Improvements on Permutation Reconstruction from Minors

Yiming Ma, Wenjie Zhong and Xiande Zhang

Abstract

We study the reconstruction problem of permutation sequences from their k-minors, which are subsequences of length k with entries renumbered by $1, 2, \ldots, k$ preserving order. We prove that the minimum number k such that any permutation of length n can be reconstructed from the multiset of its k-minors is between $\exp(\Omega(\sqrt{\ln n}))$ and $O(\sqrt{n \ln n})$. These results imply better bounds of a well-studied parameter N_d, which is the smallest number such that any permutation of length $n \geq N_d$ can be reconstructed by its $(n-d)$-minors. The new bounds are $d + \exp(\Omega(\sqrt{\ln d})) < N_d < d + O(\sqrt{d \ln d})$ asymptotically, and the previous bounds were $d + \log_2 d < N_d < d^2/4 + 2d + 4$.

1. Introduction

Reconstructing a combinatorial object from a limited amount of sub-information is a fundamental problem in computer science. Based on different combinatorial objects, different reconstruction problems have been widely studied due to their applications in bioinformatics [1], [2], information theory [3], and DNA based data storage [4]–[6].

Permutation reconstruction is a variant of the well-known graph reconstruction problem, which arose from the unsolved conjecture of Ulam [7]: any simple graph with at least three vertices can be determined up to isomorphism by the multiset of all its reduced subgraphs with one vertex deleted. For permutation reconstruction, one considers reconstructing a permutation from its k-minors, that is, subsequences of length k with entries renumbered by $1, 2, \ldots, k$ preserving order. The multiset of all its k-minors is called its k-deck. In 2006, Smith [8] introduced the notation N_d, which is the smallest number such that any permutation of length $n \geq N_d$ can be reconstructed by its $(n-d)$-deck. Raykova [9] showed the existence of N_d and gave the bounds $d + \log_2 d < N_d < d^2/4 + 2d + 4$.

In this paper, we introduce another notation $s(n)$ for given n, that is the smallest integer k such that any permutation of length n can be reconstructed from its k-deck, or equivalently, any two permutations of length n have different k-decks. We give lower and upper bounds of $s(n)$ for large n,

$$3^{0.811 \times \log_3^{1/2}(n+1)} \leq s(n) \leq 2\lceil \sqrt{(n-2)\ln(n-3)} \rceil + 2,$$

which imply much better bounds of N_d,

$$d + \exp(\Omega(\sqrt{\ln d})) < N_d < d + O(\sqrt{d \ln d})$$

for large d. We also provide a feasible algorithm which assists to determine the exact values of $s(n)$ for $n \leq 10$.

Y. Ma (mym024@mail.ustc.edu.cn) is with the School of Cyber Security, University of Science and Technology of China, Hefei, 230026, Anhui, China.

W. Zhong (zhongwj@mail.ustc.edu.cn) is with the School of Mathematical Sciences, University of Science and Technology of China, Hefei, 230026, Anhui, China.

X. Zhang (drzhangx@ustc.edu.cn) is with the School of Mathematical Sciences, University of Science and Technology of China, Hefei, 230026, Anhui, China, and with Hefei National Laboratory, University of Science and Technology of China, Hefei, 230088, Anhui, China.

The research is supported by the National Key Research and Development Programs of China 2023YFA1010201 and 2020YFA0713100, the NSFC under Grants No. 12171452 and No. 12231014, and the Innovation Program for Quantum Science and Technology (2021ZD0302902).

A. Related work

The problem of reconstructing a sequence from the multiset (i.e., the k-deck) of all its subsequences of length k, was introduced by Kalashnik [10] in an information-theoretic study about deletion channels. The main task is to determine the minimum number k such that one can reconstruct any binary sequence of length n from its k-deck. The best known bounds of the minimum k are $\exp(\Omega\sqrt{\ln n})$ [11] and $O(\sqrt{n})$ [12]. This deck problem has been extended to matrices [13] and general higher dimensions [14], for which the minimum number k such that one can reconstruct any binary d-dimensional hypermatrix of order n from its k-deck, the multiset of its sub-hypermatrices of order k, is $O(n^{\frac{d}{d+1}})$.

The problem of partition reconstruction is to ask for which n and k one can uniquely determine any partition of n from its set of k-minors [15], [16]. Here, a k-minor of a partition λ of a positive integer $n > k$ is a partition of $n-k$ whose Young diagram fits inside that of λ. Monks [17] showed that partitions of $n \geq k^2+2k$ are uniquely determined by their sets of k-minors, which is best possible. Cain and Lehtonen [18] completely characterized the standard Young tableaux that can be reconstructed from their sets or multisets of 1-minors.

For permutation reconstruction, Gouveia and Lehtonen [19] showed that every permutation of length $n \geq 5$ is reconstructible from any $\lceil n/2 \rceil + 2$ of its $(n-1)$-minors. Reconstruction of permutations from other types of minors is also well studied. Monks [20] showed that any permutation of $[n]$ can be reconstructed from its set of *cycle minors* if and only if $n \geq 6$. Lehtonen [21] showed that every permutation of a finite set with at least five elements is reconstructible from its identification minors. De Biasi [22] proved that the problem of reconstructing a permutation given the absolute differences of consecutive entries is NP-complete.

B. Organization

The paper is organized as follows. In Section 2, we give necessary definitions and notations, then give a lower bound of $s(n)$. In Section 3, we analyze the relationship between the reconstructibility and some specific functions, and give an upper bound of $s(n)$. In Section 4, we study bounds and exact values of $s(n)$ for small n by giving an efficient algorithm, and deduce much better bounds of N_d than that in [9] from our bounds of $s(n)$.

2. Notations

For positive integers n and $n_1 < n_2$, let $[n] := \{1, 2, \ldots, n\}$ and $[n_1, n_2] := \{n_1, n_1+1, \ldots, n_2\}$. For two sequences $\mathbf{x}$ and $\mathbf{y}$, let $\mathbf{x} \mid \mathbf{y}$ be the concatenation of them.

Let $\mathcal{S}_n$ be the set of all permutations on $[n]$. For a permutation $\mathbf{x} \in \mathcal{S}_n$, a *$k$-minor* of $\mathbf{x}$ is a subsequence of $\mathbf{x}$ with length k whose entries are renumbered by $1, 2, \ldots, k$ preserving order. For example, deleting the first entry of $\mathbf{x} = 25134 \in \mathcal{S}_5$, we get a subsequence 5134 of length four. Then renumber the entries by $\{1, 2, 3, 4\}$ preserving order, we have a 4-minor 4123 of $\mathbf{x}$. The *k-deck* of $\mathbf{x}$, denoted by $D_k(\mathbf{x})$, is the multiset of all k-minors of $\mathbf{x}$. If a permutation $\mathbf{x} \in \mathcal{S}_n$ can be uniquely determined by $D_k(\mathbf{x})$, we say $\mathbf{x}$ is *k-reconstructible*. Moreover, if $D_k(\mathbf{x}) = D_k(\mathbf{y})$ for two different permutations $\mathbf{x}, \mathbf{y}$ in $\mathcal{S}_n$, we say $\mathbf{x}$ and $\mathbf{y}$ are k-equivalent, and write $\mathbf{x} \overset{k}{\sim} \mathbf{y}$.

Example 2.1. *For $\mathbf{x} = 13524 \in \mathcal{S}_5$, $D_4(\mathbf{x}) = \{2413, 1423, 1324, 1243, 1342\}$. It can be verified by computer that different permutations in $\mathcal{S}_5$ have different 4-decks, so $\mathbf{x}$ is 4-reconstructible. However, one can check that $13524 \overset{3}{\sim} 14253$, that is, $\mathbf{x}$ is not 3-reconstructible.*

It is easy to see that for any two $\mathbf{x}, \mathbf{y} \in \mathcal{S}_n$, $D_l(\mathbf{x}) = D_l(\mathbf{y})$ implies that $D_k(\mathbf{x}) = D_k(\mathbf{y})$ for any $k \leq l$. That is, $\mathbf{x} \overset{l}{\sim} \mathbf{y}$ implies that $\mathbf{x} \overset{k}{\sim} \mathbf{y}$ for any $k \leq l$. So if $\mathbf{x}$ is k-reconstructible, then $\mathbf{x}$ is l-reconstructible for any $k < l \leq n$. By this fact, it is interesting to consider the following problem.

Q: Given n, determine the least integer $k \leq n$ such that any $\mathbf{x} \in \mathcal{S}_n$ is k-reconstructible. Denote this number by $s(n)$.

By Example 2.1, $s(5) = 4$. The parameter $s(n)$ is an increasing function of n by the following result.

Proposition 2.1. *For any positive integer n, we have $s(n) \leq s(n+1)$.*

Proof. It is clear that $s(1) = 1, s(2) = 2$. So $s(n) \leq s(n+1)$ holds for $n = 1$. Assume that $n \geq 2$. Since all permutations in $\mathcal{S}_n$ have the same 1-deck, we have $s(n) \geq 2$. Let $k = s(n) - 1 \geq 1$. By the definition of $s(n)$, there exist two distinct permutations $\mathbf{x}, \mathbf{y} \in \mathcal{S}_n$ such that $D_k(\mathbf{x}) = D_k(\mathbf{y})$, and hence $D_{k-1}(\mathbf{x}) = D_{k-1}(\mathbf{y})$. Then for two distinct permutations $\mathbf{x} \mid (n+1), \mathbf{y} \mid (n+1) \in \mathcal{S}_{n+1}$, we have

$$\begin{aligned} D_k(\mathbf{x} \mid (n+1)) =& D_k(\mathbf{x}) \uplus \{\mathbf{w} \mid k : \mathbf{w} \in D_{k-1}(\mathbf{x})\} \\ =& D_k(\mathbf{y}) \uplus \{\mathbf{w} \mid k : \mathbf{w} \in D_{k-1}(\mathbf{y})\} = D_k(\mathbf{y} \mid (n+1)). \end{aligned}$$

That is, $\mathbf{x} \mid (n+1)$ and $\mathbf{y} \mid (n+1)$ have the same k-deck, which implies that $s(n+1) \geq k+1 = s(n)$. This completes the proof. □

In order to provide a lower bound of $s(n)$, we establish a mapping between binary sequences and permutations, then adopt the result in [12] for sequence reconstructions. For a sequence $\mathbf{p} \in \{0,1\}^n$, let $D'_k(\mathbf{p})$ be its k-deck, i.e., the multiset of all its subsequences of length k. Let $s'(n)$ denote the smallest integer k such that any sequence in $\{0,1\}^n$ can be reconstructed from its k-deck, or equivalently, any two binary sequences of length n have different k-decks.

Theorem 2.1. *For any positive integer n, $s(n) \geq s'(n)$.*

Proof. Let $k = s'(n) - 1$. Then there exist two sequences $\mathbf{p}, \mathbf{q} \in \{0,1\}^n$ satisfying $D'_k(\mathbf{p}) = D'_k(\mathbf{q})$. It suffices to construct two distinct permutations $\mathbf{x}, \mathbf{y} \in \mathcal{S}_n$ such that $D_k(\mathbf{x}) = D_k(\mathbf{y})$, then by the definition of $s(n)$, we have $s(n) > k = s'(n) - 1$, i.e., $s(n) \geq s'(n)$.

Define a mapping Ψ from $\{0,1\}^n$ to $\mathcal{S}_n$ as follows. Suppose $\mathbf{p} \in \{0,1\}^n$ has m ones. Then Ψ maps $\mathbf{p}$ to a permutation $\mathbf{x} \in \mathcal{S}_n$ by changing the m ones in $\mathbf{p}$ into $[m]$ with increasing order and preserving index positions, and changing the $(n-m)$ zeros in $\mathbf{p}$ into $[m+1, n]$ with increasing order and preserving index positions. For example, if $\mathbf{p} = 0010011$, then $\mathbf{x} = 4516723$. It is easy to see that, every k positions in $[n]$ gives a k-subsequence $\mathbf{w}$ of $\mathbf{p}$ and a k-minor $\mathbf{z} \in \mathcal{S}_k$ of $\mathbf{x}$ with $\mathbf{z} = \Psi(\mathbf{w})$.

Let $\mathbf{y} = \Psi(\mathbf{q})$. Then $D'_k(\mathbf{p}) = D'_k(\mathbf{q})$ implies that $D_k(\mathbf{x}) = D_k(\mathbf{y})$. So we have $s(n) \geq s'(n)$. □

Known values of $s'(n)$ for small n from [11], [12] are listed below. Comparing with Table II in Section 4, we known that $s(n) > s'(n)$ in general.

n	1	2	3	4	5	6	7	8	9	10	11
$s'(n)$	1	2	2	3	3	3	4	4	4	4	4
n	12	13	14	15	16	17	18	19	20	21	22
$s'(n)$	5	5	5	5	$[5,8]$	$[5,8]$	$[5,8]$	$[5,8]$	$[5,8]$	$[5,8]$	$[5,8]$

TABLE I: Values of $s'(n)$ for small n. $[\cdot,\cdot]$ means lower and upper bounds.

In [12], the authors considered a dual parameter of $s'(n)$: Given k, let $S(k)$ denote the least integer $n > k$ such that there exist distinct binary sequences $\mathbf{p}, \mathbf{q} \in \{0,1\}^n$ with the same k-deck. By [12],

$$S(k) \leq 1.2\Gamma(\log_3 k) \times 3^{3/2 \log_3^2 k - 1/2 \log_3 k}$$

for $k \geq 5$. By assist of Mathematica, we obtain the following result for $s(n)$.

Theorem 2.2. $s(n) \geq 3^{0.811 \times \log_3^{1/2}(n+1)}$, $n \geq 16$.

Proof. Let $f(k) = \lfloor 1.2\Gamma(\log_3 k) 3^{3/2 \log_3^2 k - 1/2 \log_3 k} \rfloor$ for $k \geq 5$, which is the upper bound of $S(k)$. Then $f(k)$ is a non-decreasing function, and $f(5) = 16$. Now for any integer n, let k be the integer satisfying

$f(k) \leq n < f(k+1)$. By definition, there exist distinct sequences $\mathbf{p}, \mathbf{q} \in \{0,1\}^n$ with the same k-deck. This means $s'(n) \geq k+1$.

Let $k = 3^{0.811 \times \log_3^{1/2}(n+1)}$, we have $f(k) \leq n$ which can be verified by Mathematica. By Theorem 2.1, we obtain the lower bound

$$s(n) \geq s'(n) \geq 3^{0.811 \times \log_3^{1/2}(n+1)}.$$

$\square$

3. An Upper Bound of $s(n)$

In this section, we provide an upper bound of $s(n)$, where the idea is from [23] working on reconstruction of sequences.

For a permutation $\mathbf{z} = z_1 \ldots z_k \in \mathcal{S}_k$, define its indicator function of (i,j)-order as:

$$z_{ij} = \begin{cases} 1, & z_i < z_j \\ 0, & z_i > z_j \end{cases} \qquad 1 \leq i < j \leq k.$$

Given a permutation $\mathbf{x} = x_1 \ldots x_n \in \mathcal{S}_n$ and its k-deck $D_k(\mathbf{x})$, let $S_{ij}(\mathbf{x})$ be the total sum of (i,j)-orders of all minors in $D_k(\mathbf{x})$. That is, $S_{ij}(\mathbf{x}) = \sum_{\mathbf{z} \in D_k(\mathbf{x})} z_{ij}$, $1 \leq i < j \leq k$. The next lemma relates the value of $S_{ij}(\mathbf{x})$ to the orders of $\mathbf{x}$.

Lemma 3.1. *For any* $1 \leq i < j \leq k$*,*

$$S_{ij}(\mathbf{x}) = \sum_{1 \leq i' < j' \leq n} \binom{i'-1}{i-1} \binom{j'-i'-1}{j-i-1} \binom{n-j'}{k-j} x_{i'j'}.$$

Proof. For any $\mathbf{z} \in D_k(\mathbf{x})$, suppose that z_i, z_j are originally $x_{i'}$, $x_{j'}$ for some i' and j', respectively. Write

$$\mathbf{x} = \underbrace{\ldots}_{i'-1} \quad x_{i'} \quad \underbrace{\ldots}_{j'-i'-1} \quad x_{j'} \quad \underbrace{\ldots}_{n-j'}.$$

Then the number of such $\mathbf{z} \in D_k(\mathbf{x})$ with z_i, z_j originally from $x_{i'}$, $x_{j'}$ is

$$\binom{i'-1}{i-1} \binom{j'-i'-1}{j-i-1} \binom{n-j'}{k-j}.$$

Then $S_{ij}(\mathbf{x})$ can be obtained by summing up them multiplied by $x_{i'j'}$. $\square$

Suppose two permutations $\mathbf{x} = x_1 \ldots x_n, \mathbf{y} = y_1 \ldots y_n$ in $\mathcal{S}_n$ satisfy $\mathbf{x} \overset{k}{\sim} \mathbf{y}$. Then $D_k(\mathbf{x}) = D_k(\mathbf{y})$ and thus $S_{ij}(\mathbf{x}) = S_{ij}(\mathbf{y})$ for $1 \leq i < j \leq k$. Define $\delta_{ij} := x_{ij} - y_{ij} \in \{0, \pm 1\}$ for $1 \leq i < j \leq n$. We have the following equalities by Lemma 3.1,

$$\sum_{1 \leq i' < j' \leq n} \binom{i'-1}{i-1} \binom{j'-i'-1}{j-i-1} \binom{n-j'}{k-j} \delta_{i'j'} = 0, \quad 1 \leq i < j \leq k. \tag{3.1}$$

Consider bivariate polynomials $f_{ij}(x,y) = \binom{x-1}{i-1}\binom{y-x-1}{j-i-1}\binom{n-y}{k-j}$, $1 \leq i < j \leq k$. Note that $\deg f_{ij} = k-2$ and $f_{ij}(x,y) = 0$ for $0 \leq x < i$, or $y - x < j - i$, or $n - k + j < y \leq n$.

Lemma 3.2. *For fixed integers* $n \geq k \geq 1$*, the set* $\{f_{ij}\}_{1 \leq i < j \leq k}$ *is a basis for the space of bivariate polynomials of degree at most* $k-2$*.*

Proof. Consider $\varphi(x,y) = \sum_{1 \leq i < j \leq k} \mu_{ij} f_{ij}(x,y) = \sum_{1 \leq i < j \leq k} \mu_{ij} \binom{x-1}{i-1}\binom{y-x-1}{j-i-1}\binom{n-y}{k-j}$. It suffices to show that $\varphi(x,y)$ is not identically zero whenever the coefficients μ_{ij} are not all zero. Assume $\mu_{i_0 j_0}$ is the first nonzero coefficient in lexicographical order, that is, whenever $i < i_0$ or $i = i_0$ and $j < j_0$, $\mu_{ij} = 0$. Then

$\varphi(i_0,j_0)=\mu_{i_0j_0}\binom{n-j_0}{k-j_0}$, since $\binom{i_0-1}{i-1}=0$ for $i>i_0$ and $\binom{j_0-i_0-1}{j-i_0-1}=0$ for $i=i_0$ and $j>j_0$. Hence $\varphi(i_0,j_0)\neq 0$, that is, $\varphi(x,y)$ is not identically zero. $\square$

Combining Lemma 3.2 and Eq. (3.1), we obtain the following necessary condition for $\mathbf{x}\overset{k}{\sim}\mathbf{y}$.

Corollary 3.1. *If $\mathbf{x}\overset{k}{\sim}\mathbf{y}$ in $\mathcal{S}_n$, then for any bivariate polynomial $\varphi(x,y)$ of degree at most $k-2$,*

$$\sum_{1\le x<y\le n}\delta_{xy}\varphi(x,y)=0.$$

For easy application, we obtain the following sufficient condition for the reconstructibility by changing bivariate polynomials in Corollary 3.1 to univariate polynomials.

Theorem 3.1. *If there exists a polynomial $\phi(x)$ satisfying $\deg\phi\le k-2$ and $\phi(-1)>(n-2)\sum_{x=0}^{n-3}|\phi(x)|$, then any permutation $\mathbf{x}\in\mathcal{S}_n$ is k-reconstructible.*

Proof. Assume on the contrary that there exist $\mathbf{x}\neq\mathbf{y}\in\mathcal{S}_n$ with $\mathbf{x}\overset{k}{\sim}\mathbf{y}$. For $1\le i\le n-1$, define

$$\delta_i=\sum_{i<j\le n}\delta_{ij}=\sum_{i<j\le n}(x_{ij}-y_{ij})=\sum_{i<j\le n}x_{ij}-\sum_{i<j\le n}y_{ij}.$$

Let m be the first integer in $[n-1]$ satisfying $x_m\neq y_m$. Note that $\sum_{i<j\le n}x_{ij}$ is the number of $j>i$ satisfying $x_j>x_i$. So we have $\sum_{i<j\le n}x_{ij}=\sum_{i<j\le n}y_{ij}$ for all $i\in[m-1]$, and $\sum_{m<j\le n}x_{mj}\neq\sum_{m<j\le n}y_{mj}$. Thus $\delta_i=0$ for $i\in[m-1]$ and $\delta_m\neq 0$. Without loss of generality, assume $x_m<y_m$, then $\delta_m\ge 1$.

For $i\in[m+1,n-1]$, $|\delta_i|\le\max\{\sum_{i<j\le n}x_{ij},\sum_{i<j\le n}y_{ij}\}\le n-i\le n-2$. Define a bivariate polynomial $\varphi(x,y)=\phi(x-m-1)$. Then $\deg\varphi=\deg\phi\le k-2$ and

$$\begin{aligned}\sum_{1\le x<y\le n}\delta_{xy}\varphi(x,y)&=\sum_{x=1}^{n-1}\phi(x-m-1)\sum_{y=x+1}^{n}\delta_{xy}=\sum_{x=1}^{n-1}\phi(x-m-1)\delta_x\\&=\delta_m\phi(-1)+\sum_{x=m+1}^{n-1}\phi(x-m-1)\delta_x\ge\phi(-1)-(n-2)\sum_{x=0}^{n-3}|\phi(x)|>0.\end{aligned}$$

However, by Corollary 3.1, $\sum_{1\le x<y\le n}\delta_{xy}\varphi(x,y)=0$, which is a contradiction. $\square$

It remains to construct a polynomial $\phi(x)$ satisfying the conditions in Theorem 3.1. The construction is based on [11, Section 3].

Take $\phi(x)=p_k^2(x),\ p_k(x)=\sum_{i=0}^{k}\frac{T_i(-1)}{d_i}T_i(x)$ from [11, Theorem 3.2], where T_i is the special case of the Hahn polynomials [24], [25],

$$T_i(x)=\sum_{j=0}^{i}(-1)^j\frac{\binom{i}{j}\binom{i+j}{j}}{\binom{n}{j}}\binom{x}{j},$$

and

$$d_i=\frac{n+i+1}{2i+1}\frac{\binom{n+i}{n}}{\binom{n}{i}}.$$

Then p_k is of degree k, and $\deg\phi=2k$. For fixed $i,j\in[k]$, T_i and T_j are orthogonal on $[0,n]$, i.e.,

$$\sum_{x=0}^{n}T_i(x)\cdot T_j(x)=\delta_{i,j}d_i. \tag{3.2}$$

Define $\Delta(\phi) := (n+1)\sum_{x=0}^{n}|\phi(x)| - \phi(-1)$ and $F_k := \sum_{i=0}^{k} T_i^2(-1)/d_i$. By Eq. (3.2), we have

$$\begin{aligned}\Delta(\phi) &= (n+1)\sum_{x=0}^{n} p_k^2(x) - p_k^2(-1)\\ &= (n+1)\sum_{x=0}^{n}\left(\sum_{i=0}^{k}\frac{T_i(-1)}{d_i}T_i(x)\right)^2 - \left(\sum_{i=0}^{k}\frac{T_i^2(-1)}{d_i}\right)^2\\ &= (n+1)\sum_{i=0}^{k}\frac{T_i^2(-1)}{d_i} - \left(\sum_{i=0}^{k}\frac{T_i^2(-1)}{d_i}\right)^2\\ &= (n+1-F_k)F_k.\end{aligned} \tag{3.3}$$

Hence when $F_k > n+1$, we have $\Delta(\phi) < 0$.

Based on the above discussions, we provide an upper bound of $s(n)$ as below.

Theorem 3.2. $s(n) \le 2\lceil\sqrt{(n-2)\ln(n-3)}\rceil + 2$ *when* $n \ge 7$.

Proof. It is equivalent to show that $s(n+3) \le 2\lceil\sqrt{(n+1)\ln n}\rceil + 2$ when $n \ge 4$. Let $k = \lceil\sqrt{(n+1)\ln n}\rceil$. By Theorem 3.1, it suffices to show the existence of a polynomial ϕ of degree $2k$ satisfying $\phi(-1) > (n+1)\sum_{x=0}^{n}|\phi(x)|$. Let $\phi(x) = p_k^2(x)$, which is of degree $2k$. By Eq. (3.3), we only need to prove that $F_k > n+1$.

By the conclusion of [11, Section 3.2.1], we have

$$F_k > e^{(k+1)k/(n+1)} - 1 + 1/(n+1).$$

Since $k \ge \sqrt{(n+1)\ln n}$, we have

$$\begin{aligned}e^{(k+1)k/(n+1)} - 1 + 1/(n+1) &> e^{\ln n + \sqrt{\ln n/(n+1)}} - 1\\ &= ne^{\sqrt{\ln n/(n+1)}} - 1\\ &> n(1+\sqrt{\ln n/(n+1)}) - 1 > n+1.\end{aligned}$$

Here, the last inequality is true when $n \ge 4$. Hence $F_k > n+1$. □

4. Some specific values for small n

In this section, we consider exact values of $s(n)$ for small n by assistant of computers.

For each $\mathbf{x} \in \mathcal{S}_n$, let $S_j(\mathbf{x})$ be the total number of minors in $D_k(\mathbf{x})$ with the j-th coordinate being 1, $j \in [k]$. For each $t \in [n-k+1]$, let $i_t(\mathbf{x}) \in [n-t+1]$ be the location that t lies in $\mathbf{x}$ after deleting $1, 2, \ldots, t-1$. For example, when $\mathbf{x} = 13524$ and $k = 2$, $i_1(\mathbf{x}) = 1$, $i_2(\mathbf{x}) = 3$, $i_3(\mathbf{x}) = 1$ and $i_4(\mathbf{x}) = 2$. The following lemma relates $S_j(\mathbf{x})$ to values of $i_t(\mathbf{x})$. We write i_t instead of $i_t(\mathbf{x})$ if there is no confusion.

Lemma 4.1. *For each* $\mathbf{x} \in \mathcal{S}_n$ *and* $j \in [k]$, *we have*

$$S_j(\mathbf{x}) = \sum_{t=1}^{n-k+1}\binom{i_t-1}{j-1}\binom{n-i_t-(t-1)}{k-j}. \tag{4.1}$$

Proof. It is easy to see that only symbols $1, 2, \ldots, n-k+1$ in $\mathbf{x}$ may result in a symbol 1 in some minor in $D_k(\mathbf{x})$, since larger numbers cannot be changed to 1.

For each $t \in [n-k+1]$, the value $i_t(\mathbf{x})$ indicates that after deleting all symbols from $[t-1]$ in $\mathbf{x}$, there are exactly $i_t - 1$ symbols appear before the symbol t. Keeping the symbol t, the symbol t becomes 1 in the minor. To make this 1 in the j-th position in some minor, we need to delete $(i_t - j)$ more symbols before t, see

$$\underbrace{\cdots}_{i_t-1} \quad t \quad \underbrace{\cdots}_{n-i_t-(t-1)} .$$

So the number of minors of length k with the j-th position being 1 is $\binom{i_t-1}{i_t-j}\binom{n-i_t-(t-1)}{n-k-(i_t-j)-(t-1)} = \binom{i_t-1}{j-1}\binom{n-i_t-(t-1)}{k-j}$. Summing over $t \in [n-k+1]$, we get the conclusion. $\square$

Denote $\mathbf{i}(\mathbf{x}) := (i_1, i_2, \ldots, i_{n-k+1}) \in [n] \times [n-1] \times \cdots \times [k]$, and call it the characteristic vector of $\mathbf{x}$. By Lemma 4.1, the value $S_j(\mathbf{x})$ only depends on the characteristic vector of $\mathbf{x}$. Different permutations may have the same characteristic vector. Define the real location of t in $\mathbf{x} = x_1 x_2 \ldots x_n$ as ζ_t, that is, $x_{\zeta_t} = t$. Call the vector $\zeta(\mathbf{x}) := (\zeta_1, \zeta_2, \ldots, \zeta_{n-k+1})$ the location vector of $\mathbf{x}$. Clearly, the characteristic vector and the location vector can be deduced from each other recursively. See the following lemma.

Lemma 4.2. *For each* $t \in [n-k+1]$, *we have* $i_t = \zeta_t - \sum_{j<t} 1_{\{\zeta_t > \zeta_j\}}$ *and* $\zeta_t = i_t + \sum_{j<t} 1_{\{i_t \geq \zeta_j\}}$.

Proof. The former equalities are obvious. We only prove the latter ones. It is easy to see that $\zeta_1 = i_1$. For ζ_2, if $i_2 < \zeta_1$, deleting 1 does not affect the position of 2, so $\zeta_2 = i_2$; if $i_2 \geq \zeta_1$, deleting 1 will result in one step forward on 2, thus $\zeta_2 = i_2 + 1$. Continuing with this argument, we give $\zeta_t = i_t + \sum_{j<t} 1_{\{i_t \geq \zeta_j\}}$. $\square$

In all the above arguments, we consider the position of symbol 1 in the minors. Symmetrically, we can consider the position of symbol k in minors. Define $\bar{S}_j(\mathbf{x})$ the number of minors in $D_k(\mathbf{x})$ with the j-th coordinate being k. For each $t \in [k, n]$, let $\bar{i}_t(\mathbf{x}) \in [t]$ be the location that t lies in $\mathbf{x}$ after deleting $n, n-1, \ldots, t+1$. By the similar proof as for Lemma 4.1, we have the following result.

Corollary 4.1. *For each* $\mathbf{x} \in \mathcal{S}_n$ *and* $j \in [k]$, *we have*

$$\bar{S}_j(\mathbf{x}) = \sum_{t=k}^{n} \binom{\bar{i}_t - 1}{j-1}\binom{t - \bar{i}_t}{k-j}. \tag{4.2}$$

As in Lemma 4.2, $\bar{i}_t, t \in [k, n]$ can also determine $\zeta_t, t \in [k, n]$. In fact, $\zeta_n = \bar{i}_n$, and

$$\zeta_t = \bar{i}_t + \sum_{j>t} 1_{\{\bar{i}_t \geq \zeta_j\}} \tag{4.3}$$

for $t \in [k, n-1]$. Call $\bar{\mathbf{i}}(\mathbf{x}) = (\bar{i}_n, \bar{i}_{n-1}, \ldots, \bar{i}_k)$ the reverse characteristic vector and $\bar{\zeta}(\mathbf{x}) := (\zeta_n, \zeta_{n-1}, \ldots, \zeta_k)$ the reverse location vector. Consequently, for two permutations $\mathbf{x} \neq \mathbf{y}$, if $\mathbf{i}(\mathbf{x}) = \bar{\mathbf{i}}(\mathbf{y})$, then $\zeta(\mathbf{x}) = \bar{\zeta}(\mathbf{y})$.

Now we are ready to apply the above arguments to design an exhaustive algorithm: input n and k, and output a pair $(\mathbf{x}, \mathbf{y}) \in \mathcal{S}_n \times \mathcal{S}_n$ which meets the conditions $S_j(\mathbf{x}) = S_j(\mathbf{y})$ and $\bar{S}_j(\mathbf{x}) = \bar{S}_j(\mathbf{y})$ for all $j \in [k]$. By the fact that $\mathbf{x} \overset{k}{\sim} \mathbf{y}$ implies $S_j(\mathbf{x}) = S_j(\mathbf{y})$ and $\bar{S}_j(\mathbf{x}) = \bar{S}_j(\mathbf{y})$ for $j \in [k]$, if the algorithm returns nothing, it means that any two permutations in $\mathcal{S}_n$ are not equivalent, that is, $s(n) \leq k$. In fact, our algorithm tries to output a set of permutations in $\mathcal{S}_n$ that are possible to satisfy the above conditions.

Algorithm 1:

Step 1. For each vector $\mathbf{i} = (i_1, i_2, \ldots, i_{n-k+1}) \in [n] \times [n-1] \times \ldots \times [k]$, we compute the corresponding S_j value for each $j \in [k]$ by Lemma 4.1. These values of S_j are stored as a row in a matrix $\mathbf{S}_{l \times k}$ with row index $\mathbf{i}$, where $l = \frac{n!}{(k-1)!}$. The row $\mathbf{i}$ of $\mathbf{S}$ corresponds to a set of permutations in $\mathcal{S}_n$ with the same characteristic vector $\mathbf{i}$ and thus the same S_j values.

Step 2. By pairwise comparing all rows in $\mathbf{S}$, we collect all pairs $(\mathbf{i}, \mathbf{i}')$ of rows such that $\mathbf{S}(\mathbf{i}, j) = \mathbf{S}(\mathbf{i}', j)$ for each $j \in [k]$. We store all such pairs $(\mathbf{i}, \mathbf{i}')$ as a row in an array $\mathbf{P}$ with two columns. Note that $(\mathbf{i}', \mathbf{i})$ is also recorded in $\mathbf{P}$.

Step 3. Considering $\bar{\mathbf{i}} = (\bar{i}_n, \bar{i}_{n-1}, \dots, \bar{i}_k) \in [n] \times [n-1] \times \dots \times [k]$ and values $\bar{S}_j$, we can get exactly the same matrices $\mathbf{S}$ and $\mathbf{P}$. This means that each row $\mathbf{i}$ of $\mathbf{S}$ can be viewed as the S_j values of some permutations $\mathbf{x}$ with $\mathbf{i}(\mathbf{x}) = \mathbf{i}$, and can also be viewed as the $\bar{S}_j$ values of some other permutations $\mathbf{x}'$ with $\bar{\mathbf{i}}(\mathbf{x}') = \mathbf{i}$.

For each entry $\mathbf{i}$ in the first column of $\mathbf{P}$, by considering it as a characteristic vector $(i_1, i_2, \dots, i_{n-k+1})$, we compute a location vector $\zeta(\mathbf{x}) = (\zeta_1, \zeta_2, \dots, \zeta_{n-k+1})$ by Lemma 4.2 for some $\mathbf{x}$ with $\mathbf{i}(\mathbf{x}) = \mathbf{i}$; by considering it as a reverse characteristic vector $(\bar{i}_n, \bar{i}_{n-1}, \dots, \bar{i}_k)$, we compute a reverse location vector $\bar{\zeta}(\mathbf{x}') = (\zeta_n, \zeta_{n-1}, \dots, \zeta_k)$ by Eq. (4.3) for some $\mathbf{x}'$ with $\bar{\mathbf{i}}(\mathbf{x}') = \mathbf{i}$.

Step 4. Suppose that $(\mathbf{x}, \mathbf{y}) \in \mathcal{S}_n \times \mathcal{S}_n$ satisfies $S_j(\mathbf{x}) = S_j(\mathbf{y})$ and $\bar{S}_j(\mathbf{x}) = \bar{S}_j(\mathbf{y})$ for all $j \in [k]$. Then there exist two different rows $\mathbf{i}$ and $\bar{\mathbf{i}}$, such that $\mathbf{i}(\mathbf{x}) = \mathbf{i}$ and $\bar{\mathbf{i}}(\mathbf{x}) = \bar{\mathbf{i}}$. Then the location vector $\zeta(\mathbf{x}) = (\zeta_1, \zeta_2, \dots, \zeta_{n-k+1})$ computed by $\mathbf{i}(\mathbf{x}) = \mathbf{i}$ and the reverse location vector $\bar{\zeta}(\mathbf{x}) = (\zeta_n, \zeta_{n-1}, \dots, \zeta_k)$ computed by $\bar{\mathbf{i}}(\mathbf{x}) = \bar{\mathbf{i}}$ should match each other to form one permutation.

By definition of $\mathbf{P}$, we could simply pairwise check entries in the first column of $\mathbf{P}$, and output all possible candidates $\mathbf{x}$. Otherwise, stop and return "No solution!".

Remark 4.1. *If Algorithm 1 returns a set of permutations, we need to continue to pairwise check whether they are equivalent. However, Algorithm 1 is quite useful when it returns no solution, which implies $s(n) \le k$. In this case, it is very efficient since the data is based on all different vectors $\mathbf{i} \in [n] \times [n-1] \times \dots \times [k]$ instead of all different permutations $\mathbf{x} \in \mathcal{S}_n$, and the algorithm runs just by computing and comparing a few characteristic values, which greatly reduces the computation space and computation time.*

Lemma 4.3. *For $(n, k) \in \{(9, 5), (10, 5)\}$, Algorithm 1 returns "No solution!". So $s(9) \le 5$ and $s(10) \le 5$.*

A. A table for $s(n)$

Now we determine or bound values of $s(n)$ for small n. We first introduce a related notation studied in [8], [9]. Given d, let N_d be the smallest integer such that for any $n \ge N_d$, we can reconstruct permutations of length n from their $(n-d)$-deck. That is, for all $n \ge N_d$, $D_{n-d}(\mathbf{x}) \ne D_{n-d}(\mathbf{y})$ for different $\mathbf{x}, \mathbf{y} \in \mathcal{S}_n$. Smith [8] proved that $N_1 = 5, N_2 = 6$ and $N_3 \le 13$. Raykova [9] showed that the existence of N_d for all d, and further proved that $N_3 = 7$ and $N_4 \ge 9$, then gave upper and lower bounds by

$$d + \log_2 d < N_d < d^2/4 + 2d + 4.$$

By definition, if $N_d \le n$, that is, for any two different $\mathbf{x}, \mathbf{y} \in S_n$, $D_{n-d}(\mathbf{x}) \ne D_{n-d}(\mathbf{y})$, then $s(n) \le n-d$. So by $N_1 = 5, N_2 = 6$ and $N_3 = 7 < 8$, we have $s(5) \le 4, s(6) \le 4, s(7) \le 4$ and $s(8) \le 5$. By $N_d < d^2/4 + 2d + 4$, $s(d^2/4 + 2d + 4) \le d^2/4 + d + 4$ for every d. But this is much weaker than our upper bound $s(n) = O(\sqrt{n \ln n})$ in Theorem 3.2.

If $N_d = n + 1$, that is, there exist two different $\mathbf{x}, \mathbf{y} \in S_n$ such that $D_{n-d}(\mathbf{x}) = D_{n-d}(\mathbf{y})$, then $s(n) > n - d$, that is, $s(N_d - 1) > N_d - 1 - d$. So by $N_2 = 6 = 5 + 1$ and $N_3 = 7 = 6 + 1$, we have $s(5) > 3$, $s(6) > 3$.

Combining the above results, we have $s(5) = s(6) = 4$. By $1247356 \overset{3}{\sim} 1263475$ and $68573142 \overset{4}{\sim} 75862413$, we have $s(7) = 4$ and $s(8) = 5$. By Proposition 2.1, $s(n) \ge s(8) = 5$ when $n \ge 9$. Combining Lemma 4.3, we have $s(9) = s(10) = 5$. For $11 \le n \le 14$, since $n \ge N_3$, $s(n) \le n-3$; by Proposition 2.1, $s(n) \ge s(10) = 5$. For $15 \le n \le 19$, since $N_4 \le 4^2/4 + 2 \times 4 + 4 - 1 = 15 \le n$, we have $s(n) \le n - 4$. For $n \ge 20$, since $N_5 < 5^2/4 + 2 \times 5 + 4 = 20.25$, which means $n \ge N_5$, we have $s(n) \le n - 5$. The lower bound of $s(n)$, $16 \le n \le 22$ can be obtained from Theorem 2.2.

Finally we list the best bounds of $s(n)$ for $n \le 22$ in Table II, where the first unknown value is $s(11)$.

n	1	2	3	4	5	6	7	8	9	10	11
$s(n)$	1	2	3	4	4	4	4	5	5	5	$[5,8]$

n	12	13	14	15	16	17	18	19	20	21	22
$s(n)$	$[5,9]$	$[5,10]$	$[5,11]$	$[5,11]$	$[6,12]$	$[6,13]$	$[6,14]$	$[6,15]$	$[6,15]$	$[6,16]$	$[6,17]$

TABLE II: Bounds of $s(n)$ for small n.

B. Improvements on N_d

Based on our bounds of $s(n)$, we are able to improve bounds of N_d asymptotically.

If $s(n) > k$, that is, there exist two different $\mathbf{x}, \mathbf{y} \in S_n$ such that $D_k(\mathbf{x}) = D_k(\mathbf{y})$, then $N_{n-k} > n$. Then by Theorem 2.2, that is, $s(n) = \exp(\Omega(\sqrt{\ln n}))$, we have

$$N_d > d + \exp(\Omega(\sqrt{\ln d})), \tag{4.4}$$

when d is large, which is much stronger than $N_d > d + \log_2 d$ in [9].

For the upper bound, we improve as follows.

Theorem 4.1. *$N_d \le d + 3\sqrt{d \ln d}$ when d is large enough.*

Proof. Let $n_0 = d + 3\sqrt{d \ln d} - 2$. First, we claim that $(n_0 - d - 2)^2 > 4n_0 \ln(n_0 - 1)$ when d is large enough. This is true because $(n_0 - d - 2)^2 = 9d \ln d(1 - o(1))$ and $4n_0 \ln(n_0 - 1) = 4d \ln d(1 + o(1))$ when d goes to infinity.

For each fixed d satisfying $(n_0 - d - 2)^2 > 4n_0 \ln(n_0 - 1)$, we next claim that $(n - d - 2)^2 > 4n \ln(n - 1)$ for any $n \ge n_0$. This is true since the left hand side grows much faster than the right hand side.

Finally, we show that for each $n \ge n_0 + 2 = d + 3\sqrt{d \ln d}$, any permutation in $\mathcal{S}_n$ is $(n-d)$-reconstructible, i.e., $s(n) \le n - d$. In fact, by Theorem 3.2, we have

$$s(n) \le 2\lceil \sqrt{(n-2)\ln(n-3)} \rceil + 2 < 2\sqrt{(n-2)\ln(n-3)} + 4 < (n-2) - d - 2 + 4 = n - d.$$

This completes the proof. □

Note that Theorem 4.1 greatly improves the upper bound $N_d < d^2/4 + 2d + 4$ from [9] asymptotically. Combining Eq. (4.4) and Theorem 4.1, we have for large d,

$$d + \exp(\Omega(\sqrt{\ln d})) < N_d < d + O(\sqrt{d \ln d}).$$

5. Conclusion

By applying results for sequence reconstruction, we prove that the least integer k such that any permutation in $\mathcal{S}_n$ is k-reconstructible is between $\exp(\Omega(\sqrt{\ln n}))$ and $O(\sqrt{n \ln n})$. As a consequence, we improve the bounds significantly for the well-studied parameter N_d, the smallest integer such that any permutation in $\mathcal{S}_n$ with $n \ge N_d$ is $(n-d)$-reconstructible. The new bounds are $d + \exp(\Omega(\sqrt{\ln d})) < N_d < d + O(\sqrt{d \ln d})$ asymptotically, and the previous best bounds were $d + \log_2 d < N_d < d^2/4 + 2d + 4$ in [9].

References

[1] J. Acharya, H. Das, O. Milenkovic, A. Orlitsky, and S. Pan, "String reconstruction from substring compositions," *SIAM Journal on Discrete Mathematics*, vol. 29, no. 3, pp. 1340–1371, 2015.

[2] T. Batu, S. Kannan, S. Khanna, and A. McGregor, "Reconstructing strings from random traces," in *SODA*, vol. 4, 2004, pp. 910–918.

[3] E. Ukkonen, "Finding approximate patterns in strings," *Journal of Algorithms*, vol. 6, no. 1, pp. 132–137, 1985.

[4] R. Golm, M. Nahvi, R. Gabrys, and O. Milenkovic, "The gapped k-deck problem," in *2022 IEEE International Symposium on Information Theory (ISIT)*. IEEE, 2022, pp. 49–54.

[5] S. H. T. Yazdi, R. Gabrys, and O. Milenkovic, "Portable and error-free DNA-based data storage," *Scientific Reports*, vol. 7, no. 1, p. 5011, 2017.

[6] R. Gabrys and O. Milenkovic, "Unique reconstruction of coded strings from multiset substring spectra," *IEEE Transactions on Information Theory*, vol. 65, no. 12, pp. 7682–7696, 2019.

[7] S. M. Ulam, "A collection of mathematical problems," *Interscience*, 1960.

[8] R. Smith, "Permutation reconstruction," *The Electronic Journal of Combinatorics*, vol. 13, N11, 2006.

[9] M. Raykova, "Permutation reconstruction from minors," *The Electronic Journal of Combinatorics*, vol. 13, R66, 2006.

[10] L. Kalashnik, "The reconstruction of a word from fragments," *Numerical Mathematics and Computer Technology*, pp. 56–57, 1973.

[11] W. Foster and I. Krasikov, "An improvement of a Borwein–Erdélyi–Kós result," *Methods and Applications of Analysis*, vol. 7, no. 4, pp. 605–614, 2000.

[12] M. Dudık and L. J. Schulman, "Reconstruction from subsequences," *Journal of Combinatorial Theory, Series A*, vol. 103, no. 2, pp. 337–348, 2003.

[13] G. Kós, P. Ligeti, and P. Sziklai, "Reconstruction of matrices from submatrices," *Mathematics of Computation*, vol. 78, no. 267, pp. 1733–1747, 2009.

[14] W. Zhong and X. Zhang, "Reconstruction of hypermatrices from subhypermatrices," *Journal of Combinatorial Theory, Series A*, vol. 209, 105966, 2025.

[15] V. Mnukhin, "Combinatorial properties of partially ordered sets and group actions," *TEMPUS Lecture Notes: Discrete Mathematics and Applications*, vol. 8, 1993.

[16] P. J. Cameron, "Stories from the age of reconstruction," *Congressus Numerantium*, pp. 31–42, 1996.

[17] M. Monks, "The solution to the partition reconstruction problem," *Journal of Combinatorial Theory, Series A*, vol. 116, no. 1, pp. 76–91, 2009.

[18] A. J. Cain and E. Lehtonen, "Reconstructing young tableaux," *Journal of Combinatorial Theory, Series A*, vol. 187, 105578, 2022.

[19] M. J. Gouveia and E. Lehtonen, "Permutation reconstruction from a few large patterns," *The Electronic Journal of Combinatorics*, vol. 28, P3.41, 2021.

[20] M. Monks, "Reconstructing permutations from cycle minors," *The Electronic Journal of Combinatorics*, vol. 16, R19, 2009.

[21] E. Lehtonen, "Reconstructing permutations from identification minors," *The Electronic Journal of Combinatorics*, vol. 22, P4.20, 2015.

[22] M. De Biasi, "Permutation reconstruction from differences," *The Electronic Journal of Combinatorics*, vol. 21, P4.3, 2014.

[23] I. Krasikov and Y. Roditty, "On a reconstruction problem for sequences," *Journal of Combinatorial Theory, Series A*, vol. 77, no. 2, pp. 344–348, 1997.

[24] S. Karlin, "The hahn polynomials, formulas and an application," *Scripta Math.*, vol. 26, pp. 33–46, 1961.

[25] A. F. Nikiforov, V. B. Uvarov, S. K. Suslov, A. F. Nikiforov, V. B. Uvarov, and S. K. Suslov, *Classical orthogonal polynomials of a discrete variable*. Springer, 1991.