On the Shortest Path between Two Vertices of the Permutation Polytope.

Filip Zieliński

May 30, 2023

Abstract

After we introduce basic theory of convexity we focus on polyhedron theory, especially well-known Birkhoff-von Neumann Theorem and then Schur-Horn Theorem, both connected with special type of permutation-related polytopes. After that we introduce main theorem of this paper, which is we show our shortest path between two vertices in permutation polytope finding algorithm.

Introduction

We work over real numbers $\mathbb{R}$. Notation is to great extend adopted from [1].

1 Preliminaries

Definition 1.1. Convex combination. Let $\{x_1, ..., x_m\}$ be a finite set of points from $\mathbb{R}^d$. A Point

$$x = \sum_{i=1}^{m} \alpha_i x_i$$

where $\sum_{i=1}^{m} \alpha_i = 1$ and α_i is real non-negative number for $i = 1, ..., m$ is called a *convex combination* of $x_1, ..., x_m$.

Definition 1.2. Interval. Given two distinct points $x, y \in \mathbb{R}^d$, the set

$$[x, y] = \{\alpha x + (1 - \alpha)y : 0 \leqslant \alpha \leqslant 1\}$$

of all convex combinations of x and y is called the *interval* with endpoints x and y.

Definition 1.3. Convex set. A set $A \subset \mathbb{R}^d$ is called *convex*, provided $[x, y] \subset A$ for any two $x, y \in A$.

Definition 1.4. Convex hull. For $A \subset \mathbb{R}^d$, the set of all convex combinations of points from A is called *convex hull* and denoted $conv(A)$.

Definition 1.5. Extreme points. Let $A \subset \mathbb{R}^d$ be a set. A point $a \in A$ is called an *extreme point* of A provided for any two points $b, c \in A$ such that $\frac{b+c}{2} = a$ one must have $b = c = a$. The set of all extreme points of A is denoted $ex(A)$.

Definition 1.6. Polytope. The convex hull of a finite set of points in $\mathbb{R}^d$ is called a *polytope*.

Let $c_1, ..., c_m$ be vectors from $\mathbb{R}^d$ and let $\beta_1, ..., \beta_m$ be real numbers and additionally $\langle \cdot, \cdot \rangle$ be the standard inner product. The set:

$$P = \{x \in \mathbb{R}^d : \langle c_i, x \rangle \leqslant \beta_i, i = 1, .., m\}$$

is called *polyhedron*.

Definition 1.7. An extreme point of a polyhedron is called a *vertex*.

Let $P \subset \mathbb{R}^d$ be a polyhedron:

$$P = \{x \in \mathbb{R}^d : \langle c_i, x \rangle \leqslant \beta_i, i = 1, .., m\}$$

where $c_i \in \mathbb{R}^d$ and $\beta_i \in \mathbb{R}$ for $i = 1, .., m$. For $u \in P$ let

$$I(u) = \{i : \langle c_i, u \rangle \leqslant= \beta_i\}$$

be a set of the inequalities that are active on u. Then u is a vertex of P if and only if the set of vectors $\{c_i : i \in I(u)\}$ linearly spans the vector space $\mathbb{R}^d$. In particular if u is a vertex of P the set $I(u)$ contains at least d indices: $|I(u)| >= d$

Definition 1.8. Let σ be a permutation of a set $\{1, .., n\}$. The permutation matrix X^σ is the $n \times n$ matrix $X^\sigma = \xi_{ij}^\sigma : i, j = 1, ..., n$ defined as follows:

$$\xi_{ij}^\sigma = \begin{array}{ll} 1 & \text{if } \sigma(j) = i, \\ 0 & \text{if } \sigma(j) \neq i. \end{array}$$

Notation. In this paper permutation $\sigma = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 3 & 4 & 2 & 1 \end{pmatrix}$ will be denoted as $\sigma = (3, 4, 2, 1)$ because we will strictly associate it with a point in $\mathbb{R}^d$. If we will use notation without a comas e.g. $\sigma = (12)$ we mean a cycle or in this case a transposition of two elements. These notations may be confusing, but used in specific examples, it should be relatively easy to understand the difference.

Definition 1.9. Doubly stochastic Matrix. An $n \times n$ matrix $M = (\alpha_{ij})$ is called *doubly stochastic* if it is non-negative and the sum of elements in every row and every columns equals 1:

1. $\sum_{i=1}^{n} \alpha_{ij} = 1$ for $j = 1, ..., n$,

2. $\sum_{j=1}^{n} \alpha_{ij} = 1$ for $i = 1, ..., n$,

3. $\alpha_{ij} > 0$ for $i, j = 1, .., n$.

We have the following basic fact.

Theorem 1.10. *Every permutation matrix is a doubly stochastic matrix.*

2   Birkhoff - von Neumann Theorem

The polyhedron B_n of all $n \times n$ doubly stochastic matrices is called the *Birkhoff Polytope*. We consider an $n \times n$ matrix X as point in $\mathbb{R}^{n^2}$.

Theorem 2.1. *Birkhoff - von Neumann Theorem.* *The vertices of Birkhoff Polytope B_n are exactly $n \times n$ permutation matrices.*

Proof. First, we will prove the fact, that every permutation matrix is an extreme point of Birkhoff polytope. Let $A = (a_{ij})$ be $n \times n$ permutation matrix, and $B = (b_{ij}), C = (c_{ij})$ are any $n \times n$ doubly stochastic matrices such that: $a = \frac{b+c}{2}$. for exactly n pairs $i, j \in \{1, ..., n\}$ we have $a_{ij} = 1$. if $a_{ij} = 1$ then $b_{ij} = c_{ij} = 1$. For exactly $n^2 - n$ pairs $i, j \in \{1, ..., n\}$ we have $a_{ij} = 0$. For those pairs, $b_{ij} = c_{ij} = 0$. So A, B, C have n exact same elements equal to one and $n^2 - n$ exact same elements equal to zero. Hence A, B, C have exact same n^2 elements, therefore $A = B = C$. Now it suffices to prove that if X is an extreme point of B_n (Birkhoff polytope) $X = X^\sigma$ for some permutation σ. We prove this by induction on n. The case $n = 1$ is obvious. Suppose that $n > 1$. Let us consider the affine subspace $L \subset \mathbb{R}^{n^2}$ consisting of the $n \times n$ matrices $X = (\xi_{ij})$ such that

$\sum_{i=1}^{n} \xi_{ij} = 1$ for $j = 1, ..., n$ and $\sum_{j=1}^{n} \xi_{ij} = 1$ for $i = 1, ..., n$.

We claim that $dim L = (n-1)^2$. Indeed a point (an $n \times n$ matrix X) from L is uniquely determined by an arbitrary choice of the $(n-1)^2$ entries ξ_{ij} for $i, j = 1, ..., n$ since the remaining entries of X are found as $\xi_{in} = 1 - \sum_{j=1}^{n-1} \xi_{ij}$

for $i = 1, ..., n - 1$,
$\xi_{nj} = 1 - \sum_{i=1}^{n-1} \xi_{ij}$ for $j = 1, ..., n - 1$ and
$\xi_{nn} = (2 - n) + \sum_{i,j=1}^{n-1} \xi_{ij}$.

In the space L the polytope B_n is defined by n^2 linear inequalities $\xi_{ij} >= 0$. If X is an extreme point of B_n then some $(n-1)^2$ of these inequalities must be active on X. In other words $\xi_{ij} = 0$ for some $(n-1)^2$ entries of X. Clearly, there cannot be a row containing zeros alone and if every row contained at least two non-zero entries, the total number of zero entries would have been at most $n(n-2) < (n-1)^2$. Therefore there must be a row, say i_0 with $\xi_{ij} = 0$ for all but one $j = j_0$. Now it is clear that $\xi_{i_0 j_0} = 1$ and that all other entries in i_0-th row and j_0-th column must be zero. Crossing out the i_0-th row and j_0-th column, we get an $(n-1) \times (n-1)$ doubly stochastic matrix which must be an extreme point of B_{n-1}, so we may apply the induction hypothesis. $\square$

Definition 2.2. Let us fix a point $x = (\xi_1, ..., \xi_n)$ in $\mathbb{R}^n$. For a permutation σ of the set $\{1, ..., n\}$ let $\sigma(x)$ be the vector $y = (\eta_1, ..., \eta_n)$ where $\eta_i = \xi_{\sigma^{-1}(i)}$ Let S_n be the symmetric group of all permutations of the set $\{1, ..., n\}$. Let us define the *permutation polytope $P(x)$* by

$$P(x) = conv(\sigma(x) : \sigma \in S_n)$$

in words: we permute the coordinates of a given vector x in all possible ways and take the convex hull of resulting vectors.

Definition 2.3. Let us interpret $\mathbb{R}^{n^2}$ as the space of $n \times n$ matrices X. Let us fix a vector $a \in \mathbb{R}^n$. Consider the linear transformation $T : \mathbb{R}^{n^2} \to \mathbb{R}^n$ defined by $T(X) = Xa$.

Lemma 2.4. *$\sigma(x) = X^\sigma x$, where X^σ is the permutation matrix corresponding to σ.*

**Theorem 2.5.** *$T(B_n) = P(a)$ where B_n is a Birkhoff Polytope.*

Now we will introduce definitions and lemmas required for next theorem regarding permutation polytope. These definitions are very basic, they were put in here only as a reminder.

Definition 2.6. Real square matrix M is called *orthogonal matrix* if and only if $MM^T = M^T M = I$. Orthognal matrix can be also defined as matrix whose columns and rows are orthonormal vectors.

Lemma 2.7. *let U be real orthogonal matrix such as $U = (\xi_{ij})$. Then $M = (\xi_{ij}^2)$ is an doubly stochastic matrix.*

Definition 2.8. Real square matrix M is called *symmetric matrix* if and only if $M = M^T$

Figure 1: Permutation polytope of set $\{1, 2, 3, 4\}$
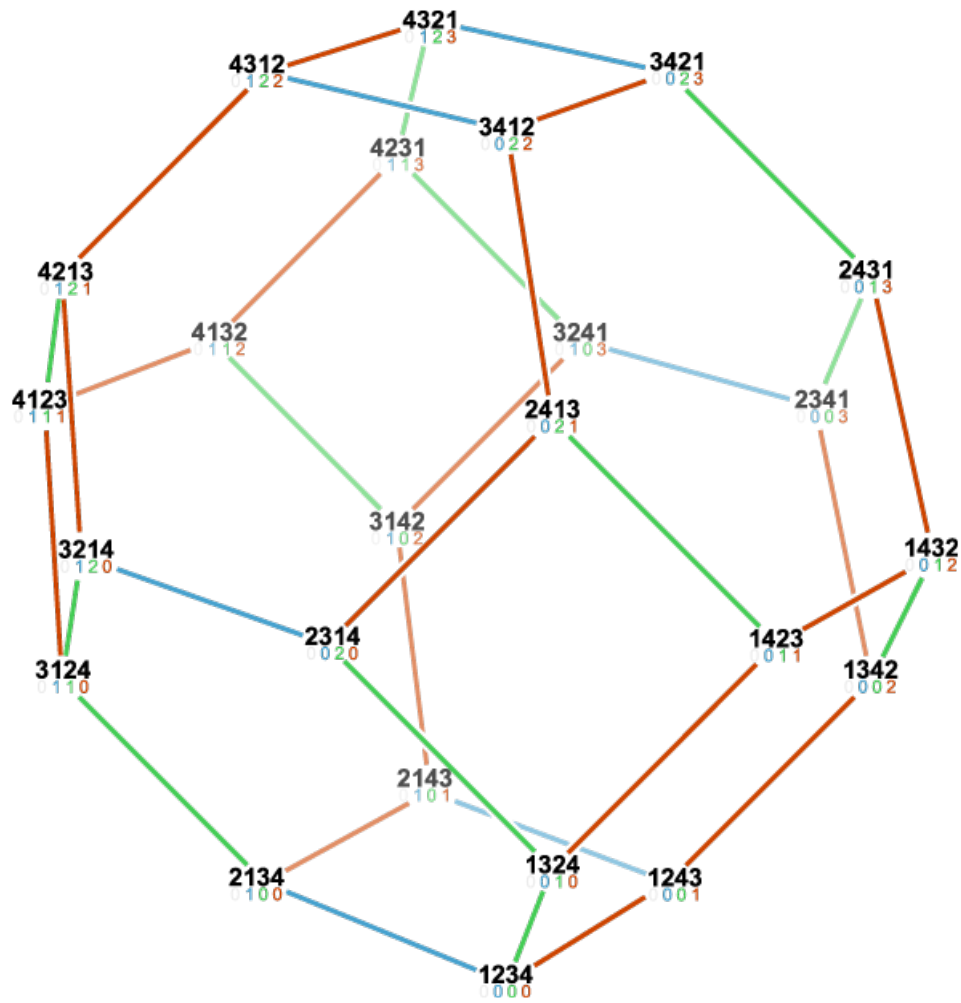
Lemma 2.9. *Symmetric matrices are always diagonalizable*

Lemma 2.10. *Eigenvectors of symmetric matrix are orthogonal vectors*

Corollary 2.11. *If matrix M is symmetric then:*

$$M = UDU^T$$

for some orthogonal matrix U.

3   Schur-Horn Theorem.

Theorem 3.1. [1] **Schur-Horn Theorem.** *Let us fix a positive integer n and real numbers $\lambda_1, ..., \lambda n$. Let $l = (\lambda_1, ..., \lambda n) \in \mathbb{R}^n$ be a vector.*

1. *Let $A = (\alpha_{ij})$ be a $n \times n$ real symmetric matrix with the eigenvalues $\lambda_1, ..., \lambda_n$. Then the diagonal $a = (\alpha_{11}, ..., \alpha_{nn}$ lies in the permutation polytope $P(l) : a \in P(l)$ (Schur's Theorem).*

2. *Let $a \in P(l)$ be a point from the permutation polytope. Then there exists an $n \times n$ real symmetric matrix $A = (\alpha_{ij})$ with the eigenvalues $\lambda_1, ..., \lambda_n$ and the diagonal $a = (\alpha_{11}, ..., \alpha_{nn})$*

We will here only present proof of part 1 of this theorem which is Schur's theorem.

Proof. Let D $= diag(\lambda_1, ..., \lambda_n)$ be diagonal matirx. Suppose that $A = (\alpha_{ij})$ is a real symmetric $n \times n$ matrix with eigenvalues $\lambda_1, ..., \lambda_n$. Then the $A = UDU^T$ for some orthogonal matrix $U = (\xi_{ij})$. Hence the diagonal entries of A can be written as

$$\alpha_{ij} = \sum_{i=1}^{n} \xi_{ki}^2 \lambda_i.$$

Let $B = (\beta_{ij})$ be the $n \times n$ matrix such that $\beta_{ij} = \xi_{ij}^2$. Hence we may write $a = Bl$ where a and l are interpreted as n-columns of real numbers. Since U is an orthogonalmatrix, the matrix B is doubly stochastic. By the Birkhoff - von Neumann Theorem, B can be written as a convex combination of permutation matrices $X^\sigma, \sigma \in S_n$. Therefore we conlude that a is a convex combination of $\sigma(l) = X^\sigma l$, that is, $a \in P(l)$ $\qquad \square$

4   Shortest path between two vertices of permutation polytope

Now we will look at a permutation polyhedron from different perspective.

Definition 4.1. Metric space. Ordered pair (M, d) where M is a set and $d : M \times M \to [0, +\infty)$ is a function such that for all $x, y, z \in M$:

1. $d(x, y) = 0 \Leftrightarrow x = y$

2. $d(x, y) \leqslant d(x, z) + d(y, z)$

is called a *metric space*.

Definition 4.2. Graph ordered pair $G = (V, E)$ is called a *graph*. $V = \{v_1, ..., v_n\}$ is a set of an elements called a *vertices* and $E = \{e_1, ...e_m\}$ where $e_i = \{v_j, v_l\} \subset V$ for some $j, l = 1, ...n$, $j \neq l$ for $i = 1, ..., m$.

Graphs are easily connected with geometrical structure, therefore we will try now to look at permutation polyhedron as a graph. The vertices and edges of polyhedron will become vertices and edges of graph.

Definition 4.3. Path. Let $G = (V, E)$ be a graph. A path P between vertices $v_1, v_{n+1} \in V$ is a sequence of edges $P = (e_1, ..., e_n), e_i \in E$ for $i = 1, .., n$ such that:

1. $e_1 = (v_1, v_2)$ for some $v_2 \in V$,

2. $e_n = (v_n, v_{n+1})$ for some v_n

3. $e_i \cap e_{i+1} = vi + 1$ for $i = 1, ...n - 1$

3. $e_i \cap ej \neq \emptyset \Rightarrow i = j$ or $i = j + 1$ or $i = j - 1$ for $i, j = 2, ..., n - 1$

length of path P is simply number of elements in sequence. Here length is equal to n.

In words: path is just possible way of getting from one vertex to another using existing edges, without using the same vertex more than once. The length of a path is the number of edges we use. If graph is connected (there exist path between every two vertices) for every pair of vertices we can find the length of the shortest path (or many paths, sharing the same length). We assume that path between vertices v and v has always the length of 0.

Definition 4.4. A graph is said to be connected if every pair of distinct vertices v_i, v_j in the graph exists a path $(e_k, ..., e_n)$ such that $v_i \in e_k$ and $v_j \in e_n$.

Definition 4.5. Let $G = (V, E)$ be a connected graph. We define $d : V \times V \to \mathbb{N} \cup \{0\}$ such that $d(v_1, v2) = n$ where n is the length of shortest path between vertices v_1, v_2 for every v_1, v_2.

Theorem 4.6. (V, d) *is a metric space.*

Proof. First condition of metric space $d(v_1, v_2) = 0 \Leftrightarrow v_1 = v_2$ is fulfilled because the implication from right to left is given by definition and implication from left to right is obvious due to the fact that any path from v_1 to v_2 for $v_1 \neq v_2$ has at least length of 1. Second condition of metric space can be replaced with two simpler ones.

2a. $d(v_1, v_2) = d(v_2, v_1)$ for every $v_1, v_2 \in V$

2b. $d(v_1, v_2) \leqslant d(v_1, v_2) + d(v_2, v_3)$ for every $v_1, v_2, v_3 \in V$.

Usual definition of metric space uses those two conditions, not the one we used, but it is easy to see that those two conditions imply the second statement of definition presented in this paper. First one is condition of being symmetric and second one is called triangle inequality.

If $d(v_1, v_2) = n$ therefore there exists path $P = (e_1, ...e_n)$ of length n. Consider path $P' = (e_n, ..., e_1)$. It meets all the requirements of path and also has the length equal to n. Hence $d(v_2, v_1) = n$.

We will prove the last condition by proof by contradiction. Assume $d(v_1, v_2) = n$ and $d(v_2, v_3) = m$ and $d(v_3, v_1) = k$ such that $n > m + k$. Therefore exist path $P_{12} = (e_1, ..., e_n)$, path $P_{13} = (e'_1, ...e'_m)$ and path $P_{32} = (e''_1, ..., e''_k)$. Hence we can create $P' = (e'_1, ..., e'_m, e''_1, ..., e''_k)$ which is either a path of length $k + m < n$ or by removing some elements of it, we can create a path of length $l < k + m < n$. Hence path P_{12} is not the shortest path between vertices v_1, v_2, which is a contradiction with our earlier assumption.

$\square$

Problem 4.7. Let v_1, v_2 be any two vertices on n-dimensional permutation polytope and d be a metric defined as earlier. What is $d(v_1, v_2)$?

First, we need to think about the problem when two vertices of permutation polytope are connected.

Theorem 4.8. *let v_1, v_2 be two vertices on n-dimensional permutation polytope represented respectfully by permutation σ_1, σ_2. There exists an edge between v_1, v_2 if and only if $\sigma_1 = \sigma_2 \circ \sigma_3$ where σ_3 is a transposition of two next elements $(ii+1)$, $i = 1, ..., n-1$.*

Proof. We will skip the proof of this fact. It can be found in [2]. $\qquad\square$

In words: consider permutation polytope created by all permutations of point $x = (1, 2, 3, 4)$. If we look at vertex $v_1 = (3, 2, 1, 4)$ it is directly connected with 3 other vertices which are v_1 composed with transpositions $(12), (23), (34)$. For example by composition with (34) we get vertex $v_2 = (4, 2, 1, 3)$ which is directly connected with vertex v_1. Now we have to count how many transpositions between two next elements we have to use to get from one vertex to another. Consider $n \times n$ permutation matrix X^σ which is a representation of permutation σ. Composition of σ with a transposition (ij), for $i, j = 1, ..., n$, might be represented with switching i-th and j-th row in X^σ.

Example 4.9. if n = 4 and $\sigma = (4, 2, 1, 3)$ then:

$$X^\sigma = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \end{bmatrix}$$

then composition of σ with (12) is $\sigma' = (4, 1, 2, 3)$ represented by matrix

$$X^{\sigma'} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \end{bmatrix}$$

We can see that $X^{\sigma'}$ is X^σ with 1-st and 2-nd row switched.

Consider two vertices v_1, v_2 which are represented respectfully by permutation of n-elements set σ_1, σ_2. σ_1 is represented by some $n \times n$ $X^{\sigma_1} = (a_{ij}), i, j = 1, .., n$ permutation matrix and σ_2 is represented by some $n \times n$ $X^{\sigma_2} = (b_{ij}), i.j = 1, ..., n$ permutation matrix. Consider first row of X^{σ_1}. There exists exactly one j from $j = 1, ..., n$ such that $a_1 j = 1$ and there exists exactly one i from $i = 1, ..., n$ such that $b_{ij} = 1$. If we want to "transform" X^{σ_1} into X^{σ_2} only by switching adjacent rows, then we must push i-th row of X^{σ_1} to the top, to match the first row of X^{σ_2}. Hence first we need to switch the i-th row with $i-1$-th row, then $i-1$-th row with $i-2$-th row, and so on until currently $i-th$ row will be first, matching the first row in second matrix. To do this we have to use total of $i-1$ transpositions. Then we can cross out the first row and j-th column and we can repeat the process again on $(n-1) \times (n-1)$ matrix.

Example 4.10. Consider $\sigma_1 = (3, 4, 2, 1)$ and $\sigma_2 = (2, 4, 1, 3)$. Therefore:

$$X^{\sigma_1} = \begin{bmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix} \text{ and } X^{\sigma_2} = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \end{bmatrix} \quad a_{14} = 1 \text{ and } b_{34} = 1 \text{ so we need to push}$$

3-rd row of X^{σ_2} to the top. To do this we have to compose X^{σ_2} first with transposition (23)

and second with transposition (12). $\sigma'_2 = \sigma_2 \circ (23) = (3, 4, 1, 2)$. $X^{\sigma'_2} = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix}$ As

you can see by composing matrix with transposition (23) we only swapped 2-nd and 3-rd row. $\sigma''_2 = \sigma'_2 \circ (12)$. $X^{\sigma''_2} = \begin{bmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix}$ Now we have after two swaps we have the same

first row in X^{σ_1} and $X^{\sigma''_2}$ therefore we can cross out the first row and fourth column and repeat algorithm.

$$X_2^{\sigma_1} = \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}, X_2^{\sigma_2} = \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$$

If the remaining matrices weren't exact same we would repeat given algorithm until two matrices become identical. Now we can count how many transposition we did have to make, in this case it was only two, $(23), (12)$. Hence there exist a path between $v_1 = (3, 4, 2, 1)$ and $v_2 = (2, 4, 1, 3)$ of length 2. With this algorithm we can easily track down how this path exactly looks, in this case we started from v_2, by transposition $(2, 3)$ we travelled to $v = (3, 4, 1, 2)$ and then using transposition $(1, 2)$ we went to $(3, 4, 2, 1) = v_1$.

Theorem 4.11. *__Main Theorem.__ The path-finding algorithm presented, finds always the path with the shortest length possible.*

Proof. Notice, that by transporting one row to the top, we do not change the order between other rows, which means that if we move i-th row to the top, then for every $k, l = 1, ..., n$ such that $i \neq k, i \neq l$ if $k > l$ then after i-th row reaches the top the row representing the old k-th row is still below the row representing the old l-th row. Using that, and the fact that if we want to transform one matrix to the other we need to somehow get first row to be the same, we can see that there is no faster way to do it. $\square$

There are several interesting problems regarding permutation polytope. Even in the shortest path finding algorithms field there are still questions unanswered, for example how many paths share exact same shortest length between two vertices. We can clearly see that we can obtain another path by alternating algorithm, to push the rows to the bottom instead to the top, but we cannot be sure that that is all what is left. Permutation polytope is very interesting structure because it shows the deep connection between algebraic structures and combinatorics.

References

[1] Alexander Barvinok. *A Course in Convexity*

[2] G. Lancia, P. Serafini. *Compact Extended Linear Programming Models*

[3] Micheal X. Goemans. *Smallest Compact Forumlation for the Permutahedron*