\max} \lambda_1(T_1).$$

\[\square\]
B.2 Proofs for Section 4

Proof of Fact 4.3. Let \( h : \Delta_t \to \mathbb{R}^d \) be the function defined by

\[
h(x) := \sum_{i \in [t]} x_i u_i.
\]

Clearly, \( f(x) = \|h(x)\|_2 \). First, we prove the statement for the case \( t = 2 \). In this case

\[
f(x) = \|x_1 u^1 + x_2 u^2\|_2.
\]

Without loss of generality, we may assume that \( \|u_1\|_2 \geq \|u_2\|_2 \). This assumption implies that the maximum of \( f \) is attained at \( x = (1, 0) \). This is a direct implication of the triangle inequality as follows.

\[
f(x_1, x_2) = \|x_1 u^1 + x_2 u^2\|_2 \leq x_1 \|u^1\|_2 + x_2 \|u^2\|_2 \leq (x_1 + x_2) \|u^1\|_2 = f(1, 0).
\]

In order to prove that \( f \) is 2-anti-concentrated we must show that

\[
\forall c > 0, \quad \Pr \left[ x \in \Delta_2 : \|x_1 u^1 + x_2 u^2\|_2 < c \|u^1\|_2 \right] < 2c.
\]

Since for all \( x_1, x_2 \geq 0 \), \( f(x) = \|x_1 u^1 + x_2 u^2\|_2 \geq |x_1 \|u^1\|_2 - x_2 \|u^2\|_2| \), we have

\[
\left\{ x \in \Delta_2 : \|x_1 u^1 + x_2 u^2\|_2 < c \|u^1\|_2 \right\} \subseteq \left\{ x \in \Delta_2 : |x_1 \|u^1\|_2 - x_2 \|u^2\|_2| < c \|u^1\|_2 \right\}.
\]

Thus, it is sufficient to show that the probability of the set in the right hand side of (30) is smaller than \( 2c \). Define \( \hat{f} : [0,1] \to \mathbb{R} \) by

\[
\hat{f}(z) := z \|u^1\|_2 - (1-z) \|u^2\|_2.
\]

Observe that for any \((x_1, x_2) \in \Delta_2\), \( f(x_1, x_2) \geq \hat{f}(x_1) \). If we are able to prove that

\[
\forall c \geq 0, \quad \Pr[|\hat{f}(z)| < c \cdot \text{OPT}] < 2c
\]

where \( \text{OPT} \) is the maximum of \(|\hat{f}| \) over \([0,1]\), then it implies that the probability of the set in the right hand side of (30) is less than \( 2c \) for every \( c \in (0,1) \). (Notice that \( \hat{f}(0) = \|u^1\|_2 \)). Without loss of generality, we may assume that \( \text{OPT} = 1 \). Then

\[
\left\{ z \in [0,1] : |\hat{f}(z)| < c \cdot \text{OPT} \right\} = \hat{f}^{-1}(-c, c) \cap [0,1].
\]

Therefore, it is enough to show that \( \lambda(\hat{f}^{-1}(-c, c) \cap [0,1]) < 2c \). We compute \( \hat{f}^{-1}(-c, c) \cap [0,1] \) for the affine function \( \hat{f} \) on the interval \([0,1]\) as following. Let \( \beta := \hat{f}(1) \).

\[
\hat{f}^{-1}(-c, c) \cap [0,1] = \begin{cases} 
\emptyset & \text{if } c < \beta, \\
\left[ \frac{1-c}{1-\beta}, 1 \right] & \text{if } -c \leq \beta \leq c, \\
\left[ \frac{1-c}{1-\beta}, \frac{1+c}{1-\beta} \right] & \text{if } -1 \leq \beta < -c.
\end{cases}
\]

In all of these cases one can check that the length of the interval never exceeds \( 2c \). Thus, the claim (31) is valid. Therefore, the measure of the right hand side of (30) is upper bounded by
2c. Consequently, the measure of the left hand side is also less than 2c. Since this holds for any $c \in (0, 1)$, by definition, $f$ is 2-anti-concentrated.

We use the special case of $t = 2$ as a tool to prove the fact for arbitrary $t$. First, we fix a number $c \in (0, 1)$. According to Lemma 4.2 the maximum of $f$ is attained at a vertex $v$ of $\Delta_t$. Without loss of generality, we may assume that $v = (1, 0, \ldots, 0)$. Define

$$T_c := \{ x \in \Delta_t : f(x) < cf(v) \}$$

Since $f$ is a continuous function, $T_c$ is an open measurable subset of $\Delta_t$. To prove that $f$ is 2-anti-concentrated it is enough to show

$$\lambda_{t-1}(T_c) < 2tc \cdot \lambda_{t-1}(\Delta_t).$$

Let

$$\Delta' := \{ x \in \Delta_t : x_1 = 0 \}.$$

For every two points $w, w' \in \mathbb{R}^d$, we call $\text{conv}\{w, w'\}$, a segment and denote it by $ww'$. $w, w'$ are called the endpoints of $ww'$. For every $v' \in \Delta'$, we call the segment $vv'$ a ray. For every $v' \in \Delta'$ define the function $h_{v'} : [0, 1] \to \mathbb{R}^d$ by

$$h_{v'}(\alpha) := \alpha h(v) + (1 - \alpha) h(v').$$

Due to the linearity of $h$, for every ray $vv'$ the value of $h$ restricted to $vv'$ is a convex combination of $h(v)$ and $h(v')$, that is

$$\forall \alpha \in [0, 1], \ h_{v'}(\alpha) = h(\alpha v + (1 - \alpha) v').$$

Thus, for every $v' \in \Delta'$ we can apply the result of $t = 2$ to $\|h_{v'}\|_2$, to conclude that $T_c(v')$ is an interval on the ray $vv'$ of Lebesgue measure less than $(1 - 2c)\lambda(vv')$ where

$$T_c(v') := T_c \cap vv'.$$

To compute the $(t - 1)$-dimensional volume of $T_c$, we proceed as follows. First, we consider an isometry $\Psi : \{ x \in \mathbb{R}^t : \sum_{i \in [t]} x_i = 1 \} \to \mathbb{R}^{t-1}$ such that $\Psi(v) = 0$. Then, we consider the hyperspherical representation of $\mathbb{R}^{t-1}$. Such a representation can be formally defined as a continuous map

$$\Phi : \mathbb{R}^{t-1} \to \left( \prod_{i \in [t-3]} [0, 1] \right) \times [0, 2\pi] \times [0, \infty)$$

given by

$$\Phi(x) = (\phi_1(x), \ldots, \phi_{t-2}(x), l(x))$$

in which

$$\phi_j(x) = \arccos \left( \frac{x_j}{\sqrt{x_{t-1}^2 + \cdots + x_{j+1}^2}} \right) \quad \forall j \in [t-3],$$

$$\phi_{t-2}(x) = 2 \cdot \arccot \left( \frac{x_{t-2} + \sqrt{x_{t-1}^2 + x_{t-2}^2}}{x_{t-1}} \right),$$

$$l(x) = \sqrt{\sum_{i \in [t-1]} x_i^2}. $$

28
From the above equations it is clear that by scaling the Euclidean coordinates of any point, all the angular coordinates of it remain unchanged. The differential element of \(dx_1 \cdots dx_{t-1}\) in the new coordinate system is equal to

\[ t^{-2} \sin^{t-3}(\phi_1) \cdots \sin(\phi_{t-3})dl \cdot d\phi_1 \cdots d\phi_{t-2}. \]

Let \(\Theta_c := \Phi(\Psi(T_c)), \hat{\Delta}_t := \Phi(\Psi(\Delta_t))\). Let \(\phi := (\phi_1, \ldots, \phi_{t-2})\). For every vector \(\phi\), let \(l(\phi)\) be the length of the ray whose angular coordinates are given by \(\phi\). Since \(\Psi\) is an isometry and \(\Phi\) is a change of variable, the volume of \(T_c\) can be computed by

\[
\lambda_{t-1}(T_c) = \int_0^{2\pi} \int_0^\pi \cdots \int_0^\pi \int_0^{l(\phi)} 1_{\Theta_c} t^{-2} \sin^{t-3}(\phi_1) \cdots \sin(\phi_{t-3})dl \cdot d\phi_1 \cdots d\phi_{t-2},
\]

and the volume of \(\Delta_t\) by

\[
\lambda_{t-1}(\Delta_t) = \int_0^{2\pi} \int_0^\pi \cdots \int_0^\pi \int_0^{l(\phi)} 1_{\hat{\Delta}_t} t^{-2} \sin^{t-3}(\phi_1) \cdots \sin(\phi_{t-3})dl \cdot d\phi_1 \cdots d\phi_{t-2}.
\]

Since our objective is to show that \(\lambda_{t-1}(T_c) < 2ct \cdot \lambda_{t-1}(\hat{\Delta}_t)\), it is enough to show that for every fixed value of \(\phi\), we have

\[
\int_0^{l(\phi)} 1_{\Theta_c} t^{-2} dl \leq 2ct \int_0^{l(\phi)} 1_{\hat{\Delta}_t} t^{-2} dl. \tag{33}
\]

Notice that the term \(1_{\Theta_c}\) is equal to 1 only on the set \(\Phi(\Psi(T_c))\) and is equal to 0, elsewhere. Furthermore, since we only consider the specific value of \(\phi\) (which corresponds to a ray), we know that the support of \(1_{\Theta_c}\) on this ray is a set of Lebesgue measure at most \(2cl(\phi)\). Thus, the inner integral is equal to

\[
\int_{J \in J} t^{-2} dl \tag{34}
\]

where \(J\) is a measurable subset of \([0, l(\phi)]\) of measure at most \(2c \cdot l(\phi)\). Since \(l^{-2}\) is an increasing function of \(l\), the maximum of \(\int_{[l(\phi) - 2c \cdot l(\phi), l(\phi)]} t^{-2}\) is attained when \(J = [l(\phi) - 2c \cdot l(\phi), l(\phi)]\). The maximum value is also equal to

\[
\frac{l(\phi)^{t-1}}{t-1} - \frac{(l(\phi) - 2c \cdot l(\phi))^{t-1}}{t-1}.
\]

For the right hand side of \(\int_{[l(\phi) - 2c \cdot l(\phi), l(\phi)]} t^{-2}\), we can simply replace the term \(1_{\hat{\Delta}_t}\) by 1. Thus, the value of the integral in the right hand side of \(\int_{[l(\phi) - 2c \cdot l(\phi), l(\phi)]} t^{-2}\) is equal to

\[
\frac{l(\phi)^{t-1}}{t-1}
\]

Using the inequality \(1 - (1 - 2c)^{t-1} \leq 2ct\) and by a simple manipulation one can conclude that

\[
\frac{l(\phi)^{t-1}}{t-1} - \frac{(l(\phi) - 2c \cdot l(\phi))^{t-1}}{t-1} \leq 2ct \frac{l(\phi)^{t-1}}{t-1}.
\]

This completes the proof. 

\[\blacksquare\]