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 \( n \times n \) matrix to within a multiplicative factor of \( e^n \).

1. INTRODUCTION

1.1 Mixed discriminant, permanent and mixed volume

Permanent:

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

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

where \( S_n \) is the symmetric group on \( n \) elements, is called the permanent of \( A \). For a \( 0,1 \) matrix \( A \), \( \text{per}(A) \) 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 \((1+\epsilon)-\text{factor}\) 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 \( 2^{O(n)}-\text{factor} \) approximation was obtained in [17]. The first efficient probabilistic algorithm that provides a \( 2^{O(n)}-\text{factor} \) approximation for arbitrary positive matrices was obtained by Barvinok in [5], [6].

A deterministic strongly polynomial algorithm achieving \( 2^{O(n)}-\text{factor} \) 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 \( \left[ \frac{1}{n}, 1 \right] \), and this solves the approximation problem. We recall, that the lower bound of \( \frac{1}{n} \) 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 \( K_1 \ldots K_n \) be convex bodies in \( \mathbb{R}^n \), and let \( V(\cdot) \) be the Euclidean volume in \( \mathbb{R}^n \). 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 \( \lambda_1 \ldots \lambda_n \), where \( + \) denotes Minkowski sum, and \( \lambda K \) denotes the dilatation of \( K \) with coefficient \( \lambda \). The coefficient \( V(K_1 \ldots K_n) \) of \( \lambda_1 \cdot \lambda_2 \cdots \lambda_n \) is called the mixed volume of \( K_1 \ldots K_n \). Alternatively,

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

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 \( \mathbb{C}^n \) 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 \((1+\epsilon)-\text{factor}\) 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 \( n^{O(n)}-\text{factor} \) approximation for arbitrary well-presented proper convex bodies was obtained by Barvinok in [5].

The question of existence of an efficient deterministic algo-
1.2 Our results

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

**Theorem 1.** There is a function $f$, such that

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

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 $\nu$ is the maximal binary representation length of the entries of $A_1, \ldots, A_n$.

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

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

holds on every $n$-tuple of proper well-presented convex bodies $K_i$ in $R^n$. 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 $A = (A_1, \ldots, A_n)$ and $B = (B_1, \ldots, B_n)$ be two $n$-tuples of $n \times n$ matrices. The tuple $B$ is a scaling of $A$, if there exist a vector $x \in \mathbb{R}^n$ and two $n \times n$ matrices $T_1, T_2$, such that $B_i = x_i T_1 A_i T_2$, for all $i = 1, \ldots, n$.

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

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

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

**Definition 5.** An $n$-tuple $A = (A_1, \ldots, A_n)$ of positive semidefinite matrices is doubly stochastic (d.s.) if

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

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

**Definition 6.** Let $A = (A_1, \ldots, A_n)$ be an $n$-tuple of $n \times n$ positive semidefinite matrices. A positive vector $x \in \mathbb{R}^n$, and a positive definite $n \times n$ matrix $S$ are scaling factors of $A$, if the $n$-tuple $B = (B_1, \ldots, B_n)$, given by $B_i = x_i S^{1/2} A_i S^{1/2}$ is doubly stochastic. \footnote{Here $S^{1/2}$ is the unique positive semidefinite matrix whose square is $S$.}

The preceding description presents a very “small scale” overview of things. In the next subsection we go into details.
1.3 An overview of the mixed discriminant approximation algorithm

- We define a notion of a fully indecomposable tuple.
- 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 aforementioned 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 \( A = (A_1, \ldots, A_n) \) 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 \( B_1 \ldots B_k \) of positive semidefinite matrices, such that

\[
D(A) = \prod_{i=1}^{k} D(B_i).
\]

The tuples \( B_1 \ldots B_k \) can be found in polynomial time.

**Theorem 9.** Let \( A = (A_1, \ldots, A_n) \) be a fully indecomposable \( n \)-tuple of positive semidefinite matrices. Then:
1. There exist scaling factors \( x \) and \( S \) such that \( B = (x_1 S^T A_1 S, \ldots, x_n S^T A_n S) \) is doubly stochastic.
2. Let there be two pairs of scaling factors \( (x, S) \) and \( (x', S') \) for \( B \), 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 \( A = (A_1, \ldots, A_n) \) be an \( n \)-tuple of \( n \times n \) positive semidefinite matrices. A positive vector \( x \in \mathbb{R}^n \), and a positive definite \( n \times n \) matrix \( S \) are \( \epsilon \)-scaling factors of \( A \), if the \( n \)-tuple \( B = (B_1, \ldots, B_n) \), given by \( B_i = x_i S^T A_i S \), is \( \epsilon \)-doubly stochastic, namely

\[
\sum_{i=1}^{n} (\text{Tr}(B_i) - 1)^2 \leq \epsilon^2,
\]

and

\[
\sum_{i=1}^{n} B_i = I.
\]

**Theorem 11.** Let \( A = (A_1, \ldots, A_n) \) be a fully indecomposable \( n \times n \) matrix of positive semidefinite \( n \times n \) matrices. Let \( \epsilon > 0 \) be a required scaling accuracy. Then \( \epsilon \)-scaling factors \( x_1 \ldots x_n \) and \( S' \) for \( A \) can be found in

\[
O\left(n^5 \log \left( \frac{\text{rand}}{\epsilon} \right) \right)
\]

arithmetic operations. Here \( \nu \) is the maximal binary representation length of the entries in \( A_1 \ldots A_n \). Moreover, if \( x_1 \ldots x_n \) and \( S \) are the proper scaling factors for \( A \), 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 \( A = (A_1, \ldots, A_n) \) be a doubly stochastic \( n \)-tuple of positive semidefinite \( n \times n \) matrices. Then

\[
\frac{n!}{n^n} \leq D(A) \leq 1.
\]

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 \( A \) be an \( n \)-tuple of positive semidefinite matrices. There is a polynomial-time algorithm which decides whether \( D(A) = 0 \) or \( D(A) > 0 \).

**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 \( A = (A_1, \ldots, A_n) \) be an \( n \)-tuple of positive semidefinite matrices. Then the following two conditions are equivalent:
1. \( D(A) > 0 \), and
2. For all \( S \subseteq \{1, \ldots, n\} \), \( 0 < |S| < n \), \( \text{Rank}(\sum_{i \in S} A_i) \geq |S| \).

Now, let for \( i = 1, \ldots, n \) the matrix \( T_i \) be the positive semidefinite square root of \( A_i \). Let \( E_i = \{t^i_1, \ldots, t^i_n\} \) be the set of columns of \( T_i \). Consider two matroids on the ground set \( E = \bigcup_{i=1}^{n} E_i = \{t^i_j\}_{i,j=1 \ldots n} \). 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 \( \{E_1, \ldots, E_n\} \). Observe, that, by the aforementioned property of the mixed discriminant, \( D(A) > 0 \) 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:
Consider now $A_j := \{v_i, v_j\}$ for all $i, j$. The proof of the lemma is essentially the same as that of proposition 14.

**Proof.** The claim follows immediately from decomposability of $A$ and theorem 14.

We now determine the minimal non-empty set $\alpha$ with this property, or to decide that $A$ is indecomposable, in which case we are done. For this purpose we consider $n(n-1)$ auxiliary $n$-tuples $A^\sigma$, where $A^\sigma$ is obtained from $A$ by substituting $A_i$ instead of $A_j$. Let $D_{ij} = D(A^\sigma)$. We define $\alpha$ as the $n \times n$ 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 $\sum_{i \in \alpha} A_i = |\alpha|$, and assume that $\alpha$ 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 $X$ as an adjacency matrix of a directed graph $G = ([n], E)$, where $e = i \to j$ belongs to $E$ iff $Z_{ij} = 1$. For $i \notin [n]$, let $V_i$ be the set of points which can be reached from $i$ by the edges of $G$. Lemma 17 implies that $\alpha$ is a minimal set with the property that $\alpha = \sum_{i \in \alpha} A_i = |\alpha|$ if for any $i \in \alpha$ holds $V_i = \alpha$, and moreover, $\alpha$ 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 $\alpha$ is indecomposable and we are done.

We set $C_1 := \alpha$ to be the first cell in our partition of $[n]$. Let $S = \{C_i\}, A_i$, and let $v_1, \ldots, v_{c_i}$ be independent columns of $S_i$, spanning $X_1$. Completing to a basis of $R^\alpha$ and orthogonalizing, we obtain an orthonormal basis $u_1, \ldots, u_{c_i}$ of $R^\alpha$ with $\text{Span}(u_1, \ldots, u_{c_i}) = \text{Span}(v_1, \ldots, v_{c_i})$. Let $U_i$ be the $n \times c_i$ matrix with columns $u_1, \ldots, u_{c_i}$ and $U^+_i$ be the $n \times (n - c_i)$ matrix with columns $u_{c_i+1}, \ldots, u_n$. We set $B_i = U^+_i A_i U_i$ for $i \in C_1$, and $A_j = (U^+_j)^T A_j U^+_j$, for $j \notin C_1$. Then, for all $i, j$. Proposition 15. Let $A = (A_1, \ldots, A_n)$ be an $n$-tuple in $N_n$ 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 $\text{rank} \sum_{i \in C_s} A_i = |C_s| := c_s$.
2. The subspaces $X_s = \text{Im} \left( \sum_{i \in C_s} A_i \right)$ decompose $R^n$ into a direct orthogonal sum $R^n = X_1 \oplus X_2 \oplus \ldots \oplus X_k$.
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 $c_i \times c_i$ matrix $B_i$ be a projection of $A_i$ onto $X_s$, namely $B_i = U^+_i A_i U_i$, where $U_i$ is a $n \times c_i$ matrix with columns $\{u_i\}_{i \in C_s}$. Then the $c_s$-tuple $B_s = (B_i)_{i \in C_s}$ is fully indecomposable.

**Lemma 18.**

$D(A_1, \ldots, A_n) = D((B_i)_{i \in C_1} \cdot D((A_j)_{j \notin C_1})$.

**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 $A_{\sigma}$ 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 $c_1 \times c_1$ 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 $A_j$. Using definition 19 of the mixed discriminant, it is not hard to see, that in this case $D(U^T A_1 U \ldots U^T A_n U) = D((B_i)_{i \in C_1}) \cdot D((A_j)_{j \notin C_1})$.

Now we inductively (on dimension) apply the claim of the proposition to the $(n - c_1)$-tuple $(A_j)_{j \notin C_1}$. We point out, that minimality of $\alpha$ 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 $A$. 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 $0 \times 1$ matrix, which entails checking $O(n^2)$ 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 $\nu$ of the components $B_i$ can not be much greater than the representation length $\nu$ of $A$. Indeed, in the course of the decomposition procedure of proposition 15, the representation length increases, basically, only in the orthogonalization procedure. A moment’s reflection gives that the multiplicative factor of the increase is at most $n^3$.

**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 $A = (A_1, \ldots, A_n)$ be a fully indecomposable $n$-tuple of positive semidefinite matrices. We define

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

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 $A$. Consider the $(n-1)$-tuples $A^{\sigma_j}$, where $A^{\sigma_j}$ is obtained from $A$ by substituting $A_i$ instead of $A_j$. Let $D_{ij} = D(A^{\sigma_j})$.

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


Proof. Assume first that \( A = (A_1, \ldots, A_n) \) is indecomposable. We claim, that for any \( 1 \leq i \neq j \leq n \) the tuple \( A'^n = (A'_1, \ldots, A'_n) \) satisfies property 2 of Theorem 14, and therefore it’s mixed discriminant is positive. Indeed, let \( R \subseteq \{1, \ldots, n\} \). Then \( \text{rank}(\sum_{k \in R} A'_k) \geq \text{rank}(\sum_{k \in R \cup \{j\}} A_k) \geq |R| \), by indecomposability of \( A \).

In the other direction, let \( A = (A_1, \ldots, A_n) \) be a decomposable tuple, namely for some subset \( R \subseteq \{1, \ldots, n\} \), the inequality \( \text{rank}(\sum_{k \in R} A_k) \leq |R| \) holds. Let \( i, j \) be a pair of indices such that \( i \in R \) and \( j \notin R \), and consider the tuple \( A'^n = (A'_1, \ldots, A'_n) \). We claim that \( D(A'^n) = 0 \). Indeed,

\[
\text{rank}(\sum_{k \in R \cup \{j\}} A_k') = \text{rank}(\sum_{k \in R} A_k') < |R| + 1 = |R \cup \{j\}| \]

which, by Theorem 14 implies \( D(A'^n) = 0 \), and we are done.

**Definition 22.** Set \( M = M_A = \min_{i \neq j} D_{ij} \).

**Lemma 23.** The function \( f \) is a convex function on \( \mathbb{R}^n \), and is strictly convex on \( \mathbb{R}^n \) to the hyperplane \( H = \{ \xi = (\xi_1, \ldots, \xi_n) \in \mathbb{R}^n \mid \sum \xi = 0 \} \).

**Proof.** Let us denote the set of \( n \)-tuples \( r = r_1, \ldots, r_n \) of nonnegative integers summing to \( n \) by \( P_n \). Recall (11) that for any \( r \in P_n \), the coefficient \( t_r \) of \( x_1^{r_1} \cdots x_n^{r_n} \) in \( \text{det}(x_1A_1 + \cdots + x_nA_n) \) is given by

\[
t_r = \frac{1}{r_1! \cdots r_n!} D \left( \frac{A_1 \cdots A_n}{r_1 \cdots r_n} \right) .
\]

Therefore,

\[
f(\xi_1, \ldots, \xi_n) = \log \det \left( e^{(\xi_1 A_1 + \cdots + e^{(\xi_n A_n)} \right) = \\
\log \sum_{r \in P_n} t_r e^{<\xi, r>} = \\
\sum_{r \in P_n} t_r e^{<\xi, r>} .
\]

\[
\sum_{r \in P_n} t_r e^{<\xi, r>} .
\]

\[
\sum_{s \in P_n} t_s e^{<\xi, s> \otimes s} - \sum_{r \in P_n} t_r e^{<\xi, r+\varepsilon> \otimes (r-s)} = \\
\frac{1}{2} \sum_{r \in P_n} t_r e^{<\xi, r+\varepsilon> \otimes (r-s) \otimes (r-s)} \geq 0,
\]

implies the convexity of \( f \).

Now, assume the tuple \( A_1, \ldots, A_n \) to be indecomposable. Set, once again, \( D_{ij} \) to be the mixed discriminant of the \( n \)-tuple obtained from \( (A_1, \ldots, A_n) \) by replacing \( A_j \) with \( A_j \). We know by Lemma 21 that \( D_{ij} > 0 \). Actually in the notation of this lemma, \( D_{ij} \) is just \( 2t_{ij} \), 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 \( M = \min_{i \neq j} D_{ij} \). We now continue the computation from (6) to obtain:

\[
\nabla^2 f \geq \frac{1}{2g^2} \sum_{r, s \in P_n} t \cdot t e^{<\xi, r+\varepsilon> \otimes (r-s) \otimes (r-s)} \\
\frac{1}{8g^2} \sum_{i \neq j, k \neq l} D_{ij} D_{kl} e^{<\xi, r_{ij} + r_{kl} \otimes (r_{ij} - r_{kl}) \geq M^2 \sum_{i \neq j, k \neq l} (r_{ij} - r_{kl}) \otimes (r_{ij} - r_{kl}),
\]

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

\[
(r_{ij} - r_{kl}) \otimes (r_{ij} - r_{kl}) = (e_i - e_j - e_k + e_l) \otimes (e_i - e_j - e_k + e_l) = \\
E_{ii} + E_{jj} + E_{kk} + E_{ll} + 2(E_{il} + E_{lj} - E_{ij} - E_{ik} - E_{jk} - E_{kl}).
\]

Let

\[
S = \sum_{i \neq j, k \neq l} (r_{ij} - r_{kl}) \otimes (r_{ij} - r_{kl}).
\]

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 \((n_1 = n(n-1))(n-2)(n-3)\). Accordingly, \( S_{ii} = (n-1)(n-2)(n-3) \) and \( S_{ij} = (n-2)(n-3) \) for \( i \neq j \). Note that \( S \) is simply the orthogonal projection on the hyperplane \( H = \{ \xi = (\xi_1, \ldots, \xi_n) \in \mathbb{R}^n \mid \sum \xi = 0 \} \), multiplied by \((n-1)\)’

Therefore, the projection of \( \nabla^2 f \) onto \( H \) is greater or equal to \( \frac{M^2}{8g^2} S \), \( I_{n-1} \), where \( I_{n-1} \) is a \((n-1) \times (n-1)\) identity matrix, implying \( f \) is strictly convex on \( H \).

**Lemma 24.** For any \( \xi \in \mathbb{R}^n \), the gradient \( \langle \nabla f \rangle_\xi \) of \( f \) at \( \xi \) is \( \left( \text{Tr}(e^{\xi} S^T A_i S^T) \right)_\xi \), 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 \( \xi^* \) is a minimum of \( f \) on \( H \) if and only if the gradient \( \nabla f \) at \( \xi^* \) is a constant multiple of \((1, 1, \ldots, 1)\), the vector of all ones.

\[
\nabla f \mid_{\xi^*} = c \cdot (1, 1, \ldots, 1) \text{ if and only if } x_i = e^{\xi_i}, \ i = 1, \ldots, n, \text{ and } S = \left( \sum_{i=1}^n e^{\xi_i} A_i \right)^{-1}, \text{ are the scaling factors for } A.
\]
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 $f(\xi) \leq f(0)$, then
\[
\|\xi\|_2 \leq n^{\frac{1}{2}} \cdot \log \frac{2 \det(A_1 + \ldots + A_n)}{M}.
\]

**Proof.** Let $\xi$ be a point in $H$ with $f(\xi) \leq f(0) = \log \det(A_1 + \ldots + A_n)$. Then, in the notation of the proof of lemma 23:
\[
\det(A_1 + \ldots + A_n) \geq \det(e^{\xi_1} A_1 + \ldots + e^{\xi_n} A_n) = \sum_{\tau \in \mathcal{P}_n} t_\tau e^{\sum_{i,j} D_{ij} e^{\xi_i - \xi_j}} \geq \frac{1}{2} \sum_{i \neq j} D_{ij} e^{\xi_i - \xi_j} \geq \frac{1}{2} M \sum_{i \neq j} e^{\xi_i - \xi_j} \geq \frac{1}{2} M e^{\max_{i \neq j} \xi_i - \xi_j}.
\]
The last inequality uses the fact that since $\sum_i \xi_i = 0$, we have $\langle \xi, \xi \rangle = 0$. Therefore
\[
\|\xi\|_2 \leq n^{\frac{1}{2}} \cdot \|\xi\|_\infty \leq n^{\frac{1}{2}} \cdot \log \frac{2 \det(A_1 + \ldots + A_n)}{M}.
\]

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 $\xi^*$, 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.

**Remark 28.** We observe that, in the notation of lemma 25,
\[
f(\xi^*) = \log \det \left( e^{\xi_1} A_1 + \ldots + e^{\xi_n} A_n \right) = \log \det(S) = \log \left( \det(S) \prod_{i=1}^n x_i \right)
\]
is the product of the scaling factors of $A$.

### 4. Finding the Minimum

In the previous section we have seen, that finding the point of minimum of the function $f = f_A$ on the hyperplane $H$ is equivalent to computing the scaling factors of $A$. This is interesting if we want to scale $A$. We have also seen, that finding the value of the minimum is equivalent to computing the product of the scaling factors of $A$. 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 $\delta$-minimizer of a continuous convex function $f$ in a ball $B$, that is a point $x_\delta \in B$ with $f(x_\delta) \leq \min_B f + \delta$, in no more than
\[
O \left( n^2 \ln \left( \frac{2\delta + \text{Var}_B(f)}{\delta} \right) \right)
\]
iterations, where $\text{Var}_B(f) = \max_B f - \min_B f$. 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: $R \leq n^{\frac{1}{2}} \cdot \log \frac{2 \det(A_1 + \ldots + A_n)}{M}$.

**Lemma 29.**
\[
\text{Var}_B(f) \leq O \left( n^{\frac{3}{2}} (\nu + \log n) \right),
\]
where $\nu$ is the binary representation length of entries in $A$.

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

On the other hand, the proof of lemma 26 demonstrates that for any $\xi \in H$ holds $f(\xi) \geq \log \left( \frac{1}{M} \right) \geq -1$. Therefore
\[
\text{Var}_B(f) \leq O \left( n^{\frac{3}{2}} (\nu + \log n) \right).
\]

**Proposition 30.** Let $A = (A_1, \ldots, A_n)$ be a fully indecomposable $n$-tuple of positive semidefinite matrices, and let $0 < \epsilon < 1$. Let $\xi$ be an $\epsilon^2/10$-minimizer of $f$ on $H$. Then $x_i = e^{\xi_i}$, for $i = 1, \ldots, n$ and $S = (\sum_{i=1}^n e^{\xi_i} A_i)^{-1}$ scale $A$ to an $\epsilon$-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 $A$ is doubly stochastic. We know that
\[
\log \det \left( e^{\xi_1} A_1 + \ldots + e^{\xi_n} A_n \right) \leq \log \det \left( e^{\xi_1} A_1 + \ldots + e^{\xi_n} A_n \right) + \delta.
\]
Taking exponents and observing that for a small $\delta$ holds $e^\delta \leq 1 + 2\delta$, we get:
\[
\det \left( e^{\xi_1} A_1 + \ldots + e^{\xi_n} A_n \right) \leq \det \left( e^{\xi_1} A_1 + \ldots + e^{\xi_n} A_n \right) \leq \det \left( e^{\xi_1} A_1 + \ldots + e^{\xi_n} A_n \right) \cdot (1 + 2\delta).
\]
Setting $S' = \left( \sum_{i=1}^{n} e_i^e A_i \right)^{-1}$, $B_i = (S')^{\frac{1}{2}} e_i^e A_i (S')^{\frac{1}{2}}$, and $(\Delta \xi)_j = \xi - \xi^*$, we get

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

Observe that $B = (B_1, \ldots, B_n)$ is a doubly stochastic tuple, by lemma 25, and that the claim of the proposition is equivalent to proving that the tuple $B'_1 := (S')^{\frac{1}{2}} e^{(\Delta \xi)}_1 B_i (S')^{\frac{1}{2}}$ is $\epsilon$-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 $WSW^T$ is diagonal, namely the columns $w_1, \ldots, w_n$ of $W$ are eigenvectors of $S'$. Let $b_{ij} = \langle w_i, B_j w_j \rangle$. Then the matrix $B = (b_{ij})$ is doubly stochastic, and (8) reduces to

$$1 \leq \prod_{i=1}^{n} (By_i) \leq 1 + 2\delta, \quad (9)$$

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

$$\prod_{i=1}^{n} y_i = e^{\sum_{i=1}^{n} \xi^* - \sum_{i=1}^{n} t^*_i} = 1. \quad (10)$$

Our claim amounts to showing, given double-stochasticity of $B$, (9) and (10) that the matrix $C = (c_{ij}) = \left( \frac{b_{ij} y_i}{\sum_{j=1}^{n} b_{ij} y_j} \right)$ is 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 $\sum_{j=1}^{n} (c_j - 1)^2 \leq \epsilon^2$. Note that $\sum_{j=1}^{n} c_j = 1$, and since $B$ is doubly stochastic $\prod_{j=1}^{n} c_j = \prod_{j=1}^{n} b_{ij} y_j \geq \frac{1}{(1+2\delta)^n} \geq 1 - 2\delta$.

Now we are in a familiar situation. Lemma 3.10 of [18] states that for nonnegative numbers $z_1, \ldots, z_n$ summing to $n$, and for a sufficiently small $\Delta$, $(0 \leq \Delta \leq 1/10$ is enough) holds $\sum_{j=1}^{n} (z_j - 1)^2 = \Delta \implies \prod_{j=1}^{n} z_j \leq 1 - \frac{\Delta}{2}$. We deduce that in our case

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

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 $A = (A_1, \ldots, A_n)$ be a doubly stochastic $n$-tuple of positive semidefinite matrices, then

$$D(A) \leq 1,$$

and this bound is tight.

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

$$t_r = \frac{1}{r_1! \cdots r_n!} \det \left( \begin{array}{ccc} A_1 & \cdots & A_n \\ r_1 & \cdots & r_n \end{array} \right). \quad (11)$$

Since mixed discriminant of positive semidefinite matrices is nonnegative, (11) implies that all the coefficients of the polynomial $\det(x_1 A_1 + \ldots + x_n A_n)$ are nonnegative. Substituting $x_1 = 1$, for $i = 1, \ldots, n$, we obtain $D(A_1, \ldots, A_n) \leq \det(A_1 + \cdots + A_n) = D(I) = 1$.

Recall, that if the matrices $A_1, \ldots, A_n$ are diagonal, namely $A_j = \text{diag}(b_{1j}, \ldots, b_{nj})$, for $j = 1, \ldots, n$, and $B = (b_{ij})$, then $\text{per}(B) = D(A_1, \ldots, A_n)$. It is easily seen, that the $n$-tuple $A_1, \ldots, A_n$ 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 $A = (A_1, \ldots, A_n)$ be a doubly stochastic $n$-tuple of positive semidefinite matrices, then

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

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 permanent 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 $F : \mathbb{R}^{n^3} \to \mathbb{R}$, where $F(X) = F(X_1, \ldots, X_n) = D(X_1, \ldots, X_n)$ is the mixed discriminant of $n \times n$ matrices $X_i$. The constraints are:

1. $\langle X_i v, v \rangle \geq 0$ for all $1 \leq i \leq n$ and $v \in \mathbb{R}^n$, with $\|v\|_2 = 1$.
2. $\text{tr}(X_i) = 1$, for $1 \leq i \leq n$.
3. $\sum_{i=1}^{n} X_i = I$.

We would like to write out necessary conditions for a point $A = (A_1, \ldots, A_n)$ 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
Here are unit vectors, such that

\[ Q \]

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

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

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

\[ \nabla F = \sum_{i=1}^{s} \sum_{k,l=1}^{s_i} \lambda_{i,k,l} \cdot \nabla < X_i v_{ikl}, v_{ikl} > + \sum_{i=1}^{n} \mu_i \cdot \nabla (\text{tr}(X_i) - 1) + \sum_{i,j=1}^{n} \rho_{ij} \cdot \nabla \left( \sum_{k=1}^{n} X_k(i,j) - I(i,j) \right). \]

(12)

Here \( s_1, \ldots, s_k \) are nonnegative integers, and the coefficients \( \lambda_{i,k,l}, \mu_i, \rho_{ij} \) are nonnegative real numbers. The vectors \( v_{ikl} \) are unit vectors, such that \( < v_{ikl}, A_i v_{ikl} > = 0 \), for \( i = 1, \ldots, n \) and \( k = 1, \ldots, s_i \).

Since we view points \( X \in \mathbb{R}^{n^3} \) as \( n \)-tuples of \( n \times n \) matrices, \( X = (X_1, \ldots, X_n) \), the gradients above may also be naturally viewed as \( n \)-tuples. We will use the following notation \( \nabla f = \left( \frac{\partial f}{\partial X(1,1)}, \ldots, \frac{\partial f}{\partial X(n,n)} \right) \), where \( \frac{\partial f}{\partial X(k,l)} \) is an \( n \times n \) matrix with the \( k,l \) coordinate given by \( \frac{\partial f}{\partial X(k,l)} \).

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 \( Q_i := \frac{\partial}{\partial X_i} \). Then

\[ Q_i(k,l) = D(X_1, \ldots, X_{i-1}, E_{kl}, X_{i+1}, \ldots, X_n). \]

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

\[ \text{tr}(X_i Q_i) = D(X_1, \ldots, X_n), \]

for any \( 1 \leq i \leq n \). And

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

for any unit vector \( w \in \mathbb{R}^n \).

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

\[ \text{tr}(X_1 Q_1) = \sum_{k,l} X_1(k,l)Q_1(k,l) = \sum_{k,l} X_1(k,l)D(E_{kl}, X_2, \ldots, X_n) = D(X_1, \ldots, X_n). \]

The second part is only a little more difficult. By the definition of \( Q_i \), it is easily checked that \( < X_i w, Q_i w > = D(X_1, \ldots, X_i, X_i w, \ldots, X_n) \), and therefore we need to prove

\[ \sum_{i=1}^{n} D(X_1, \ldots, X_{i-1}, X_i w, w, X_{i+1}, \ldots, X_n) = D(X_1, \ldots, X_n) \quad (14) \]

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:

\[ \sum_{i=1}^{n} D(X_1, \ldots, X_{i-1}, X_i e_1 \otimes e_1, X_{i+1}, \ldots, X_n) = \sum_{i=1}^{n} \sum_{\sigma \in S_n} \det(X_\sigma) = \sum_{\sigma \in S_n} \det(X_\sigma) = D(X_1, \ldots, X_n). \]

For a general unit vector \( w \), let \( W \) be an orthogonal matrix with \( W w = e_1 \). Then

\[ \sum_{i=1}^{n} D(X_1, \ldots, X_{i-1}, X_i w \otimes w, X_{i+1}, \ldots, X_n) = \sum_{i=1}^{n} D(W^t X_1 W, \ldots, W^t X_i w \otimes w, \ldots, W^t X_n W) = \sum_{i=1}^{n} D(W^t X_1 W, \ldots, W^t X_i W e_1 \otimes e_1, \ldots, W^t X_n W) = D(W^t X_1 W, \ldots, W^t X_n W) = D(X_1, \ldots, X_n). \]

\[ \square \]

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 \( A^{ij} \) is the \( n \)-tuple obtained from \( A \) by replacing \( A_i \) with \( A_j \). We claim

\[ D(A^{ij}) = \text{tr}(A_j Q_i). \]

(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 \( \delta_{ij} \) is the Kronecker’s \( \delta \)-function. Also

\[ \frac{\partial \text{tr}(X_i)}{\partial X_i} = \delta_{ij} \cdot I, \]

where \( I \) is the \( n \times n \) identity matrix, and

\[ \frac{\partial \sum_{k,l} X_k(i,l)}{\partial X_i} = E_{ki}. \]

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

\[ Q_i = R + \mu_i I + P_i, \]

(16)

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

$$\text{tr}(A_i (R + \mu_i I)) = D(A_i, \ldots, A_n),$$

(17)

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

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

(18)

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

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

for all $i = 1, \ldots, n$, and

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

for all $i = 1, \ldots, n$. These equations are, of course, encountered also in the matrix case. Proceeding in exactly the same way, we easily deduce that $\mu = (\mu_1, \ldots, \mu_n)$ is an eigenvector of $BB^T$ with eigenvalue 1. Similarly $\theta$ 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 $\mu$ and $\theta$ are multiples of the vector $j$ 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 $A$. There are several possible ways to show indecomposability of $A$. We choose may be not the best, but the simplest one: Note, that (by induction) we may assume from indecomposability of the minimizing tuple $B$, of dimensions $m_1$ and $m_2$ correspondingly, this would imply that

$$D(A) = D(B_1) D(B_2) \geq \frac{m_1}{m_2} \cdot \frac{m_1}{m_2} > \frac{n}{n}.$$  

which is clearly wrong, since $D(A) \leq D(I/n, \ldots, I/n) = \frac{n}{n}$. Thus both $\mu$ and $\theta$ are multiples of $j$, and $\mu + \theta = \delta \cdot j$.

Substituting in (16) we get

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

for all $i = 1, \ldots, n$. The $\geq$ 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 $A = (A_1, \ldots, A_n)$ be an $n$-tuple of positive semidefinite matrices, and let $A^{\nu}$ be the $n$-tuple obtained from $A$ by replacing $A_j$ with $A_j$. Then

$$D^2(A) \geq D(A^{\nu}) \cdot D(A^{\nu}).$$

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

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

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

$$\delta^2 = D^2(A) \geq D(A^{\nu}) D(A^{\nu}) = \text{tr}(A_j Q_i) \text{tr}(A_j Q_i) > \delta^2,$$

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

**REFERENCES**


