Notes on Egoritsjev's proof of the van der Waerden conjecture

Citation for published version (APA):

Document status and date:
Published: 01/01/1981

Document Version:
Publisher’s PDF, also known as Version of Record (includes final page, issue and volume numbers)

Please check the document version of this publication:
• A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
• The final author version and the galley proof are versions of the publication after peer review.
• The final published version features the final layout of the paper including the volume, issue and page numbers.

Link to publication

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

• Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
• You may not further distribute the material or use it for any profit-making activity or commercial gain
• You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the “Taverne” license above, please follow below link for the End User Agreement:
www.tue.nl/taverne

Take down policy
If you believe that this document breaches copyright please contact us at:
openaccess@tue.nl
providing details and we will investigate your claim.
Notes on Egoritsjev's proof of the van der Waerden conjecture

by

J.H. van Lint
Notes on Egoritsjev's proof of the van der Waerden conjecture

by J.H. van Lint

1. Introduction

If A is an \( n \times n \) matrix with entries \( a_{ij} \) \((i = 1, \ldots, n; j = 1, \ldots, n)\), then the \textit{permanent} of A (notation: \( \text{per} A \)) is defined by

\[
\text{per} A := \sum_{\sigma \in S_n} a_{\sigma(1)} a_{\sigma(2)} \cdots a_{\sigma(n)},
\]

where \( S_n \) denotes the symmetric group on \( n \) symbols. In the following we shall often consider the columns of A as vectors in \( \mathbb{R}^n \) and we write

\[
\text{per} A = \text{per} (\vec{a}_1, \vec{a}_2, \ldots, \vec{a}_n),
\]

where

\[
\vec{a}_j = (a_{1j}, a_{2j}, \ldots, a_{nj}), \quad (j = 1, \ldots, n).
\]

From (1.1) it is clear that \( \text{per} A \) is a linear function of \( \vec{a}_j \) (for each \( j \)).

If we denote the matrix obtained from A by deleting the \( i^{\text{th}} \) row and \( j^{\text{th}} \) column by \( A(i|j) \) then it follows from (1.1) that

\[
\text{per} A = \sum_{i=1}^{n} a_{ij} \text{per} A(i|j), \quad \text{(for any } j\text{)}.
\]
We shall use this same notation (in an obvious way) when more rows and columns are deleted.

If all entries of \( A \) are non-negative and for each row of \( A \) and for each column of \( A \) the sum of the entries is 1, then \( A \) is called a \textit{doubly stochastic matrix}. The class of all such matrices is denoted by \( \Omega_n \). By Birkhoff's theorem \( \Omega_n \) is a convex polyhedron with permutation matrices as vertices (cf. [4], Theorem 3.3). In the interior of \( \Omega_n \) the simplest matrix is the matrix for which every entry is \( n^{-1} \). This matrix is denoted by \( J_n \). Clearly \( \text{per } J_n = n! / n^n \). The following statement is known as the van der Waerden conjecture (cf. [4],[6]):

\[
(1.3) \text{ If } A \in \Omega_n \text{ and } A \neq J_n \text{ then } \text{per } A > \text{per } J_n.
\]

We shall call a matrix \( A \in \Omega_n \) such that \( \text{per } A = \min \{ \text{per } S | S \in \Omega_n \} \) a \textit{minimizing matrix}.

Recently the conjecture (1.3) was proved by G.P. Egoritsjev (cf. [2]). The proof is based on an inequality for permanents which follows from a result of A.D. Alexandroff on positive definite quadratic forms (cf. [1]). The paper by Alexandroff is not easily accessible and it is quite difficult to read. Furthermore the result which he proves is much more general than what is needed for a proof of (1.3). It seems useful to present a direct proof of the special case. This is done in section 2. The proof resembles the proof given by Alexandroff but, following a suggestion by J.J. Seidel, we have chosen a presentation using the concept of a Lorentz space, which makes it easier to understand the inequality.
The other main tool in Egoritsjev's proof is a theorem due to D. London (cf. [3]). In section 3 we give the very short proof of this theorem which was given by H. Minc (cf. [5]). The fact that it may take a while before Egoritsjev's paper is generally accessible and the arguments given above are the motivation for the publication of these notes.

2. Alexandroff's inequality

The following inequality for permanents can be obtained as a special case of a theorem due to A.D. Alexandroff ([1]).

(2.1) Theorem: Let \( \mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_{n-1} \) be vectors in \( \mathbb{R}^n \) with positive coordinates and let \( \mathbf{b} \in \mathbb{R}^n \). Then

\[
\text{(2.2) } \left( \text{per}(\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_{n-1}, \mathbf{b}) \right)^2 \geq \text{per}(\mathbf{a}_1, \ldots, \mathbf{a}_{n-1}, \mathbf{a}_{n-1}) \cdot \text{per}(\mathbf{a}_1, \ldots, \mathbf{a}_{n-2}, \mathbf{b}, \mathbf{b})
\]

and equality holds if and only if \( \mathbf{b} = \lambda \mathbf{a}_{n-1} \) for some constant \( \lambda \).

(2.3) Remark: Clearly the inequality (2.2) is also true if we only require that the coordinates of the vectors \( \mathbf{a}_i \) are non-negative. In that case the claim about the consequence of equality cannot be made.

We shall prove Theorem 2.1 using the concept of a Lorentz space. In the following we consider \( \mathbb{R}^n \) with the standard basis.

(2.4) Definition: The space \( \mathbb{R}^n \) is called a Lorentz space if a symmetric inner product \( \langle \mathbf{x}, \mathbf{y} \rangle = \mathbf{x}^T \mathbf{Q} \mathbf{y} \) has been defined such that \( \mathbf{Q} \) has one positive eigenvalue and \( n - 1 \) negative eigenvalues.
We call a vector $\mathbf{x}$ positive (resp. negative) if $\langle \mathbf{x}, \mathbf{x} \rangle$ is positive (resp. negative) and isotropic if $\langle \mathbf{x}, \mathbf{x} \rangle = 0$. By Sylvester's theorem there is no plane such that $\langle \mathbf{x}, \mathbf{x} \rangle$ is positive on this plane ($\mathbf{x} \neq 0$).

The following lemma is a consequence of this fact.

(2.5) Lemma: If $\mathbf{a}$ is a positive vector in a Lorentz space and $\mathbf{b}$ is arbitrary, then $\langle \mathbf{a}, \mathbf{b} \rangle^2 \geq \langle \mathbf{a}, \mathbf{a} \rangle \cdot \langle \mathbf{b}, \mathbf{b} \rangle$ and equality holds iff $\mathbf{b} = \lambda \mathbf{a}$ for some constant $\lambda$.

Proof: If $\mathbf{b}$ is not a multiple of $\mathbf{a}$ then the plane spanned by $\mathbf{a}$ and $\mathbf{b}$ contains an isotropic vector and a negative vector. Consider $\langle \mathbf{a} + \lambda \mathbf{b}, \mathbf{a} + \lambda \mathbf{b} \rangle$ as a quadratic form in $\lambda$. Since this form is 0 resp. negative for suitable values of $\lambda$ it has a positive discriminant.

Consider vectors $\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_{n-2}$ in $\mathbb{R}^n$ with positive coordinates. Let $\mathbf{e}_i$ denote the $i^{th}$ basis vector ($1 \leq i \leq n$). We define an inner product on $\mathbb{R}$ by

\begin{equation}
\langle \mathbf{x}, \mathbf{y} \rangle := \text{per} \ (\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_{n-2}, \mathbf{x}, \mathbf{y}),
\end{equation}

i.e. $\langle \mathbf{x}, \mathbf{y} \rangle = \mathbf{x}^T \mathbf{Q} \mathbf{y}$,

where $\mathbf{Q}$ is given by

\begin{equation}
q_{ij} := \text{per} \ (\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_{n-2}, \mathbf{e}_i, \mathbf{e}_j)
= \text{per} \ A(i,j \ | \ n - 1, n),
\end{equation}
where $A$ is a matrix with columns $a_1, \ldots, a_n$. Note that at this point we do not use $a_{n-1}$ and $a_n$.

\[ (2.8) \text{ Theorem: } \mathbb{R}^n \text{ with the inner product given by } (2.6) \text{ is a Lorentz space.} \]

**Proof:** We use induction. For $n = 2$ we have $Q = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$ and the assertion is true. Now assume that the assertion is true for $\mathbb{R}^{n-1}$. We shall first show that $Q$ does not have the eigenvalue $0$. Suppose $Qc = 0$. Then by (1.2)

\[ (2.9) \text{ per } (a_1, a_2, \ldots, a_{n-2}, c, e_j) = 0 \text{ for } 1 \leq j \leq n. \]

Consider the $n - 1$ by $n - 1$ matrices

\[ (a_1, \ldots, a_{n-3}, x, y, e_j) \text{ (j|n)} \]

and apply the induction hypothesis and Lemma 2.5. Since $a_{n-2}$ has positive coordinates it follows from (2.9) that for $1 \leq j \leq n$

\[ \text{per } (a_1, \ldots, a_{n-3}, c, c, e_j) \leq 0 \]

and for each $j$ equality holds iff all coordinates of $c$ except $c_j$ are 0. The assumption $Qc = 0$ therefore implies $c = 0$.

Let $\mathbf{1} := (1, 1, \ldots, 1)^T$. We consider the inner product $\mathbf{x}^T Q_{\mathbf{c}} \mathbf{y}$ defined by

\[ \mathbf{x}^T Q_{\mathbf{c}} \mathbf{y} := \text{per}(1 - 3) \mathbf{1} + 3a_1, \ldots, (1 - 3) \mathbf{1} + 3a_{n-2}, x, y). \]
For every $\delta$ in $[0,1]$ this satisfies the condition of the theorem. Hence $Q_0$ does not have the eigenvalue 0. Hence $Q = Q_1$ has the same number of positive eigenvalues as $Q_0$ and since $Q_0$ is a multiple of $nJ_n - I$ this number is one.

The proof of Theorem 2.1 is nothing but the observation that the theorem is a combination of Theorem 2.8 and Lemma 2.5.

3. Earlier results concerning van der Waerden's conjecture

In this section we mention a number of theorems on minimizing matrices which lead to London's theorem, which is the second main tool in the proof of Section 4. Most of these results will be stated without proof since proofs are easily accessible, e.g. in [4].

(3.1) If $A$ is an $n \times n$ matrix with non-negative entries then $\text{per} A = 0$ if and only if $A$ contains an $s \times t$ zero submatrix such that $s + t = n + 1$.

(3.2) $A$ is called **partly decomposable** if it contains a $k \times (n - k)$ zero submatrix. Otherwise, $A$ is called **fully indecomposable**.

(3.3) If $A \in \mathcal{Q}_n$ and $A$ is partly decomposable then there exist permutation matrices $P$ and $Q$ such that $PAQ$ is a direct sum of an element of $\mathcal{Q}_k$ and an element of $\mathcal{Q}_{n-k}$ (for some $k$).

(3.4) If $A \in \mathcal{Q}_n$ then $\text{per} A > 0$.

(3.5) If $A \in \mathcal{Q}_n$ is a minimizing matrix then $A$ is fully indecomposable.

(3.6) If $A \in \mathcal{Q}_n$ is a minimizing matrix and $a_{hk} > 0$ then $\text{per} A(h|k) = \text{per} A$.

In the proof of (3.6) the fact that $a_{hk}$ is positive is exploited to define a set for which $A$ is an interior point and then use Lagrange multipliers (i.e. standard differential calculus). The boundary of $\mathcal{Q}_n$ is considerably more difficult to handle. In fact for the interior of $\mathcal{Q}_n$ the conjecture
has been proved:

(3.7) If \( A \in \Omega_n \) is a minimizing matrix and \( a_{hk} > 0 \) for all \( h \) and \( k \) then \( A = J_n \).

(3.8) **Theorem**: (D. London [3], cf. [4]). If \( A \in \Omega_n \) is a minimizing matrix then \( \text{per} A(i|j) \geq \text{per} A \) for all \( i \) and \( j \).

**Proof**: Let \( P \) be a permutation matrix corresponding to the permutation \( \sigma \). For \( 0 \leq \theta \leq 1 \) define \( f_p(\theta) := \text{per} ((1 - \theta)A + \theta P) \).

By definition we must have \( f_p'(0) \geq 0 \). Since every entry of \( (1 - \theta)A + \theta P \) is a linear function of \( \theta \) we find from (1.2) that

\[
f_p'(0) = \sum_{i=1}^{n} \sum_{j=1}^{n} (-a_{ij} + p_{ij}) \text{per} A(i|j)
\]

\[= \sum_{s=1}^{n} \text{per} A(s|\sigma(s)) - n \text{ per} A.
\]

Therefore we have, for every permutation \( \sigma \),

\[
(3.9) \quad \sum_{s=1}^{n} \text{per} A(s|\sigma(s)) \geq n \text{ per} A.
\]

From (3.5) and (3.1) it follows that for every pair \( i,j \) there is a permutation \( \sigma \) such that \( j = \sigma(i) \) and \( a_{s,\sigma(s)} > 0 \) for \( 1 \leq s \leq n, s \neq i \). This implies (using 3.6) that in (3.9) the terms on the LHS with \( s \neq i \) are equal to \( \text{per} A \) and the result follows. \( \square \)
(3.10) Remark:

It is also known (cf. [4]) that a proof of: "A ∈ Ωₙ a minimizing matrix ⇒ per A(i|j) = per A for all i and j" would imply that (1.3) is true. However, Egoritsjev does not use this fact since it is relatively easy to complete the proof without this statement.

We remark that many of the ideas of this section, e.g. the important result (3.6), are due to M. Marcus and M. Newman.

4. Proof of the van der Waerden conjecture

This section is essentially a translation of the argument given by Egoritsjev in [2]. We first prove a theorem which is known to be sufficient to prove (1.3). (cf. Remark 3.10).

(4.1) Theorem: If A ∈ Ωₙ is a minimizing matrix then per A(i|j) = per A

Proof: Suppose the statement is false. Then by Theorem 3.8 there is a pair r,s such that per A(r|s) > per A. For this r: there is a t such that aᵣₜ > 0.

We now apply Theorem 2.1 (using Remark 2.3)

\[
\begin{align*}
(\text{per } A)^2 &= \text{per } (a_1, \ldots, a_s, \ldots, a_t, \ldots, a_n)^2 \\
&\geq \text{per } (a_1, \ldots, a_s, \ldots, a_t, \ldots, a_n), \text{ per } (a_1, \ldots, a_t, \ldots, a_n) \\
&= \left( \sum_{k=1}^{n} a_{ks} \text{ per } A(k|t) \right) \left( \sum_{k=1}^{n} a_{kt} \text{ per } A(k|s) \right).
\end{align*}
\]

On the RHS every subpermanent is at least per A and per A(r|s) > per A. Since per A(r|s) is multiplied by aᵣₜ, which is positive, the RHS is larger than (per A)^2, a contradiction. □
(4.2) **Lemma:** If $A = (a_1, \ldots, a_n) \in \Omega_n$ is a minimizing matrix and $A'$ is obtained from $A$ by replacing both $a_i$ and $a_j$ by $\frac{1}{2}(a_i + a_j)$ then $A'$ is again a minimizing matrix in $\Omega_n$ and hence Theorem 4.1 applies to $A'$.

**Proof:** It is trivial that $A' \in \Omega_n$. By (1.2) and Theorem 4.1 we have

$$
\operatorname{per} A' = \frac{1}{2} \operatorname{per} A + \frac{1}{4} \operatorname{per}(a_1, \ldots, a_1, \ldots, a_n) + \frac{1}{4} \operatorname{per}(a_1, \ldots, a_j, \ldots, a_j, \ldots, a_n)
$$

$$
= \frac{1}{2} \operatorname{per} A + \frac{1}{4} \sum_{k=1}^{n} a_{k1} \operatorname{per} A(k|j) + \frac{1}{4} \sum_{k=1}^{n} a_{kj} \operatorname{per} A(k|i)
$$

$$
= \operatorname{per} A. \quad \square
$$

Now let $A$ be a minimizing matrix in $\Omega_n$. We consider an arbitrary column of $A$, say $a_n$. From (3.5) it follows that in every row of $A$ there is a positive element in one of the remaining columns. Hence a finite number of applications of Lemma 4.2 yields a minimizing matrix $A'$ which also has $a_n$ as final column and which has as other columns $a_1', \ldots, a_{n-1}'$, each with positive coordinates. We apply Theorem 2.1 to $\operatorname{per}(a_1', \ldots, a_{n-1}', a_n')$. By expanding the permanents on both sides using Theorem 4.1 we see that equality holds. It follows that $a_n$ is a multiple of $a_{n-1}'$ and similarly we find that $a_n$ is a multiple of $a_i'$ for every $i \leq n - 1$. Since $a_1' + \ldots + a_{n-1}' + a_n = 1$ this means that $a_n = n^{-1}$. Since we had taken an arbitrary column of $A$ the proof of (1.3) is now complete.
5. Acknowledgements

The author thanks A. Korlaar for his assistance with the translation of [1] and [2] and J.J. Seidel for the observation that the use of Lorentz spaces makes the proof of Alexandroff's inequality more lucid.

References


