Log-Concave Polynomials, Entropy, and a Deterministic Approximation Algorithm for Counting Bases of Matroids

Nima Anari
Computer Science Department
Stanford University
anari@cs.stanford.edu

Shayan Oveis Gharan
Computer Science and Engineering
University of Washington
shayan@cs.washington.edu

Cynthia Vinzant
Department of Mathematics
North Carolina State University
clvinzan@ncsu.edu

***Abstract*—We give a deterministic polynomial time $2^{O(r)}$-approximation algorithm for the number of bases of a given matroid of rank r and the number of common bases of any two matroids of rank r. To the best of our knowledge, this is the first nontrivial deterministic approximation algorithm that works for arbitrary matroids. Based on a lower bound of Azar, Broder, and Frieze [7] this is almost the best possible assuming oracle access to independent sets of the matroid.**

There are two main ingredients in our result: For the first, we build upon recent results of Adiprasito, Huh, and Katz [1] and Huh and Wang [23] on combinatorial hodge theory to derive a connection between matroids and log-concave polynomials. We expect that several new applications in approximation algorithms will be derived from this connection in future. Formally, we prove that the multivariate generating polynomial of the bases of any matroid is log-concave as a function over the positive orthant. For the second ingredient, we develop a general framework for approximate counting in discrete problems, based on convex optimization. The connection goes through subadditivity of the entropy. For matroids, we prove that an approximate superadditivity of the entropy holds by relying on the log-concavity of the corresponding polynomials.

***Keywords*-matroid; deterministic counting; entropy; log-concave polynomial;**

I. Introduction

Efficient algorithms for optimizing linear functions over convex sets, convex programming, are one of the pinnacles of algorithm design. Convex sets yield easy instances of optimization in the continuous world. Much the same way, matroids yield easy optimization instances in the discrete world. Going beyond, computing the volume of convex sets is well understood algorithmically; however there has not been an analogous progress on counting problems involving matroids. In this work, we try to address this by designing nearly tight deterministic approximate counting algorithms for discrete structures involving matroids and their intersections. We introduce a general optimization-based algorithm for approximate counting involving discrete objects, and show that it performs well for matroids and their intersections.

A matroid $M = (E, \mathcal{I})$ is a structure consisting of a finite ground set E and a non-empty collection $\mathcal{I}$ of *independent* subsets of E satisfying:

i) If $S \subseteq T$ and $T \in \mathcal{I}$, then $S \in \mathcal{I}$.
ii) If $S, T \in \mathcal{I}$ and $|T| > |S|$, then there exists an element $i \in T \setminus S$ such that $S \cup \{i\} \in \mathcal{I}$.

The *rank* of a matroid is the size of the largest independent set of that matroid. If M has rank r, any set $S \in \mathcal{I}$ of size r is called a *basis* of M. Let $\mathcal{B}_M \subset \mathcal{I}$ denote the bases of M.

Many optimization problems are well understood on matroids. Matroids are exactly the class of objects for which an analogue of Kruskal's algorithm works and gives the smallest weight basis.

One can associate to any matroid M a polytope $\mathcal{P}_M$, defined by exponentially many constraints, called the *matroid base polytope*. The vertices of $\mathcal{P}_M$ are the indicator vectors of all bases of M, i.e., $\mathcal{P}_M = \text{conv}\{\mathbb{1}_B \mid B \in \mathcal{B}_M\}$. Furthermore, using the duality of optimization and separation, one can design a separation oracle for $\mathcal{P}_M$ in order to minimize any convex function over $\mathcal{P}_M$ [15].

More difficult problems associated to matroids come from counting. For example, given a matroid M, is there a polynomial time algorithm to count the number of bases of M? This problem is #P-hard in the worst case even if the matroid is representable over a finite field [35], so the next natural question is: How well can we approximate the number of bases of a given matroid M in polynomial time? This is the main question addressed in this paper.

Note that the number of bases of any matroid M of rank r is at most $\binom{|E|}{r} \approx |E|^r$, so there is a simple $|E|^r$ approximation to the number of bases of M.

We also address counting problems on the intersection of two matroids. Given two matroids $M = (E, \mathcal{I}_M), N = (E, \mathcal{I}_N)$ of rank r on the same ground set E, the matroid intersection problem is to optimize a (linear) function over all bases B common to both M and N. This problem can also be solved in polynomial time because $\mathcal{P}_M \cap \mathcal{P}_N$ is exactly the convex hull of the indicator vectors of $\mathcal{B}_M \cap \mathcal{B}_N$ [see, e.g., 32]. Perhaps, the most famous example of matroid intersection is perfect matchings in bipartite graphs. Since we can optimize over the intersection of two matroids, it is natural to ask if one can approximate the number of bases common to two rank-r matroids M and N. This is the second problem that we address in this paper.

2575-8454/18/$31.00 ©2018 IEEE
DOI 10.1109/FOCS.2018.00013

Note that there are NP-hard problems involving the intersection of just three matroids, e.g., the Hamiltonian path problem. It is NP-hard to test if there is a single basis in the intersection of three matroids. We will not discuss intersections of more than two matroids in this paper, since any multiplicative approximation would be NP-hard.

A. Previous Work

There is a long line of work on designing approximation algorithms to count the number of bases of a matroid. One idea is to use the Markov Chain Monte Carlo technique. For any matroid M, there is a well-known chain, called the basis exchange walk, which mixes to the uniform distribution over all bases. Mihail and Vazirani conjectured, about three decades ago, that the chain mixes in polynomial time, and hence one can approximate the number of bases of any matroid on n elements within $1+\epsilon$ factor in time $\mathrm{poly}(n, 1/\epsilon)$. To this date, this conjecture has been proved only for a special class of matroids known as *balanced matroids* [16, 30]. Balanced matroids are special classes of matroids where the uniform distribution over the bases of the matroid, and any of its minors, satisfies the pairwise negative correlation property. Unfortunately, many interesting matroids are not balanced. An important example is the matroid of all acyclic subsets of edges of a graph $G = (V, E)$ of size at most k (for some $k < |V| - 1$) [16].

Many of the results in this area [3, 12, 13, 17, 25, 27, 28] study approximation algorithms for a limited class of matroids, and not much is known beyond balanced matroids.

Most of the classical results in approximate counting rely on randomized algorithms based on the Markov Chain Monte Carlo technique. There are also a few results in the literature that exploit tools from convex optimization [8, 9, 24]. To the best of our knowledge, the only non-trivial approximation algorithm that works for any matroid is a *randomized* algorithm of Barvinok and Samorodnitsky [9] that gives, roughly, a $\log(|E|)^r$ approximation factor to the number of bases of a given matroid of rank r and the number of common bases of any two given matroids of rank r, in the worst case. We remark that this algorithm works for any family of subsets, not just matroids and their intersections, assuming access to an optimization oracle. The approximation factor gets better if the number of bases of the given matroid(s) are significantly less than $\binom{|E|}{r} \simeq |E|^r$.

On the negative side, Azar, Broder, and Frieze [7] showed that any deterministic polynomial time algorithm that has access to the matroid M on n elements only through an independence oracle can only approximate the number of bases of M up to a factor of $2^{O(n/\log(n)^2)}$. They actually showed the stronger result that any deterministic algorithm making k queries to the independence oracle must have an approximation factor of at least $2^{\Omega(n/\log(k)^2)}$, as long as $k = 2^{o(n)}$. An immediate corollary is a *rank-dependent* lower bound, namely that any deterministic algorithm making polynomially many independence queries to a matroid of rank r must have an approximation ratio of $2^{\Omega(r/\log(n)^2)}$ as long as $r \gg \log(n)$. This is because we can always start with a matroid on $\simeq r$ elements and add loops to get n elements without changing the number of bases or the rank.

Approximating the number of bases in the intersection of two matroids M, N is poorly understood. Jerrum, Sinclair, and Vigoda [26] give a randomized polynomial time approximation to the number of perfect matchings of a bipartite graph, a special case of matroid intersection, up to a factor of $1+\epsilon$. As for deterministic algorithms, for this special case of bipartite perfect matchings, a $2^{O(r)}$-approximation was first introduced by Linial, Samorodnitsky, and Wigderson [29], relying on the Van der Waerden conjecture, and later an improvement in the base of the exponent was achieved by Gurvits and Samorodnitsky [21]. Recently, a subset of the authors [2] have shown how to approximate the number of bases in the intersection of two *real stable* matroids, having oracle access to each of their generating polynomials, up to a $2^{O(r)}$ multiplicative error [also, cf. 36]. These are a special case of balanced matroids. See the techniques in the full version of this paper and the prior work [2] for more details.

B. Our Results

The main result of this paper is the following.

Theorem 1. *Let $M = ([n], \mathcal{I})$ be a matroid of rank r given by an independence oracle. There is a* deterministic *polynomial time algorithm that outputs a number β satisfying*

$$\max\{2^{-O(r)}\beta, \sqrt{\beta}\} \leq |\mathcal{B}_M| \leq \beta,$$

where $\mathcal{B}_M$ is the set of bases of M.

Our algorithm can be implemented with only oracle access to the independent sets of the matroid. Therefore, by the work of Azar, Broder, and Frieze [7], this is almost the best we can hope for any deterministic algorithm.

As an immediate corollary of the above result we can count the number of independent sets of any given size k. Independent sets of size k form the bases of the truncation of the original matroid, which itself is a matroid.

Corollary 2. *Let M be a matroid given by an independence oracle. There is a deterministic polynomial time algorithm that for any given integer k outputs a number β such that*

$$\max\{2^{-O(k)}\beta, \sqrt{\beta}\} \leq |\mathcal{I}_M^k| \leq \beta,$$

where $\mathcal{I}_M^k$ is the set of independence sets of size exactly k.

Going further, we show that one can approximate the number of bases in the intersection of *any* two matroids.

Theorem 3. *Let M and N be matroids of rank r on the set $[n]$ given by independence oracles. There is a* deterministic *polynomial time algorithm that outputs β satisfying*

$$2^{-O(r)}\beta \leq |\mathcal{B}_M \cap \mathcal{B}_N| \leq \beta,$$

where $\mathcal{B}_M, \mathcal{B}_N$ are the bases of M and N, respectively.

Counting common bases of two matroids is a self-reducible problem. Roughly speaking, this means that if we want to count common bases that include some given elements $i_1, \ldots, i_k \in [n]$ and exclude some other elements $j_1, \ldots, j_l \in [n]$, then we get an instance of the same problem; we just have to replace the input matroids by their minors.

Sinclair and Jerrum [33] showed that for self-reducible counting problems, the approximation factor of any algorithm can be boosted at the expense of increased running time and using randomization. As a corollary of Theorem 3 and the results of Sinclair and Jerrum [33] we get the following.

Corollary 4. *Let M and N be two matroids of rank r on the set $[n]$ given by independence oracles. There is a* randomized *algorithm that for any $\epsilon, \delta > 0$ outputs a number β approximating the number of common bases of M and N within a factor of $1-\epsilon$ with probability at least $1-\delta$:*

$$\mathbb{P}\Big[(1-\epsilon)\beta \leq |\mathcal{B}_M \cap \mathcal{B}_N| \leq \beta\Big] \geq 1-\delta.$$

The running time of this algorithm is $2^{O(r)} \operatorname{poly}(n, \frac{1}{\epsilon}, \log\frac{1}{\delta})$.

Counting bases of a single matroid is a special case obtained by letting $M = N$. Observe that this algorithm becomes a fully polynomial time randomized approximation scheme (FPRAS) when $r = O(\log n)$. A generalization of our algorithmic framework provides handles weighted counting:

Theorem 5. *Let M and N be two matroids of rank r on the set $[n]$ given by independence oracles. There is a deterministic polynomial time algorithm that for given weights $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_n) \in \mathbb{R}^n_{\geq 0}$, outputs β which $2^{O(r)}$-approximates the $\boldsymbol{\lambda}$-weighted intersection of M and N:*

$$2^{-O(r)}\beta \leq \sum_{B \in \mathcal{B}_M \cap \mathcal{B}_N} \prod_{i \in B} \lambda_i \leq \beta.$$

C. Techniques

In this section we discuss the main ideas of our proof. We rely on the *basis generating polynomial* of a matroid M,

$$g_M(z_1, \ldots, z_n) = \sum_{B \in \mathcal{B}_M} \prod_{i \in B} z_i. \tag{1}$$

For some matroids, such as partition matroids and graphic matroids, the polynomial g_M has a special property called real stability. A multivariate polynomial $g \in \mathbb{R}[z_1, \ldots, z_n]$ is *real stable* if $g(z_1, \ldots, z_n) \neq 0$ whenever $\operatorname{Im}(z_i) > 0$ for all $i = 1, \ldots, n$. Real stable polynomials have been used for numerous counting and sampling tasks [2–5, 19, 31, 36].

Matroids with real stable basis generating polynomials are called *real stable matroids*. Many properties of these matroids can be derived from those of their generating polynomials. If g_M is real stable, then the uniform distribution over the bases of the matroid, as well as its minors, satisfy pairwise negative correlation. Then M is a balanced matroid and one can count the number of bases of M within a $1+\epsilon$ multiplicative error by a polynomial time randomized algorithm.

On the counting side, roughly speaking, real stable matroids are almost all we know how to handle. However, it is known that many matroids are not real stable. Even if we allow arbitrary positive coefficients in front of the monomials in the generating polynomial, instead of uniform coefficients, for some matroids we can never get a real stable polynomial [11]. Here we define a more general class of polynomials, namely log-concave and completely log-concave polynomials, to be able to study all matroids with analytical techniques.

Given a real stable polynomial $g(z_1, z_2, \ldots, z_n)$, its univariate restriction $g(z, z, \ldots, z)$ is real-rooted, and it follows that its coefficients form a *log-concave* sequence. Recently, Adiprasito, Huh, and Katz [1] proved that certain univariate polynomials associated with matroids have log-concave coefficients, for *any* matroid. Their work resolved several long standing open problems in combinatorics. Note that such results are unlikely to follow from the theory of real stability because not all matroids support real stable polynomials.

The first ingredient of our paper is that we exploit some of the results and theory developed by Adiprasito, Huh, and Katz [1] to show that $\log(g_M(z_1, \ldots, z_n))$ is concave as a function on the positive orthant (see Theorem 25). Any real stable polynomial with nonnegative coefficients is log-concave on the positive orthant but the converse is not necessarily true. In this paper we study properties of log-concave polynomials with nonnegative coefficients as a generalization of real stable polynomials and show that they satisfy many of the closure properties of real stable polynomials. Here, we mainly focus on applications in approximate counting, but we hope that these techniques can be used for many other applications in algorithm design, operator theory, and combinatorics.

Our second ingredient is a general framework for approximate counting based on convex optimization. We consider probability distributions $\mu : 2^{[n]} \to \mathbb{R}_{\geq 0}$ on subsets of $[n]$. Firstly, we show that if μ has a log-concave generating polynomial, we can approximate its entropy using the marginal probabilities of the underlying elements (see Theorem 31). The marginal probability μ_i of an element $i \in [n]$ is the probability that i is included in a random sample of μ. We show that $\sum_{i=1}^n (\mu_i \log\frac{1}{\mu_i} + (1-\mu_i)\log\frac{1}{1-\mu_i})$ gives a "good" approximation of the entropy $\mathcal{H}(\mu)$ of μ.

This is particularly interesting when μ is the uniform distribution over the bases of a matroid $M = ([n], \mathcal{I})$, in which case $\mathcal{H}(\mu)$ equals $\log(|\mathcal{B}_M|)$. From the marginals μ_i, one can approximate $\mathcal{H}(\mu)$, but finding the marginal probabilities is no easier than estimating $\mathcal{H}(\mu)$. Instead, we observe that the vector of marginals $(\mu_1, \ldots, \mu_n)$ must lie in $\mathcal{P}_M$. So instead of trying to find μ_i's, we use a convex program to find a point $\boldsymbol{p} = (p_1, \ldots, p_n) \in \mathcal{P}_M$ maximizing the sum of marginal entropies, $\sum_{i=1}^n (p_i \log\frac{1}{p_i} + (1-p_i)\log\frac{1}{1-p_i})$. Using properties of maximum entropy convex programs (see Theorem 16), we show that this also gives a "good" approximation of $\mathcal{H}(\mu) = \log(|\mathcal{B}_M|)$.

To prove Theorem 3, we exploit some of the tools that a subset of authors developed [2] to approximate the number of bases in the intersection of two real stable matroids. In this paper we generalize these techniques to all matroids. In the process we show that for any matroid M the polynomial g_M is *completely log-concave*, meaning that taking directional derivatives of the polynomial g_M with respect to directions in $\mathbb{R}^n_{\geq 0}$ results in a polynomial that as a function is log-concave on the positive orthant (see Theorem 25).

D. Algorithmic Framework

Our algorithms in the cases of a single matroid and the intersection of two matroids are actually instantiations of the same framework that could be applied to more general discrete structures. We use a general framework based on convex programming. Take an arbitrary family of sets $\mathcal{B} \subseteq 2^{[n]}$. For us $\mathcal{B}$ will be the bases of a matroid, or common bases of two matroids, but our framework could be applied to any family. We assume that we can optimize linear functions over $\mathcal{P}_\mathcal{B} = \operatorname{conv}\{\mathbb{1}_B \mid B \in \mathcal{B}\}$. This is true in both cases involving matroid(s), when we have access to the independence oracle(s).

The key observation is that the entropy of the uniform distribution μ over the elements of $\mathcal{B}$ equals $\log(|\mathcal{B}|)$ and that using subadditivity of the entropy this entropy can be related to the points in $\mathcal{P}_\mathcal{B}$. If μ_i is the marginal probability of element i being in a uniformly random element of $\mathcal{B}$, then $(\mu_1, \ldots, \mu_n)$ is a point of $\mathcal{P}_\mathcal{B}$; it is the average of all vertices. Together with subadditivity of the entropy, this gives us

$$\begin{aligned}\log(|\mathcal{B}|) = \mathcal{H}(\mu) &\leq \sum_{i=1}^n (\mu_i \log\frac{1}{\mu_i} + (1-\mu_i)\log\frac{1}{1-\mu_i}) \\ &\leq \max\left\{\begin{array}{c} \sum_{i=1}^n (p_i \log\frac{1}{p_i} + (1-p_i)\log\frac{1}{1-p_i}) \\ \text{for } \boldsymbol{p} = (p_1, \ldots, p_n) \in \mathcal{P}_\mathcal{B} \end{array}\right\}.\end{aligned}$$

We can efficiently compute the last quantity, because we can optimize over $\mathcal{P}_\mathcal{B}$. Therefore, we have a convex-programming-based way of obtaining an upper bound for $\log(|\mathcal{B}|)$. Exponentiating the result, we get an upper bound on $|\mathcal{B}|$; this is the output of our algorithm, β in Theorems 1 and 3. We show that for matroids and intersections of two matroids, this bound becomes an approximation, in that there is a complementary lower bound. We leave finding more discrete structures for which this algorithm provides a "good" approximation open. We make the following concrete conjecture.

Conjecture 6. *Let $G = (V, E)$ be a graph with an even number of nodes, and let $\mathcal{B} \subseteq 2^E$ be the set of all perfect matchings in G. Then*

$$\max\left\{\sum_{i\in E}(p_i \log\frac{1}{p_i} + (1-p_i)\log\frac{1}{1-p_i}) \;\middle|\; \boldsymbol{p} \in \mathcal{P}_\mathcal{B}\right\} \leq O(|V|) + \log(|\mathcal{B}|).$$

If Conjecture 6 is correct, we get a deterministic polynomial time $2^{O(|V|)}$-approximation algorithm for counting perfect matchings, because we can efficiently optimize over $\mathcal{P}_\mathcal{B}$. To the best of our knowledge, for nonbipartite graphs, no such result is known as of this writing.

Finally we remark that this framework can handle weighted counting. For weight vector $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_n) \in \mathbb{R}^n_{\geq 0}$, the $\boldsymbol{\lambda}$-weighted count of a family $\mathcal{B} \subseteq 2^{[n]}$ is simply $\sum_{B\in\mathcal{B}} \prod_{i\in B} \lambda_i$. For $\lambda_1, \ldots, \lambda_n = 1$, this quantity becomes $|\mathcal{B}|$. To handle $\boldsymbol{\lambda}$-weighted counts we can simply change our concave program slightly to

$$\max\left\{\begin{array}{c} \sum_{i=1}^n (p_i \log\frac{\lambda_i}{p_i} + (1-p_i)\log\frac{1}{1-p_i}) \\ \text{for } \boldsymbol{p} = (p_1, \ldots, p_n) \in \mathcal{P}_\mathcal{B} \end{array}\right\}.$$

We defer more details to the full version of this work.

E. Organization

The rest of the paper is organized as follows. In Section II, we go over the necessary preliminaries from matroid theory, convex entropy programs, linear algebra and develop some of the theory of log-concave polynomials. In Section III we review some tools from combinatorial Hodge theory and derive a connection with the application of differential operators to the basis generating polynomial of a matroid. Then, in Section IV we prove that the generating polynomial of *any* matroid is log-concave. In Section V we prove that one can approximate the entropy of a log-concave distribution from its marginals. Finally, in Section VI we prove Theorem 1 and we defer the proof of Theorem 3 to the full version of this paper. Some proofs are omitted here and can be found in the full version of the paper.

II. Preliminaries

Let us establish some notational conventions. Unless otherwise specified, all logs are in base e. We use bold letters to emphasize symbols representing a vector, array, or matrix. All vectors are assumed to be column vectors. For $\boldsymbol{v}, \boldsymbol{w} \in \mathbb{R}^n$, we denote the standard dot product between $\boldsymbol{v}$ and $\boldsymbol{w}$ by $\langle \boldsymbol{v}, \boldsymbol{w}\rangle = \boldsymbol{v}^\intercal \boldsymbol{w}$. We use $\mathbb{R}_{>0}$ and $\mathbb{R}_{\geq 0}$ to denote positive and nonnegative real numbers, respectively, and $[n]$ to denote $\{1, \ldots, n\}$. When n is clear from context, for a set $S \subseteq [n]$, we let $\mathbb{1}_S \in \mathbb{R}^n$ denote the indicator vector of S; $(\mathbb{1}_S)_i = 1$ if $i \in S$, and is 0 otherwise. We let $\mathbb{1}_i = \mathbb{1}_{\{i\}}$ be the i-th element of the standard basis. For vectors $\boldsymbol{z}, \boldsymbol{p} \in \mathbb{R}^n$ we use $\boldsymbol{z}^{\boldsymbol{p}}$ to denote $\prod_{i=1}^n z_i^{p_i}$. Similarly we let $e^{\boldsymbol{p}}$ denote $\prod_{i=1}^n e^{p_i} = e^{\sum_{i=1}^n p_i}$. For set S we let $\boldsymbol{z}^S = \boldsymbol{z}^{\mathbb{1}_S}$.

We use ∂_{z_i} or ∂_i to denote the partial differential operator $\partial/\partial z_i$. Given $\boldsymbol{v} = (v_1, \ldots, v_n) \in \mathbb{R}^n$, we use $\partial_{\boldsymbol{v}}$ to denote $\sum_{i=1}^n v_i \partial_i$. For a collection of vectors $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k \in \mathbb{R}^n$, we use $D_{\boldsymbol{v}_1,\ldots,\boldsymbol{v}_k}$ to denote $\prod_{i=1}^k \partial_{\boldsymbol{v}_i}$. For a matrix $\boldsymbol{V} = [\boldsymbol{v}_1 \mid \ldots \mid \boldsymbol{v}_k] \in \mathbb{R}^{n\times k}$, viewed as a collection of column vectors, we use $D_{\boldsymbol{V}}$ to denote $D_{\boldsymbol{v}_1,\ldots,\boldsymbol{v}_k}$, i.e., $D_{\boldsymbol{V}} = \prod_{j=1}^k \sum_{i=1}^n V_{ij}\partial_i$. We denote the gradient of a polynomial g by ∇g and the Hessian of g by $\nabla^2 g$. For a polynomial $g \in \mathbb{R}[z_1, \ldots, z_n]$ and $c \in \mathbb{R}$ we write $g(z_1, \ldots, z_n)|_{z_1=c}$ to denote the restricted polynomial in $z_2, \ldots, z_n$ obtained by setting $z_1 = c$. For a polynomial $g(z_1, \ldots, z_n)$, we define $\mathrm{supp}(g) \subset \mathbb{Z}^n_{\geq 0}$ as the set of vectors $\boldsymbol{\kappa} = (\kappa_1, \ldots, \kappa_n) \in \mathbb{Z}^n_{\geq 0}$ such that the coefficient of the monomial $\prod_i z_i^{\kappa_i}$ is nonzero. The convex hull of $\mathrm{supp}(g)$ is known as the Newton polytope of g. We call a polynomial multiaffine if the degree of each variable is at most one.

A. Log-Concave Polynomials

A polynomial $g \in \mathbb{R}[z_1, \ldots, z_n]$ with nonnegative coefficients is *log-concave* if for any two points $\boldsymbol{v}, \boldsymbol{w} \in \mathbb{R}^n_{\geq 0}$ and $\lambda \in [0, 1]$ we have $g(\lambda \boldsymbol{v} + (1-\lambda)\boldsymbol{w}) \geq g(\boldsymbol{v})^{\lambda} \cdot g(\boldsymbol{w})^{1-\lambda}$.

A polynomial g with nonnegative coefficients is log-concave if and only if the Hessian of $\log(g)$ is negative semidefinite at all points of $\mathbb{R}^n_{\geq 0}$ where it is defined. The set of log-concave polynomials is closed in the space of polynomials of degree $\leq d$. To see this, note that the nonnegativity of the coefficients of g ensure that if $g \neq 0$, then $\log(g)$ is defined at all points of $\mathbb{R}^n_{>0}$. The entries of $\nabla^2 \log(g)$ at a point in $\mathbb{R}^n_{>0}$ are continuous functions in the coefficients of g. The closed-ness of the set of log-concave polynomials follows from the closed-ness of the cone of negative semidefinite matrices. The basic operations that preserve log-concavity are affine transformations.

Lemma 7. *Let $g \in \mathbb{R}[z_1, \ldots, z_n]$ be a log-concave polynomial with nonnegative coefficients. Then for any affine transformation $T : \mathbb{R}^m \to \mathbb{R}^n$ with $T(\mathbb{R}^m_{\geq 0}) \subseteq \mathbb{R}^n_{\geq 0}$, $g(T(y_1, \ldots, y_m)) \in \mathbb{R}[y_1, \ldots, y_m]$ has nonnegative coefficients and is log-concave.*

The following operations will be useful for us.

Proposition 8. *The following operations preserve log-concavity:*

1. *Permutation: $g \mapsto g(z_{\pi(1)}, \ldots, z_{\pi(n)})$ for $\pi \in S_n$.*
2. *Specialization: $g \mapsto g(a, z_2, \ldots, z_n)$, where $a \in \mathbb{R}_{\geq 0}$.*
3. *Scaling: $g \mapsto c \cdot g(\lambda_1 z_1, \ldots, \lambda_n z_n)$, where $c, \lambda_1, \ldots, \lambda_n \in \mathbb{R}_{\geq 0}$.*
4. *Expansion: $g(z_1, \ldots, z_n) \mapsto g(y_1 + y_2 + \cdots + y_m, z_2, \ldots, z_n) \in \mathbb{R}[y_1, \ldots, y_m, z_2, \ldots, z_n]$.*
5. *Multiplication: $g, h \mapsto g \cdot h$ where g, h are log-concave.*

Unlike real stability, log-concavity is not preserved under taking derivatives. In Section IV, we remedy this by considering completely log-concave polynomials, for which log-concavity is preserved under differentiation.

B. Matroids

Let $M = (E, \mathcal{I})$ be a matroid, as defined in Section I. For any set $S \subseteq E$, the rank of S, denoted $\mathrm{rank}(S)$, is the size of the largest subset $A \subseteq S$ such that $A \in \mathcal{I}$. The rank of the matroid is the rank of the set E, and a set $B \subseteq E$ is a basis of M if and only if $B \in \mathcal{I}$ and $\mathrm{rank}(B) = \mathrm{rank}(E)$.

We say a matroid M is *simple* if it has no loops and no parallel elements, meaning that for all pairs $i \neq j \in E$, $\mathrm{rank}(\{i, j\}) = 2$. The *dual* matroid of M is the matroid $M^* = (E, \mathcal{I}^*)$ on the same set of elements E whose bases are the complements $E \setminus B$ of bases B of M (and whose independent are subsets of those bases).

The matroid base polytope $\mathcal{P}_M \subset \mathbb{R}^E$ of M is the convex hull of the indicator vectors of its bases. If M has rank r, it can also be defined by the following system of inequalities:

$$\mathcal{P}_M = \left\{ \boldsymbol{p} \in \mathbb{R}^E \;\middle|\; \begin{array}{ll} \langle \mathbb{1}_E, \boldsymbol{p} \rangle = r, & \\ \langle \mathbb{1}_S, \boldsymbol{p} \rangle \leq \mathrm{rank}(S) & \forall S \subseteq E, \\ \langle \mathbb{1}_i, \boldsymbol{p} \rangle \geq 0 & \forall i \in E. \end{array} \right\}. \tag{2}$$

While this description requires exponentially many constraints, because of the matroidal structure, one can optimize linear functions over $\mathcal{P}_M$ in polynomial time, assuming access to an independence oracle. Thereby one can construct a separation oracle for $\mathcal{P}_M$ [15] and therefore minimize any convex function over $\mathcal{P}_M$ in polynomial time. See Boyd and Vandenberghe [10] for background on convex optimization.

For $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_n) \in \mathbb{R}^n_{\geq 0}$, we define $\boldsymbol{\lambda}$-weight of a subset of E as $\boldsymbol{\lambda}^S = \prod_{i \in S} \lambda_i$. If g_M is the basis generating polynomial of M, as in Eq. (1), the $\boldsymbol{\lambda}$-weight of a basis is the coefficient of the corresponding monomial in g_M after scaling the variables by $\lambda_1, \ldots, \lambda_n$,

$$g_M(\lambda_1 z_1, \ldots, \lambda_n z_n) = \sum_{B \in \mathcal{B}_M} \boldsymbol{\lambda}^B \prod_{i \in B} z_i.$$

C. Linear Algebra

We say that a sequence of real numbers $\beta_1 \geq \cdots \geq \beta_n$ *interlaces* $\alpha_1 \geq \alpha_2 \geq \cdots \geq \alpha_n$ if

$$\alpha_1 \geq \beta_1 \geq \alpha_2 \geq \cdots \geq \beta_{n-1} \geq \alpha_n \geq \beta_n.$$

The following is the useful Cauchy's Interlacing Theorem:

Theorem 9 (Cauchy's Interlacing Theorem I [see 22, Corollary 4.3.9])**.** *For a symmetric matrix $A \in \mathbb{R}^{n \times n}$ and $v \in \mathbb{R}^n$, the eigenvalues of A interlace the eigenvalues of $A + vv^{\intercal}$.*

The following is an immediate consequence:

Lemma 10. *Let $A \in \mathbb{R}^{n \times n}$ be a symmetric matrix and let $P \in \mathbb{R}^{m \times n}$. If A has at most one positive eigenvalue, then $PAP^{\intercal}$ has at most one positive eigenvalue.*

Another version of Cauchy's Interlacing Theorem can be stated as follows:

Theorem 11 (Cauchy's Interlacing Theorem II, [see 22, Theorem 4.3.17])**.** *Let $A \in \mathbb{R}^{n \times n}$ be symmetric and $B \in \mathbb{R}^{(n-1) \times (n-1)}$ a principal submatrix of A. Then the eigenvalues of B interlace the eigenvalues of A. That is,*

$$\alpha_1 \geq \beta_1 \geq \alpha_2 \geq \cdots \geq \beta_{n-1} \geq \alpha_n,$$

where $\alpha_1, \ldots, \alpha_n$ and $\beta_1, \ldots, \beta_{n-1}$ are the eigenvalues of A and B, respectively.

This has the immediate corollary:

Corollary 12. *Let $A \in \mathbb{R}^{n\times n}$ be a symmetric matrix. If there is an $(n-1)$-dimensional linear subspace on which the quadratic form $z \mapsto z^\intercal Az$ is nonpositive, then A has at most one positive eigenvalue.*

We will also need the following lemma inspired by arguments of Huh and Wang [23].

Lemma 13. *Let $A \in \mathbb{R}^{n\times n}$ be a real symmetric matrix with nonnegative entries and at most one positive eigenvalue. Then for every $v \in \mathbb{R}^n_{\geq 0}$, the $n\times n$ matrix $(v^\intercal Av)\cdot A - t(Av)(Av)^\intercal$ is negative semidefinite for all $t \geq 1$.*

Proof: By limiting arguments, it suffices to prove this for $v \in \mathbb{R}^n_{>0}$ and $t = 1$. If A is the zero matrix, the claim is immediate. Otherwise, $v^\intercal Av > 0$. Let $w \in \mathbb{R}^n$ and consider the $2 \times n$ matrix P with rows $v^\intercal$ and $w^\intercal$. Then

$$PAP^\intercal = \begin{bmatrix} v^\intercal Av & v^\intercal Aw \\ w^\intercal Av & w^\intercal Aw \end{bmatrix}.$$

By Lemma 10, $PAP^\intercal$ has at most one positive eigenvalue. On the other hand, the diagonal entry $v^\intercal Av$ is positive, so Theorem 11 implies that $PAP^\intercal$ has a positive eigenvalue, meaning that it must have exactly one. It follows that

$$\det(PAP^\intercal) = v^\intercal Av \cdot w^\intercal Aw - w^\intercal Av \cdot v^\intercal Aw \leq 0. \quad (3)$$

Thus $w^\intercal(v^\intercal Av \cdot A - (Av)(Av)^\intercal)w \leq 0$ for all $w \in \mathbb{R}^n$. ■

D. Entropy and External Fields

For a probability distribution μ supported on a finite set Ω, its *entropy*, $\mathcal{H}(\mu)$, is defined to be $\sum_{\omega\in\Omega} \mu(\omega)\log\frac{1}{\mu(\omega)}$. For a number $p \in [0,1]$, we also use the shorthand $\mathcal{H}(p) = p\log\frac{1}{p} + (1-p)\log\frac{1}{1-p}$ to denote the entropy of the Bernoulli distribution with parameter p. See Cover and Thomas [14] for background on entropy and its properties. A basic fact we will use about entropy is subadditivity.

Fact 14. *If X and Y are finitely supported random variables and the joint distribution of (X,Y) is μ, and μ_X and μ_Y denote the marginal distributions of X and Y, then*

$$\mathcal{H}(\mu) \leq \mathcal{H}(\mu_X) + \mathcal{H}(\mu_Y),$$

with equality if and only if X and Y are independent.

Let $\mu : 2^{[n]} \to \mathbb{R}_{\geq 0}$ be a nonnegative function; we say μ is a *probability distribution* if $\sum_{S\subseteq[n]} \mu(S) = 1$. The *support* of μ, denoted $\mathrm{supp}(\mu)$, is the collection of $S \subseteq [n]$ for which $\mu(S) \neq 0$, and the *Newton polytope* $\mathcal{P}_\mu \subset \mathbb{R}^n$ of μ is the convex hull of the support, i.e., $\mathcal{P}_\mu = \mathrm{conv}\{\mathbb{1}_S \mid S \in \mathrm{supp}(\mu)\}$. The entropy of μ equals $\mathcal{H}(\mu) = \sum_{S\in\mathrm{supp}(\mu)} \mu(S)\log\frac{1}{\mu(S)}$. To use entropy for approximate counting, we use the following fact:

Proposition 15. *If $u : 2^{[n]} \to \mathbb{R}_{\geq 0}$ is the uniform distribution over $S \in \mathrm{supp}(u)$, then $\mathcal{H}(u)$ equals the* log *of the number of elements in the support of u and this is an upper bound for the entropy of any distribution μ with $\mathrm{supp}(\mu) \subseteq \mathrm{supp}(u)$. That is, for any distribution $\mu : 2^{[n]} \to \mathbb{R}_{\geq 0}$, $\mathcal{H}(\mu) \leq \log(|\mathrm{supp}(\mu)|)$, with equality when μ is the uniform distribution over its support.*

We define the generating polynomial of μ to be the multiaffine polynomial $g_\mu(z_1,\ldots,z_n) = \sum_{S\subseteq[n]} \mu(S)\prod_{i\in S} z_i$. The nonnegative function or probability distribution μ is said to be *log-concave* if its generating polynomial g_μ is log-concave on the positive orthant. The marginal of an element i, μ_i, is the probability that i is in a random sample from μ,

$$\mu_i = \mathbb{P}_{S\sim\mu}[i \in S] = \partial_{z_i} g_\mu(z_1,\ldots,z_n)\big|_{z_1=\cdots=z_n=1}.$$

For a collection of positive numbers $\boldsymbol{\lambda} = (\lambda_1,\ldots,\lambda_n)$, the $\boldsymbol{\lambda}$-external field applied to μ is a probability distribution $\boldsymbol{\lambda} \star \mu : 2^{[n]} \to \mathbb{R}_{\geq 0}$ where for every S,

$$\mathbb{P}_{\boldsymbol{\lambda}\star\mu}[S] \propto \boldsymbol{\lambda}^S \cdot \mu(S) = (\prod_{i\in S}\lambda_i)\cdot\mu(S). \quad (4)$$

As with matroid weights, we note that $g_{\boldsymbol{\lambda}\star\mu}(z_1,\ldots,z_n) \propto \sum_{S\subseteq[n]} \boldsymbol{\lambda}^S \mu(S)\prod_{i\in S} z_i = g_\mu(\lambda_1 z_1,\ldots,\lambda_n z_n)$.

The following theorem has been rediscovered many times:

Theorem 16 ([6, 34]). *Let $\mu : 2^{[n]} \to \mathbb{R}_{\geq 0}$ be a function. For any point $\boldsymbol{p}$ in the polytope $\mathcal{P}_\mu$ and for any $\epsilon > 0$, there exist weights $\lambda_1,\ldots,\lambda_n \in \mathbb{R}_{>0}$ such that the marginal probabilities of $\boldsymbol{\lambda}\star\mu$ are within ϵ of $\boldsymbol{p}$, i.e., for all $i \in [n]$,*

$$|p_i - \mathbb{P}_{S\sim\boldsymbol{\lambda}\star\mu}[i \in S]| \leq \epsilon.$$

If $\boldsymbol{p}$ is in the relative interior of the polytope $\mathcal{P}_\mu$, it turns out that we can take $\epsilon = 0$ in the above theorem. In either case though, we have the following.

Corollary 17. *Let $\mu : 2^{[n]} \to \mathbb{R}_{\geq 0}$ be a nonnegative function and let $\boldsymbol{p} \in \mathcal{P}_\mu$. There is a probability distribution $\tilde{\mu}$ with marginals $\boldsymbol{p}$ such that $\mathrm{supp}(\tilde{\mu}) \subseteq \mathrm{supp}(\mu)$. Moreover $\tilde{\mu}$ can is a limit of distributions of the form $\boldsymbol{\lambda}\star\mu$ for $\boldsymbol{\lambda} \in \mathbb{R}^n_{>0}$.*

The following follows from Proposition 8 and Theorem 16.

Corollary 18. *Let $\mu : 2^{[n]} \to \mathbb{R}_{\geq 0}$ be log-concave and let $\boldsymbol{p} \in \mathcal{P}_\mu$. There is a log-concave probability distribution $\tilde{\mu}$ with marginals $\boldsymbol{p}$ such that $\mathrm{supp}(\tilde{\mu}) \subseteq \mathrm{supp}(\mu)$. Moreover, $\tilde{\mu}$ is the limit of distributions $\boldsymbol{\lambda}\star\mu$ for $\boldsymbol{\lambda} \in \mathbb{R}^n_{>0}$.*

Entropy has a relationship with geometric programs.

Lemma 19 ([34]). *Let $\mu : 2^{[n]} \to \mathbb{R}_{\geq 0}$ have generating polynomial $g_\mu \in \mathbb{R}[z_1,\ldots,z_n]$. Let $\boldsymbol{p}$ be a point in the Newton polytope of μ. Then*

$$\log\left(\inf_{\boldsymbol{z}\in\mathbb{R}^n_{>0}} \frac{g_\mu(\boldsymbol{z})}{\prod_i z_i^{p_i}}\right) = \sum_S \tilde{\mu}(S)\log\frac{\mu(S)}{\tilde{\mu}(S)},$$

where $\tilde{\mu}$ is the probability distribution given by Corollaries 17 and 18. If μ is the indicator function of a family $\mathcal{B} \subseteq 2^{[n]}$, then the above quantity is the same as the entropy $\mathcal{H}(\tilde{\mu})$.

III. Hodge Theory for Matroids

In this section we review several recent developments on combinatorial Hodge theory by Adiprasito, Huh, and Katz [1] and Huh and Wang [23]. The main result we prove in this section is Theorem 23. Later in Section IV, we use this to prove that the generating polynomial of the bases of any matroid is a log-concave function over the positive orthant.

In this section, we take all matroids to be simple. To describe the algebraic tools used [1, 23], we introduce a little more matroid terminology, namely the theory of flats. A subset $F \subseteq E$ is a flat of $M = (E, \mathcal{I})$ if it is a maximal set with rank equal to $\mathrm{rank}(F)$, i.e., for any $i \notin F$, $\mathrm{rank}(F \cup \{i\}) = \mathrm{rank}(F) + 1$. In particular, $F = E$ is the unique flat of rank equal to $\mathrm{rank}(M)$.

We say that a flat F is *proper* if $F \neq E$. Flats F_1, F_2 are *comparable* if $F_1 \subseteq F_2$ or $F_2 \subseteq F_1$ and they are *incomparable* otherwise. A *flag* of M is a strictly monotonic sequence of nonempty proper flats of M, $F_1 \subsetneq F_2 \subsetneq \cdots \subsetneq F_l$. Note that any flag of M has at most $\mathrm{rank}(M) - 1$ flats.

A. The Chow ring

Here we go through some of the commutative algebra used by Adiprasito, Huh, and Katz [1] and explain a special case of their main theorem. Following the set up of Adiprasito, Huh, and Katz [1], for a matroid M of rank $r+1$ on the ground set E, define the *Chow ring* to be the ring

$$A^*(M) = \mathbb{R}[x_F \mid F \text{ is a nonempty proper flat of } M]$$

whose generators x_F satisfy the relations

$$x_{F_1} x_{F_2} = 0 \text{ for all incomparable } F_1, F_2, \text{ and}$$
$$\sum_{F \ni i} x_F - \sum_{F \ni j} x_F = 0 \text{ for all } i, j \in E.$$

$A^*(M)$ is a graded ring, and we use $A^d(M)$ to denote homogeneous polynomials in $A^*(M)$ of degree d. The top degree part, $A^r(M)$, is a one-dimensional vector space over $\mathbb{R}$, and we write "deg" for the isomorphism $A^r(M) \simeq \mathbb{R}$ determined uniquely by requiring $\deg(x_{F_1} \dots x_{F_r}) = 1$ for any flag $F_1 \subsetneq F_2 \subsetneq \cdots \subsetneq F_r$ of M.

A function f on the set of nonempty proper subsets of E is said to be *strictly submodular* if $f(S) + f(T) > f(S \cap T) + f(S \cup T)$ for any two incomparable subsets S, T of E, where we take $f(\emptyset) = f(E) = 0$. We say f is *submodular* if it satisfies the weak form of the above inequality. We remark that our notion of submodularity differs slightly from the conventional notion, in that we additionally require that f takes a value of 0 on $\emptyset, E$.

Define the open convex cone in $A^1(M)$

$$K(M) = \{\sum_F f(F) x_F \mid f \text{ is strictly submodular}\},$$

where the sum is over all flats of M. We let $\overline{K}(M)$ denote the Euclidean closure of this cone, namely those elements of $A^1(M)$ obtained from submodular functions.

The following is a special case of one of the main theorems of Adiprasito, Huh, and Katz [1].

Theorem 20 ([1, Theorem 8.9]). *Let M be a simple matroid of rank $r+1$. For any $\ell_0, \ell_1, \dots, \ell_{r-2}$ in $K(M)$, consider the symmetric bilinear form $Q_{\ell_1,\dots,\ell_{r-2}} : A^1(M) \times A^1(M) \to \mathbb{R}$ defined by*

$$Q_{\ell_1,\dots,\ell_{r-2}}(v, w) = \deg(v \cdot \ell_1 \cdot \ell_2 \cdot \dots \ell_{r-2} \cdot w).$$

Then, as a quadratic form, $Q_{\ell_1,\dots,\ell_{r-2}}$ is negative definite on the kernel of $Q_{\ell_1,\dots,\ell_{r-2}}(\ell_0, \cdot)$, i.e., on $\{v \in A^1(M) \mid Q_{\ell_1,\dots,\ell_{r-2}}(\ell_0, v) = 0\}$.

The kernel of $Q_{\ell_1,\dots,\ell_{r-2}}(\ell_0, \cdot)$ has codimension one in $A^1(M)$. So the operator $Q_{\ell_1,\dots,\ell_{r-2}}$ has at most one nonnegative eigenvalue. Observe that the above result naturally extends to taking $\ell_0, \ell_1, \dots, \ell_{r-2}$ in the closure $\overline{K}(M)$ at the expense of having the slightly weaker guarantee that the operator $Q_{\ell_1,\dots,\ell_{r-2}}$ will be negative *semidefinite* on the kernel of $Q_{\ell_1,\dots,\ell_{r-2}}(\ell_0, \cdot)$. In this case, $Q_{\ell_1,\dots,\ell_{r-2}}$ has at most one positive eigenvalue.

B. Graded Möbius Algebra

To connect the Chow ring with the basis generating polynomial of a matroid, we introduce another algebra used by Huh and Wang [23]. Here we take M to be a simple matroid of rank r on the ground set $[n]$. For flats F_1, F_2, define $F_1 \vee F_2$ to be the inclusion-minimal flat containing $F_1 \cup F_2$.

Let $B^*(M)$ denote the ring $\mathbb{R}[y_F | F \text{ is a flat of } M]$ whose multiplication rule is given by $y_{F_1} y_{F_2} =$

$$\begin{cases} y_{F_1 \vee F_2} & \text{if } \mathrm{rank}(F_1) + \mathrm{rank}(F_2) = \mathrm{rank}(F_1 \vee F_2), \\ 0 & \text{otherwise}, \end{cases} \tag{5}$$

for all pairs of flats F_1, F_2. These relations imply that for any flat F and any basis I_F of F, y_F equals the product $\prod_{i \in I_F} y_i$, where $y_i = y_{\{i\}}$. So $y_1, \dots, y_n$ generate $B^*(M)$. Then $B^*(M)$ is a graded algebra in terms of the degree in $y_1, \dots, y_n$ and we use $B^d(M)$ to denote the homogeneous polynomials of degree d in $B^*(M)$.

Huh and Wang [23] relate this to the Chow ring as follows. Let M_0 denote the matroid of rank $r+1$ on ground set $\{0, 1, \dots, n\}$ obtained by adding 0 as a coloop. Its independent sets have the form I or $\{0\} \cup I$, where I is independent in M. In the Chow ring of M_0, for each $i = 1, \dots, n$, define the degree one element $\beta_i = \sum_{F : i \in F, 0 \notin F} x_F \in A^1(M_0)$, where the sum is taken over flats F of M_0 for which $i \in F$ and $0 \notin F$. Since the indicator function of the condition $i \in F$ and $0 \notin F$ is submodular, β_i belongs to $\overline{K}(M_0)$. Huh and Wang use this to establish the following relationship between $A^*(M_0)$ and $B^*(M)$.

Theorem 21 ([23, Prop 9]). *There is a unique injective graded $\mathbb{R}$-algebra homomorphism*

$$\varphi : B^*(M) \mapsto A^*(M_0) \qquad \text{with} \qquad \varphi(y_i) = \beta_i.$$

Note that for any basis B of M, $\prod_{i\in B} y_i = y_{[n]}$ is nonzero in $B^r(M)$. On the other hand, if $S \subseteq [n]$ is a dependent set of the matroid M, then $\prod_{i\in S} y_i$ is zero. From the existence and injectivity of this map, it follows that for any $B \subseteq [n]$ with $|B| = r$, up to global scaling by a positive real number,

$$\deg(\prod_{i\in B} \beta_i) = \begin{cases} 1 & \text{if } B \text{ is a basis of } M, \\ 0 & \text{otherwise.} \end{cases}$$

$\deg : A^r(M_0) \to \mathbb{R}$ is the isomorphism from Section III-A.

This is particularly useful for us because of the following connection with differential operators on the basis generating polynomial $g_M(z_1, \ldots, z_n)$.

Proposition 22. *For a matrix $\boldsymbol{V} \in \mathbb{R}^{n\times r}$ with columns $\boldsymbol{v}_1 \ldots \boldsymbol{v}_r \in \mathbb{R}^n$,*

$$\deg(\prod_{j=1}^{r}\sum_{i=1}^{n} V_{ij}\beta_i) = \partial_{\boldsymbol{v}_1}\cdots\partial_{\boldsymbol{v}_r} g_M(\boldsymbol{z}).$$

Furthermore, for $0 \le k \le r$ and $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_n) \in \mathbb{R}^n$,

$$\begin{aligned}\deg((\sum_{i=1}^{n}\lambda_i\beta_i)^{r-k} \cdot \prod_{j=1}^{k}\sum_{i=1}^{n} V_{ij}\beta_i) \\ = (r-k)! \cdot \partial_{\boldsymbol{v}_1}\cdots\partial_{\boldsymbol{v}_k} g_M(\boldsymbol{z})|_{\boldsymbol{z}=\boldsymbol{\lambda}}.\end{aligned}$$

Proof: For the first claim, both sides are linear in each $\boldsymbol{v}_i$, so it suffices to prove the case when the columns are standard basis vectors $[\boldsymbol{v}_1 \mid \ldots \mid \boldsymbol{v}_r] = [\mathbb{1}_{i_1} \mid \ldots \mid \mathbb{1}_{i_r}]$.

$$\begin{aligned}\deg(\beta_{i_1}\cdots\beta_{i_r}) &= \partial_{i_1}\cdots\partial_{i_r} g_M(\boldsymbol{z}) \\ &= \begin{cases} 1 & \text{if } \{i_1, \ldots, i_r\} \text{ is a basis of } M, \\ 0 & \text{otherwise.} \end{cases}\end{aligned}$$

The general case then follows from linearity in each column.

For the second claim, again we can consider $[\boldsymbol{v}_1|\ldots|\boldsymbol{v}_k] = [\mathbb{1}_{i_1} \mid \ldots \mid \mathbb{1}_{i_k}]$. Let $I = \{i_1, \ldots, i_k\}$. If $\operatorname{rank}(I) < k$, then $\prod_{i\in I}\beta_i$ is zero in $A^*(M_0)$ and similarly $(\prod_{i\in I}\partial_i)g_M(\boldsymbol{z})$ is zero. Otherwise, we find that

$$\begin{aligned}\deg((\sum_{i=1}^{n}\lambda_i\beta_i)^{r-k}\prod_{i\in I}\beta_i) &= (r-k)! \cdot \sum_{B\supseteq I}\lambda^{B\setminus I} \\ &= (r-k)! \cdot (\prod_{i\in I}\partial_i) g_M(\boldsymbol{z})|_{\boldsymbol{z}=\boldsymbol{\lambda}}.\end{aligned}$$

■

We can translate Theorem 20 to a statement about $g_M(\boldsymbol{z})$.

Theorem 23. *Let M be a simple matroid of rank r on the ground set $[n]$. For any $0 \le k \le r-2$, matrix of nonnegative real numbers $\boldsymbol{V} \in \mathbb{R}_{\ge 0}^{n\times k}$, and any $\boldsymbol{\lambda} \in \mathbb{R}^n_{\ge 0}$, the symmetric bilinear form $q_{\boldsymbol{V},\boldsymbol{\lambda}} : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ given by*

$$q_{\boldsymbol{V},\boldsymbol{\lambda}}(\boldsymbol{a}, \boldsymbol{b}) = \partial_{\boldsymbol{a}}\partial_{\boldsymbol{b}} D_{\boldsymbol{V}} g_M(\boldsymbol{z})|_{\boldsymbol{z}=\boldsymbol{\lambda}}$$

is negative semidefinite on the kernel of $q_{\boldsymbol{V},\boldsymbol{\lambda}}(\boldsymbol{\lambda}, \cdot)$. In particular, the Hessian of $D_{\boldsymbol{V}} g_M(\boldsymbol{z})$ evaluated at $\boldsymbol{z} = \boldsymbol{\lambda}$ has at most one positive eigenvalue.

Proof: For $1 \le j \le k$, define $\ell_j = \sum_{i=1}^{n} V_{ij}\beta_i$ and for $k < j \le r-2$, define $\ell_j = \sum_{i=1}^{n}\lambda_i\beta_i$. For each i, β_i belongs to $\overline{K}(M_0)$, so by the nonnegativity of V_{ij} and λ_i, so does each ℓ_j. By Proposition 22, $q_{\boldsymbol{V},\boldsymbol{\lambda}}$ equals the restriction of $Q_{\ell_1,\ldots,\ell_{r-2}}$ to the subspace of $A^1(M_0)$ spanned by $\{\beta_1, \ldots, \beta_n\}$. That is, for all $\boldsymbol{a}, \boldsymbol{b} \in \mathbb{R}^n$,

$$q_{\boldsymbol{V},\boldsymbol{\lambda}}(\boldsymbol{a}, \boldsymbol{b}) = \frac{1}{(r-k-2)!} Q_{\ell_1,\ldots,\ell_{r-2}}(\sum_i a_i\beta_i, \sum_i b_i\beta_i).$$

Let $\ell_0 = \sum_i \lambda_i\beta_i$. By Theorem 20, $Q_{\ell_1,\ldots,\ell_{r-2}}$ is negative semidefinite on the kernel of $Q_{\ell_1,\ldots,\ell_{r-2}}(\ell_0, \cdot)$, implying that $q_{\boldsymbol{V},\boldsymbol{\lambda}}$ is negative semidefinite on the kernel of $q_{\boldsymbol{V},\boldsymbol{\lambda}}(\boldsymbol{\lambda}, \cdot)$.

Finally, note that $\nabla^2 D_{\boldsymbol{V}} g_M(\boldsymbol{z})$ evaluated at $\boldsymbol{z} = \boldsymbol{\lambda}$ is the $n \times n$ matrix representing $q_{\boldsymbol{V},\boldsymbol{\lambda}}$ in the standard basis. Since it is negative semidefinite on a linear subspace of dimension $n-1$, namely the kernel of $q_{\boldsymbol{V},\boldsymbol{\lambda}}(\boldsymbol{\lambda}, \cdot)$, Corollary 12 implies that it has at most one positive eigenvalue. ■

In the next section we use the above statement to show that generating polynomials of matroids are log-concave and remain log-concave under directional derivatives along directions in the positive orthant.

IV. Completely Log-Concave Polynomials

We call a polynomial $g \in \mathbb{R}[z_1, \ldots, z_n]$ *completely log-concave* if for every $k \ge 0$ and nonnegative matrix $\boldsymbol{V} \in \mathbb{R}^{n\times k}_{\ge 0}$, $D_{\boldsymbol{V}} g(\boldsymbol{z})$ is nonnegative and log-concave as a function over $\mathbb{R}^n_{>0}$, where $D_{\boldsymbol{V}} g(\boldsymbol{z}) = (\prod_{j=1}^{k}\sum_{i=1}^{n} V_{ij}\partial_i) g(\boldsymbol{z})$.

Note that for $k = 0$, this condition implies log-concavity of g itself. We call a distribution $\mu : 2^{[n]} \to \mathbb{R}_{\ge 0}$ completely log-concave if and only if g_μ is completely log-concave.

Remark 24. Related notions of "strongly log-concave" and "Alexandrov-Fenchel" polynomials were studied in the work of Gurvits [18, 20] to design approximation algorithms for mixed volume of polytopes and to show Newton-like inequalities for coefficients of these polynomials. Gurvits [20] also states that because a positive combination of convex polytopes is a convex polytope, the stronger property we call complete log-concavity is satisfied for the volume polynomial. Unlike strong log-concavity, complete log-concavity is readily seen to be preserved under many useful operations.

Note that complete log-concavity implies nonnegativity of the coefficients of g. This is because the coefficient of $\prod_i z_i^{\kappa_i}$ in g is a positive multiple of $\partial_1^{\kappa_1}\cdots\partial_n^{\kappa_n} g(\boldsymbol{z})|_{\boldsymbol{z}=0}$.

To verify complete log-concavity for a polynomial with nonnegative coefficients, we only have to check log-concavity of order k derivatives for $k \le r-2$. For $k \ge r$, $D_{\boldsymbol{V}} g(\boldsymbol{z})$ is a nonnegative constant and for $k = r-1$, it is a linear function with nonnegative coefficients in $z_1, \ldots, z_n$.

The main result of this section is that the basis generating polynomial of any matroid is completely log-concave.

Theorem 25. *For any matroid M, $g_M(\boldsymbol{z})$ is completely log-concave over the positive orthant.*

First, we show that the statement holds when M is a simple matroid. To do this, we use a corollary of Euler's identity, which states that if a polynomial $g(\boldsymbol{z})$ is homogeneous of degree d then

$$\langle \nabla g, \boldsymbol{z}\rangle = \sum_{i=1}^{n} z_i \partial_i g = d \cdot g(\boldsymbol{z}). \tag{6}$$

Corollary 26 (Euler's identity)**.** *If $g \in \mathbb{R}[z_1, \ldots, z_n]$ is homogeneous of degree d, then*

$$\nabla^2 g \cdot \boldsymbol{z} = (d-1) \cdot \nabla g \qquad \text{and} \qquad \boldsymbol{z}^\intercal \cdot \nabla^2 g \cdot \boldsymbol{z} = d(d-1) \cdot g.$$

Proof: The i-th entry of the vector $\nabla^2 g \cdot \boldsymbol{z}$ equals $\sum_{j=1}^{n} z_j \cdot \partial_i \partial_j g$. Since $\partial_i g$ is homogeneous of degree $d-1$, it follows by Euler's identity, Eq. (6), that this equals $(d-1)\partial_i g$. Multiplying by $\boldsymbol{z}^\intercal$ and using Eq. (6) again gives the second claim. ■

Lemma 27. *If M is a simple matroid, then $g_M(\boldsymbol{z})$ is completely log-concave.*

Proof: Let M be a simple matroid of rank r on the set $[n]$. Let $0 \leq k \leq r-2$ and take a nonnegative matrix $\boldsymbol{V} \in \mathbb{R}_{\geq 0}^{n \times k}$. We will show that for any $\boldsymbol{\lambda} \in \mathbb{R}_{>0}^n$, $\nabla^2 \log(D_{\boldsymbol{V}} g_M(\boldsymbol{z}))$ is negative semidefinite at the point $\boldsymbol{z} = \boldsymbol{\lambda}$. Note that for $h \in \mathbb{R}[z_1, \ldots, z_n]$, we have $h^2 \cdot \nabla^2 \log(h) = \left[h \cdot \partial_i \partial_j h - \partial_i h \cdot \partial_j h\right]_{1 \leq i,j \leq n} = h \cdot \nabla^2 h - (\nabla h)(\nabla h)^\intercal$.

Now let $h = D_{\boldsymbol{V}} g_M(\boldsymbol{z})$ and consider the quadratic form $q_{\boldsymbol{V},\boldsymbol{\lambda}}(\boldsymbol{a}, \boldsymbol{b}) = \partial_{\boldsymbol{a}} \partial_{\boldsymbol{b}} h(\boldsymbol{\lambda})$ as in Theorem 23. This is represented by the Hessian matrix of h at $\boldsymbol{z} = \boldsymbol{\lambda}$: $\nabla^2 h\big|_{\boldsymbol{z}=\boldsymbol{\lambda}} = \left[\partial_i \partial_j h(\boldsymbol{\lambda})\right]_{1 \leq i,j \leq n}$.

By Theorem 23 this matrix has at most one positive eigenvalue. Since it also has nonnegative entries, we can apply Lemma 13 with $A = \nabla^2 h\big|_{\boldsymbol{z}=\boldsymbol{\lambda}}$ and $v = \boldsymbol{\lambda}$. Since $h(\boldsymbol{z})$ is homogeneous of degree $r-k$, Corollary 26 implies that

$$v^\intercal A v = (r-k)(r-k-1) h(\boldsymbol{\lambda}), \text{ and}$$
$$(Av)(Av)^\intercal = (r-k-1)^2 \cdot (\nabla h(\boldsymbol{\lambda}))(\nabla h(\boldsymbol{\lambda}))^\intercal.$$

Then Lemma 13 states that the matrix $(v^\intercal A v) \cdot A - t(Av)(Av)^\intercal =$

$$(r-k)(r-k-1)(h \nabla^2 h - \frac{t(r-k-1)}{r-k} (\nabla h)(\nabla h)^\intercal)|_{\boldsymbol{z}=\boldsymbol{\lambda}}$$

is negative semidefinite for all $t \geq 1$. Taking $t = \frac{r-k}{r-k-1}$ then shows that $h(\boldsymbol{\lambda})^2 \cdot \nabla^2 \log(h)\big|_{\boldsymbol{z}=\boldsymbol{\lambda}}$ is negative semidefinite. Thus $h = D_{\boldsymbol{V}} g_M$ is log-concave on $\mathbb{R}_{>0}^n$. ■

Next we show that similar to Lemma 7, affine transformations preserve complete log-concavity.

Lemma 28. *Let $T : \mathbb{R}^m \to \mathbb{R}^n$ be an affine transformation such that $T(\mathbb{R}_{\geq 0}^m) \subseteq \mathbb{R}_{\geq 0}^n$, and let $g \in \mathbb{R}[z_1, \ldots, z_n]$ be a completely log-concave polynomial. Then $g(T(y_1, \ldots, y_m)) \in \mathbb{R}[y_1, \ldots, y_m]$ is completely log-concave.*

Proof: We must have $T(\boldsymbol{y}) = \boldsymbol{A}\boldsymbol{y} + \boldsymbol{b}$ where $\boldsymbol{A} \in \mathbb{R}_{\geq 0}^{n \times m}$ and $\boldsymbol{b} \in \mathbb{R}_{\geq 0}^n$. It follows that $g(T(\boldsymbol{y}))$ has nonnegative coefficients. Therefore for any nonnegative matrix $\boldsymbol{V}$, $D_{\boldsymbol{V}} g(T(\boldsymbol{y}))$ has nonnegative coefficients and is nonnegative over $\mathbb{R}_{\geq 0}^m$, so we just need to check that it is log-concave.

The Jacobian of T at every point is given by $\boldsymbol{A}$. Then the chain rule yields for any $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k \in \mathbb{R}^m$

$$\partial_{\boldsymbol{v}_1} \cdots \partial_{\boldsymbol{v}_k} g(T(\boldsymbol{y})) = \left(\partial_{\boldsymbol{A}\boldsymbol{v}_1} \cdots \partial_{\boldsymbol{A}\boldsymbol{v}_k} g(\boldsymbol{z})\right)\big|_{\boldsymbol{z}=T(\boldsymbol{y})}.$$

So for any $k \geq 0$ and nonnegative matrix of directions $\boldsymbol{V} \in \mathbb{R}_{\geq 0}^{m \times k}$, we have

$$D_{\boldsymbol{V}} g(T(\boldsymbol{y})) = \left(D_{\boldsymbol{A}\boldsymbol{V}} g(\boldsymbol{z})\right)\big|_{\boldsymbol{z}=T(\boldsymbol{y})}.$$

Since $\boldsymbol{A}, \boldsymbol{V}$ have nonnegative entries, so does $\boldsymbol{A}\boldsymbol{V}$. From complete log-concavity of g it follows that $D_{\boldsymbol{A}\boldsymbol{V}} g(\boldsymbol{z})$ is log-concave over $\mathbb{R}_{>0}^n$. Now Lemma 7 implies that the composition with T remains log-concave. ■

We have the following corollary.

Lemma 29. *The following operations on polynomials preserve complete log-concavity:*

1. *Permutation: $g \mapsto g(z_{\pi(1)}, \ldots, z_{\pi(n)})$ for $\pi \in S_n$.*
2. *Specialization: $g \mapsto g(a, z_2, \ldots, z_n)$, where $a \in \mathbb{R}_{\geq 0}$.*
3. *Scaling $g \mapsto c \cdot f(\lambda_1 z_1, \ldots, \lambda_n z_n)$, where $c, \lambda_1, \ldots, \lambda_n \in \mathbb{R}_{\geq 0}$.*
4. *Expansion: $g(z_1, \ldots, z_n) \mapsto g(y_1 + y_2 + \cdots + y_m, z_2, \ldots, z_n) \in \mathbb{R}[y_1, \ldots, y_m, z_2, \ldots, z_n]$.*
5. *Differentiation: $g \mapsto \partial_{\boldsymbol{v}} g = \sum_{i=1}^n v_i \partial_i g$ for $\boldsymbol{v} \in \mathbb{R}_{\geq 0}^n$.*

Now, we are ready to prove Theorem 25.

Proof of Theorem 25: Let M be a matroid of rank r on ground set $[n]$. If M is simple, then the result follows from Lemma 27.

Otherwise let $\tilde{M} = (\tilde{E}, \tilde{\mathcal{I}})$ be the simple matroid obtained by deleting loops and identifying each set of parallel elements of M. Say each non-loop $i \in [n]$ gets mapped to the element $\psi(i) \in \tilde{E}$. Consider the generating polynomial $g_{\tilde{M}}(\boldsymbol{z}) \in \mathbb{R}[z_e \mid e \in \tilde{E}]$. Each basis of M uses at most one of a set of parallel elements, meaning that the basis generating polynomial of M is obtained from that of $\tilde{M}$ by substituting $z_e \mapsto \sum_{i \in \psi^{-1}(e)} y_i$. That is, if $T : \mathbb{R}^n \to \mathbb{R}^{\tilde{E}}$ is the linear map defined by

$$T(\mathbb{1}_i) = \begin{cases} 0 & \text{if } i \text{ is a loop,} \\ \mathbb{1}_{\psi(i)} & \text{otherwise,} \end{cases}$$

then $g_M(y_1, \ldots, y_n) = g_{\tilde{M}}(T(y_1, \ldots, y_n))$. By Lemma 28, it follows that g_M is completely log-concave. ■

V. Entropy of Log-Concave Distributions

Let $\mu : 2^{[n]} \to \mathbb{R}_{\geq 0}$ be a probability distribution on the subsets of the set $[n]$. In other words, $\forall S \subseteq [n]$: $\mu(S) \geq 0$ and, $\sum_{S \subseteq [n]} \mu(S) = 1$. As in Section II-D, we consider the multiaffine *generating polynomial* of μ, $g_\mu(\boldsymbol{z}) = \sum_{S \subseteq [n]} \mu(S) \cdot \prod_{i \in S} z_i$.

We say μ is *d-homogeneous* if g_μ is a degree d homogeneous polynomial, i.e., $g_\mu(\alpha \boldsymbol{z}) = \alpha^d g_\mu(\boldsymbol{z})$ for any $\alpha \in \mathbb{R}$ and that μ is *log-concave* and *completely log-concave* if the generating polynomial $g_\mu(\boldsymbol{z})$ is log-concave and completely log-concave, respectively. In this section we prove a bound on the entropy of log-concave probability distributions.

Recall that the marginal probability of an element i, μ_i is the probability that i is in a random sample from μ,

$$\mu_i = \mathbb{P}_{S\sim\mu}[i \in S] = \partial_{z_i} g_\mu(z_1, \dots, z_n)\big|_{z_1=\dots=z_n=1}.$$

Given marginal probabilities $\mu_1, \dots, \mu_n$, it is easy to derive an upper bound on the entropy of μ by using the subadditivity of entropy, Fact 14.

Fact 30. *For any probability distribution $\mu : 2^{[n]} \to \mathbb{R}_{\geq 0}$ with marginals $\mu_1, \dots, \mu_n$, we have $\mathcal{H}(\mu) \leq \sum_{i=1}^n \mathcal{H}(\mu_i)$.*

This inequality is tight if components of μ are independent, i.e., if for all sets $S \subseteq [n]$, $\mu(S) = \prod_{i\in S} \mu_i \prod_{i\notin S}(1 - \mu_i)$. The main result of this section is a lower bound on the entropy of log-concave distributions, which will imply that the inequality in Fact 30 is tight within a factor of 2 under certain further restrictions.

Theorem 31. *For any log-concave probability distribution $\mu : 2^{[n]} \to \mathbb{R}_{\geq 0}$ with marginal probabilities $\mu_1, \dots, \mu_n \geq 0$, we have*

$$\mathcal{H}(\mu) \geq \sum_{i=1}^n \mu_i \log \frac{1}{\mu_i}.$$

To prove, Theorem 31, we use Jensen's inequality in order to exploit log-concavity.

Lemma 32 (Jensen's Inequality)**.** *If $f : \mathbb{R}^n_{\geq 0} \to \mathbb{R} \cup \{-\infty\}$ is a concave function, and X is a $(\mathbb{R}^n_{\geq 0})$-valued random variable with finite support, then $f(\mathbb{E}[X]) \geq \mathbb{E}[f(X)]$.*

Proof of Theorem 31: In order to apply Lemma 32, we have to specify the concave function f and the random variable X. We let X be $\mathbb{1}_S$, where S is chosen randomly according to the distribution μ. In other words, for every S, we let $\mathbb{P}[X = \mathbb{1}_S] = \mu(S)$. For the function f, we will use

$$f(z_1, \dots, z_n) = \log g_\mu\left(\frac{z_1}{\mu_1}, \dots, \frac{z_n}{\mu_n}\right).$$

Note that even though μ_i could be zero for some i, the above expression is still well-defined, since if $\mu_i = 0$, then g_μ does not depend on z_i. By Proposition 8, Part 3, the function f is concave over the positive orthant.

First, note that $\mathbb{E}[X] = \mathbb{E}_{S\sim\mu}[\mathbb{1}_S] = (\mu_1, \dots, \mu_n)$, so the left hand side of Lemma 32 is

$$f(\mathbb{E}[X]) = \log g_\mu\left(\frac{\mu_1}{\mu_1}, \dots, \frac{\mu_n}{\mu_n}\right) = 0. \tag{7}$$

For the right hand side, note that for any $S \in \text{supp}(\mu)$, by the definition of f and g_μ,

$$\begin{aligned} f(\mathbb{1}_S) = \log\Big(\sum_{T\subseteq S} \mu(T) \prod_{i\in T} \frac{1}{\mu_i}\Big) &\geq \log\Big(\mu(S) \prod_{i\in S} \frac{1}{\mu_i}\Big) \\ &= \log \mu(S) + \sum_{i\in S} \log \frac{1}{\mu_i}, \end{aligned}$$

where the inequality follows from monotonicity of log.

Now $\mathbb{E}[f(X)] = \sum_S \mu(S) f(\mathbb{1}_S) \geq \sum_S \mu(S) \log \mu(S) + \sum_S \mu(S) \sum_{i\in S} \log \frac{1}{\mu_i} = -\mathcal{H}(\mu) + \sum_{i=1}^n (\sum_{S\ni i} \mu(S)) \cdot \log \frac{1}{\mu_i} = -\mathcal{H}(\mu) + \sum_{i=1}^n \mu_i \log \frac{1}{\mu_i}$. By Lemma 32 and Eq. (7), the above quantity is ≤ 0. Rearranging yields the desired inequality. ■

Next, we discuss several corollaries of Theorem 31.

Corollary 33. *If μ is r-homogeneous and log-concave, then $\sum_{i=1}^n \mathcal{H}(\mu_i)$ gives an additive r-approximation of $\mathcal{H}(\mu)$, i.e.,*

$$\sum_{i=1}^n \mathcal{H}(\mu_i) - r \leq \mathcal{H}(\mu) \leq \sum_{i=1}^n \mathcal{H}(\mu_i).$$

Proof: The second inequality is simply Fact 30. For the first inequality, we use the fact that $(1-p)\log\frac{1}{1-p} \leq p$ for all $0 \leq p \leq 1$, which means that

$$\sum_{i=1}^n (1 - \mu_i) \log \frac{1}{1 - \mu_i} \leq \sum_{i=1}^n \mu_i = \mathbb{E}_{S\sim\mu}[|S|], \tag{8}$$

which for an r-homogeneous distribution equals r. Now Theorem 31 implies $\mathcal{H}(\mu) \geq \sum_{i=1}^n \mu_i \log \frac{1}{\mu_i} = \sum_{i=1}^n \mathcal{H}(\mu_i) - \sum_{i=1}^n (1-\mu_i) \log \frac{1}{1-\mu_i} \geq \sum_{i=1}^n \mathcal{H}(\mu_i) - r$, as desired. ■

For a probability distribution $\mu : 2^{[n]} \to \mathbb{R}_{\geq 0}$, define its *dual*, $\mu^* : 2^{[n]} \to \mathbb{R}_{\geq 0}$, to be the distribution for which the probability of a set is equal to the probability of its complement under μ, i.e. $\mu^*(S) = \mu([n] \setminus S)$ for all $S \subseteq [n]$. Then for $1 \leq i \leq n$, the i-th marginal of μ^* is $\mu^*_i = 1 - \mu_i$.

Corollary 34. *If μ, μ^* are both log-concave probability distributions then $\sum_{i=1}^n \mathcal{H}(\mu_i)$ gives a multiplicative 2-approximation to $\mathcal{H}(\mu)$, i.e.,*

$$\frac{1}{2} \sum_{i=1}^n \mathcal{H}(\mu_i) \leq \mathcal{H}(\mu) \leq \sum_{i=1}^n \mathcal{H}(\mu_i).$$

Proof: Applying Theorem 31 to μ and μ^* gives $\mathcal{H}(\mu) \geq \sum_{i=1}^n \mu_i \log \frac{1}{\mu_i}$, and $\mathcal{H}(\mu^*) \geq \sum_{i=1}^n (1-\mu_i) \log \frac{1}{1-\mu_i}$. Since $\mathcal{H}(\mu) = \mathcal{H}(\mu^*)$, averaging these inequalities gives

$$\mathcal{H}(\mu) \geq \frac{1}{2} \sum_{i=1}^n \left(\mu_i \log \frac{1}{\mu_i} + (1 - \mu_i) \log \frac{1}{1 - \mu_i}\right),$$

as desired. The other inequality follows from Fact 30. ■

Let μ be the uniform distribution over the bases of a matroid M. It follows from Theorem 25 that μ is a log-concave distribution. Furthermore, the dual probability distribution μ^* is the uniform distribution over the bases of the dual matroid M^*, meaning that it is also log-concave. Then Corollary 33 and Corollary 34 immediately yield:

Corollary 35. *Let M be an arbitrary matroid of rank r on ground set $[n]$ and let μ be the uniform distribution over its bases. Then $\sum_{i=1}^n \mathcal{H}(\mu_i)$ is both an additive r-approximation and multiplicative 2-approximation to $\mathcal{H}(\mu) = \log(|\mathcal{B}_M|)$:*

$$\max\left\{\frac{1}{2}\sum_{i=1}^n \mathcal{H}(\mu_i), \sum_{i=1}^n \mathcal{H}(\mu_i) - r\right\} \leq \mathcal{H}(\mu) \leq \sum_{i=1}^n \mathcal{H}(\mu_i).$$

We will also use the following fact, which enables us to apply Theorem 31 to distributions other than the uniform distribution on $\mathcal{B}_M$.

Lemma 36. *Let M be a matroid on ground set $[n]$ and let $\boldsymbol{p}$ be a point in $\mathcal{P}_M$. Then there is a distribution $\tilde{\mu}$ supported on $\mathcal{B}_M$ with marginals $\boldsymbol{p}$, i.e., $\tilde{\mu}_i = p_i$, such that both $\tilde{\mu}$ and $\tilde{\mu}^*$ are completely log-concave. Furthermore $\tilde{\mu}$ and $\tilde{\mu}^*$ can be obtained as the limit of external fields applied to μ and μ^*, where μ is the uniform distribution on $\mathcal{B}_M$.*

Proof: If μ is the uniform distribution over $\mathcal{B}_M$, then $g_\mu(\boldsymbol{z}) = g_M(\boldsymbol{z})$, which is completely log-concave by Theorem 25. Similarly, μ^* is the uniform distribution on the bases of the dual matroid, so μ^* is also completely log-concave. Furthermore, since μ and μ^* are homogeneous distributions, for any $\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_n) \in \mathbb{R}^n_{>0}$, $((\lambda_1, \dots, \lambda_n) \star \mu)^* = (\lambda_1^{-1}, \dots, \lambda_n^{-1}) \star \mu^*$, where $\star$ is the external field operation described in Eq. (4) of Section II-D.

By Lemma 29, Part 3, both $\boldsymbol{\lambda} \star \mu$ and $(\boldsymbol{\lambda} \star \mu)^*$ are completely log-concave. Now take $\tilde{\mu}$ to be the distribution promised by Corollary 18 with marginals $\boldsymbol{p}$. Then $\tilde{\mu}$ is a limit of distributions $\boldsymbol{\lambda} \star \mu$, and $\tilde{\mu}^*$ is the limit of $(\boldsymbol{\lambda} \star \mu)^*$. It follows that both $\tilde{\mu}$ and $\tilde{\mu}^*$ are completely log-concave. ∎

VI. Max Entropy Convex Programs and Counting Bases of a Matroid

In this section we prove Theorem 1. Let M be a matroid of rank r on ground set $[n]$. Let $\mu : 2^{[n]} \to \mathbb{R}_{\geq 0}$ be the uniform distribution over the bases of M. By Corollary 35, it would be enough to compute the marginals of μ, but it can be seen that computing marginals is no easier than counting bases.

Instead, we use the convex programming framework described in Section I-D. We claim that the optimum solution of the following concave program gives an additive r-approximation to $\mathcal{H}(\mu) = \log(|\mathcal{B}_M|)$ as well as a multiplicative 2-approximation:

$$\tau = \max\left\{\sum_{i=1}^n \mathcal{H}(p_i) \;\middle|\; \boldsymbol{p} = (p_1, \dots, p_n) \in \mathcal{P}_M\right\}. \tag{9}$$

The objective function is a concave function of $\boldsymbol{p}$, so we can solve the above program using, e.g., the ellipsoid method.

Proof of Theorem 1: Let $\boldsymbol{p} = (p_1, \dots, p_n) \in \mathcal{P}_M$ be a vector achieving the maximum in Eq. (9). The output of our algorithm will simply be $\beta = e^\tau = \exp(\sum_{i=1}^n \mathcal{H}(p_i))$.

By Proposition 15, the entropy $\mathcal{H}(\mu)$ equals $\log(|\mathcal{B}_M|)$. Therefore to prove Theorem 1, it suffices to show that $\tau = \log(\beta)$ is an additive r-approximation and also a multiplicative 2-approximation of $\mathcal{H}(\mu)$, i.e., $\max\{\frac{1}{2}\tau, \tau - r\} \leq \mathcal{H}(\mu) \leq \tau$.

Firstly, note that since $(\mu_1, \dots, \mu_n) \in \mathcal{P}_M$, we have

$$\tau \geq \sum_{i=1}^n \mathcal{H}(\mu_i) \geq \mathcal{H}(\mu), \tag{10}$$

where the first inequality follows from the definition, Eq. (9), and the second inequality follows from the subadditivity of entropy, Fact 30.

Secondly, since $\boldsymbol{p}$ is in the polytope $\mathcal{P}_M = \mathcal{P}_\mu$, by Lemma 36, there is a probability distribution $\tilde{\mu}$ on the bases of M such that for all i, $\tilde{\mu}_i = p_i$, and both $\tilde{\mu}$ and $\tilde{\mu}^*$ are log-concave. Applying Corollaries 33 and 34 to $\tilde{\mu}$, we get

$$\mathcal{H}(\tilde{\mu}) \geq \max\left\{\frac{1}{2}\sum_{i=1}^n \mathcal{H}(\tilde{\mu}_i), \sum_{i=1}^n \mathcal{H}(\tilde{\mu}_i) - r\right\}.$$

But note that $\sum_{i=1}^n \mathcal{H}(\tilde{\mu}_i) = \sum_{i=1}^n \mathcal{H}(p_i) = \tau$, so $\mathcal{H}(\tilde{\mu}) \geq \max\{\frac{1}{2}\tau, \tau - r\}$.

Since μ is the uniform distribution over its support, and $\mathrm{supp}(\tilde{\mu}) \subseteq \mathrm{supp}(\mu)$, its follows from Proposition 15 that $\mathcal{H}(\mu) \geq \mathcal{H}(\tilde{\mu})$. So we find that $\mathcal{H}(\mu) \geq \max\{\frac{1}{2}\tau, \tau - r\}$, which together with Eq. (10) finishes the proof. ∎

Acknowledgements

We would like to thank Matt Baker, June Huh, and Josephine Yu for their comments on an early draft of this paper. Part of this work was done while the first and third authors were visiting the Simons Institute for the Theory of Computing. It was partially supported by the DIMACS/Simons Collaboration on Bridging Continuous and Discrete Optimization through NSF grant CCF-1740425. Shayan Oveis Gharan is supported by the NSF grant CCF-1552097 and ONR-YIP grant N00014-17-1-2429. Cynthia Vinzant was partially supported by the National Science Foundation grant DMS-1620014.

References

[1] Karim Adiprasito, June Huh, and Eric Katz. "Hodge theory for combinatorial geometries". In: *arXiv preprint arXiv:1511.02888* (2015).

[2] Nima Anari and Shayan Oveis Gharan. "A generalization of permanent inequalities and applications in counting and optimization". In: *STOC*. 2017, pp. 384–396.

[3] Nima Anari, Shayan Oveis Gharan, and Alireza Rezaei. "Monte Carlo Markov Chain Algorithms for Sampling Strongly Rayleigh Distributions and Determinantal Point Processes". In: *COLT*. 2016, pp. 103–115.

[4] Nima Anari et al. “Nash Social Welfare for Indivisible Items under Separable, Piecewise-Linear Concave Utilities”. In: *SODA*. 2018, pp. 2274–2290.

[5] Nima Anari et al. “Nash Social Welfare, Matrix Permanent, and Stable Polynomials”. In: *ITCS*. 2017, 36:1–36:12.

[6] Arash Asadpour et al. “An $O(\log n/ \log\log n)$-approximation Algorithm for the Asymmetric Traveling Salesman Problem”. In: *SODA*. 2010, pp. 379–389.

[7] Y. Azar, A.Z. Broder, and A.M. Frieze. “On the problem of approximating the number of bases of a matriod”. In: *Information Processing Letters* 50.1 (1994), pp. 9–11. ISSN: 0020-0190.

[8] Alexander I Barvinok. “Approximate counting via random optimization”. In: *Random Structures and Algorithms* 11.2 (1997), pp. 187–198.

[9] Alexander Barvinok and Alex Samorodnitsky. “Random weighting, asymptotic counting, and inverse isoperimetry”. In: *Israel Journal of Mathematics* 158.1 (Mar. 2007), pp. 159–191. ISSN: 1565-8511.

[10] Stephen Boyd and Lieven Vandenberghe. *Convex Optimization*. Cambridge University Press, Mar. 2004.

[11] Petter Branden. “Polynomials with the half-plane property and matroid theory”. In: *Advances in Mathematics* 216.1 (2007), pp. 302–320.

[12] Brian D. Cloteaux. “Approximating the Number of Bases for Almost All Matroids”. In: *Congressus Numerantium* 202 (2010), pp. 149–154.

[13] Emma Cohen, Prasad Tetali, and Damir Yeliussizov. “Lattice path matroids: negative correlation and fast mixing”. In: *arXiv preprint arXiv:1505.06710* (2015).

[14] Thomas M Cover and Joy A Thomas. *Elements of information theory*. John Wiley & Sons, 2012.

[15] William H Cunningham. “Testing membership in matroid polyhedra”. In: *Journal of Combinatorial Theory, Series B* 36.2 (1984), pp. 161–188. ISSN: 0095-8956.

[16] Tomás Feder and Milena Mihail. “Balanced matroids”. In: *STOC*. 1992, pp. 26–38.

[17] Anna Gambin. “On approximating the number of bases of exchange preserving matroids”. In: *International Symposium on Mathematical Foundations of Computer Science*. Springer. 1999, pp. 332–342.

[18] Leonid Gurvits. “A polynomial-time algorithm to approximate the mixed volume within a simply exponential factor”. In: *Discrete & Computational Geometry* 41.4 (2009), pp. 533–555.

[19] Leonid Gurvits. “Hyperbolic polynomials approach to Van der Waerden/Schrijver-Valiant like conjectures: sharper bounds, simpler proofs and algorithmic applications”. In: *STOC*. 2006, pp. 417–426.

[20] Leonid Gurvits. “On multivariate Newton-like inequalities”. In: *Advances in Combinatorial Mathematics*. Springer, 2009, pp. 61–78.

[21] Leonid Gurvits and Alex Samorodnitsky. “Bounds on the permanent and some applications”. In: *arXiv preprint arXiv:1408.0976* (2014).

[22] Roger A Horn, Roger A Horn, and Charles R Johnson. *Matrix analysis*. Cambridge university press, 1990.

[23] June Huh and Botong Wang. “Enumeration of points, lines, planes, etc”. In: *arXiv preprint arXiv:1609.05484* (2016).

[24] Hiroshi Imai et al. “Combinatorial and geometric approaches to counting problems on linear matroids, graphic arrangements, and partial orders”. In: *International Computing and Combinatorics Conference*. Springer. 1996, pp. 68–80.

[25] Mark Jerrum. “Two Remarks Concerning Balanced Matroids”. In: *Combinatorica* 26.6 (2006), pp. 733–742.

[26] Mark Jerrum, Alistair Sinclair, and Eric Vigoda. “A Polynomial-time Approximation Algorithm for the Permanent of a Matrix with Nonnegative Entries”. In: *J. ACM* 51.4 (July 2004), pp. 671–697.

[27] Mark Jerrum and Jung Bae Son. “Spectral gap and log-Sobolev constant for balanced matroids”. In: *FOCS*. 2002, pp. 721–729.

[28] Mark Jerrum et al. “Elementary bounds on Poincaré and log-Sobolev constants for decomposable Markov chains”. In: *Annals of Applied Probability* (2004), pp. 1741–1765.

[29] Nathan Linial, Alex Samorodnitsky, and Avi Wigderson. “A deterministic strongly polynomial algorithm for matrix scaling and approximate permanents”. In: *STOC*. 1998, pp. 644–652.

[30] Milena Mihail and Madhu Sudan. *Connectivity Properties of Matroids*. Tech. rep. EECS Department, University of California, Berkeley, Dec. 1991.

[31] Aleksandar Nikolov and Mohit Singh. “Maximizing determinants under partition constraints”. In: *STOC*. 2016, pp. 192–201.

[32] Alexander Schrijver. “Matroid intersection”. In: *Combinatorial Optimization: Polyhedra and Efficiency*. Vol. B. Springer, 2003.

[33] Alistair Sinclair and Mark Jerrum. “Approximate counting, uniform generation and rapidly mixing Markov chains”. In: *Information and Computation* 82.1 (1989), pp. 93–133.

[34] Mohit Singh and Nisheeth K. Vishnoi. “Entropy, optimization and counting”. In: *STOC*. 2014, pp. 50–59.

[35] Michael Snook. “Counting bases of representable matroids”. In: *The Electronic Journal of Combinatorics* 19.4 (2012).

[36] Damian Straszak and Nisheeth K. Vishnoi. “Real stable polynomials and matroids: optimization and counting”. In: *STOC*. 2017, pp. 370–383.