Order-isomorphic twins in permutations

Boris Bukh*    Oleksandr Rudenko†

Abstract

Let \(a_1, \ldots, a_n\) be a permutation of \([n]\). Two disjoint order-isomorphic subsequences are called twins. We show that every permutation of \([n]\) contains twins of length \(\Omega(n^{3/5})\) improving the trivial bound of \(\Omega(n^{1/2})\). We also show that a random permutation contains twins of length \(\Omega(n^{2/3})\), which is sharp.

In this paper we regard permutations as sequences of symbols, devoid of any group-theoretic meaning. So, for us a permutation on a finite set \(\Sigma\) is a sequence of elements of \(\Sigma\) in which each element of \(\Sigma\) appears exactly once. We call a subsequence of a permutation subpermutation. For instance, 135642 is a permutation of \([6]\), and 1562 is a subpermutation inside, which itself is a permutation of \(\{1, 2, 5, 6\}\). We denote permutations by bold letters.

Throughout the paper we consider only the permutations of finite sets of natural numbers. We say that permutations \(a = (a_1, \ldots, a_L)\) and \(b = (b_1, \ldots, b_L)\) are order-isomorphic if \((a_i < a_j) \iff (b_i < b_j)\). For example, 1562 is order-isomorphic to 1342.

We call a pair of subpermutation \(a, b\) of \(c\) twins if \(a\) and \(b\) are order-isomorphic and disjoint (do not contain the same symbol). For example, 152 and 364 are twins in 135642 of length 3. We denote by \(t(n)\) the largest integer such that every permutation of \([n]\) contains a pair of twins of length \(t(n)\).

The problem of estimating \(t(n)\) was raised by Gawron [5], who observed that \(t(n) \geq (n^{1/2} - 1)/2\) follows from the Erdős–Szekeres theorem, and that \(t(n) = O(n^{2/3})\) follows from the first moment method. He further conjectured that \(t(n) = \Omega(n^{2/3})\). This is not known even for random permutations: the best result is due to Dudek, Grytczuk, Ruciński [4] who showed that a random permutation almost surely contains twins of length \(\Omega(n^{2/3}/\log^{1/3} n)\).

In this short note, we give a first non-trivial lower bound on \(t(n)\), and remove the logarithmic factor from the Dudek–Grytczuk–Ruciński result.

**Theorem 1.** For \(n \geq 2\), every permutation of \([n]\) contains twins of length at least \(\frac{1}{5} n^{3/5}\).

**Theorem 2.** A random permutation of \([n]\) almost surely contains twins of length at least \(\frac{1}{80} n^{2/3}\), as \(n \to \infty\).

In view of Gawron’s result, **Theorem 2** is sharp up to the constant factor.

*Department of Mathematical Sciences, Carnegie Mellon University, Pittsburgh, PA 15213, USA. Supported in part by Sloan Research Fellowship and by U.S. taxpayers through NSF CAREER grant DMS-1555149.

†Department of Mathematical Sciences, Carnegie Mellon University, Pittsburgh, PA 15213, USA. Supported in part by U.S. taxpayers through NSF CAREER grant DMS-1555149.
Proof of Theorem 1

The proof relies on a result of Beame and Huynh-Ngoc [1, Lemma 5.9], which previously was used by Bukh and Zhou [2] to study a related notion of twins in words.

Lemma 3. Among any three permutations $c^{(0)}, c^{(1)}, c^{(2)}$ of $[m]$ we may find two distinct, say, $c^{(k)}$ and $c^{(t)}$, that contain the same subpermutation of length at least $m^{1/3}$.

Call twin subpermutations $a_1, \ldots, a_L$ and $b_1, \ldots, b_L$ close if $|a_i - b_i| \leq n^{2/5}$ for all $i$. Let $t'(m, n)$ be the largest integer so that whenever $\Sigma \subset [n]$ is any set of at least $m$ elements, every permutation of $\Sigma$ contains close twins of length at least $t'(m, n)$.

Claim 4. If $m \geq 7n^{2/5}$, then $t'(m, n) \geq t'(m - 7n^{3/5}, n) + n^{1/5}$.

From $t(n) \geq t'(n, n)$ and $\lceil \frac{1}{7} n^{2/5} \rceil$ many uses of Claim 4 we infer that $t(n) \geq \lceil \frac{1}{7} n^{2/5} \rceil n^{1/5}$, implying Theorem 1 for $n \geq 565/2$. When $2 \leq n \leq 565/2$, Theorem 1 follows from $t(n) \geq (n^{1/2} - 1)/2 \geq \frac{1}{8} n^{3/5}$.

We now prove the claim. We can clearly assume that $|\Sigma| = m$. Let $a = (a_1, \ldots, a_m)$ be an arbitrary permutation of $\Sigma$. Consider its first $3r$ elements, where $r = \lceil 2n^{3/5} \rceil$. Say $a_1, \ldots, a_{3r}$ is a permutation of the set $\{b_0, \ldots, b_{3r-1}\}$, where $b_0 < \cdots < b_{3r-1}$. Consider the triples $(b_0, b_1, b_2), (b_3, b_4, b_5), \ldots, (b_{3r-3}, b_{3r-2}, b_{3r-1})$. Since $\sum_{i=0}^{r}(b_{3i+2} - b_{3i}) \leq n$, the set $I_0 = \{i : b_{3i+2} - b_{3i} \leq 2n/r\}$ has at least $r/2$ elements. For each $j = 0, 1, 2$ let $c^{(j)}(i)$ be the subpermutations of $a_1, \ldots, a_m$ obtained by keeping only the elements $b_{3i+j}$ with $i \in I_0$. Note that this choice ensures that $|c^{(j)}(i) - c^{(k)}(i)| \leq n^{2/5}$.

By Lemma 3, there is $I \subset I_0$ of size $|I| \geq |I_0|^{1/3} \geq n^{1/5}$ and $k < \ell$ such that the subpermutations $c^{(k)}(i) = (c^{(k)}(i) : i \in I)$ and $c^{(\ell)}(i) = (c^{(\ell)}(i) : i \in I)$ are order-isomorphic. By deleting some elements of $I$ if necessary, we may assume that $|I| = [n^{1/5}]$. Note that $c^{(k)}(i)$ and $c^{(\ell)}(i)$ is a pair of close twins.

Let $d$ be the subpermutation of $a$ obtained by deleting the first $3r$ elements, and also deleting all elements that are contained in the intervals of the form $[c^{(k)}(i), c^{(k)}(i) + n^{2/5}]$ for $i \in I$. Since in total these intervals contain no more than $(n^{2/5} + 1)|I|$ elements, and each interval contains at least two elements among the first $3r$, the permutation $d$ is of length at least $m - 3r - (n^{2/5} - 1)|I| \geq m - 7n^{3/5}$.

Given a pair of close twins $e, f$ in $d$, we may obtain a pair of close twins in $a$ by concatenating $c^{(k)}(i)$ with $e$ and concatenating $c^{(\ell)}(i)$ with $f$. Indeed, let $i$ and $j$ be arbitrary, and consider two pairs of elements $c^{(k)}(i), c^{(\ell)}(i)$ and $e_j, f_j$. Because neither of $e_j, f_j$ is contained in the interval $T = [c^{(k)}(i), c^{(k)}(i) + n^{2/5}]$, and $|c^{(k)}(i) - d_j| \leq n^{2/5}$, it follows that $e_j, f_j$ are either both smaller than $\min T$ or both larger than $\max T$. As both $c^{(k)}(i)$ and $c^{(\ell)}(i)$ are contained in $T$, we deduce that $(c^{(k)}(i) < e_j) \iff (c^{(\ell)}(i) < f_j)$. Hence, the two concatenations indeed form a pair of twins.

Proof of Theorem 2

We modify the argument of Dudek–Grytczuk–Ruciński. They construct a certain bipartite graph $B$ such that the matchings in $B$ correspond to twins in the original permutation. They note that $B$ contains a matching of size $v(B)/2\Delta(B)$, where $v(B)$ and $\Delta(B)$ denote the number of vertices and the maximum degree respectively. The logarithmic factor is lost because of the union bound to bound $\Delta(B)$. In our proof, instead of the maximum degree, we effectively work with the typical vertex degrees. To help with this, we gain more independence by first Poissonizing the random process.
Let \( t(p) \) be the length of the longest twin in a permutation \( p \). We consider two ways of generating a random permutation. First, we may sample \( p \) uniformly from all permutations of \([n]\). Denote this probability distribution by \( S_n \). Second, we may consider a Poisson process of intensity \( n \) on the unit square, and consider relative positions of resulting points. Denote this probability distribution by \( \mathcal{S}_n \).

Consider an infinite sequence \( p_1, p_2, \ldots \) of independent points in \([0, 1]^2\). We may regard its prefix \( p_1, \ldots, p_m \) of length \( m \) as a permutation \( p^{(m)} \) of length \( m \). We clearly have \( t(p^{(m)}) \leq t(p^{(\ell)}) \) whenever \( m \leq \ell \). Note that we may sample from \( \mathcal{S}_n \) by sampling a number \( m \) from the Poisson distribution of mean \( n \) and returning \( p^{(m)} \).

Since \( \Pr[\text{Poisson}(n/2) \geq n] \leq \exp(-cn) \) (see, for example [3]), we infer that to show that \( t(p^{(n)}) \geq 1/40 \) \( n^{2/3} \) a.a.s., it suffices to establish \( t(S_{n/2}) \geq 1/40 n^{2/3} \) a.a.s.

Partition \([0, 1]\) into \( r \) equal intervals of length \( 1/r \) each, denoted \( A_1, \ldots, A_r \). This induces a partition of \([0, 1]^2\) into \( r^2 \) smaller squares of the form \( A_i \times A_j \). Sample a set \( P \) from a Poisson process of intensity \( n/2 \) on \([0, 1]^2\). Make a bipartite graph \( B \) whose parts are two copies of \([r]\), with \((i, j)\) being an edge if \( A_i \times A_j \) contains at least two points of \( P \). The edges are independent with probability \( p = \Pr[\text{Poisson}(n/2r^2) \geq 2] \geq \frac{1}{10} n^{-2/3} \), for large \( n \). Clearly, every matching in \( B \) corresponds to a pair of twins in the associated permutation.

Theorem 2 follows once we show that \( B \) is likely to contain a large matching. This is well-known in the (very similar) context of the \( G(n, p) \) model. We include such a proof for completeness.

Claim 5. Let \( p \leq \frac{1}{6r} \). Then a random bipartite graph \( G(r + r, p) \) contains a matching of size \( pr^2/7 \) a.a.s.

Proof. Let \( L \cup R \) be the bipartition. As long as \( |L| = |R| \geq r/2 \), do the following. Pick any vertex \( v \in L \). It has a neighbor with probability \( \geq p|L| - p^2 \binom{|L|}{2} \geq pr/3 \). Let \( u \in R \) be a neighbor, match \( u \) to \( v \). Else, let \( u \) be any vertex in \( R \). Remove \( v \) from \( L \) and \( u \) from \( R \). This way, we match \( \text{Binom}(r/2, pr/3) \) edges, which is \( \geq pr^2/7 \) a.a.s. \( \square \)

Acknowledgment. We thank Andrzej Dudek and Andrzej Ruciński for feedback on earlier drafts.

References

[1] Paul Beame and Dang-Trinh Huynh-Ngoc. On the value of multiple read/write streams for approximating frequency moments. Electronic Colloquium on Computational Complexity (ECCC), May 2008. Technical Report TR08-024.

[2] Boris Bukh and Lidong Zhou. Twins in words and long common subsequences in permutations. Israel J. Math., 213(1):183–209, 2016. arXiv:1307.0088.

[3] Clément Cannone. A short note on Poisson tail bounds. http://www.cs.columbia.edu/~ccanonne/files/misc/2017-poissonconcentration.pdf.

[4] Andrzej Dudek, Jarosław Grytczuk, and Andrzej Ruciński. Variations on twins in permutations. arXiv:2001.05589, January 2020.

[5] Maciej Gawron. Izomorficzne podstruktury w słowach i permutacjach. Master’s thesis, Uniwersytet Jagielloński, 2014.