A deterministic polynomial-time algorithm for approximating mixed discriminant and mixed volume

Leonid Gurvits
NECI, Princeton.
Work partly done at Technion, Haifa, Israel.
gurvits@ias.edu

Alex Samorodnitsky
Institute for Advanced Study and DIMACS.
asamor@ias.edu

ABSTRACT

We present a deterministic polynomial algorithm that computes the mixed discriminant of an n-tuple of positive semidefinite matrices to within a multiplicative factor of e^n.
To this end we extend the notion of doubly stochastic matrix scaling to a larger class of n-tuples of positive semidefinite matrices, and provide a polynomial-time algorithm for this scaling.
We obtain tight upper and lower bounds on the mixed discriminant of doubly stochasic n-tuples, proving a conjecture of Bapat, and generalizing the van der Waerden - Falikman - Egorychev theorem.
As a corollary, we obtain a deterministic polynomial algorithm that computes the mixed volume of n convex bodies in $\mathbf{R}^n$ to within a multiplicative factor of $n^{O(n)}$. This answers a question of Dyer, Gritzmann and Hufnagel.

1. INTRODUCTION

1.1 Mixed discriminant, permanent and mixed volume

Permanent:
Let $A = (a_{ij})$ be an $n \times n$ matrix. The number

$$per(A) = \sum_{\sigma \in S_n} \prod_{i=1}^{n} a_{i\sigma(i)},$$

where S_n is the symmetric group on n elements, is called the permanent of A. For a $0,1$ matrix A, $per(A)$ counts the number of perfect matchings in G, the bipartite graph represented by A.
It is $\#P$-hard to compute the permanent of a nonnegative (even $0,1$) matrix [22], and so it is unlikely to be efficiently computable exactly for all matrices. The realistic goal, then, is to try and *approximate* the permanent efficiently as well as possible, for large classes of matrices.
Efficient poly-time probabilistic algorithms that approximate the permanent extremely tightly ($(1+\epsilon)$-factor) were developed for several classes of graphs [15], and others.
How well can the permanent be approximated in polynomial time? For arbitrary $0,1$ matrices, a poly-time probabilistic algorithm giving a $2^{O(n)}$-factor approximation was obtained in [17]. The first efficient probabilistic algorithm that provides a $2^{O(n)}$-factor approximation for *arbitrary positive matrices* was obtained by Barvinok in [5], [6].
A deterministic strongly polynomial algorithm achieving $2^{O(n)}$-factor approximation for arbitrary positive matrices was constructed in [18]. The algorithm uses *matrix scaling* to reduce the problem to estimation of the permanent of a doubly stochastic matrix. For these matrices the permanent is known to lie in the interval $[\frac{n!}{n^n}, 1]$, and this solves the approximation problem. We recall, that the lower bound of $\frac{n!}{n^n}$ on the permanent of a doubly stochastic matrix was conjectured by van der Waerden and proven by Egorychev [10] and Falikman[11] fifty years later. (A slightly weaker, but sufficient for the purposes of [18], bound of e^{-n} was proven by Friedland [12]).
Mixed volume
Let $K_1...K_n$ be convex bodies in the Euclidean space $\mathbf{R}^n$, and let $V(\cdot)$ be the Euclidean volume in $\mathbf{R}^n$. It is well known (see for instance [21]), that the value of $V(\lambda_1 K_1 + \cdots \lambda_n K_n)$ is a homogeneous polynomial of degree n in nonnegative variables $\lambda_1...\lambda_n$, where $''+''$ denotes Minkowski sum, and λK denotes the dilatation of K with coefficient λ. The coefficient $V(K_1...K_n)$ of $\lambda_1 \cdot \lambda_2 \cdots \lambda_n$ is called the *mixed volume* of $K_1...K_n$. Alternatively,

$$V(K_1...K_n) = \frac{\partial^n}{\partial\lambda_1...\partial\lambda_n} V(\lambda_1 K_1 + \cdots \lambda_n K_n).$$

The problem of computing the mixed volume of convex bodies is important for combinatorics and algebraic geometry [8]. For instance, the number of toric solutions to a generic system of n polynomial equations on $\mathbf{C}^n$ is equal to the mixed volume of the Newton polytopes of the equations.
This problem is also $\#P$-complete, since volume is a special case of the mixed volume, and computing the volume is $\#P$-complete [7]. Therefore the reasonable goal, once again, is to seek approximate solutions.
Efficient poly-time probabilistic algorithms that approximate the mixed volume extremely tightly ($(1+\epsilon)$-factor) were developed for some classes of well-presented convex bodies [8].
How well can the mixed volume be approximated in polynomial time? The first efficient probabilistic algorithm that provides a $n^{O(n)}$-factor approximation for *arbitrary well-presented proper convex bodies* was obtained by Barvinok in [5].
The question of existence of an efficient *deterministic* algo-

rithm for approximating mixed volume of arbitrary well-presented proper convex bodies with an error depending only on the dimension was posed in [8]. They quote a lower bound [3] of $\left(\frac{cn}{\log n}\right)^{\frac{n}{2}}$ for the approximation factor of such an algorithm. (For some constant c).
A deterministic polynomial time algorithm that approximates the mixed volume with a factor of $n^{O(n)}$ was given, for a certain class of proper convex bodies in [5].

Mixed discriminant

Let $A_1...A_n$ be $n \times n$ real symmetric matrices. It is well known (and easily seen) that the value of $\det(x_1A_1 + \cdots x_nA_n)$ is a homogeneous polynomial of degree n in nonnegative variables $x_1...x_n$. The number

$$D(A_1...A_n) = \frac{\partial^n}{\partial x_1...\partial x_n} \det(x_1A_1 + \cdots x_nA_n)$$

is called the *mixed discriminant* of $A_1...A_n$.
We will be interested in the special case where the matrices $A_1...A_n$ are positive semidefinite. In this case the mixed discriminant $D(A_1...A_n)$ is known [4] to be nonnegative.
Mixed discriminants generalize permanents: If the matrices $A_1...A_n$ are diagonal, namely $A_j = diag(b_{1j}...b_{nj})$, for $j = 1...n$, let $B = (b_{ij})$. Then $per(B) = D(A_1...A_n)$. It follows that the computation of a mixed discriminant of n positive semidefinite matrices is $\#P$-hard, since it is at least as hard as computing the permanent of a nonnegative matrix.
A positive definite $n\times n$ matrix A defines an ellipsoid in $\mathbf{R}^n$, by setting $\mathcal{E}_A = \{x \in \mathbf{R}^n \ : \ \langle x,\ Ax\rangle \leq 1\}$. The following relation between the mixed discriminant of positive definite matrices and mixed volume of ellipsoids was established in [5]:

$$3^{-\frac{n+1}{2}} v_n \leq \frac{V(\mathcal{E}_{A_1}...\mathcal{E}_{A_n})}{D^{\frac{1}{2}}(A_1^{-1}...A_n^{-1})} \leq v_n. \tag{1}$$

Here v_n is the volume of the unit ball in $\mathbf{R}^n$.
Recall, that for any convex body K in $\mathbf{R}^n$ there exists [16] an ellipsoid $\mathcal{E}_k$, such that (after translating it's center to the origin): $\mathcal{E}_k \subseteq K \subseteq n\mathcal{E}_k$. For a well-presented convex body K, an ellipsoid $\mathcal{E}'_k$ such that $\mathcal{E}'_k \subseteq K \subseteq n\sqrt{n+1}\mathcal{E}'_k$ can be constructed efficiently [13].
Barvinok [5], [6] gives efficient poly-time probabilistic algorithm for approximating mixed discriminant of n positive semidefinite matrices with a $2^{O(n)}$-factor. Using the above relations between mixed discriminant and mixed volume of ellipsoids, and the fact that every well-presented convex body can be efficiently approximated by an ellipsoid, he obtains his permanent and mixed volume approximation results.
Apart from their connection to permanents and mixed volumes, mixed discriminants also have some independent applications to computationally hard problems of combinatorial counting such as counting the 'coloured spanning trees' [6].

1.2 Our results

We achieve $O(e^n)$-factor polynomial-time approximation *deterministically*.

THEOREM 1. *There is a function f, such that*

$$D(A_1...A_n) \leq f(A_1...A_n) \leq O(e^n)D(A_1...A_n)$$

holds on every n-tuple of positive semidefinite $n\times n$ matrices A_i. The function f is computable in time polynomial in n and $\log \nu$, where ν is the maximal binary representation length of the entries of $A_1...A_n$.

Similarly to [5], we obtain mixed volume approximation results, using theorem 1, (1) and efficient approximation of convex bodies by ellipsoids.

THEOREM 2. *There is a function g, such that*

$$V(K_1...K_n) \leq g(K_1...K_n) \leq n^{O(n)}V(K_1...K_n)$$

holds on every n-tuple of proper well-presented convex bodies K_i in $\mathbf{R}^n$. The function g is computable in time polynomial in n and the presentation size of the bodies.

Our approach to this problem follows the approach of [18]. In short, we reduce the question of approximating mixed discriminants of n-tuples to that of approximating mixed discriminants on a smaller class of *doubly stochastic* n-tuples. The reduction technique is *n-tuple scaling*. We then prove combinatorial upper and lower bounds on the mixed discriminant of doubly stochastic n-tuples, and use them to obtain the desired approximation.

DEFINITION 3. *Let $\mathbf{A} = (A_1...A_n)$ and $\mathbf{B} = (B_1...B_n)$ be two n-tuples of $n \times n$ matrices. The tuple $\mathbf{B}$ is a scaling of $\mathbf{A}$, if there exist a vector $x \in \mathbf{R}^n$ and two $n \times n$ matrices T_1, T_2, such that $B_i = x_iT_1A_iT_2$, for all $i = 1...n$.*

The important (for us) property of scaling is, that we know, how it changes the mixed discriminant.

LEMMA 4. *In the notation of the preceding definition*

$$D(\mathbf{B}) = \prod_{i=1}^{n} x_i \cdot \det T_1 \cdot \det T_2 \cdot D(\mathbf{A}).$$

PROOF. The claim is a straightforward consequence of the definition of mixed discriminant, and the multiplicative property of the determinant: $\det(AB) = \det(A)\det(B)$. $\square$

DEFINITION 5. *An n-tuple $\mathbf{A} = (A_1...A_n)$ of positive semidefinite matrices is doubly stochastic (d.s.) if*

$$\forall i \quad Tr(A_i) = 1 \quad \textit{and} \quad \sum_i A_i = I, \tag{2}$$

here (and from now on) I is the identity matrix.

DEFINITION 6. *Let $\mathbf{A} = (A_1...A_n)$ be an n-tuple of $n\times n$ positive semidefinite matrices. A positive vector $x \in \mathbf{R}^n$, and a positive definite $n\times n$ matrix S are scaling factors of $\mathbf{A}$, if the n-tuple $\mathbf{B} = (B_1, ..B_n)$, given by $B_i = x_iS^{\frac{1}{2}}A_iS^{\frac{1}{2}}$ is doubly stochastic* [1].

The preceding description presents a very "small scale" overview of things. In the next subsection we go into details.

[1] Here $S^{\frac{1}{2}}$ is the unique positive semidefinite matrix whose square is S.

1.3 An overview of the mixed discriminant approximation algorithm

- We define a notion of a *fully indecomposable* tuple

  Definition 7. *An n-tuple* $\mathbf{A} = (A_1, ...A_n)$ *of positive semidefinite* $n \times n$ *matrices is fully indecomposable if for all* $S \subseteq \{1, ..., n\}$, $0 < |S| < n$, $Rank(\sum_{i \in S} A_i) > |S|$.

  and show a reduction of the problem to the case of fully indecomposable tuples. This is done in section 2.

- We show that the problem of existence and computation of scaling factors for an indecomposable tuple can be translated to determining whether an explicitly given convex function obtains a minimum over a specific convex set, and to finding this minimum. We deduce existence (and uniqueness) of scaling factors for an indecomposable tuple. This is done in section 3.

- We give an approximate solution of the forementioned convex optimization problem using the Ellipsoid method. This, together with lemma 4, gives a reduction of the problem to the case of doubly stochastic tuples. This is done in section 4.

- We prove tight upper and lower bounds on the mixed discriminant of doubly stochastic tuples. This is done in section 5.

We prove four theorems, corresponding to the four clauses above.

Theorem 8. *Let* $\mathbf{A} = (A_1, \ldots, A_n)$ *be an n-tuple of positive semidefinite matrices with positive mixed discriminant. Then there is an integer* $1 \leq k \leq n$, *and fully indecomposable tuples* $\mathbf{B}_1 ... \mathbf{B}_k$ *of positive semidefinite matrices, such that*

$$D(\mathbf{A}) = \prod_{s=1}^{k} D(\mathbf{B}_s).$$

The tuples $\mathbf{B}_1 ... \mathbf{B}_k$ *can be found in polynomial time.*

Theorem 9. *Let* $\mathbf{A} = (A_1, ...A_n)$ *be a fully indecomposable n-tuple of positive semidefinite matrices. Then:*
(1) There exist scaling factors x *and* S *such that*
$\mathbf{B} = (x_1 S^{\frac{1}{2}} A_1 S^{\frac{1}{2}}, \ldots, x_n S^{\frac{1}{2}} A_n S^{\frac{1}{2}})$ *is doubly stochastic.*
(2) Let there be two pairs of scaling factors (x, S) *and* (x', S') *for* $\mathbf{B}$, *and assume a normalization* $\prod_{i=1}^{n} x_i = \prod_{i=1}^{n} x'_i = 1$. *Then* $x_i = x'_i$ *for all* $1 \leq i \leq n$ *and* $S' = S$.

Definition 10. *Let* $\mathbf{A} = (A_1 ... A_n)$ *be an n-tuple of* $n \times n$ *positive semidefinite matrices. A positive vector* $x \in \mathbf{R}^n$, *and a positive definite* $n \times n$ *matrix* S *are* ϵ*-scaling factors of* $\mathbf{A}$, *if the n-tuple* $\mathbf{B} = (B_1, ..B_n)$, *given by* $B_i = x_i S^{\frac{1}{2}} A_i S^{\frac{1}{2}}$ *is* ϵ*-doubly stochastic, namely*

$$\sum_{i=1}^{n} (Tr(B_i) - 1)^2 \leq \epsilon^2, \tag{3}$$

and

$$\sum_{i=1}^{n} B_i = I. \tag{4}$$

Theorem 11. *Let* $\mathbf{A} = (A_1, ...A_n)$ *be a fully indecomposable n-tuple of positive semidefinite* $n \times n$ *matrices. Let* $\epsilon > 0$ *be a required scaling accuracy. Then* ϵ*-scaling factors* $x'_1 ... x'_n$ *and* S' *for* $\mathbf{A}$ *can be found in*

$$O\left(n^5 \log\left(\frac{n\nu}{\epsilon}\right)\right)$$

arithmetic operations. Here ν *is the maximal binary representation length of the entries in* $A_1 ... A_n$. *Moreover, if* $x_1 ... x_n$ *and* S *are the proper scaling factors for* $\mathbf{A}$, *then*

$$\prod_{i=1}^{n} x_i \det S \leq \prod_{i=1}^{n} x'_i \det S' \leq (1 + \epsilon^2) \prod_{i=1}^{n} x_i \det S.$$

Theorem 12. *Let* $\mathbf{A} = (A_1, ...A_n)$ *be a doubly stochastic n-tuple of positive semidefinite* $n \times n$ *matrices. Then*

$$\frac{n!}{n^n} \leq D(\mathbf{A}) \leq 1. \tag{5}$$

Observe, that since the class of doubly stochastic n-tuples is a generalization of the class of doubly stochastic matrices, theorem 12 proves a *generalization of the van der Waerden conjecture* to this class. This generalization was posed as a conjecture by Bapat [4] in 1989.
Theorem 1 is a result of a combination of theorems 8 through 12.

2. REDUCTION TO THE FULLY INDECOMPOSABLE CASE

The main result in this section is theorem 8. Before that, we point out that one can check in polynomial time, whether the mixed discriminant of a given n-tuple of positive semidefinite matrices is 0. (Recall, that it is always nonnegative).

Lemma 13. *Let* $\mathbf{A}$ *be an n-tuple of positive semidefinite matrices. There is a polynomial-time algorithm which decides whether* $D(\mathbf{A}) = 0$ *or* $D(\mathbf{A}) > 0$.

Proof. We follow the argument of [8] (Theorem 8), who solved a similar problem of determining whether a mixed volume of n convex well-presented bodies is zero.
We need a following property of the mixed discriminant:

Theorem 14. *[20] Let* $\mathbf{A} = (A_1, .., A_n)$ *be an n-tuple of positive semidefinite matrices. Then the following two conditions are equivalent:*

1. $D(\mathbf{A}) > 0$, *and*
2. *For all* $S \subseteq \{1, ..., n\}$, $0 < |S| < n$,
   $Rank(\sum_{i \in S} A_i) \geq |S|$.

Now, let for $i = 1, ..., n$ the matrix T_i be the positive semidefinite square root of A_i. Let $E_i = \{t_i^1, ..., t_i^n\}$ be the set of columns of T_i. Consider two matroids on the ground set $E = \cup_{i=1}^{n} E_i = \{t_j^i\}_{i,j=1...n}$. The first is the linear matroid in which the independent sets are the linear independent subsets of E. The second is the transversal matroid, the bases of which are the transversals of the family $\{E_1, ...E_n\}$. Observe, that, by the forementioned property of the mixed discriminant, $D(\mathbf{A}) > 0$ if and only if the two matroids share a common base. To determine this, we have to solve a 2-matroid intersection problem. The complexity of this problem is known [9] to be polynomial, and we are done. □

Theorem 8 is a consequence of the following proposition:

PROPOSITION 15. *Let $\mathbf{A} = (A_1, \ldots, A_n)$ be an n-tuple in $\mathcal{N}_n$ with positive mixed discriminant. Then there is a partition $C_1 \cup \ldots \cup C_k$ of $[n]$ such that*

1. *For all $1 \leq s \leq k$ holds rank $\sum_{i \in C_s} A_i = |C_s| := c_s$.*
2. *The subspaces $X_s = Im\left(\sum_{i \in C_s} A_i\right)$ decompose $\mathbf{R}^n$ into a direct orthogonal sum $\mathbf{R}^n = X_1 \oplus X_2 \oplus \ldots \oplus X_k$.*
3. *For $1 \leq s \leq k$, let $\{u_i\}_{i \in C_s}$ be an orthonormal basis of X_s. For $i \in C_s$ let a $c_s \times c_s$ matrix B_i be a projection of A_i onto X_s, namely $B_i = U_s^t A_i U_s$, where U_s is a $n \times c_s$ matrix with columns $\{u_i\}_{i \in C_s}$. Then the c_s-tuple $\mathbf{B}_s = (B_i)_{i \in C_s}$ is fully indecomposable.*
4. $$D(\mathbf{A}) = \prod_{s=1}^{k} D(\mathbf{B}_s).$$

The partition $[n] = C_1 \cup \ldots \cup C_k$, and the matrices $B_1, \ldots, B_n$ can be found in polynomial time.

PROOF. (Of the proposition)
Let $\mathbf{A}$ be a decomposable n-tuple with positive mixed discriminant. In this case we have the following

LEMMA 16. *There exists $\alpha \subsetneq [n]$, with rank $\sum_{i \in \alpha} A_i = |C_\alpha|$.*

PROOF. The claim follows immediately from decomposability of $\mathbf{A}$ and theorem 14. $\square$

Our first step is to find a minimal non-empty set α with this property, or to decide that $\mathbf{A}$ is indecomposable, in which case we are done. For this purpose we consider $n(n-1)$ auxiliary n-tuples $\mathbf{A}^{ij}$, where $\mathbf{A}^{ij}$ is obtained from $\mathbf{A}$ by substituting A_i instead of A_j. Let $D_{ij} = D(\mathbf{A}^{ij})$. We define an $n \times n$ matrix Z by $Z_{ij} = 0$ if $D_{ij} = 0$, and $Z_{ij} = 1$ otherwise. By lemma 13 the matrix Z is constructible in polynomial time. The next lemma explains how this matrix highlights the sets we are looking for.

LEMMA 17. *Let $\emptyset \neq \alpha \subseteq [n]$, with rank $\left(\sum_{i \in \alpha} A_i\right) = |\alpha|$, and assume that α contains no proper non-empty subsets with this property. Then $Z_{ij} = 1$ for all $i, j \in \alpha$ and $Z_{ij} = 0$ for all $i \in \alpha$ and $j \notin \alpha$.*

The proof of the lemma is essentially the same as that of lemma 21. We refer to the forthcoming proof of that lemma.

Consider now Z as an adjacency matrix of a directed graph $G = ([n], E)$, where $e = i \to j$ belongs to E iff $Z_{ij} = 1$. For $i \in [n]$, let V_i be the set of points which can be reached from i by the edges of G. Lemma 17 implies that α is a minimal set with the property rank $\left(\sum_{i \in \alpha} A_i\right) = |\alpha|$ iff for any $i \in \alpha$ holds $V_i = \alpha$, and moreover, α is a clique of G. We compute the sets V_i for all $i \in [n]$ and check whether V_i is a clique. If it holds for some $i \in [n]$, and $V_i \neq G$, we set $\alpha = V_i$. Otherwise $\mathbf{A}$ is indecomposable and we are done.
We set $C_1 := \alpha$ to be the first cell in our partition of $[n]$. Let $S = \sum_{i \in C_1} A_i$, and let $v_1, \ldots, v_{c_1}$ be independent columns of S, spanning X_1. Completing to a basis of $\mathbf{R}^n$ and orthogonalizing, we obtain an orthonormal basis $u_1, \ldots, u_n$ of $\mathbf{R}^n$ with $\mathrm{Span}(u_1, .., u_{c_1}) = \mathrm{Span}(v_1, .., v_{c_1})$. Let U_1 be the $n \times c_1$ matrix with columns $u_1, \ldots u_{c_1}$ and $U_1^\perp$ be the $n \times (n - c_1)$ matrix with columns $u_{c_1+1}, \ldots u_n$. We set $B_i = U_1^t A_i U_1$ for $i \in C_1$, and $\tilde{A}_j = (U_1^\perp)^t A_j U_1^\perp$, for $j \notin C_1$. Then,

LEMMA 18.
$$D(A_1, \ldots A_n) = D((B_i)_{i \in C_1}) \cdot D((\tilde{A}_j)_{j \notin C_1}).$$

PROOF. We will use the following, alternative definition of the mixed discriminant [4]

DEFINITION 19. *Given $A_1, \ldots A_n$ and a permutation $\sigma \in S_n$, let A_σ be the $n \times n$ matrix whose i'th column is the i'th column of $A_{\sigma(i)}$. Then:*
$$D(A_1, \ldots A_n) = \sum_{\sigma \in S_n} \det(A_\sigma).$$

Let U be the orthogonal matrix with columns $u_1 \ldots u_n$. By lemma 4, we have $D(A_1 \ldots A_n) = D(U^t A_1 U \ldots U^t A_n U)$. Observe, that for $i \in C_1$, the matrix $U^t A_i U$ is zero everywhere, but on a $c_1 \times c_1$ upper left submatrix, and this submatrix is B_i. Observe also, that for $j \notin C_1$, the lower right $(n - c_1) \times (n - c_1)$ submatrix of $U^t A_j U$ is precisely $\tilde{A}_j$.
Using definition 19 of the mixed discriminant, it is not hard to see, that in this case $D(U^t A_1 U \ldots U^t A_n U) = D((B_i)_{i \in C_1}) \cdot D((\tilde{A}_j)_{j \notin C_1})$. $\square$

Now we inductively (on dimension) apply the claim of the proposition to the $(n - c_1)$-tuple $(\tilde{A}_j)_{j \notin C_1}$. We point out, that minimality of α implies indecomposability of the tuple $(B_i)_{i \in C_1}$.
It remains to estimate the cost that we pay for decomposing the n-tuple $\mathbf{A}$. We perform $O(n)$ steps in which the consecutive cells $C_1, C_2 \ldots$ of partition are determined. In each of these steps the heaviest part by far is constructing the $0-1$ matrix Z, which entails checking $\Omega(n^2)$ mixed discriminants for being zero. The total cost is, therefore, polynomial, and we are done.

The only thing which has yet to be pointed out, is that the representation length ν_i of the components $\mathbf{B}_i$ can not be much greater than the representation length ν of $\mathbf{A}$. Indeed, in the course of the decomposition procedure of proposition 15, the representation length increases, essentially, only in the orthogonalization procedure. A moment's reflection gives that the multiplicative factor of the increase is at at most n^n.

3. CONVEX MINIMIZATION PROBLEM. EXISTENCE AND UNIQUENESS OF SCALING FACTORS

Given a fully indecomposable n-tuple of positive semidefinite matrices, we will now define a convex function, whose minima will correspond to the scaling factors of the tuple.

DEFINITION 20. *Let $\mathbf{A} = (A_1 \ldots A_n)$ be a fully indecomposable n-tuple of positive semidefinite matrices. We define*
$$f(\xi_1, \ldots \xi_n) = f_{\mathbf{A}}(\xi_1, \ldots \xi_n) = \log\det\left(e^{\xi_1} A_1 + \ldots + e^{\xi_n} A_n\right).$$

Before we describe the properties of f and its minima, we need a technical lemma and a definition which attempt to "quantify" the indecomposability of $\mathbf{A}$. Consider the $n(n-1)$ n-tuples $\mathbf{A}^{ij}$, where $\mathbf{A}^{ij}$ is obtained from $\mathbf{A}$ by substituting A_i instead of A_j. Let $D_{ij} = D(\mathbf{A}^{ij})$.

LEMMA 21. *$\mathbf{A}$ is indecomposable if and only if $D_{ij} > 0$, for all i, j.*

Proof. Assume first that $\mathbf{A} = (A_1, \dots A_n)$ is indecomposable. We claim, that for any $1 \le i \ne j \le n$ the tuple $\mathbf{A}^{ij} = (A'_1, \dots, A'_n)$ satisfies property 2 of theorem 14, and therefore it's mixed determinant is positive. Indeed, let $R \subseteq \{1, \dots, n\}$. Then $rank(\sum_{k\in R} A'_k) \ge rank(\sum_{k\in R\setminus\{j\}} A_k) \ge |R|$, by indecomposability of $\mathbf{A}$.
In the other direction, let $\mathbf{A} = (A_1, \dots A_n)$ be a decomposable tuple, namely for some subset $R \subset \{1, \dots, n\}$, the inequality $rank(\sum_{k\in R} A_k) \le |R|$ holds. Let i, j be a pair of indices such that $i \in R$ and $j \notin R$, and consider the tuple $\mathbf{A}^{ij} = (A'_1, \dots, A'_n)$. We claim that $D(\mathbf{A}^{ij}) = 0$. Indeed,

$$rank(\sum_{k\in R\cup\{j\}} A'_k) = rank(\sum_{k\in R} A_k) < |R| + 1 = |R \cup \{j\}|$$

which, by theorem 14 implies $D(\mathbf{A}^{ij}) = 0$, and we are done. $\square$

Definition 22. *Set*

$$M = M_{\mathbf{A}} = \min_{i\ne j} D_{ij}.$$

Lemma 23. *The function f is a convex function on $\mathbf{R}^n$, and is strictly convex* [2] *on the hyperplane* $H = \{\xi = (\xi_1, ..., \xi_n) \in \mathbf{R}^n \mid \sum \xi_i = 0\}$.

Proof. Let us denote the set of n-tuples $r = r_1, ... r_n$ of nonnegative integers summing to n by P_n. Recall (11) that for any $r \in P_n$ the coefficient t_r of $x_1^{r_1} ... x_n^{r_n}$ in $\det(x_1 A_1 + ... + x_n A_n)$ is given by

$$t_r = \frac{1}{r_1! ... r_n!} D\left(\underbrace{A_1 ... A_1}_{r_1} \dots \underbrace{A_n .. A_n}_{r_n}\right).$$

Therefore

$$f(\xi_1, ... \xi_n) = \log\det\left(e^{\xi_1} A_1 + ... + e^{\xi_n} A_n\right) =$$

$$= \log \sum_{r\in P_n} t_r e^{<\xi, r>},$$

where the coefficients t_r are nonnegative since they are given by a mixed discriminant of positive semidefinite matrices multiplied by a multinomial coefficient, and $<,>$ is the usual inner product in $\mathbf{R}^n$. It is well known [2] that the R.H.S. represents a convex function of ξ. Nonetheless, we provide another proof of this fact, which will also imply that f is strictly convex on H, if the tuple $A_1 ... A_n$ is indecomposable. Let $g(\xi) = \det\left(e^{\xi_1} A_1 + ... + e^{\xi_n} A_n\right)$, namely $f = \log g$. To show that f is convex we have to show that it's Hessian $\bigtriangledown^2 f$ is positive semidefinite. Clearly $\bigtriangledown^2 f = \frac{1}{g^2}(g \bigtriangledown^2 g - (\bigtriangledown g)(\bigtriangledown g)^t)$. Therefore, we have to show $g \bigtriangledown^2 g \ge (\bigtriangledown g)(\bigtriangledown g)^t$, where the inequality is in the cone of positive semidefinite matrices. For a vector $v \in \mathbf{R}^n$, let $v \otimes v$ denote the one dimensional projection vv^t on the span of v, which, by definition, is positive semidefinite. Observe that for any $v \in \mathbf{R}^n$ holds $\bigtriangledown e^{<\xi, v>} = e^{<\xi, v>} \cdot v$ and $\bigtriangledown^2 e^{<\xi, v>} = e^{<\xi, v>} \cdot v \otimes v$, and therefore

$$g \bigtriangledown^2 g - (\bigtriangledown g)(\bigtriangledown g)^t =$$

$$\sum_{r\in P_n} t_r e^{<\xi, r>} \cdot \sum_{s\in P_n} t_s e^{<\xi, s>} s \otimes s - \sum_{r,s\in P_n} t_r t_s e^{<\xi, r+s>} r \otimes s =$$

$$\frac{1}{2} \sum_{r,s\in P_n} t_r t_s e^{<\xi, r+s>} (r - s) \otimes (r - s) \ge 0, \qquad (6)$$

implying the convexity of f.
Now, assume the tuple $A_1, ..., A_n$ to be indecomposable. Set, once again, D_{ij} to be the mixed discriminant of the n-tuple obtained from $(A_1, ... A_n)$ by replacing A_j with A_i. We know by lemma 21 that $D_{ij} > 0$. Actually in the notation of this lemma, D_{ij} is just $2t_{r_{ij}}$, where r_{ij} is the vector with 1 in every coordinate but i, j, and with 2 in the i'th and 0 in the j'th coordinates. Recall (definition 22) the notation $M = min_{i\ne j} D_{ij}$. We now continue the computation from (6) to obtain:

$$\bigtriangledown^2 f \ge \frac{1}{2g^2} \sum_{r,s\in P_n} t_r t_s e^{<\xi, r+s>} (r - s) \otimes (r - s) \ge$$

$$\frac{1}{8g^2} \sum_{i\ne j, k\ne l} D_{ij} D_{kl} e^{<\xi, r_{ij} + r_{kl}>} (r_{ij} - r_{kl}) \otimes (r_{ij} - r_{kl})$$

$$\ge \frac{M^2}{8g^2} \sum_{i\ne j\ne k\ne l} (r_{ij} - r_{kl}) \otimes (r_{ij} - r_{kl}),$$

where the last summation is over distinct indices i, j, k, l. Let e_i be the i'th unit vector, and let E_{ij} be the $n \times n$ matrix with 1 in the (i, j)'th coordinate and zero everywhere else. Then

$$(r_{ij} - r_{kl}) \otimes (r_{ij} - r_{kl}) = (e_i - e_j - e_k + e_l) \otimes (e_i - e_j - e_k + e_l) =$$

$$E_{ii} + E_{jj} + E_{kk} + E_{ll} + 2\left(E_{il} + E_{jk} - E_{ij} - E_{ik} - E_{jl} - E_{kl}\right).$$

Let

$$S = \sum_{i\ne j\ne k\ne l} (r_{ij} - r_{kl}) \otimes (r_{ij} - r_{kl}).$$

Finding S might seem like a mess, but actually it is not. By symmetry considerations, the entries of S have only two distinct values, on the main diagonal and every else, and moreover they sum to zero. Therefore we have only to find the trace of S and it's easily seen to be $(n)_4 = n(n-1)(n-2)(n-3)$. Accordingly, $S_{ii} = (n-1)_3 = (n-1)(n-2)(n-3)$ and $S_{ij} = (n-2)(n-3)$ for $i \ne j$. Note that S is simply the orthogonal projection on the hyperplane $H = \{\xi = (\xi_1, ..., \xi_n) \in \mathbf{R}^n \mid \sum \xi_i = 0\}$, multiplied by $(n-1)_3$. Therefore, the projection of $\bigtriangledown^2 f$ onto H is greater or equal to $\frac{M^2 \cdot (n-1)_3}{8g^2} \cdot I_{n-1}$, where I_{n-1} is a $(n-1) \times (n-1)$ identity matrix, implying f is strictly convex on H. $\square$

Lemma 24. *For any $\xi \in \mathbf{R}^n$, the gradient $(\bigtriangledown f)_\xi$ of f at ξ is $\left(Tr(e^{\xi_i} S^{\frac{1}{2}} A_i S^{\frac{1}{2}})\right)_i$, where $S = \left(\sum_{i=1}^n e^{\xi_i} A_i\right)^{-1}$.*

The proof is a simple limit computation, which we omit.

Lemma 25. *A point ξ^* is a minimum of f on H if and only if the gradient $\bigtriangledown f$ at ξ^* is a constant multiple of $(1...1)$, the vector of all ones.*
$(\bigtriangledown f)_{\xi^*} = c \cdot (1...1)$ if and only if $x_i = e^{\xi_i^*}$, $i = 1...n$, and $S = \left(\sum_{i=1}^n e^{\xi_i^*} A_i\right)^{-1}$, are the scaling factors for $\mathbf{A}$.

[2] Namely, for all ξ, ξ', and $0 < \lambda < 1$ holds $f(\lambda\xi + (1-\lambda)\xi') < \lambda f(\xi) + (1-\lambda) f(\xi')$.

PROOF. The first claim: Since f is convex, ξ^* is a point of minimum for f on H if and only if $(\bigtriangledown f)_{\xi^*}$ is a convex combination of the gradients at ξ^* of the defining equations for H.
The second claim follows immediately from the first claim and lemma 24. $\square$

The value of $M = \min_{i\neq j} D_{ij}$ plays a role in the following lemma as well.

LEMMA 26. *Let $\xi \in H$ be such that $f(\xi) \le f(0)$, then*

$$\|\xi\|_2 \le n^{\frac{1}{2}} \cdot \log \frac{2\det(A_1 + ... + A_n)}{M},$$

PROOF. Let ξ be a point in H with $f(\xi) \le f(0) = \log\det(A_1 + ... + A_n)$. Then, in the notation of the proof of lemma 23:

$$\det(A_1 + ... + A_n) \ge \det\left(e^{\xi_1}A_1 + ... + e^{\xi_n}A_n\right) =$$

$$\sum_{r\in P_n} t_r e^{<\xi,r>} \ge \frac{1}{2}\sum_{i\neq j} D_{ij}e^{<\xi,r_{ij}>} \ge \frac{1}{2}M\sum_{i\neq j} e^{<\xi,r_{ij}>} \ge$$

$$\frac{1}{2}Me^{\max_{i\neq j}\xi_i-\xi_j}.$$

The last inequality uses the fact that since $\sum_i \xi_i = 0$, we have $<\xi, r_{ij}>= \xi_i - \xi_j$.
Therefore

$$\|\xi\|_2 \le n^{\frac{1}{2}} \cdot \|\xi\|_\infty \le n^{\frac{1}{2}} \cdot \log \frac{2\det(A_1 + ... + A_n)}{M}.$$

$\square$

Theorem 9 is a simple consequence of lemma 25 and the following lemma which describes the behaviour of minima of f on H.

LEMMA 27. *The function f attains a unique minimum on the hyperplane H.*

PROOF. By lemma 23, f is strictly convex on H. Therefore, the minimum ξ^*, if attained, is unique. On the other hand, by lemma 26, the minimum of f on H is the minimum of f on a ball with finite radius. Since this ball is compact, the minimum is attained. $\square$

REMARK 28. *We observe that, in the notation of lemma 25,*

$$f(\xi^*) = \log\det\left(e^{\xi_1^*}A_1 + \cdots + e^{\xi_n^*}A_n\right) =$$

$$\log\det(S) = \log\left(\det(S)\cdot\prod_{i=1}^n x_i\right)$$

is the product of the scaling factors of $\mathbf{A}$.

4. FINDING THE MINIMUM

In the previous section we have seen, that finding the point of minimum of the function $f = f_{\mathbf{A}}$ on the hyperplane H is equivalent to computing the scaling factors of $\mathbf{A}$. This is interesting if we want to scale $\mathbf{A}$. We have also seen, that finding the value of the minimum is equivalent to computing the product of the scaling factors of $\mathbf{A}$. This is sufficient for the reduction of the mixed discriminant approximation question to the doubly stochastic case. In this section we do both.
Our main tool is the following property of the ellipsoid algorithm [19]: For a prescribed accuracy $\delta > 0$, it finds a δ-minimizer of a continuous convex function f in a ball B, that is a point $x_\delta \in B$ with $f(x_\delta) \le \min_B f + \delta$, in no more than

$$O\left(n^2 \ln\left(\frac{2\delta + \mathrm{Var}_B(f)}{\delta}\right)\right) \tag{7}$$

iterations, where $\mathrm{Var}_B(f) = \max_B f - \min_B f$. Each iteration requires a single computation of value and of a gradient of f at a given point, plus $O(n^2)$ elementary operations to run the algorithm itself. In our case, this is easily seen to cost at most $O(n^3)$ elementary operations.
Recall, that the radius R of the ball B is given by lemma 26: $R \le n^{\frac{1}{2}} \cdot \log \frac{2\det(A_1+...+A_n)}{M}$.

LEMMA 29.

$$Var_B(f) \le O\left(n^{\frac{3}{2}}(\nu + \log n)\right),$$

where ν is the binary representation length of entries in $\mathbf{A}$.

PROOF. We may, without loss of generality, assume that all the matrices A_i in $\mathbf{A}$ have *integer entries.* Note, that since the binary representation length of entries in $\mathbf{A}$ is ν, the maximal size of an entry will not exceed 2^ν. By (19), since M is greater than zero, it is at least 1. On the other hand, by Hadamard inequality, $\det(A_1+...+A_n) \le (n2^\nu)^n = n^n 2^{\nu n}$. Therefore, $R \le n^{\frac{1}{2}} \cdot \log n^n 2^{\nu n} = n^{\frac{3}{2}}(\nu + \log n)$.
We conclude that $\max_B f \le \log(R^n \det(A_1 + ... + A_n)) \le O\left(n^{\frac{3}{2}}(\nu + \log n)\right)$.
On the other hand, the proof of lemma 26 demonstrates that for any $\xi \in H$ holds $f(\xi) \ge \log(\frac{M}{2}) \ge -1$. Therefore

$$Var_B(f) \le O\left(n^{\frac{3}{2}}(\nu + \log n)\right).$$

$\square$

PROPOSITION 30. *Let $\mathbf{A} = (A_1, ...A_n)$ be a fully indecomposable n-tuple of positive semidefinite matrices, and let $0 < \epsilon < 1$. Let ξ be an $\epsilon^2/10$-minimizer of f on H. Then $x_i = e^{\xi_i}$, for $i = 1...n$ and $S = \left(\sum_{i=1}^n e^{\xi_i}A_i\right)^{-1}$ scale $\mathbf{A}$ to an ϵ-doubly stochastic tuple.*

PROOF. Let $\delta := \epsilon^2/10$. We prove the proposition by a sequence of easy reductions to simpler cases. First we show, that in effect we may assume $\mathbf{A}$ is doubly stochastic. We know that

$$\log\det\left(e^{\xi_1^*}A_1 + .. + e^{\xi_n^*}A_n\right) \le \log\det\left(e^{\xi_1}A_1 + .. + e^{\xi_n}A_n\right)$$

$$\le \log\det\left(e^{\xi_1^*}A_1 + ... + e^{\xi_n^*}A_n\right) + \delta.$$

Taking exponents and observing that for a small δ holds $e^\delta \le 1 + 2\delta$, we get:

$$\det\left(e^{\xi_1^*}A_1 + ... + e^{\xi_n^*}A_n\right) \le \det\left(e^{\xi_1}A_1 + ... + e^{\xi_n}A_n\right) \le$$

$$\det\left(e^{\xi_1^*}A_1 + ... + e^{\xi_n^*}A_n\right)\cdot(1 + 2\delta).$$

Setting $S^* = \left(\sum_{i=1}^n e^{\xi^*} A_i\right)^{-1}$, $B_i = (S^*)^{\frac{1}{2}} e^{\xi^*} A_i (S^*)^{\frac{1}{2}}$, and $(\Delta\xi)_i = \xi - \xi^*$, we get

$$1 \le \det\left(e^{(\Delta\xi)_1} B_1 + ... + e^{(\Delta\xi)_n} B_n\right) \le 1 + 2\delta. \quad (8)$$

Observe that $\mathbf{B} = (B_1, ..., B_n)$ is a doubly stochastic tuple, by lemma 25, and that the claim of the proposition is equivalent to proving that the tuple $B_i' := (S')^{\frac{1}{2}} e^{(\Delta\xi)_i} B_i (S')^{\frac{1}{2}}$ is ϵ-doubly stochastic, where $S' := \left(\sum_{i=1}^n e^{(\Delta\xi)_i} B_i\right)^{-1}$.
Next, we go from dealing with doubly stochastic n-tuples to an easier case of doubly stochastic matrices. Let W be an orthogonal matrix such that $W^t S' W$ is diagonal, namely the columns $w_1, ...w_n$ of W are eigenvectors of S'. Let $b_{ij} =< w_j B_i w_j >$. Then the matrix $B = (b_{ij})$ is doubly stochastic, and (8) reduces to

$$1 \le \prod_{i=1}^n (By)_i \le 1 + 2\delta, \quad (9)$$

where $y \in \mathbf{R}^n$ is given by $y_i = e^{(\Delta\xi)_i}$, for $i = 1, ..., n$. Note that

$$\prod_{i=1}^n y_i = e^{\sum_i \xi_i - \sum_i \xi_i^*} = 1. \quad (10)$$

Our claim amounts to showing, given double-stochasticity of B, (9) and (10) that the matrix $C = (c_{ij}) = \left(\frac{b_{ij} y_j}{\sum_{k=1}^n b_{ik} y_k}\right)$ is ϵ-doubly stochastic. Clearly C is row-normalized. Setting $c_j = \sum_{i=1}^n c_{ij}$ to be the column sums of C we have to show $\sum_{j=1}^n (c_j - 1)^2 \le \epsilon^2$. Note that $\sum_{j=1}^n c_j = 1$, and since B is doubly stochastic $\prod_{j=1}^n c_j = \frac{\prod_{i=1}^n y_i}{\prod_{i=1}^n (By)_i} \ge \frac{1}{1+2\delta} \ge 1 - 2\delta$.
Now we are in a familiar situation. Lemma 3.10 of [18] states that for nonnegative numbers $z_1, ..., z_n$ summing to n, and for a sufficiently small Δ, $(0 \le \Delta \le 1/10$ is enough) holds $\sum_{j=1}^n (z_j - 1)^2 = \Delta \Longrightarrow \prod_{j=1}^n z_j \le 1 - \frac{\Delta}{3}$. We deduce that in our case

$$\sum_{j=1}^n (c_j - 1)^2 \le 6\delta < \epsilon^2$$

and we are done. $\square$

Theorem 11 follows from lemma 29, proposition 30, and the described properties of the ellipsoid method.

5. BOUNDS ON THE MIXED DISCRIMINANT OF DOUBLY STOCHASTIC TUPLES

LEMMA 31. *Let $\mathbf{A} = (A_1, ..., A_n)$ be a doubly stochastic n-tuple of positive semidefinite matrices, then*

$$D(\mathbf{A}) \le 1,$$

and this bound is tight.

PROOF. By the definition, the mixed discriminant $D(A_1...A_n)$ is the coefficient of $x_1 x_2 ... x_n$ in the polynomial $\det(x_1 A_1 + ... + x_n A_n)$. It turns out [4], that all the coefficients of this polynomial can be expressed through mixed discriminants. Let $r_1, ...r_n$ be nonnegative integers adding to n. Then the coefficient of $x_1^{r_1} \cdots x_n^{r_n}$ is equal to

$$t_r = \frac{1}{r_1!...r_n!} D\left(\underbrace{A_1...A_1}_{r_1} ... \underbrace{A_n..A_n}_{r_n}\right). \quad (11)$$

Since mixed discriminant of positive semidefinite matrices is nonnegative, (11) implies that all the coefficients of the polynomial $\det(x_1 A_1 + ... + x_n A_n)$ are nonnegative. Substituting $x_i = 1$, for $i = 1...n$, we obtain $D(A_1...A_n) \le \det(A_1 + \cdots + A_n) = D(I) = 1$.
Recall, that if the matrices $A_1...A_n$ are diagonal, namely $A_j = diag(b_{1j}...b_{nj})$, for $j = 1...n$, and $B = (b_{ij})$, then $per(B) = D(A_1...A_n)$. It is easily seen, that the n-tuple $A_1...A_n$ of diagonal matrices is doubly stochastic if and only if the matrix B is doubly stochastic. Therefore, the class of doubly stochastic tuples includes the class of doubly stochastic matrices, and the mixed discriminant on this subclass behaves as a permanent. It follows, that the upper bound of the lemma is tight, since it is tight for the doubly stochastic matrices (take $B = I$). $\square$

THEOREM 32. *Let $\mathbf{A} = (A_1, ..., A_n)$ be a doubly stochastic n-tuple of positive semidefinite matrices, then*

$$D(\mathbf{A}) \ge \frac{n!}{n^n},$$

and this bound is tight.

By the argument at the end of the preceding lemma, the lower bound in the theorem is tight, since it is obtained for the permanent on the doubly stochastic matrix $J = (\frac{1}{n})$.
By the same argument, this lower bound extends the permanental lower bound to the larger class of doubly stochastic n-tuples, and therefore implies the van der Waerden conjecture as a special case.

PROOF. (See also [14]) Naturally enough, the proof tries to emulate the proofs of Falikman [11] and Egorychev [10] for the van der Waerden conjecture. It means that we have to translate the language of nonnegative matrices to the language of n-tuples of positive semidefinite matrices. Although all the ingredients, including the crucial one - the Alexandrov-Fenchel inequalities, are there in the new setting, the translation turns out to be rather difficult.
We are looking for a constrained minimum of a smooth function $F : \mathbf{R}^{n^3} \to \mathbf{R}$, where $F(\mathbf{X}) = F(X_1, ..., X_n) = D(X_1, ...X_n)$ is the mixed discriminant of n $n \times n$ matrices X_i. The constraints are:

1. $< X_i v, v > \quad \ge 0$ for all $1 \le i \le n$ and $v \in \mathbf{R}^n$, with $\|v\|_2 = 1$.

2. $tr(X_i) = 1$, for $1 \le i \le n$.

3. $\sum_{i=1}^n X_i = I$.

We would like to write out necessary conditions for a point $\mathbf{A} = (A_1, ..., A_n)$ to be a constrained (local) minimum. Note that there is an infinite number of constraints, however this does not constitute a difficulty, due to a theorem of John [16] which says that in this case there is a *finite* number of constraints of the form $C_i(\mathbf{X}) \geq 0$ or $C_i(\mathbf{X}) = 0$, for $i = 1, ..., k$, such that the function

$$\Phi(\mathbf{X}) = \lambda_0 F(\mathbf{X}) - \sum_{i=1}^{s} \lambda_i C_i(\mathbf{X})$$

has an *unconstrained minimum* at $\mathbf{A}$, and the constrains C_i are tight at $\mathbf{A}$, namely $C_i(\mathbf{A}) = 0$ for $i = 1, ..., k$. The coefficients λ_i have the following properties: $\lambda_0 \geq 0$ and $\lambda_i > 0$, whenever C_i is an *inequality*.

Therefore, the linear combination of gradients $\lambda_0 \bigtriangledown F - \sum_{i=1}^{s} \lambda_i \bigtriangledown C_i$ at $\mathbf{A}$ is zero. It's easy to see that in our case this can't hold if λ_0 were zero. Dividing out by λ_0, we get, with some abuse of notation:

$$\bigtriangledown F = \sum_{i=1}^{n} \sum_{k_i=1}^{s_i} \lambda_{i,k_i} \cdot \bigtriangledown(< X_i v_{ik_i}, v_{ik_i} >) +$$

$$\sum_{i=1}^{n} \mu_i \cdot \bigtriangledown(tr(X_i) - 1) +$$

$$\sum_{i,j=1}^{n} \rho_{ij} \cdot \bigtriangledown \left(\sum_{k=1}^{n} X_k(i,j) - I(i,j) \right). \quad (12)$$

Here $s_1, ..., s_k$ are nonnegative integers, and the coefficients λ_{i,k_i}, μ_j are nonnegative real numbers. The vectors v_{ik_i} are unit vectors , such that $< v_{ik_i}, A_i v_{ik_i} >= 0$, for $i = 1...n$ and $k_i = 1, ..., s_i$.

Since we view points $\mathbf{X} \in \mathbf{R}^{n^3}$ as n-tuples of $n \times n$ matrices, $\mathbf{X} = (X_1, ..., X_n)$, the gradients above also may be naturally viewed as n-tuples. We will use the following notation $\bigtriangledown f = \left(\frac{\partial f}{\partial X_1}, ..., \frac{\partial f}{\partial X_1} \right)$, where $\frac{\partial f}{\partial X_i}$ is an $n \times n$ matrix with the k, l coordinate given by $\frac{\partial f}{\partial X_i(k,l)}$.

The gradients in (12) are easily computable. Let E_{ij} denote the matrix with 1 in the i, j'th entry and 0 everywhere else. We start with the gradient on the RHS of (12). Let $Q_i := \frac{\partial F}{\partial X_i}$. Then

$$Q_i(k,l) = D(X_1, ...X_{i-1}, E_{kl}, X_{i+1}, ...X_n). \quad (13)$$

In the next lemma we point out two crucial properties of the Q_i, which correspond to the expansion of permanent by the row and by the column in the matrix case.

Lemma 33. *Let $\mathbf{X} = (X_1, ..., X_n)$ be an n-tuple of symmetric real matrices, and let $Q_i = \frac{\partial D(X_1,...,X_n)}{\partial X_i}$ at $\mathbf{X}$. Then:*

$$tr(X_i Q_i) = D(X_1, ..., X_n),$$

for any $1 \leq i \leq n$. And

$$\sum_{i=1}^{n} < X_i w, Q_i w >= D(X_1, ..., X_n),$$

for any unit vector $w \in \mathbf{R}^n$.

Proof. The first part of the lemma is very simple. Assume, without loss of generality, that $i = 1$. Then

$$tr(X_1 Q_1) = \sum_{k,l} X_1(k,l) Q_1(k,l) =$$

$$\sum_{k,l} X_1(k,l) D(E_{kl}, X_2, ...X_n) = D(X_1, ...X_n).$$

The second part is only a little more difficult. By the definition of Q_i, it is easily checked that $< X_i w, Q_i w >= D(X_1, ..., X_{i-1}, X_i w \otimes w, X_{i+1}, ...X_n)$, and therefore we need to prove

$$\sum_{i=1}^{n} D(X_1..X_{i-1}, X_i w \otimes w, X_{i+1}..X_n) = D(X_1...X_n) \quad (14)$$

Let us first prove (14) for $w = e_1$, the first unit vector. Note that for any matrix X holds $X e_1 \otimes e_1$ is the matrix whose first column is the first column of X, and all the other columns are zero. Therefore, using definition 19:

$$\sum_{i=1}^{n} D(X_1, ..., X_{i-1}, X_i e_1 \otimes e_1, X_{i+1}, ...X_n) =$$

$$\sum_{i=1}^{n} \sum_{\sigma \in S_n, \ \sigma^{-1}(i)=1} \det(X_\sigma) = \sum_{\sigma \in S_n} \det(X_\sigma) = D(X_1, ..., X_n).$$

For a general unit vector w, let W be an orthogonal matrix with $Ww = e_1$. Then

$$\sum_{i=1}^{n} D(X_1, ..., X_{i-1}, X_i w \otimes w, X_{i+1}, ...X_n) =$$

$$\sum_{i=1}^{n} D(W^t X_1 W, ..., , W^t X_i w \otimes w W, , ...W^t X_n W) =$$

$$\sum_{i=1}^{n} D(W^t X_1 W, ..., W^t X_i W e_1 \otimes e_1, ...W^t X_n W) =$$

$$D(W^t X_1 W, ..., W^t X_n W) = D(X_1, ..., X_n).$$

$\square$

We also point out another easy fact, which will become useful later, and which may be proven exactly as the first part of the preceding lemma. Recall, that for any two indices $1 \leq i, j \leq n$, the tuple $\mathbf{A}^{ji}$ is the n-tuple obtained from $\mathbf{A}$ by replacing A_i with A_j. We claim

$$D(\mathbf{A}^{ji}) = tr(A_j Q_i). \quad (15)$$

Moving to the LHS of (12), we have:

$$\frac{\partial < X_i v, v >}{\partial X_j} = \delta_{ij} \cdot v \otimes v,$$

where δ_{ij} is the Kronecker's δ-function. Also

$$\frac{\partial tr(X_i)}{\partial X_i} = \delta_{ij} \cdot I,$$

where I is the $n \times n$ identity matrix, and

$$\frac{\partial \sum_{s=1}^{n} X_s(k,l)}{\partial X_i} = E_{kl}.$$

We go back to the linear dependency (12) between the gradients at the point of minimum $\mathbf{A}$. Let $R := (\rho_{ij})$ and $P_i := \sum_{k_i=1}^{s_i} \lambda_{i,k_i} v_{ik_i} \otimes v_{ik_i}$. Note, that P_i are positive semidefinite matrices such that $A_i P_i = 0$, for $i = 1...n$. The conditions (12) translate to

$$Q_i = R + \mu_i I + P_i, \quad (16)$$

for all $1 \leq i \leq n$.

Combining this with the fact that $A_i P_i = 0$, for all $1 \le i \le n$, implies that we can substitute $R + \mu_i I$ instead of Q_i in lemma 33, obtaining

$$tr(A_i(R + \mu_i I)) = D(A_1, ..., A_n), \tag{17}$$

for any $1 \le i \le n$. And

$$\sum_{i=1}^{n} < A_i w, (R + \mu_i I)w >= D(A_1, ..., A_n), \tag{18}$$

for any unit vector $w \in \mathbf{R}^n$.
Now, note that for any $1 \le i \le n$, the matrix Q_i is symmetric. Indeed $Q_i(k,l) = D(A_1, ..., A_{i-1}, E_{kl}, A_{i+1}, ...A_n) = D(A_1, ..., A_{i-1}, E_{lk}, A_{i+1}, ...A_n) = Q_i(l,k)$, since the matrices $A_1, ..., A_n$ are symmetric, and replacing the matrices with their transposes does not change the mixed discriminant. Consequently R is a symmetric matrix by (16), and there exists an orthogonal matrix W such that $W^t R W = diag(\theta_1, ..., \theta_n)$, for some real $\theta_1, ..., \theta_n$. Let $w_1, ..., w_n$ be the columns of W. We define a *doubly stochastic* matrix $B = (b_{ij})$, by setting $b_{ij} =< w_j, A_i w_j >$. Going over $i = 1...n$ in (17) and over $w = w_1, ..., w_n$ in (17), and setting $\delta := D(A_1, ..., A_n)$ we obtain the following equations on $\mu_1, ...\mu_n$ and $\theta_1, ..., \theta_n$:

$$\mu_i + \sum_{j=1}^{n} b_{ij}\theta_j = \delta,$$

for all $i = 1...n$, and

$$\theta_j + \sum_{i=1}^{n} b_{ij}\mu_i = \delta,$$

for all $i = 1...n$. These equations are, of course, encountered also in the matrix case. Proceeding in exactly the same way, we easily deduce that $\mu = (\mu_1, ..., \mu_n)$ is an eigenvector of BB^t with eigenvalue 1. Similarly θ is an eigenvector of $B^t B$ with eigenvalue 1.
Now we would like to use the fact that B is an indecomposable matrix, implying that BB^t and $B^t B$ are indecomposable, and that both μ and θ are multiples of the vector $\mathbf{j}$ of all ones. But first we have to show that B is indecomposable. The indecomposability of B would easily follow (we omit the easy proof) from indecomposability of the minimizing n-tuple $\mathbf{A}$. There are several possible ways to show indecomposability of $\mathbf{A}$. We choose may be not the best, but the simplest one: Note, that (by induction) we may assume the claim of the theorem to be true for $m < n$. Therefore is a minimizing tuple $\mathbf{A}$ would decompose into two tuples $\mathbf{B}_1$ and $\mathbf{B}_2$, of dimensions m_1 and m_2 correspondingly, this would imply that

$$D(\mathbf{A}) = D(\mathbf{B}_1)D(\mathbf{B}_2) \ge \frac{m_1!}{m_1^{m_1}} \cdot \frac{m_2!}{m_2^{m_2}} > \frac{n!}{n^n},$$

which is clearly wrong, since $D(\mathbf{A}) \le D(I/n, ...I/n) = \frac{n!}{n^n}$. Thus both μ and θ are multiples of $\mathbf{j}$, and $\mu + \theta = \delta \cdot \mathbf{j}$. Substituting in (16) we get

$$Q_i = \delta I + P_i \succeq \delta I$$

for all $i = 1, ..., n$. The $\succeq$ sign means just that $P_i = Q_i - \delta I$ is a positive semidefinite matrix.
Now is the time to use the Alexandrov-Fenchel inequalities:

THEOREM 34. *[1] Let $\mathbf{A} = (A_1, .., A_n)$ be an n-tuple of positive semidefinite matrices, and let $\mathbf{A}^{ij}$ be the n-tuple obtained from $\mathbf{A}$ by replacing A_j with A_i. Then*

$$D^2(\mathbf{A}) \ge D(\mathbf{A}^{ij}) \cdot D(\mathbf{A}^{ji}).$$

We intend to show, via the Alexandrov-Fenchel inequalities that actually

$$Q_i = \delta \cdot I \tag{19}$$

holds for any $1 \le i \le n$. Indeed assume that this is not so, and there is an index i such that $Q_i \ne \delta \cdot I$. Then $tr(Q_i) = tr(IQ_i) = tr(\sum_{j=1}^{n} A_j Q_i) > \delta n$, namely there is an index j such that $tr(A_j Q_i) > \delta$. Therefore

$$\delta^2 = D^2(\mathbf{A}) \ge D(\mathbf{A}^{ij})D(\mathbf{A}^{ji}) = tr(A_i Q_j)tr(A_j Q_i) > \delta^2,$$

contradiction. The first inequality is the Alexandrov-Fenchel inequality, the next equality is given by (15) and the last one follows since $tr(A_j Q_i) > \delta$, and $tr(A_i Q_j) \ge tr(A_i \cdot \delta I) = \delta$. Given (19) we conclude the proof by observing, as in the matrix case, that if $\mathbf{A} = A_1, ..., A_n$ is a minimizing tuple then so is the tuple $\mathbf{A}'$, obtained by replacing any two matrices in $\mathbf{A}$ by (two copies) of their arithmetic mean. Performing this averaging procedure iteratively, and passing to the limit, we obtain, that the tuple $\mathbf{E} = (E_1, ...E_n)$ with $E_i = \frac{1}{n}\sum_{i=1}^{n} A_i = I/n$ is a minimizing tuple, and we are done, since obviously $D(\mathbf{E}) = \frac{n!}{n^n}$.

6. REFERENCES

[1] A. Aleksandrov, On the theory of mixed volumes of convex bodies, IV, Mixed discriminants and mixed volumes (in Russian), *Mat. Sb. (N.S.)* 3:227-251, 1938.

[2] M. Avriel, *Nonlinear programming: analysis and methods*, Prentice-Hall, 1976.

[3] I. Barany and Z. Furedi, Computing the volume is difficult, *Discrete & Computational Geometry*, 2:319-326, 1987.

[4] R. Bapat, Mixed discriminants of positive semidefinite matrices, *Linear Algebra and its Applications* 126:107-124, 1989.

[5] A. I. Barvinok, Computing Mixed Discriminants, Mixed Volumes, and Permanents, *Discrete & Computational Geometry*, 18:205-237 (1997).

[6] A. I. Barvinok, A simple polynomial time algorithm to approximate permanents and mixed discriminants within a simply exponential factor, *Random Structures and Alg.*, 14:29-61, 1999.

[7] M. Dyer and A. Frieze, The complexity of computing the volume of a polyhedron, *SIAM J. Comput.*, 17:967-994, 1988.

[8] M. Dyer, P. Gritzmann and A. Hufnagel, On the complexity of computing mixed volumes, *SIAM J. Comput.*, 27(2):356-400, 1998.

[9] J. Edmonds, Submodular functions, matroids, and certain polyhedra, in *Combinatorial structures and their applications*, R. Guy, H. Hanani, N. Sauer and J. Schönheim, eds., Gordon and Breach, New York, 69-87, 1970.

[10] G.P. Egorychev, The solution of van der Waerden's problem for permanents, *Advances in Math.*, 42:299-305, 1981.

[11] D. I. Falikman, Proof of the van der Waerden's conjecture on the permanent of a doubly stochastic matrix, *Mat. Zametki* 29(6):931-938, 957, 1981, (in Russian).

[12] S. Friedland, A lower bound for the permanent of a doubly stochastic matrix, *Annals of Mathematics*, 110:167-176, 1979.

[13] M. Grötschel, L. Lovasz and A. Schrijver, *Geometric Algorithms and Combinatorial Optimization*, Springer-Verlag, Berlin, 1988.

[14] L. Gurvits, Proving Generalized van der Waerden Conjecture - A proof of Bapat's conjecture on mixed discriminants, preprint, 1999.

[15] M. Jerrum and A. Sinclair, Approximating the permanent, *SIAM J. Comput.*, 18:1149-1178, 1989.

[16] F. John, Extremum problems with inequalities as subsidiary conditions, *Studies and Essays, presented to R. Courant on his 60th birthday*, Interscience, New York, 1948.

[17] N. Karmarkar, R. Karp, R. Lipton, L. Lovasz and M. Luby, A Monte-Carlo algorithm for estimating the permanent, *SIAM J. Comput.*, 22(2):284-293, 1993.

[18] N. Linial, A. Samorodnitsky and A. Wigderson, A deterministic strongly polynomial algorithm for matrix scaling and approximate permanents, *Proc. 30 ACM Symp. on Theory of Computing*, ACM, New York, 1998.

[19] Y. Nesterov and A. Nemirovskii, *Interior-Point Polynomial Algorithms in Convex Programming*, SIAM, Philadelphia, PA, 1994.

[20] A. Panov, On mixed discriminants connected with positive semidefinite quadratic forms, *Soviet Math. Dokl.* 31 (1985).

[21] R. Schneider, *Convex bodies: The Brunn-Minkowski Theory*, Encyclopedia of Mathematics and Its Applications, vol. 44, Cambridge University Press, New York, 1993.

[22] L. G. Valiant, The complexity of computing the permanent, *Theoretical Computer Science*, 8(2):189-201, 1979.