Maximizing Determinants under Partition Constraints

Aleksandar Nikolov [*1] and Mohit Singh [†2]

[1]Department of Computer Science, University of Toronto.
[2]Microsoft Research, Redmond.

December 11, 2015

Abstract

Given a positive semidefinte matrix L whose columns and rows are indexed by a set U, and a partition matroid $\mathcal{M} = (U, \mathcal{I})$, we study the problem of selecting a basis B of $\mathcal{M}$ such that the determinant of submatrix of L, restricted to rows and columns of B, is maximized. This problem appears in diverse areas including determinantal point processes in machine learning [KT12], experimental design, geographical placement problem [Lee98, KLQ95], discrepancy theory and computational geometry [Nik14]. Our main result is to give a *geometric concave program* for the problem which approximates the optimum value within a factor of $e^{r+o(r)}$, where r denotes the rank of the partition matroid $\mathcal{M}$.

We bound the integrality gap of the geometric concave program by giving a polynomial time randomized rounding algorithm. To analyze the rounding algorithm, we relate the solution of our algorithm as well the objective value of the relaxation to a certain stable polynomial. To prove the approximation guarantee, we utilize a general inequality about stable polynomials proved by Gurvits [Gur08] in the context of estimating the permanent of a doubly stochastic matrix.

---

[*]Email:anikolov@cs.toronto.edu
[†]Email:mohits@microsoft.com.

1   Introduction

We consider the problem of picking a submatrix of a positive semi-definite (PSD) matrix of maximum determinant where the rows/columns of the submatrix must satisfy combinatorial constraints. Given a $n \times n$ PSD matrix L of rank d, let U index the columns as well as the rows. We are also given a partition matroid $\mathcal{M} = (U, \mathcal{I})$ defined by the partition $U = P_1 \cup \ldots \cup P_k$ with bounds $b_1, \ldots, b_k$ where a set $S \in \mathcal{I}$ if $|S \cap P_i| \leq b_i$ for each i. The set $\mathcal{B}$ of bases of $\mathcal{M}$ includes those $S \in \mathcal{I}$ s.t. $|B \cap P_i| = b_i$ for each i. The goal is to find a basis $B \in \mathcal{B}$ such that $\det(L_{B,B})$ is maximized where $L_{B,B}$ is the submatrix restricted to columns and rows of B. We refer to this problem as the *maximum determinant problem under partition constraints*. The problem of finding a submatrix of a positive semi-definite matrix of maximum determinant under combinatorial constraints appears in many diverse areas, and we discuss some of these applications below.

In a determinantal point process (DPP)[Lyo03], we are given a probability distribution over subsets of U, where $\Pr[S] \propto \det(L_{S,S})$ for each subset $S \subseteq U$ where L is a PSD matrix. DPPs display various negative dependence properties [Lyo03, BBL09]; informally, subsets containing *dissimilar* elements have higher probability. Thus, they have been successfully used in information retrieval and machine learning as a model where diversity is a natural criterion (we refer the reader to the excellent survey [KT10] and references therein). A natural problem of interest in these applications is to find the most likely configuration (the MAP inference problem) under certain constraints on the set S. Thus the MAP inference problem can be formulated as $\max_{S \in \mathcal{C}} \det(L_{S,S})$ where constraints are defined by set family $\mathcal{C}$. Gillenwater et al [GKT12] consider the MAP inference problem with matroid constraints and knapsack constraints. It is easy to see that the maximum determinant problem under partition constraints can also be recast as a MAP inference problem.

A closely related problem of constrained maximum entropy sampling appears in the area of *experimental design* where the goal is to select the most *informative* subset from a collection of random variables subject to certain constraints on the set. In many settings, the collection of random variables are assumed to be distributed jointly as a Gaussian and typically the information is measured by the entropy. In such a case, the problem reduces to maximizing the logarithm of the determinant of a principal submatrix of the covariance matrix, where the columns picked must satisfy certain constraints [KLQ95, Lee98]. Lee [Lee98] considers cardinality constraints as well more general linear constraints.

In the maximum volume simplex problem [GKL95], we are given a collection of vectors and the goal is pick a subset of particular size such that the simplex formed by the picked vectors has maximum volume. The maximum volume simplex problem reduces to the maximum determinant problem under a cardinality constraint and thus is a special case of the maximum determinant problem under partition constraints [Nik14]. Apart from applications in computational geometry, the maximum volume simplex problem also finds applications in discrepancy theory and we refer the reader to Nikolov [Nik14] for a detailed discussion on this connection.

1.1   Our Results and Techniques

Our main result is to give an efficiently computable mathematical programming formulation which provably estimates the value of the optimum. We use a "geometric concave program" as a relaxation for the problem. In a geometric program [DPZ67, BV04], while the natural formulation is not a convex program, an appropriate change of variables leads to an equivalent convex program. Our formulation is a $\max - \min$ saddle point formulation where the inner minimization is a geometric program and the outer maximization is a concave program. We prove the following theorem.

Theorem 1.1 *There is a geometric concave relaxation for the maximum determinant problem under partition constraints whose optimum value is at least $\frac{r!}{r^r} = \frac{1}{e^{r+o(r)}}$ of the optimum solution to the problem. Here r denotes the rank of the matroid which equals the size of any basis. We also give a randomized polynomial time algorithm that returns an integral solution whose expected value is at least $\frac{r!}{r^r}$ of the optimum.*

We remark that it is NP-hard to approximate the problem better than a factor of c^r for some constant $c > 1$; this follows even for the cardinality constraints [ÇM13, DEFM14]. From an approximation algorithm perspective, previous results have focused on the cardinality constraint and Nikolov [Nik14] gave a $e^{r+o(r)}$-approximation.

Techniques. Previous works [Kha95, DEFM14, Nik14], for the special case of cardinality constraints, have implicitly utilized the fact that $f(x) = \det(\sum_{i=1}^{n} x_i A_i)$ is a log-concave function of x where $A_1, \ldots, A_n$ are positive semi-definite matrices. This allows naturally to formulate a concave relaxation for the problem by maximizing $f(x)$ over linear constraints.[1] Although for the simple case of cardinality constraints such an approach is successful, for the general partition constraints, such an approach does not suffice. The natural relaxation has an unbounded integrality gap; indeed, it cannot even detect whether the answer is zero or not.

Our main contribution is to give a new mathematical formulation for the problem. We define a function $f(x, y) := \det(\sum_{i=1}^{n} x_i y_i A_i)$ and formulate a saddle point problem of the form $\sup_{x \in P} \inf_{y \in Q} f(x, y)$. While the constraints on x, represented by P, are still linear, we place *polynomial* constraints on y, which are represented by Q. We show that our formulation can be optimized efficiently using the fact that $f(x, y)$ is *log-convex* in $\log y$ where $\log y$ is defined by taking logarithms coordinate wise. Thus the inner minimization problem is a geometric problem and can be reduced to a convex program by taking the logarithm of the objective and using $\log y$ as variables.[2] The details on the relaxation appear in Section 2.1.

To bound the integrality gap of the formulation, we analyze a natural randomized rounding algorithm using the theory of *stable polynomials*. In recent years, the theory of stable polynomials has been used to prove remarkable results in combinatorics, probability and theoretical computer science [Gur08, Pem12, BBL09, MSS13]. Our analysis leverages general inequalities regarding stable polynomials given by Gurvits [Gur08] who used these inequalities to obtain a new proof of the Van der Waerden conjecture regarding the permanent of doubly stochastic matrices. Let $\bar{x}$ denote the optimal solution to the geometric concave relaxation. For our case, we consider the polynomial $f(\bar{x}, y)$ in variables y which is known to be stable. The geometric relaxation can be simply seen as a constrained minimization of $f(\bar{x}, y)$. We relate the expected value of the algorithm to certain coefficients of the polynomial $f(\bar{x}, y)$. Since the inequalities developed by Gurvits are not directly applicable to $f(\bar{x}, y)$, we apply some technical transformations to obtain a polynomial $q(s)$ in new variables s. To proceed, we keep track of two things. Firstly, to control the objective of the randomized algorithm, we keep track of the correspondence between coefficients of $f(\bar{x}, y)$ and of $q(s)$. Secondly, we formulate an optimization problem over $q(s)$ whose optimal value equals the optimal value of the geometric concave relaxation. Applying Gurvits's inequalities to the polynomial $q(s)$ then gives us the result.

The above discussion applies to the special case when the rank r of the matroid $\mathcal{M}$ equals the rank d of L. When the dimension of L is more than r, the approach needs to be generalized to be applicable. In particular, we appeal to *elementary symmetric polynomials* of eigenvalues of matrices of the form $\sum_i x_i y_i A_i$ to represent our objective function. While the determinant can also be represented as an elementary symmetric polynomial in this way, the log-concavity

---

[1]In fact, these papers use the dual of this natural concave relaxation.

[2]A natural question is to ask why we did not formulate our program with variables as $\log y$ rather than y? There are two reasons. Firstly, the constraints are naturally described in the y variables and, secondly, the analysis of our rounding algorithm crucially uses the fact that the objective $f(x, y)$ is a polynomial in y.

and log-convexity properties discussed above generalize to all elementary symmetric polynomials and thus we still get a geometric concave relaxation. For the analysis of the algorithm, we note that the elementary symmetric polynomials can be obtained via differentiation of the polynomial $\det(\sum_i x_i y_i A_i + tI)$ with respect to the new variable t. Since differentiation preserves stability, this allows us to show that the objective function is still a stable polynomial and thus the inequalities on stable polynomials apply, giving us the desired result.

Even in the special case of cardinality constraints, our approach in the case $r < d$ provides a significant simplification compared to the algorithm and analysis in [Nik14].

1.2 More Related Work

The special case of a uniform matroid, i.e. maximizing the determinant under cardinality constraints, has been extensively studied. Khachiyan [Kha95] gave a simple greedy algorithm based on computing a minimum volume enclosing ellipsoid and showed that it gives a d^d approximation to the maximum determinant problem in the full-dimensional case, i.e. when the rank of the uniform matroid is equal to the rank d of the input matrix. Recently Di Summa et al. [DEFM14] gave a new analysis of Khachiyan's algorithm, and showed that it gives an approximation factor on the order of $(c \log d)^d$ for a constant c. When the rank of the uniform matroid is $r < d$, until recently the best known approximation was $r!^2$, due to Çivril and Magdon-Ismail [ÇM09]. Nikolov [Nik14] gave an algorithm with approximation factor $r^r/r!$ for any $r \le d$. Çivril and Magdon-Ismail [ÇM13] showed that it is NP-hard to approximate the maximum determinant problem under uniform matroid constraints better than a factor c^r, where $c > 1$ is a constant. Their reduction showed hardness for matroids of rank $r = \alpha d$ with $0 < \alpha < 1$ a constant. Recently, Di Summa et al. showed an analogous hardness result for the full dimensional case $r = d$.

The constrained maximum determinant problem is also closely connected with constrained submodular maximization problems. Indeed, the set function $g : 2^U \to \mathbb{R} \cup \{-\infty\}$ defined by $g(S) = \ln \det(L_{S,S})$ is known to be submodular. Thus our problem can be reduced to a special case of submodular function maximization subject to matroid constraints; a problem that has received much attention lately [CCPV11, FNS11, LMNS09, NWF78]. Since the function g defined as above is not necessarily non-negative, these results are not applicable in our setting. Moreover, notice that to get a multiplicative approximation to our problem, we need an additive approximation to the maximum of g over $\mathcal{B}$.

For the applications of constrained determinant problem in machine learning, we refer the reader to the survey [KT12] and the works of Gillenwater et al [GKT12]. Ko et al [KLQ95] detail applications of the constrained entropy maximization problem in experimental design. We refer to this paper and references therein. Nikolov [Nik14] gives a survey of the work and applications of the maximum volume simplex problem in different areas.

2   A Simple Case: $r = d$

We first assume that $r = d$ and thus the rank of the matroid $\mathcal{M}$ equals the rank of the matrix L. Although this simple case is subsumed in the general case considered in Section 3, our main ideas can be conveyed easily for this special case. Let $L = V^T V$ where V is a d by n matrix and d is the rank of L (so V is full rank). For any set $S \subseteq U$, we let V_S denote the d by $|S|$ matrix formed by picking columns indexed by S. Let v_e denote the column of V indexed by $e \in U$. For any $x \in \mathbb{R}^U$ and $S \subseteq U$, we also have $x^S := \prod_{e \in S} x_e$.

2.1   Geometric Concave Program

We first give a concave geometric relaxation for the problem. We describe our concave program for a general matroid since it leads to a cleaner exposition. While the relaxation is valid for any

matroid, the approximation algorithm utilizes the structure of the partition matroid. We refer the reader to Section 4 for remarks on the general matroid setting. Let $P(\mathcal{B})$ denote the convex hull of indicator vectors of bases of $\mathcal{M}$. Let $Q(\mathcal{B}) = \{y \in \mathbb{R}_+^U : y^B = \prod_{e \in B} y_e = 1 \ \forall \ B \in \mathcal{B}\}$. Let $f(x, y) = \det\left(\sum_{e \in U} x_e y_e v_e v_e^T\right)$. We have the following relaxation for the problem.

Lemma 2.1 *We have*

$$\max_{S \in \mathcal{B}} \det(L_{S,S}) \leq \sup_{x \in P(\mathcal{B})} \inf_{y \in Q(\mathcal{B})} f(x, y). \tag{1}$$

Moreover we can compute a (near) optimal solution $\bar{x}$ to $\sup_{x \in P(\mathcal{B})} \inf_{y \in Q(\mathcal{B})} f(x, y)$ in polynomial time.

Proof: Let $B \in \mathcal{B}$ denote the maximizer of the left hand side. Let $\hat{x} = 1_B$. We show that $f(\hat{x}, y) = \det(L_{B,B})$ for any $y \in Q(\mathcal{B})$. Indeed, for any $y \in Q(\mathcal{B})$, we have

$$f(\hat{x}, y) = \det\left(\sum_{e \in B} y_e v_e v_e^T\right) \quad = (\prod_{e \in B} y_e) \cdot \det\left(V_B V_B^T\right) = 1 \cdot \det\left(V_B^T V_B\right) \quad = \det(L_{B,B}) \tag{2}$$

where we use the fact that $\det(AC) = \det(CA)$ and $y^B = 1$ since $y \in Q(\mathcal{B})$ which proves the first claim of the lemma. Now, we observe that $\log f(x, y)$ is concave in x and convex in $\log y$ where $\log y$ is the vector obtained by taking logarithms coordinate wise. The proof is standard, see Section 3.1, page 74 [BV04] for the former and Lemma 23 [GS00] for the latter. For an alternate proof, we refer to the general case of elementary symmetric polynomials discussed in Section 3.1, Lemma 3.2, of which the determinant is a special case.

Claim 2.2 *For any $y \in \mathbb{R}_+^U$, let $z \in \mathbb{R}^U$ such that $z_e = \log y_e$ for each $e \in U$. Let $g(x, z) = \log \det\left(\sum_{e \in U} x_e \exp(z_e) v_e v_e^T\right)$. For any $z \in \mathbb{R}^U$, g is a concave function of x and for any $x \in \mathbb{R}^U$, g is a convex function of z.*

Let $\tilde{Q}(\mathcal{B}) = \{z \in \mathbb{R}^U : \sum_{e \in B} z_e = 0 \ \forall B \in \mathcal{B}\}$. Then we have

$$\sup_{x \in P(\mathcal{B})} \inf_{y \in Q(\mathcal{B})} \log f(x, y) = \sup_{x \in P(\mathcal{B})} \inf_{z \in \tilde{Q}(\mathcal{B})} g(x, z)$$

Observe that separation over $P(\mathcal{B})$ can be implemented using an independence oracle. The separation oracle for $\tilde{Q}(\mathcal{B})$ needs to check whether both the maximum weight basis and the minimum weight basis have weight zero and thus can also be implemented using an independence oracle. Since g is a convex function of z, for any x, given a first order oracle for $g(x, .)$, we can find in polynomial time $\inf_{z \in \tilde{Q}(\mathcal{B})} g(x, z)$. Since $g(x, z)$ is concave for any z, we obtain that $\inf_{z \in \tilde{Q}(\mathcal{B})} g(x, z)$ is also a concave function of x and thus $\sup_{x \in P(\mathcal{B})} \inf_{z \in \tilde{Q}(\mathcal{B})} g(x, z)$ can be optimized in polynomial time using the ellipsoid algorithm. Observe that first order oracle computation amounts to computing basic operations on involved matrices and thus can be implemented efficiently. Finally, we note that the ellipsoid algorithm (or the interior point algorithm) will return a near optimal solution to the problem to a desired accuracy. For the rest of the paper, we omit the error introduced by this accuracy parameter which will lead to a small loss of $(1 + \epsilon)$ factor in the approximation ratio. $\quad \square$

2.2 Stable Polynomials

Before, we describe our algorithm, we describe stable polynomials and inequalities regarding them that will be utilized in the proof. We refer the reader to the surveys [Pem12, Vis13] for various applications of stable polynomials in combinatorics, probability theory and theoretical computer science. We let $\mathbb{R}_+[x_1, \ldots, x_n]$ denote the set of polynomials in variables $x_1, \ldots, x_n$ with non-negative real coefficients.

Definition 2.3 *A homogeneous degree d polynomial $p \in \mathbb{R}_+[x_1, \ldots, x_n]$ is called $\mathcal{H}$-stable (stable for short), if $p(x) \neq 0$ for each $x \in \mathbb{C}$ such that $Re(x) > 0$ (coordinate wise). Thus $p(x)$ has no roots all of whose coordinates are in the right half plane.*

Typically, the stable polynomials are also defined by not allowing roots in the upper half plane. Since we restrict our attention to homogeneous polynomials, the two definitions coincide. We use the following two theorems about stable polynomials.

Theorem 2.4 *[Pem12, Prop. 6.18] Given PSD matrices $A_1, \ldots, A_n$ such that $\sum_{i=1}^n A_i$ is positive definite, the polynomial $p(x) = \det(\sum_{i=1}^n x_i A_i)$ is stable and homogeneous of degree n.*

Theorem 2.5 *[Pem12, pp 34-35] Let $p(x)$ be a stable polynomial in d variables. Then p remains stable under the following operations: (a) setting some $x_i = c$ for some i and some constant c with $Re(c) \geq 0$; (b) replacing p by $\frac{\partial p}{\partial x_i}$ for some i.*

The following theorem of Gurvits[Gur08] was crucial in the alternate proof of the Van der Waerden conjecture [Ego81, Fal81] about permanent of doubly stochastic matrices (also see Laurent and Schrjiver [LS10] for an exposition). Given a polynomial $h \in \mathbb{R}_+[x_1, \ldots, x_n]$, we define $cap(h) = \inf_{\{x \in \mathbb{R}_+^n : \prod_{i=1}^n x_i = 1\}} h(x)$. We also define $h'(x_1, \ldots, x_{n-1}) = \frac{\partial h}{\partial x_n}|_{x_n=0}$.

Theorem 2.6 *[Gur08] Given a n-variate degree n homogeneous stable polynomial $h \in \mathbb{R}_+[x_1, \ldots, x_n]$, either we have $h' = 0$ or h' is stable. Moreover,*

$$cap(h') \geq \left(\frac{n-1}{n}\right)^{n-1} cap(h). \tag{3}$$

Thus, if $\frac{\partial^n h}{\partial x_1 \ldots \partial x_n} \neq 0$ then we have

$$\frac{\partial^n h}{\partial x_1 \ldots \partial x_n} \geq \frac{n!}{n^n} \inf_{\{x \in \mathbb{R}_+^n : \prod_{i=1}^n x_i = 1\}} h(x). \tag{4}$$

We remark that the Van der Waerden conjecture is immediate from inequality (4) as shown in Gurvits [Gur08]. Indeed, given a non-negative $n \times n$ doubly stochastic matrix A, let $h(x) = \prod_{i=1}^n \sum_{j=1}^n (A_{ij} x_j)$ which can be easily seen to be stable. A simple check shows that the LHS of Inequality (4) equals the permanent and the infimum of right hand side equals one. Thus we obtain that the permanent of a $n \times n$ doubly stochastic matrix is at least $\frac{n!}{n^n}$. We also note that tighter estimates of Inequality (4) are shown in Gurvits [Gur08] depending on the degree of each x_i in h and useful in estimating permanent of matrices in other restricted classes.

2.3 Rounding Algorithm

Given an optimal solution $\bar{x}$ to $\max_{x \in P(\mathcal{B})} \min_{y \in Q(\mathcal{B})} f(x, y)$, we select a basis $B \in \mathcal{B}$ as follows. Independently for each $1 \leq i \leq k$, we select b_i elements from P_i, with replacement, where an element e in P_j is selected with probability $\frac{\bar{x}_e}{b_j}$. We reject the solution if $B \notin \mathcal{B}$, that is, we reject B if all sampled elements are not distinct. In case of rejection, we assume that the objective value achieved by the algorithm is 0. Let B denote the random basis obtained in the above sampling.

To analyze the performance of the algorithm, we consider the polynomial $p(y) = f(\bar{x}, y) = \det(\sum_{e \in U} \bar{x}_e y_e v_e v_e^T)$. The objective value of the concave program, which is an upper bound on the optimum, is the infimum value of the polynomial over the set $Q(\mathcal{B})$. We relate this bound to the value of the solution returned by the algorithm using inequalities from stable polynomials. In Lemma 2.7, we relate the expected value of the algorithm to the sum of certain

coefficients of this polynomial. From Theorem 2.4, we have that $p(y)$ is a stable polynomial. We intend to apply Theorem 2.6 but since p does not satisfy the conditions of the theorem, we apply it to a closely related polynomial.

Lemma 2.7 *For any subset $S \subseteq U$, we let c_S denote the coefficient of y^S in $p(y)$. The expected value of algorithm is*

$$E[\det(L_{B,B})] = \left(\prod_{i=1}^{k} \frac{b_i!}{b_i^{b_i}}\right) \sum_{S \in \mathcal{B}} c_S. \tag{5}$$

Proof: For any element $e \in P_i$, let $p_e = \frac{\bar{x}_e}{b_i}$ denote the probability with which e is included in the random sample. Notice that for any $S \in \mathcal{B}$, there are $\prod_{i=1}^{k} b_i!$ permutation of the sampled elements that allow S to be selected. Namely, we can arbitrarily permute the b_i elements of S in P_i for each i. Thus

$$E[\det(M_{B,B})] = \sum_{S \in B(\mathcal{M})} \left(\prod_{i=1}^{k} b_i!\right) \det(L_{S,S}) \prod_{e \in S} p_e = \left(\prod_{i=1}^{k} \frac{b_i!}{b_i^{b_i}}\right) \sum_{S \in B(\mathcal{M})} \bar{x}^S \det(L_{S,S})$$

Now, the Cauchy-Binet formula shows that

$$\det(\sum_{e \in U} \bar{x}_e y_e v_e v_e^T) = \sum_{S \subseteq U : |S| = d} \bar{x}^S \bar{y}^S \det(\sum_{e \in S} v_e v_e^T) = \sum_{S \subseteq U : |S| = d} y^S \bar{x}^S \det(L_{S,S})$$

We then have that the coefficient of y^S is $c_S = \bar{x}^S \det(L_{S,S})$, giving us the lemma. $\square$

Consider $d = \sum_{i=1}^{k} b_i$ variables $s_{i,j}$ for $1 \le i \le k$ and $1 \le j \le b_i$. We construct a new polynomial $q(s)$ in these d variables defined as follows. Given any s, we let $y(s)$ be $|U|$ variables where for any $e \in U$ such that $e \in P_i$, we let $y(s)_e = \frac{1}{b_i} \sum_{j=1}^{b_i} s_{i,j}$. We then let $q(s) = p(y(s))$. Observe that q is a degree d homogeneous polynomial with d variables. We first relate the optimization problem over p to an optimization problem over q. For the next lemma, we assume without loss of generality, that $|P_i| > b_i$. Else, we can reduce the problem by selecting all elements in P_i in the basis and solving the reduced problem.

Lemma 2.8 *The polynomial $q(s)$ is a d-variate homogeneous degree d stable polynomial and we have*

$$\inf_{\{s \in \mathbb{R}_+^d : \prod_{i=1}^{k} \prod_{j=1}^{b_i} s_{ij} = 1\}} q(s) = \inf_{y \in Q(\mathcal{B})} p(y).$$

Proof: The polynomial $q(s) = \det(\sum_{i=1}^{k} \sum_{j=1}^{b_i} s_{ij} A^i)$ for the PSD matrices $A^i := \frac{1}{b_i} \sum_{e \in P_i} x_e v_e v_e^T$. Also it follows from Lemma 2.1 that, unless $max_{B \in \mathcal{B}} \det(L_{B,B}) = 0$, $\sum_i A^i$ is positive definite. Thus q is degree d homogeneous and stable from Theorem 2.4.

For the second claim, first let s be a minimizer of the left hand side. Observe that for any $1 \le i \le k$, $q(s)$ is symmetric in the variables $s_{i,1}, \ldots, s_{i,b_i}$. Moreover $\log(q(s))$ is convex in $\log s$ (defined coordinate-wise), because $\log q(y)$ is convex and monotone in $\log y(s)$, and each coordinate of $\log y(s)$ is convex in $\log s$, as can be shown by a direct computation of the Hessian, or using Hölder's inequality. Finally, the constraint set $\{s \in \mathbb{R}_+^d : \prod_{i=1}^{k} \prod_{j=1}^{b_i} s_{ij} = 1\}$ is linear in $\log s$ and symmetric under permutations of the variables. Thus there exists an optimum solution such that for each $1 \le i \le k$, we have $s_{i,j} = s_{i,j'}$ for any $1 \le j, j' \le b_i$. Now we construct a feasible $y \in Q(\mathcal{B})$. For any $e \in U$ such that $e \in P_i$, we let $y_e = s_{i1}$. Then for any $B \in \mathcal{B}$, we have

$$\prod_{e \in B} y_e = \prod_{i=1}^{k} s_{i1}^{b_i} = \prod_{i=1}^{k} \prod_{j=1}^{b_i} s_{ij} = 1.$$

and therefore $y \in Q(\mathcal{B})$. But a simple check shows that $p(y) = q(s)$ and therefore we obtain that the left hand side is at least the right hand side.

Now consider any $y \in Q(\mathcal{B})$. We claim that $y_e = y_f$ if $e, f \in P_i$ for some i. Indeed consider a basis $B \in \mathcal{B}$ such that $e \in B$ but $f \notin B$. Such a basis exists since $|P_i| > b_i$ by our assumption. Then we also have $B' = B \cup \{f\} \setminus \{e\} \in \mathcal{B}$. Since $y^B = y^{B'} = 1$ we obtain that $y_e = y_f$. Now, for any $1 \le i \le k$ and $1 \le j \le k$, we let $s_{ij} = y_e$ where $e \in P_i$. Observe that $\prod_{i=1}^{k} \prod_{j=1}^{b_i} s_{ij} = \prod_{e \in B} y_e$ for any basis B and therefore s satisfies the constraints. Moreover, $q(s) = p(y(s)) = p(y)$ where we use the fact that $y(s)_e = \frac{1}{b_i} \sum_{j=1}^{b_i} s_{i,j} = y_e$ for any $e \in P_i$. Thus we obtain the desired equality. $\qquad\square$

Now we apply Theorem 2.6 and prove the required guarantee.

Lemma 2.9 *Let $B \in \mathcal{B}$ be the random basis returned by the algorithm. We have*

$$E[\det(L_{B,B})] \ge \frac{d!}{d^d} \inf_{y \in Q(\mathcal{B})} p(y)$$

Proof: From Lemma 2.7, we have that

$$E[\det(L_{B,B})] = \left(\prod_{i=1}^{k} \frac{b_i!}{b_i^{b_i}} \right) \sum_{S \in \mathcal{B}} c_S.$$

where c_S is coefficient of y^S in $p(y)$. Now, observe that the coefficient of $\prod_{i=1}^{k} \prod_{j=1}^{b_i} s_{ij}$ in $q(s) = p(y(s))$ equals $\prod_{i=1}^{k} \frac{b_i!}{b_i^{b_i}} \sum_{S \in \mathcal{B}} c_S$. Applying Theorem 2.6, we obtain that coefficient of $\prod_{i=1}^{k} \prod_{j=1}^{b_i} s_{ij}$ in $q(s)$ which equals

$$\frac{\partial^d q}{\partial s_{1,1} \ldots \partial s_{k,b_k}} \ge \frac{d!}{d^d} \inf_{\{s \in \mathbb{R}_+^d : \prod_{i=1}^{k} \prod_{j=1}^{b_i} s_{ij}\}} q(s).$$

From the second claim in Lemma 2.8, we have that the right hand side is equal to $\frac{r!}{r^r} \inf_{y \in Q(\mathcal{B})} p(y)$, and we obtain the desired inequality. $\qquad\square$

The special case of Theorem 1.1 for $r = d$ now follows directly from Lemmas 2.1 and 2.9.

3 General Rank

We now consider the general case when the rank r of $\mathcal{M}$ is at most d, but not necessarily equal to d. (Notice that the case $r > d$ is trivial, as $\det(L_{B,B})$ would be 0 for any basis B of $\mathcal{M}$.) We keep the notation from before.

3.1 Elementary Symmetric Polynomials

Recall the definition of the degree k elementary symmetric polynomial σ_r:

$$\sigma_r(x_1, \ldots, x_n) = \sum_{S \subseteq [n]:|S|=r} x^S.$$

We define the function sym_r on the space of d by d positive semidefinite matrices as

$$\text{sym}_r(A) = \sigma_r(\lambda_1, \ldots, \lambda_d),$$

where $\lambda_1, \ldots, \lambda_d$ are the eigenvalues of the matrix A. Notice that $\mathrm{sym}_1(A)$ is the trace of A, and $\mathrm{sym}_d(A)$ is the determinant. In general,

$$\mathrm{sym}_r(A) = \frac{1}{(d-r)!} \frac{\partial^{d-r}}{\partial t^{d-r}} \det(A+tI)|_{t=0}. \tag{6}$$

Equation (6) gives a method for efficiently computing $\mathrm{sym}_r(A)$: we can compute in polynomial time $\det(A+tI)$ as a degree d polynomial in t. Then $\mathrm{sym}_r(A)$ is exactly the coefficient of t^{d-r} in this polynomial.

Suppose we have a rank r matrix, e.g. $A = \sum_{i=1}^r a_i a_i^T$. Then $\mathrm{sym}_r(A)$ is equal to the product of eigenvalues of the matrix A. A in turn has the same non-zero eigenvalues as the Gram matrix $G = (\langle a_i, a_j \rangle)_{i,j}$, so we have $\mathrm{sym}_r(A) = \det(G)$. It follows that $\mathrm{sym}_r(A)$ satisfies the usual scaling properties of the determinant, e.g.

$$\mathrm{sym}_r \left(\sum_{i=1}^r c_i a_i a_i^T \right) = \det(C^{1/2} G C^{1/2}) = \det(G) \prod_{i=1}^r c_i = \mathrm{sym}_r(A) \prod_{i=1}^r c_i, \tag{7}$$

where C is the diagonal matrix with the numbers $c_1, \ldots, c_r$ on the diagonal.

We have the following well-known formula for any collection of vectors $a_1, \ldots, a_n \in \mathbb{R}^d$:

$$\mathrm{sym}_r(\sum_{i=1}^n a_i a_i^T) = \sum_{S \subseteq [n]:|S|=r} \mathrm{sym}_r(\sum_{i \in S} a_i a_i^T) \tag{8}$$

Equation (8) generalizes the Cauchy-Binet formula, which is the special case for $r = d$. For completeness, we include a proof in the appendix.

The following theorem has been proved several times, e.g. in [ML57, McL59], and is also a consequence of the Aleksandrov-Fenchel inequality for mixed discriminants (see e.g. [Sch93, Thm 6.8.1.] and [Gar02, Sect. 17]).

Theorem 3.1 *The function* $\log \mathrm{sym}_r(A)$ *is concave over the space of positive semidefinite matrices.*

See the appendix for the relationship between Theorem 3.1 and the Aleksandrov-Fenchel inequality.

The derivatives of σ_r, and, therefore, $\log \sigma_r$ are easy to compute directly:

$$\frac{\partial}{\partial x_i} \log \sigma_r(x_1, \ldots, x_n) = \frac{\sigma_{r-1}(x_1, \ldots, x_{i-1}, x_{i+1}, \ldots x_n)}{\sigma_r(x)}.$$

Since $\log \sigma_r$ is a symmetric concave function (the concavity is the special case of Theorem 3.1 for diagonal matrices), by [Lew95, Thm 3.1] the gradient of $\log \mathrm{sym}_r(A)$ for a PSD matrix A with singular value decomposition $A = W \Lambda W^T$ and vector of eigenvalues $\lambda = (\lambda_1, \ldots, \lambda_n)$ is given by

$$\nabla \log \mathrm{sym}_r(A) = W \mathrm{diag}(\nabla \log \sigma_r(\lambda)) W^T,$$

where $\mathrm{diag}(x)$ is the diagonal matrix with the entries of the vector x on the diagonal. Therefore, $\nabla \log \mathrm{sym}_r(A)$ can be computed from the SVD decomposition of A in polynomial time.

3.2 Geometric Concave Program

Our approach for matroids of general rank is analogous to the full rank case. The main difference is that, in formulating our geometric concave programming relaxations we replace determinants by sym_r. Let, to this end, $f(x, y) = \mathrm{sym}_r(\sum_{e \in U} x_e y_e v_e v_e^T)$ for the rest of this section.

It is a standard fact that $f(x, y)$ is log-convex in $\log y$. We include a proof in the appendix for completeness.

Lemma 3.2 *The function* $\log f(x, y)$ *is convex in* $\log y$, *where* $\log y$ *is the vector obtained from* $y \in \mathbb{R}_+^U$ *by taking logarithms coordinate wise.*

We can now prove an analogue of Lemma 2.1. Let, as before $P(\mathcal{B})$ denote the convex hull of the indicator vectors of bases of $\mathcal{M}$, and $Q(\mathcal{B}) = \{y \in \mathbb{R}_+^U : y^B = 1 \forall B \in \mathcal{B}\}$.

Lemma 3.3 *We have*

$$\max_{B \in \mathcal{B}} \det(L_{B,B}) \leq \sup_{x \in P(\mathcal{B})} \inf_{y \in Q(\mathcal{B})} f(x, y).$$

Moreover can compute an (near) optimal solution $\bar{x}$ *to* $\sup_{x \in P(\mathcal{B})} \inf_{y \in Q(\mathcal{B})} f(x, y)$ *in polynomial time.*

Proof: Let $B \in \mathcal{B}$ be a maximizer of the left hand side, and let $\hat{x} = 1_B$. We show that $f(\hat{x}, y) = \det(L_{B,B})$ for any $y \in Q(\mathcal{B})$. Indeed, for any $y \in Q(\mathcal{B})$ we have

$$f(\hat{x}, y) = \mathrm{sym}_r \left(\sum_{e \in B} y_e v_e v_e^T \right) = y^B \det(V_B^T V_B) = \det(L_{B,B}),$$

where we used (7) together with the fact that $V_B^T V_B$ is the Gram matrix of the vectors $\{v_e\}_{e \in B}$, and also we used that $y^B = 1$ for $y \in Q(\mathcal{B})$ and $B \in \mathcal{B}$.

To show that the optimization problem on the right hand side can be solved in polynomial time, we formulate it as an equivalent convex optimization problem, similarly to the proof of Lemma 2.1. Define $g(x, z) = \log \mathrm{sym}_r(\sum_{e \in U} x_e e^{z_e} v_e v_e^T)$. By Theorem 3.1, g is a concave function of x, and by Lemma 3.2, g is a convex function of z. Moreover, for $\tilde{Q}(\mathcal{B}) = \{z \in \mathbb{R}^U : \sum_{e \in B} z_e = 0 \ \forall B \in \mathcal{B}\}$, we have

$$\sup_{x \in P(\mathcal{B})} \inf_{y \in Q(\mathcal{B})} \log f(x, y) = \sup_{x \in P(\mathcal{B})} \inf_{z \in \tilde{Q}(\mathcal{B})} g(x, z).$$

As argued in the proof of Lemma 2.1, optimization over $\tilde{Q}(B)$ can be implemented efficiently given an independence oracle for $\mathcal{M}$. Then, because $g(x, z)$ is convex in z, for any x we can find in polynomial time $\inf_{z \in \tilde{Q}(\mathcal{B})} g(x, z)$ using a first-order oracle for $g(x, z)$. As argued in Subsection 3.1, the first-order oracle can be implemented in polynomial time by computing singular value decompositions. Because $g(x, z)$ is concave in x for any z, $\inf_{z \in \tilde{Q}(\mathcal{B})} g(x, z)$ is also concave in x, and $\sup_{x \in P(\mathcal{B})} \inf_{z \in \tilde{Q}(\mathcal{B})} g(x, z)$ can be approximated to any desired accuracy in polynomial time using the ellipsoid algorithm. $\square$

3.3 Rounding Algorithm

The rounding algorithm is identical to the full rank case, and the analysis is analogous. Given an optimal solution $\bar{x}$ to $\sup_{x \in P(\mathcal{B})} \inf_{y \in Q(\mathcal{B})} f(x, y)$, we select a basis $B \in \mathcal{B}$ by picking b_j elements from P_j with replacement, and each element $e \in P_j$ is selected with probability $\frac{\bar{x}_e}{b_j}$. We reject the solution if $B \notin \mathcal{B}$, which happens exactly when some element is sampled more than once. In case of rejection, we assume that the objective value achieved by the algorithm is 0.

Once again we define the polynomial $p(y) = f(\bar{x}, y)$. The analysis follows similar lines to the full rank case and the proof appears in the appendix.

Lemma 3.4 *For any subset* $S \subseteq U$, *we* c_S *denote the coefficient of* y^S *in* $p(y)$. *The expected value of algorithm is*

$$E[\det(L_{B,B})] = \left(\prod_{i=1}^k \frac{b_i!}{b_i^{b_i}} \right) \sum_{S \in \mathcal{B}} c_S. \tag{9}$$

Define $r = \sum_{i=1}^{k} b_i$ variables $s_{i,j}$ for $1 \le i \le k$ and $1 \le j \le b_i$. We define a new polynomial q in these r variables analogously to the way we defined q in the full rank case. Let $y(s)$ be a vector of variables indexed by U such that for any $e \in P_i$, $y(s)_e = \frac{1}{b_i} \sum_{j=1}^{b_i} s_{i,j}$. We then let $q(s) = p(y(s))$; q is a degree r homogeneous polynomial with r variables.

Lemma 3.5 *The polynomial $q(s)$ is a r-variate homogeneous degree r stable polynomial and we have*

$$\inf_{\{s \in \mathbb{R}_+^d : \prod_{i=1}^{k} \prod_{j=1}^{b_i} s_{ij} = 1\}} q(s) = \inf_{y \in Q(\mathcal{B})} p(y).$$

Proof: The function $\log(q(s)) = \log p(y(s))$ is convex in $\log s$, defined coordinatewise, since $p(y)$ is monotone, convex in $\log(y(s))$ by Lemma 3.2, and each coordinate of $\log(y(s))$ is a convex function of $\log s$. The proof of the second claim is then analogous to the proof of the second claim of Lemma 2.8. Next we prove the first claim.

The polynomial q is given by $q(s) = \mathrm{sym}_r(\sum_{i=1}^{k} \sum_{j=1}^{b_i} s_{ij} A^i)$ for the PSD matrices $A^i := \frac{1}{b_i} \sum_{e \in P_i} x_e v_e v_e^T$. By Theorem 2.4, the degree d homogeneous polynomial

$$\tilde{q}(s,t) = \det(\sum_{i=1}^{k} \sum_{j=1}^{b_i} s_{ij} A^i + tI)$$

is stable. By equation (6),

$$q(s) = \frac{1}{(d-r)!} \frac{\partial^{d-r} \tilde{q}(s,t)}{\partial t^{d-r}} \Bigg|_{t=0}.$$

Moreover, it follows from Lemma 3.3 that q cannot be identically zero, so, by Theorem 2.5, q is stable. $\square$

We are now ready to apply Theorem 2.6 and finish the proof.

Lemma 3.6 *Let $B \in \mathcal{B}$ be the random basis returned by the algorithm. We have*

$$E[\det(L_{B,B})] \ge \frac{r!}{r^r} \inf_{y \in Q(\mathcal{B})} p(y)$$

Proof: From Lemma 3.4, we have that

$$E[\det(L_{B,B})] = \left(\prod_{i=1}^{k} \frac{b_i!}{b_i^{b_i}} \right) \sum_{S \in \mathcal{B}} c_S.$$

where c_S is coefficient of y^S in $p(y)$. Now, observe that the coefficient of $\prod_{i=1}^{k} \prod_{j=1}^{b_i} s_{ij}$ in $q(s)$ equals $\prod_{i=1}^{k} \frac{b_i!}{b_i^{b_i}} \sum_{S \in \mathcal{B}} c_S$. Using the fact, proved in Lemma 3.5, that q is stable and homogeneous of degree r, and applying Theorem 2.6, we obtain that the coefficient of $\prod_{i=1}^{k} \prod_{j=1}^{b_i} s_{ij}$ in $q(s)$ satisfies

$$\frac{\partial^d q}{\partial s_{1,1} \dots \partial s_{k,b_k}} \ge \frac{r!}{r^r} \inf_{\{s \in \mathbb{R}_+^d : \prod_{i=1}^{k} \prod_{j=1}^{b_i} s_{ij}\}} q(s).$$

From the second claim in Lemma 3.5, we have that the right hand side is equal to $\frac{r!}{r^r} \inf_{y \in Q(\mathcal{B})} p(y)$, and we obtain the desired inequality. $\square$

Theorem 1.1 follows directly from Lemmas 3.3 and 3.6.

4   Conclusion

In this paper, we gave a geometric concave program for the maximum determinant problem with partition constraints and showed a bound of $\frac{r^r}{r!}$ on the integrality gap where r is the rank of the matroid. The upper bound was shown via a randomized algorithm whose analysis used the properties of stable polynomials. A natural open question is whether there is a deterministic algorithm achieving the same guarantee? This question achieves significance since the randomized algorithm can have significantly large variance and thus cannot guarantee a solution with the same or slightly worse guarantee with high probability.

The other natural question is whether the geometric concave program outlined in the paper can approximate the maximum determinant problem under general matroid constraints? As discussed in Section 2.1, the computational question of solving the geometric concave program is resolved in the presence of an independence oracle for the matroid. Unfortunately, the geometric program cannot detect whether the answer is zero or not even in the special case of graphic matroid constraints. Thus it cannot be used to give any multiplicative guarantee. However, we note that the problem of detecting whether the optimum solution is zero even under arbitrary matroid constraints can be reduced to the matroid intersection problem. This leaves open the possibility of achieving similar bounds for general matroid constraints.

Acknowledgements

References

[BBL09]   Julius Borcea, Petter Brändén, and Thomas Liggett. Negative dependence and the geometry of polynomials. *Journal of the American Mathematical Society*, 22(2):521–567, 2009.

[BV04]   Stephen Boyd and Lieven Vandenberghe. *Convex optimization*. Cambridge university press, 2004.

[CCPV11]   Gruia Calinescu, Chandra Chekuri, Martin Pál, and Jan Vondrák. Maximizing a monotone submodular function subject to a matroid constraint. *SIAM Journal on Computing*, 40(6):1740–1766, 2011.

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

[ÇM13]   Ali Çivril and Malik Magdon-Ismail. Exponential inapproximability of selecting a maximum volume sub-matrix. *Algorithmica*, 65(1):159–176, 2013.

[DEFM14]   Marco Di Summa, Friedrich Eisenbrand, Yuri Faenza, and Carsten Moldenhauer. On largest volume simplices and sub-determinants. *CoRR*, abs/1406.3512, 2014. To appear in SODA 2015.

[DPZ67]   Richard James Duffin, Elmor L Peterson, and Clarence Zener. *Geometric programming: theory and application*. Wiley New York, 1967.

[Ego81]   GP Egorychev. The solution of van der waerden's problem for permanents. *Advances in Mathematics*, 42(3):299–305, 1981.

[Fal81]   Dmitry I Falikman. Proof of the van der waerden conjecture regarding the permanent of a doubly stochastic matrix. *Mathematical Notes*, 29(6):475–479, 1981.

[FNS11]   Moran Feldman, Joseph Naor, and Roy Schwartz. A unified continuous greedy algorithm for submodular maximization. In *Foundations of Computer Science (FOCS), 2011 IEEE 52nd Annual Symposium on*, pages 570–579. IEEE, 2011.

[Gar02]   R. J. Gardner. The Brunn-Minkowski inequality. *Bull. Amer. Math. Soc. (N.S.)*, 39(3):355–405, 2002.

[GKL95]   Peter Gritzmann, Victor Klee, and David Larman. Largestj-simplices inn-polytopes. *Discrete & Computational Geometry*, 13(1):477–515, 1995.

[GKT12]   Jennifer Gillenwater, Alex Kulesza, and Ben Taskar. Near-optimal map inference for determinantal point processes. In *Advances in Neural Information Processing Systems*, pages 2735–2743, 2012.

[GS00]   Leonid Gurvits and Alex Samorodnitsky. A deterministic polynomial-time algorithm for approximating mixed discriminant and mixed volume. In *Proceedings of the thirty-second annual ACM symposium on Theory of computing*, pages 48–57. ACM, 2000.

[Gur08]   Leonid Gurvits. Van der waerden/schrijver-valiant like conjectures and stable (aka hyperbolic) homogeneous polynomials: one theorem for all. *Electron. J. Combin*, 15(1), 2008.

[Kha95]   Leonid Khachiyan. On the complexity of approximating extremal determinants in matrices. *Journal of Complexity*, 11(1):138–153, 1995.

[Kho84]   A. G. Khovanskii. Analogues of Aleksandrov-Fenchel inequalities for hyperbolic forms. *Dokl. Akad. Nauk SSSR*, 276(6):1332–1334, 1984.

[KLQ95]   Chun-Wa Ko, Jon Lee, and Maurice Queyranne. An exact algorithm for maximum entropy sampling. *Operations Research*, 43(4):684–691, 1995.

[KT10]   Alex Kulesza and Ben Taskar. Structured determinantal point processes. In *Advances in neural information processing systems*, pages 1171–1179, 2010.

[KT12]   Alex Kulesza and Ben Taskar. Determinantal point processes for machine learning. *arXiv preprint arXiv:1207.6083*, 2012.

[Lee98]   Jon Lee. Constrained maximum-entropy sampling. *Operations Research*, 46(5):655–664, 1998.

[Lew95]   A. S. Lewis. The convex analysis of unitarily invariant matrix functions. *J. Convex Anal.*, 2(1-2):173–183, 1995.

[LMNS09]   Jon Lee, Vahab S Mirrokni, Viswanath Nagarajan, and Maxim Sviridenko. Non-monotone submodular maximization under matroid and knapsack constraints. In *Proceedings of the forty-first annual ACM symposium on Theory of computing*, pages 323–332. ACM, 2009.

[LS10]   Monique Laurent and Alexander Schrijver. On leonid gurvitss proof for permanents. *The American Mathematical Monthly*, 117(10):903–911, 2010.

[Lyo03]   Russell Lyons. Determinantal probability measures. *Publications Mathématiques de l'IHÉS*, 98:167–212, 2003.

[McL59]   J. B. McLeod. On four inequalities in symmetric functions. *Proc. Edinburgh Math. Soc.*, 11:211–219, 1958/1959.

[ML57]   M. Marcus and L. Lopes. Inequalities for symmetric functions and Hermitian matrices. *Canad. J. Math.*, 9:305–312, 1957.

[MSS13]   Adam Marcus, Daniel A Spielman, and Nikhil Srivastava. Interlacing families ii: Mixed characteristic polynomials and the kadison-singer problem. *arXiv preprint arXiv:1306.3969*, 2013.

[Nik14]   Aleksandar Nikolov. Randomized rounding for the largest *j*-simplex problem. *arXiv preprint arXiv:1412.0036*, 2014.

[NWF78]   George L Nemhauser, Laurence A Wolsey, and Marshall L Fisher. An analysis of approximations for maximizing submodular set functionsi. *Mathematical Programming*, 14(1):265–294, 1978.

[Pem12]   Robin Pemantle. Hyperbolicity and stable polynomials in combinatorics and probability. *arXiv preprint arXiv:1210.3231*, 2012.

[Sch93]   Rolf Schneider. *Convex bodies: the Brunn-Minkowski theory*, volume 44 of *Encyclopedia of Mathematics and its Applications*. Cambridge University Press, Cambridge, 1993.

[Vis13]   Nisheeth K Vishnoi. Zeros of polynomials and their applications to theory: a primer. 2013.

A   Mixed Discriminants and Symmetric Polynomials

Here we give some background on the function sym_r and its relationship to mixed discriminants. We also sketch a proof of equation (8) based on elementary properties of mixed discriminants and outline how Theorem 3.1 follows from the Alexandrov-Fenchel inequality.

Definition A.1 *Given $d \times d$ PSD matrices $A_1, \ldots, A_d$, the mixed discriminant is defined as*

$$D(A_1; \ldots; A_d) = \frac{1}{d!} \sum_{\sigma \in S_d} \det(A_{\sigma(1)}^{(1)}, A_{\sigma(2)}^{(2)}, \ldots, A_{\sigma(d)}^{(d)})$$

where $A^{(j)}$ denotes the jth column of the matrix A. We denote

$$D\left(A_1, d_1; A_2, d_2; \ldots; A_k, d_k\right) = D(A_1, \ldots, A_1, A_2, \ldots, A_2, \ldots, A_k, \ldots, A_k)$$

where A_i appears d_i times.

The following lemma is quite useful.

Lemma A.2 *Given $d \times d$ PSD matrices $A_1, \ldots, A_m$, the mixed discriminant is symmetric:*

$$D(A_1, \ldots, A_d) = D(A_{\sigma(1)}, \ldots, A_{\sigma(d)}) \tag{10}$$

for any permutation $\sigma \in S_d$. Also, the mixed discriminant is linear in each argument:

$$D(sA + tB, A_2, \ldots, A_d) = tD(A, A_2, \ldots, A_d) + sD(B, A_2, \ldots, A_d). \tag{11}$$

Moreover, $D(A_1, \ldots, A_d) \geq 0$ and $D(A, d) = \det(A)$.

Finally, the following polynomial identity follows from the above properties:

$$\det\left(\lambda_1 A_1 + \lambda_2 A_2 + \ldots \lambda_m A_m\right) = \sum_{i_1, \ldots, i_d = 1}^{m} \lambda_{i_1} \lambda_{i_2} \ldots \lambda_{i_d} D\left(A_{i_1}, \ldots, A_{i_d}\right). \tag{12}$$

Recall that, by equation (6), $\text{sym}_r(A)$ is the coefficient of t^{d-r} in the polynomial $\det(A+tI)$, or, equivalently, the coefficient of $s^r t^{d-r}$ in $\det(sA + tI)$. Then, by the identity (12),

$$\text{sym}_r(A) = \binom{d}{r} D(A, r; I, d - r). \tag{13}$$

It is easy to verify that $D(aa^T, 2; a_2 a_2^T, \ldots, a_{r-1} a_{r-1}^T; I, d - r) = 0$ for any sequence of vectors $a, a_2, \ldots, a_{r-1} \in \mathbb{R}^d$. Then, from equation (13) and the symmetry and linearity properties (10), (11) it follows that

$$
\begin{aligned}
\text{sym}_r\left(\sum_{i=1}^n a_i a_i^T\right) &= D\left(\sum_{i=1}^n a_i a_i^T, r; I, d - r\right) \\
&= \sum_{i_1, \ldots, i_r = 1}^n D(a_{i_1} a_{i_1}^T, \ldots, a_{i_r} a_{i_r}^T; I, d - r) \\
&= \sum_{i_1=1}^{n-r+1} \sum_{i_2=i_1+1}^{n-r+2} \cdots \sum_{i_r=i_{r-1}+1}^n r! D(a_{i_1} a_{i_1}^T, \ldots, a_{i_r} a_{i_r}^T; I, d - r)
\end{aligned}
$$

But, by an analogous calculation, each term on the right hand side is equal to $D(\sum_{i \in S} a_i a_i^T, r; I, d-r) = \text{sym}_r(\sum_{i \in S} a_i a_i^T)$ where $S = \{i_1, \ldots, i_r\}$. This proves equation (8).

The Aleksandrov-Fenchel inequality for mixed discriminants is equivalent to the following theorem.

Theorem A.3 *For any PSD matrices $A, B, A_{i+1}, \ldots, A_d$, the function $f(\alpha) = D((1-\alpha)A + \alpha B, i; A_{i+1}, \ldots, A_d)^{1/i}$ is concave on $0 \le t \le 1$.*

For more background and references about the Aleksandrov-Fenchel inequalities, see the book [Sch93, Sect 6.8.], and the survey [Gar02, Sect. 17]. A short proof of (a more general form of) Theorem A.3 using stable polynomials can be found in [Kho84].

Theorem 3.1 follows directly from Theorem A.3. By (13), Theorem A.3 and the AM-GM inequality, for any PSD matrices A, B we have

$$
\begin{aligned}
\text{sym}_r((1-\alpha)A + \alpha B)^{1/r} &= D((1-\alpha)A + \alpha B, r; I, d - r)^{1/r} \\
&\ge (1-\alpha)D(A, r; I, d - r)^{1/r} + \alpha D(B, r; I, d - r)^{1/r} \\
&\ge D(A, r; I, d - r)^{(1-\alpha)/r} D(B, r; I, d - r)^{\alpha/r} \\
&= \text{sym}_r(A)^{(1-\alpha)/r} \text{sym}_r(B)^{\alpha/r}.
\end{aligned}
$$

Taking logs on both sides and multiplying through by r gives Theorem 3.1.

B Omitted Proofs

Proof of Lemma 3.2: Let $w, y \in \mathbb{R}_+^U$ and $\alpha \in [0, 1]$ be arbitrary, and let $z \in \mathbb{R}_+^U$ be defined by $z_i = w_i^\alpha y_i^{1-\alpha}$. We need to prove $f(x, z) \le f(x, w)^\alpha f(x, y)^{1-\alpha}$. Using the formulas (7) and (8),

and applying Hölder's inequality, we have

$$
\begin{aligned}
f(x,z) &= \sum_{S \subseteq U : |S|=r} \mathrm{sym}_r \Big(\sum_{e \in S} x_e z_e v_e v_e^T \Big) \\
&= \sum_{S \subseteq U : |S|=r} z^S \mathrm{sym}_r \Big(\sum_{e \in S} x_e v_e v_e^T \Big). \\
&= \sum_{S \subseteq U : |S|=r} (w^S)^\alpha (y^S)^{1-\alpha} \mathrm{sym}_r \Big(\sum_{e \in S} x_e v_e v_e^T \Big) \\
&\leq \left(\sum_{S \subseteq U : |S|=r} w^S \mathrm{sym}_r \Big(\sum_{e \in S} x_e v_e v_e^T \Big) \right)^\alpha \left(\sum_{S \subseteq U : |S|=r} y^S \mathrm{sym}_r \Big(\sum_{e \in S} x_e v_e v_e^T \Big) \right)^{1-\alpha} \\
&= \left(\sum_{S \subseteq U : |S|=r} \mathrm{sym}_r \Big(\sum_{e \in S} x_e w_e v_e v_e^T \Big) \right)^\alpha \left(\sum_{S \subseteq U : |S|=r} \mathrm{sym}_r \Big(\sum_{e \in S} x_e y_e v_e v_e^T \Big) \right)^{1-\alpha} \\
&= f(x,w)^\alpha f(x,y)^{1-\alpha}.
\end{aligned}
$$

This completes the proof. □

Proof of Lemma 3.4: For any element $e \in P_i$, let $p_e = \frac{\bar{x}_e}{b_i}$ denote the probability with which e is included in the random sample. Then

$$
E[\det(L_{B,B})] = \sum_{S \in B(\mathcal{M})} \left(\prod_{i=1}^k b_i! \right) \det(L_{S,S}) \prod_{e \in S} p_e = \left(\prod_{i=1}^k \frac{b_i!}{b_i^{b_i}} \right) \sum_{S \in B(\mathcal{M})} \bar{x}^S \det(L_{S,S})
$$

From (7) it follows that that for any set S of size r, $\det(L_{S,S}) = \mathrm{sym}_r(\sum_{e \in S} v_e v_e^T)$, as $L_{S,S}$ is the Gram matrix of the vectors $\{v_e\}_{e \in S}$. Now, formula (8) shows that

$$
p(y) = \mathrm{sym}_r \Big(\sum_{e \in U} \bar{x}_e y_e v_e v_e^T \Big) = \sum_{S \subseteq U : |S|=r} \bar{x}^S y^S \mathrm{sym}_r \Big(\sum_{e \in S} v_e v_e^T \Big) = \sum_{S \subseteq U : |S|=d} y^S \bar{x}^S \det(L_{S,S}).
$$

Thus we have that the coefficient of y^S in p is $c_S = \bar{x}^S \det(L_{S,S})$, giving us the lemma. □