(This is a sample cover image for this issue. The actual cover is not yet available at this time.)

On the Lipschitz constant of the RSK correspondence

Nayantara Bhatnagar [1], Nathan Linial

Department of Computer Science, Hebrew University of Jerusalem, Israel

A R T I C L E   I N F O

A B S T R A C T

We view the RSK correspondence as associating to each permutation $\pi \in S_n$ a Young diagram $\lambda = \lambda(\pi)$, i.e. a partition of n. Suppose now that π is left-multiplied by t transpositions, what is the largest number of cells in λ that can change as a result? It is natural refer to this question as the search for the Lipschitz constant of the RSK correspondence.

We show upper bounds on this Lipschitz constant as a function of t. For $t = 1$, we give a construction of permutations that achieve this bound exactly. For larger t we construct permutations which come close to matching the upper bound that we prove.

© 2011 Elsevier Inc. All rights reserved.

1. Introduction

The Robinson–Schensted–Knuth (RSK) correspondence [7,10,11] maps an arbitrary permutation $\pi \in S_n$ bijectively to an ordered pair of Young tableaux of the same shape $\lambda = \lambda(\pi)$. How much can $\lambda(\pi)$ change as we mildly vary π? For example, if we left-multiply π by t transpositions, to what extent can λ change [2]? We begin with the case when $t = 1$ and show that the resulting Young diagram can differ from λ on at most $\sqrt{n/2}$ cells. We show that this bound is tight by giving explicit constructions of permutations π for which this bound is attained where the diagrams differ in at least $(1 - o(1))\sqrt{n/2}$ cells. We then turn to consider the same question for larger t and show that the corresponding diagram changes in at most $O(\sqrt{nt \log t})$ cells. The best constructions we know nearly match this bound and yield, e.g., $(1 - o(1))\sqrt{nt/2}$ changes for $t = o(n)$.

The outline of this paper is as follows. In the remainder of this section we recall some definitions and properties of Young tableaux and the RSK correspondence. In Section 2 we prove upper bounds

---

on the Lipschitz constant when $t = 1$ and show a matching construction. In Section 3 we give upper bounds and extend our constructions for the case of general t. We conclude with some directions for further research in Section 4.

1.1. Notation and preliminaries

We recall some definitions and background on Young tableaux and the RSK algorithm here. For more detailed expositions refer to [4,9] or [12].

Let $n \in \mathbb{N}$ be a positive integer. A vector $\lambda = (\lambda_1, \lambda_2, \ldots)$ of positive integers is a *partition* of n (denoted by $\lambda \vdash n$) if

$$\lambda_1 \geqslant \lambda_2 \geqslant \cdots > 0 \quad \text{and} \quad \sum_i \lambda_i = n.$$

The *Young diagram* (or diagram) of a partition λ is a left-justified array of cells with λ_i cells in the i-th row for each $i \geqslant 1$. For example, the diagram of the partition $(5, 5, 3, 2)$ is
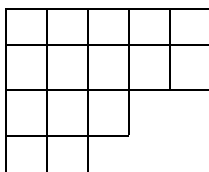
The cell in the i-th row and j-th column is referenced by its *coordinate* (i, j). Thus $(1, 1)$ is the top leftmost cell of the diagram.

The *conjugate* of a partition λ, denoted by λ' is the partition whose diagram is the transpose of the diagram of λ.

A *standard Young tableau* (SYT or tableau) of size n with entries from $[n]$ is a diagram whose cells are filled with the elements of $[n]$ in such a way that the entries are strictly increasing from left to right along a row as well as from top to bottom down a column. The *shape* of a tableau T, denoted $\mathrm{sh}(T)$ is the partition corresponding to the diagram of T. For example,

1	2	4	7
3	6		
5			
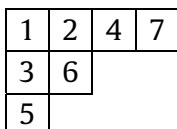
is a tableau of size 7 of shape $(4, 2, 1)$. Note that the elements in the cells of a SYT are distinct integers. Let $\mathcal{T}_n$ denote the set of SYT of size n.

1.2. The Robinson–Schensted–Knuth (RSK) correspondence

The RSK correspondence discovered by Robinson [10], Schensted [11] and further extended by Knuth [7] is a bijection between the set of permutations S_n and pairs of tableau of size n of the same shape. This bijection is intimately related to the representation theory of the symmetric group [3,6], the theory of symmetric functions [12, Chapter 7], and the theory of partitions [1].

The bijection can be defined through a *row-insertion* algorithm first defined by Schensted [11] in order to study the longest increasing subsequence of a permutation. Suppose that we have a tableau T. The row-insertion procedure below inserts a positive integer x that is distinct from all entries of T, into T and results in a tableau denoted by $T \leftarrow x$.

1. Let y be the smallest number larger than x in the first row of T. Replace the cell containing y with x. If there is no such y, add a cell containing x to the end of the row.
2. If y was removed from the first row, attempt to insert it into the next row by the same procedure as above. If there is no row to add y to, create a new row with a cell containing y.
3. Repeat this procedure on successive rows until either a number is added to the end of a row or added in a new row at the bottom.

The RSK correspondence from S_n to $\{(P, Q) \in \mathcal{T}_n \times \mathcal{T}_n : \mathrm{sh}(P) = \mathrm{sh}(Q)\}$ can now be defined as follows. Let $\pi \in S_n$ and let π_i denote the element of $[n]$ in position i in π. Let P_1 be the tableau with a single cell containing π_1. Let $P_j = P_{j-1} \leftarrow \pi_j$ for all $1 < j \leqslant n$ and set $P = P_n$. The tableau Q is defined recursively in terms of tableaux Q_i of size i as follows. Let Q_1 be the tableau with one cell containing the integer 1. The equality of shapes $\mathrm{sh}(Q_i) = \mathrm{sh}(P_i)$ is maintained throughout the process. The cell of Q_i containing i is the (unique) cell of P_i that does not belong to P_{i-1}. The remaining cells of Q_i are identical to those of Q_{i-1}. Finally, set $Q = Q_n$. We refer to P as the *insertion tableau* and Q is the *recording tableau*.

Let $\pi \in S_n$ and let (P, Q) be the corresponding tableaux under the RSK correspondence. The *shape of π* is $\mathrm{sh}(P) = \mathrm{sh}(Q)$ and will be denoted by $\lambda = \lambda(\pi)$. The RSK correspondence has numerous interesting properties (see [4,8,9] or [12]). Some that will be useful in particular are as follows.

Proposition 1.1. *Let $\lambda = \lambda(\pi)$. Then the diagram corresponding to π^R, the reversal of π, is λ', the conjugate of λ.*

Proposition 1.2. *Let (P, Q) be the tableaux corresponding to a permutation π under the RSK correspondence. Then the tableaux corresponding to the inverse permutation π^{-1} are (Q, P). Thus the shape remains invariant upon inversion, i.e., $\lambda(\pi^{-1}) = \lambda(\pi)$.*

1.3. Motivation and related work

In view of the important role of the RSK correspondence, it is natural to investigate various aspects of it. Thus Fomin's appendix in [12, Chapter 7] starts with the following two motivating questions:

(1) Given a partition λ, characterize those permutations π for which $\lambda(\pi) = \lambda$.
(2) Given a tableau P, characterize the permutations π which have P as their insertion tableau.

We consider an *approximate* version of such questions and ask to what extent λ changes as π changes slightly. Question (1) is answered by the following theorem of Greene.

Theorem 1.3. *(See Greene [5].) Let π be a permutation, and suppose that the largest cardinality of the union of j increasing subsequences in π is μ_j, then $\lambda(\pi) = \lambda_1, \ldots, \lambda_k$, where $\lambda_1 = \mu_1$ and $\lambda_j = \mu_j - \mu_{j-1}$ for all $j \geqslant 2$.*

In his study of the RSK correspondence, Knuth discovered certain equivalence relations that are key to the solution of Question (2) above. Two permutations are *Knuth equivalent* if one can be obtained from the other by certain restricted sequences of adjacent transpositions. Knuth equivalent permutations are the equivalency classes of permutations that have the same insertion tableau. For more on the subject, see [12].

In order to make our question concrete, we need to specify two measures of distance: One between permutations and the other between diagrams. A natural metric on permutations is *left-multiplication* by adjacent transpositions. An *adjacent transposition* is a permutation of the form $(i, i + 1)$. Left-multiplying π by an adjacent transposition is denoted by $(i, i + 1) \circ \pi$ and means that first, the permutation π is applied and then the transposition. We denote the least number of adjacent transpositions that transform the permutation π to τ by $d(\pi, \tau)$. Recall that $d(\cdot, \cdot)$ is the graph metric in the Cayley graph of S_n w.r.t. the generating set of adjacent transpositions $(1, 2), (2, 3), \ldots, (n, n - 1)$. We will say that two permutations π and τ are *at distance t* if $d(\pi, \tau) = t$. If λ and μ are two diagrams, define their distance to be

$$\Delta = \Delta(\lambda, \mu) := \frac{1}{2} \sum_{i=1}^{n} |\lambda_i - \mu_i|.$$

Let π and τ be any two permutations. We are interested in the Lipschitz constant of this mapping, i.e.,

$$L(n, t) := \max \frac{\Delta(\lambda(\pi), \lambda(\tau))}{d(\pi, \tau)}$$

where the maximum is over all $\pi, \tau \in S_n$ with $d(\pi, \tau) = t$.

The choice of left-multiplication above is in fact without loss of generality. By Proposition 1.2 the shape of a permutation and its inverse under the RSK correspondence are the same. Our results thus all follow immediately for right-multiplication since $\tau = (i, i + 1) \circ \pi$ is equivalent to $\pi^{-1} = \tau^{-1} \circ (i, i + 1)$.

In general, although d and Δ are natural metrics to study for permutations and diagrams respectively, the same question can be asked for other metrics. We discuss the extension of our results to other metrics on permutations in Section 4.

2. Exact bounds on the Lipschitz constant for a single transposition

In this section we will show upper bounds on the Lipschitz constant when the number of transpositions $t = 1$. We also give a construction of a family of permutations which achieve this bound asymptotically.

2.1. Upper bounds

The first step of the proof is to show that left-multiplying a permutation by a transposition can result in only a bounded number of cells being different in each row of the diagram.

Proposition 2.1. *Let $\pi, \tau \in S_n$ and let λ, μ be the respective diagrams. Suppose that $\tau = (i, i + 1) \circ \pi$, and $\pi_i < \pi_{i+1}$. Then,*

$$\forall 1 \leqslant j \leqslant n, \quad \sum_{i=1}^{j} \mu_i \leqslant \sum_{i=1}^{j} \lambda_i \leqslant \sum_{i=1}^{j} \mu_i + 1. \tag{2.1}$$

Proof. Suppose that the largest cardinality of the union of j increasing subsequences in π is ℓ. Suppose there is a subsequence which includes the pair that is being transposed in π. By deleting one of the elements of the pair we obtain a set of j increasing subsequences of τ whose cardinality is at least $\ell - 1$. If there is no such subsequence, then the same j subsequences are also increasing in τ. By Greene's Theorem 1.3 this implies $\sum_{i=1}^{j} \lambda_i \leqslant \sum_{i=1}^{j} \mu_i + 1$.

For the lower bound, consider the largest cardinality of the union of j increasing sequences in τ. No subsequence in this union can contain both of the elements involved in the transposition. Since the pair involved in the transposition have no other elements between them in both π and τ the subsequences are also increasing in π. We conclude in the same way that $\sum_{i=1}^{j} \mu_i \leqslant \sum_{i=1}^{j} \lambda_i$. $\quad\square$

Fig. 1 will be useful in the following discussion. It depicts the union of two diagrams λ and μ, which is also a Young diagram. The symmetric difference consists of the cells marked by a dot. The remaining set of cells of the diagram labeled W is the intersection of λ and μ and this is a Young diagram as well.

Corollary 2.2. *Let $\pi, \tau \in S_n$ where $d(\pi, \tau) = 1$ and let λ, μ be the corresponding diagrams. Then at most one cell in each row and each column of the union of λ and μ can be in the symmetric difference.*

Proof. To see this for a row r, consider inequality (2.1) for $j = r$ and for $j = r - 1$ and take their difference. A similar argument applied to the reversed permutations implies the claim for columns (see Proposition 1.1). $\quad\square$
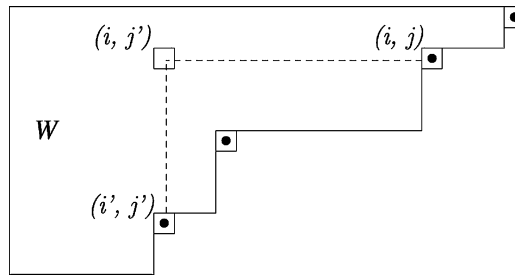
Fig. 1. The union of λ and μ with cells of the symmetric difference marked.

Theorem 2.3. *Let π and τ be permutations in $\mathcal{S}_n$ with respective Young diagrams λ and μ, and suppose that $d(\pi, \tau) = 1$. Then*

$$\Delta = \Delta(\lambda, \mu) \leqslant \sqrt{\frac{n}{2}}.$$

Proof. As shown in Fig. 1, let (i, j) and (i', j') be the coordinates of two distinct cells in the symmetric difference. By Corollary 2.2, $i \neq i'$ and $j \neq j'$, and $(\min(i, i'), \min(j, j')) \in W$. This gives a $1 : 1$ map from unordered pairs of cells in the symmetric difference into W. Therefore,

$$\binom{2\Delta}{2} \leqslant n - \Delta$$

implying the required bound

$$\Delta \leqslant \sqrt{\frac{n}{2}}. \quad \square$$

2.2. Construction

In this section we construct pairs of permutations in S_n which differ by a single transposition whose corresponding Young diagrams differ by at least $(1 - o(1))\sqrt{n/2}$ cells, matching the upper bound in Theorem 3.2 asymptotically. The following lemma characterizes the shape of a permutation by the cardinalities of increasing and decreasing subsequences.

Lemma 2.4. *Let $\pi \in S_n$ be a permutation whose elements can be decomposed in the following two ways:* (i) *into increasing subsequences of cardinalities $\lambda_1, \lambda_2, \ldots$, and* (ii) *into decreasing subsequences of cardinalities $\lambda'_1, \lambda'_2, \ldots$, where the partitions λ and λ' are conjugate. Then $\lambda = \lambda(\pi)$.*

Proof. By Greene's Theorem 1.3 it suffices to show that for each r, the largest cardinality of the union of r increasing sequences in π is $\sum_{i \leqslant r} \lambda_i$. By assumption we know it is at least this number and we need to show the opposite inequality. Namely, that if $s_1, \ldots, s_j$ is a collection of disjoint increasing sequences in π, $\sum_{i=1}^{r} |s_i| \leqslant \sum_{i \leqslant r} \lambda_i$.

By assumption, there is a decomposition $d_1, d_2, \ldots$ of π into disjoint decreasing subsequences of cardinalities $\lambda'_1, \lambda'_2, \ldots$. But each s_i and d_j can have at most one element in common, so that

$$\sum_{i=1}^{r} |s_i| = \sum_{r \geqslant i \geqslant 1, \, j} |s_i \cap d_j| \leqslant \sum_{j} \min\{|d_j|, r\} = \sum_{j} \min\{\lambda'_j, r\} = \sum_{i=1}^{r} \lambda_i$$

where the last equality follows because the partitions $\lambda_1, \lambda_2, \ldots$ and $\lambda'_1, \lambda'_2, \ldots$ are conjugate. $\quad \square$

Theorem 2.5. *For every n there are permutations $\pi, \tau \in S_n$ with $d(\pi, \tau) = 1$ and respective shapes λ, μ such that $\Delta(\lambda, \mu) \geqslant (1 - o(1))\sqrt{n/2}$.*

Proof. Our proof says, in fact, a little more than what is stated. Namely for $n = (k+1)^2/2$ with k an odd integer, we will construct two permutations π and τ of shapes $\lambda = (k+1, k-1, k-1, \ldots, 2, 2)$ and $\mu = (k, k, k-2, k-2, \ldots, 1, 1)$ which differ by exactly one cell in each row and column, giving $\Delta = \sqrt{n/2}$. Thus it can be verified that together with Theorem 2.3 this gives a complete answer to our question for n of this form. For other values of n we get the result by padding this basic construction. In the discussion that follows we decompose these permutations into monotone subsequences. The decompositions we exhibit are not necessarily unique, but for our purpose any decomposition suffices.

The construction can, perhaps, be best understood by observing alongside with the general discussion a concrete special case. So we intersperse our general constructions with an illustration that shows how things work for $n = 18$ ($k = 5$). We start by dividing the elements of $[n]$ into three categories according to their magnitude. The "small" elements are those in the interval $[1, n/2 - \frac{k+1}{2}]$. The next $k+1$ elements, i.e., interval $[n/2 - \frac{k-1}{2}, n/2 + \frac{k+1}{2}]$ are "intermediate" and members of the interval $[n/2 + \frac{k+3}{2}, n]$ are "big".

We further subdivide the big elements (in order) into blocks $b_1, \ldots, b_{(k-1)/2}$. The small elements are split (in order) into blocks $s_{(k-1)/2}, \ldots, s_1$. Both s_i and b_i have cardinality $2i$.

$$\underbrace{\overset{s_2}{(1\ 2\ 3\ 4)}\ \overset{s_1}{(5\ 6)}}_{\text{small}}\ \underbrace{7\ 8\ 9\ 10\ 11\ 12}_{\text{intermediate}}\ \underbrace{\overset{b_1}{(13\ 14)}\ \overset{b_2}{(15\ 16\ 17\ 18)}}_{\text{big}}.$$

The permutation π is constructed by spreading out the intermediate elements with $n/2$ and $n/2 + 1$ remaining fixed points (see below). The blocks of big elements are then inserted in the order $b_{(k-1)/2}, \ldots, b_1$ in the spaces between the smaller intermediate elements while the blocks of small elements are inserted in the order $s_1, \ldots, s_{(k-1)/2}$ in the spaces between the larger intermediate elements. To obtain τ we apply the transposition $(n/2, n/2 + 1)$ to π. The permutations are defined in this manner with a view to decomposing them into increasing and decreasing sequences of desired cardinalities.

$$\pi = 7 (15\ 16\ 17\ 18)\ 8\ (13\ 14)\ \underline{9\ 10}\ (5\ 6)\ 11\ (1\ 2\ 3\ 4)\ 12,$$

$$\tau = 7 (15\ 16\ 17\ 18)\ 8\ (13\ 14)\ \underline{10\ 9}\ (5\ 6)\ 11\ (1\ 2\ 3\ 4)\ 12.$$

From the construction we claim that π and τ can be decomposed into a disjoint union of increasing subsequences of cardinalities $(k+1, k-1, k-1, \ldots, 2, 2)$ and $(k, k, k-2, k-2, \ldots, 1, 1)$ respectively. For π the increasing sequences consist of (i) the intermediate elements, which in our example is $7, 8, 9, 10, 11, 12$; (ii) the blocks of small elements, i.e., $1, 2, 3, 4$ and $5, 6$, and (iii) the blocks of big elements, i.e., $15, 16, 17, 18$ and $13, 14$.

The permutation τ can be decomposed into the increasing subsequences of the following three types: (i) an intermediate element and the block of big elements following it, which in the example are $7, 15, 16, 17, 18$ and $8, 13, 14$; (ii) a block of small elements and the following intermediate element, i.e., $5, 6, 11$ and $1, 2, 3, 4, 12$, and (iii) the two subsequences of length one consisting of one of the two middle intermediate elements, i.e. 10 and 9.

The proof that π and τ have the shapes $\lambda = (k+1, k-1, k-1, \ldots, 2, 2)$ and $\mu = (k, k, k-2, k-2, \ldots, 1, 1)$ respectively uses Lemma 2.4. It is enough to decompose π and τ into a union of decreasing sequences whose cardinalities are given by the respective conjugate sequences. Note that as it happens, the shapes λ and μ are conjugates.

We assign the elements of π to decreasing subsequences $d_1, \ldots, d_{k+1}$ of cardinalities $k, k, \ldots, 1, 1$ as follows. Since the d_i are subsequences, elements in them appear in the same order as in the permutation. Secondly, the assignment is made so that each subsequence has exactly one of the intermediate elements, and it appears after any of the big elements and before any of the small elements. (There is more than one way to do this.) We first see how this is done in the example.

We first construct d_1 and d_2 which are both decreasing sequences of length 5. The largest elements in the blocks b_i, the largest elements in the blocks s_i and one of the middle intermediate elements are assigned to d_1. Then we choose d_2 similarly from among the remaining elements.

$$d_1 : 7(15 \ 16 \ 17 \ \boxed{18}) \ 8 \ (\ 13 \ \boxed{14}) \ \boxed{9} \ 10 \ (\ 5 \ \boxed{6}) \ 11 \ (1 \ 2 \ 3 \ \boxed{4}) \ 12,$$

$$d_2 : 7(15 \ 16 \ \boxed{17} \ \cancel{18} \ ) \ 8 \ (\boxed{13} \ \cancel{14} \ ) \ \cancel{9} \ \boxed{10} \ (\boxed{5} \ \cancel{6}) \ 11 \ (1 \ 2 \ \boxed{3} \ \cancel{4} \ ) \ 12.$$

The remaining elements can be seen to have the same structure recursively (the remaining elements appear in the same relative order as would the elements of the permutation for $n = 8$), where the brackets indicate blocks of big and small elements as before.

$$7(15 \ 16) \ 8 \ 11 \ (1 \ 2) \ 12.$$

To assign elements to d_3 and d_4, we want to continue with the strategy of choosing the largest elements that remain in the blocks. Note that since the big elements 13 and 14 have been assigned, there are no big elements that follow the element 8, and it now becomes "available". Thus d_3 and d_4 are constructed by assigning the largest elements that remain in the small and big blocks and one of the remaining intermediate elements in the middle of the blocks.

$$d_3 : 7(15 \ \boxed{16} \ \cancel{17} \ \cancel{18}) \ \boxed{8} \ (\cancel{13} \ \cancel{14}) \ \cancel{9} \ \cancel{10} \ (\cancel{5} \ \cancel{6}) \ 11 \ (\ 1 \ \boxed{2} \ \cancel{3} \ \cancel{4}) \ 12,$$

$$d_4 : 7(\boxed{15} \ \cancel{16} \ \cancel{17} \ \cancel{18}) \ \cancel{8} \ (\cancel{13} \ \cancel{14}) \ \cancel{9} \ \cancel{10} \ (\cancel{5} \ \cancel{6}) \ \boxed{11} \ (\boxed{1} \ \cancel{2} \ \cancel{3} \ \cancel{4}) \ 12.$$

Proceeding the same way, we obtain the subsequences: $d_1 = 18, 14, 9, 6, 4$, $d_2 = 17, 13, 10, 5, 3$, $d_3 = 16, 8, 2$, $d_4 = 15, 11, 1$, $d_5 = 7$, $d_6 = 12$. In general, the assignment is done as follows.

- The i-th largest element in each block of big elements, is assigned to the subsequence d_i.
- The i-th largest element in each block of small elements, is assigned to the subsequence d_i.
- For the intermediate elements, assign the lower $(k + 1)/2$ elements to the subsequences $d_k, d_{k-2}, \ldots, d_1$ (in that order) and the top $(k+1)/2$ elements to $d_2, \ldots, d_{k-1}, d_{k+1}$ (in that order).

Clearly, this is a decomposition of $[n]$ with exactly $k - 2\lfloor(i - 1)/2\rfloor$ elements in d_i. It remains to show that each d_i is a decreasing subsequence. By the construction of the permutation, the big and small elements in d_i form a decreasing subsequence since each of them is from a different block. Secondly, the intermediate element in d_i appears after all the big elements and before any of the small ones.

Similarly, for the permutation τ, we define the decreasing subsequences $f_1, \ldots, f_k$ of cardinalities $k+1, k-1, k-1, \ldots, 2, 2$, where $|f_i| = k + 1 - 2\lfloor i/2 \rfloor$. As before, the assignment is made so that each sequence but for one (which has the two middle intermediate elements) has at most one intermediate element, and at most one element from each of the small and the big blocks. In our example, we construct f_1, a subsequence of length 6, by taking the largest element from each block and the two middle intermediate elements.

$$f_1 : 7(15 \ 16 \ 17 \ \boxed{18}) \ 8 \ (13 \ \boxed{14}) \ \boxed{10} \ \boxed{9} \ (5 \ \boxed{6}) \ 11 \ (1 \ 2 \ 3 \ \boxed{4}) \ 12.$$

Next, we choose f_2 and f_3 which are both subsequences of length 4. At this point, we cannot continue to follow the strategy of assigning the largest elements from each block to f_2 (by choosing $17, 13, 5, 3$) as in the next step we would fail to construct f_3 of length 4. Instead, note that when only one element remains in a block of small elements, the intermediate element which follows that block has not yet been assigned and it does not follow any other small elements. Thus the strategy for f_2 is to assign to it the largest elements from all blocks except from s_1 in which only one element remains, and to assign the intermediate element following s_1 to f_2. To construct f_3, we take the largest remaining elements in all the blocks, and the intermediate element that precedes the block of big elements whose smallest element was assigned to f_2. Diagrammatically, we have

$$f_2 : 7(15 \ 16 \ \boxed{17} \ \cancel{18}) \ 8 \ (\boxed{13} \ \cancel{14}) \ \cancel{10} \ \cancel{9} \ (\ 5 \ \cancel{6}) \ \boxed{11} \ (1 \ 2 \ \boxed{3} \ \cancel{4}) \ 12,$$

$$f_3 : 7(15 \ \boxed{16} \ \cancel{17} \ \cancel{18}) \ \boxed{8} \ (\cancel{13} \ \cancel{14}) \ \cancel{10} \ \cancel{9} \ (\boxed{5} \ \cancel{6}) \ \cancel{11} \ (1 \ \boxed{2} \ \cancel{3} \ \cancel{4}) \ 12.$$

Repeating the same arguments for the remaining elements, we obtain the following subsequences for the example: $f_1 = 18, 14, 10, 9, 6, 4$, $f_2 = 17, 13, 11, 3$, $f_3 = 16, 8, 5, 2$, $f_4 = 15, 12$, $f_5 = 7, 1$. In general, the subsequences can be defined as follows.

- The i-th largest element in each block of big elements is assigned to f_i.
- The smallest element in a block of small elements s_j is assigned to f_{2j+1}. Among the remaining elements, the i-th largest element goes to f_i.
- The lower $(k-1)/2$ of the intermediate elements go to $f_k, f_{k-2}, \ldots, f_3$ (in that order). The top $(k-1)/2$ elements to $f_2, \ldots, f_{k-1}$ (in that order). The two middle intermediate elements are in f_1.

As before, the f_i constitute a decomposition and they have the appropriate sizes. By construction, the big and small elements in any subsequence f_i form a decreasing subsequence. Lastly, for $i \neq 1$ there is at most one intermediate element in f_i and if one exists, it appears after all the big elements and before all the small ones. For $i = 1$, the two intermediate elements appear consecutively in decreasing order, after all big elements and before all small ones. Thus, π and τ have the claimed shapes and it follows that

$$\Delta = \frac{k+1}{2} = \sqrt{\frac{n}{2}}.$$

For n not of the form $(k+1)^2/2$, we construct two permutations as follows. Let $n_0 < n$ be the largest integer such that $n_0 = (k+1)^2/2$ for odd k. The first n_0 elements of π and τ are set according to the construction above on n_0 elements. The last $n - n_0$ elements of both π and τ are $n_0 + 1, \ldots, n$. Then, we have that

$$\Delta = \sqrt{\frac{n_0}{2}} \geqslant (1 - o(1)) \sqrt{\frac{n}{2}}. \qquad \square$$

We have carried out computer simulations and found other pairs of permutations for which the bound holds with equality. Several mysteries remain here, a few of which we mention in Section 4.

3. Bounds on the Lipschitz constant for $t > 1$

In this section we show bounds on the Lipschitz constant for $t > 1$. Extending the arguments from the previous section for both the upper and lower bound gives bounds that are tight up to constant factors for $t = O(1)$. In the latter half of this section we give a more complicated argument that yields an improved upper bound for general t.

3.1. A construction for permutations at linear distance t

The construction for the case of one transposition can be extended to the case of more than one transposition as follows.

Theorem 3.1. *Let $t \leqslant n/2$. For every n there are permutations $\pi, \tau \in S_n$ with $d(\pi, \tau) = t$ and respective shapes λ, μ such that $\Delta(\lambda, \mu) \geqslant (1 - \sqrt{t/2n})\sqrt{nt/2}$.*

Proof. Let $k = \lfloor \sqrt{2n/t} - 1 \rfloor$, and $m = (k+1)^2/2$ so that $mt \leqslant n$. Divide the first mt elements of $[n]$ into t blocks of length m each. To construct the permutations, in each block we permute the elements as in the construction for one transposition, and then concatenate the blocks with the remaining $n - mt$ elements following. Then, it is not difficult to see that the RSK algorithm on this pair of permutations will result in a shape with t of the smaller Young diagrams corresponding to each block being pasted one after the other, with an additional $n - mt$ boxes in the top row of each diagram. Then, $\Delta = t\sqrt{m/2} \geqslant \sqrt{nt/2}(1 - \sqrt{t/2n})$. Thus when $t = o(n)$, $\Delta \geqslant (1 - o(1))\sqrt{nt/2}$. $\square$

3.2. Upper bounds
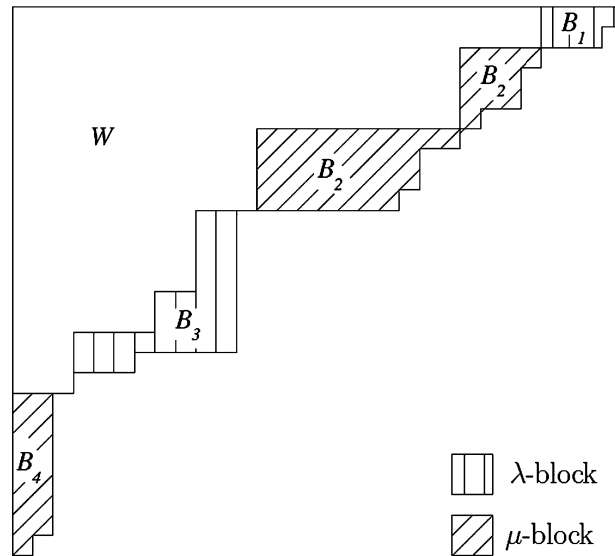
We start with an easy observation:

Fig. 2. The union of λ and μ and the symmetric difference split into blocks.

Theorem 3.2. *Let π and τ be permutations in $\mathcal{S}_n$ such that $d(\pi, \tau) = t$. Let λ and μ be the respective Young diagrams. Then*

$$\Delta = \Delta(\lambda, \mu) \leqslant t\sqrt{\frac{n}{2}}.$$

Proof. Since $d(\pi, \tau) = t$, there is a sequence of permutations $\pi = \sigma_0, \sigma_1, \ldots, \sigma_t = \tau$ such that for each $0 \leqslant i < t$, σ_i and σ_{i+1} differ by an adjacent transposition. The distance $\Delta(\cdot, \cdot)$ is a metric on diagrams and hence the bound follows by the triangle inequality from Theorem 2.3. $\square$

We do not see how to appropriately adapt the bijective argument of Theorem 2.3. However, the following argument yields a near-optimal bound.

Theorem 3.3. *Let $\pi, \tau \in \mathcal{S}_n$ be such that $d(\pi, \tau) = t$. Let λ, μ be the corresponding diagrams. Then*

$$\Delta(\lambda, \mu) \leqslant O(\sqrt{nt \log t}).$$

We start by showing some preliminary results that will be useful in the proof. Suppose π and τ are two permutations such that $d(\pi, \tau) = t$. Let $\pi = \sigma_0, \sigma_1, \ldots, \sigma_t = \tau$ be a sequence of permutations such that for each $0 \leqslant i < t$, σ_i and σ_{i+1} differ by an adjacent transposition. Say that s (resp. r) of the transpositions put the relevant pair in decreasing (resp. increasing) order, where $t = r + s$. Let λ and μ be the diagrams corresponding to π and τ respectively.

Lemma 3.4. *Let π, τ be as above. Then,*

$$\forall 1 \leqslant j \leqslant n, \quad \sum_{i=1}^{j} \mu_i - r \leqslant \sum_{i=1}^{j} \lambda_i \leqslant \sum_{i=1}^{j} \mu_i + s. \tag{3.1}$$

Proof. For each pair σ_i, σ_{i+1} in the sequence of permutations $\pi = \sigma_0, \sigma_1, \ldots, \sigma_t = \tau$, by Proposition 2.1 the inequality (2.1) holds for the diagrams corresponding to σ_i and σ_{i+1}. The result is obtained by adding up all these inequalities. $\square$

In Fig. 2 we depict the union of two diagrams. Their intersection is labeled W as before. We split the symmetric difference of the two diagrams into *blocks*. We say that j indexes a λ-row if $\lambda_j > \mu_j$.

A maximal interval of λ-rows determines a λ-*pre-block*. A maximal collection of consecutive λ-pre-blocks constitutes a λ-*block*. We likewise define μ-blocks. Blocks are labeled B_i as in the figure. The number of cells in a set S will be denoted by $A(S)$. We use the following fact about the sizes of the blocks.

Proposition 3.5. *Let $d(\pi, \tau) = t$ with corresponding diagrams λ, μ and let B be a block in the union of the diagrams, then $A(B) \leqslant t$.*

Proof. This bound is obtained from Lemma 3.4 as follows. Let B reside in the set of rows I of the diagram. Assuming it exists, let i_0 be the row just preceding I, and $i_1 = \max I$. Then the bound is obtained by subtracting the inequality (3.1) corresponding to $j = i_0$ from the inequality corresponding to $j = i_1$, and using the fact that $r + s = t$. If there is no row i_0, then the bound is immediate from the inequality for $j = i_1$. $\quad\square$

The main step in the proof of Theorem 3.3 is the following lemma about two sequences of integers.

Lemma 3.6. *Let $k \geqslant 2$, $T \geqslant 3$ and let $a_1, \ldots, a_k$ and $b_1, \ldots, b_k$ be two sequences of positive integers. Denote $\Delta = \sum_{i=1}^{k} a_i b_i$ and $N = \sum_{1 \leqslant i \leqslant j \leqslant k} a_i b_j$. If*

$$a_1 = b_k = 1 \quad and \quad \forall i, \quad a_i b_i \leqslant T,$$

then

$$\Delta \leqslant \sqrt{32 N T \log T}.$$

This bound is tight up to constants.

We first show how to derive the theorem from Lemma 3.6.

Let λ and μ be two diagrams of size n (not necessarily corresponding to permutations at distance t). For the union of these diagrams, define the blocks of the symmetric difference $\{B_i\}$ and W as before. Suppose that for each block B, $A(B) \leqslant t$. To prove Theorem 3.3, it is sufficient to show that for these diagrams,

$$\frac{1}{2} \sum_i A(B_i) \leqslant O\left(\sqrt{t \log t \left[A(W) + \frac{1}{2} \sum_i A(B_i) \right]} \right). \tag{3.2}$$

With this formulation in mind, we can make the following assumptions about the pair of diagrams. The aim is to make a number of transformations and show that the pair of diagrams can be assumed to be of the form shown in Fig. 6.

Reduction 1. For any row i, $\lambda_i \neq \mu_i$ and similarly, for any column j, $\lambda'_j \neq \mu'_j$. If this is not the case (as in the shaded part of Fig. 3), we delete such rows or columns from both λ and μ. Consequently, $A(W)$ decreases, whereas $\sum_i A(B_i)$ remains unchanged. Thus, if the bound holds for the new pair of diagrams, it holds as well for the old pair.

Reduction 2. In general, each block is a skew-diagram (the set theoretic difference of a diagram and another contained in it). However, as we show, we may assume it is a Young diagram. The dotted lines in Fig. 4 mark the "shade" of a block in W determined by its top row and leftmost column. If a block is not a (left-aligned) tableau, we can change it to one by removing the cells of W in its shade and replacing it with a Young diagram of area $A(B)$ contained in the union of the block and its shade.

This transformation decreases $A(W)$ and keeps the size of the block fixed. Secondly, we may assume that the transformation is done so that all rows of a block, with the possible exception of the
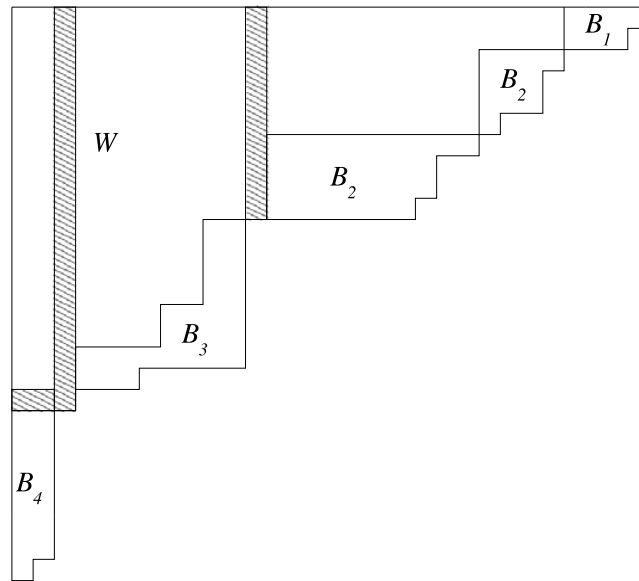
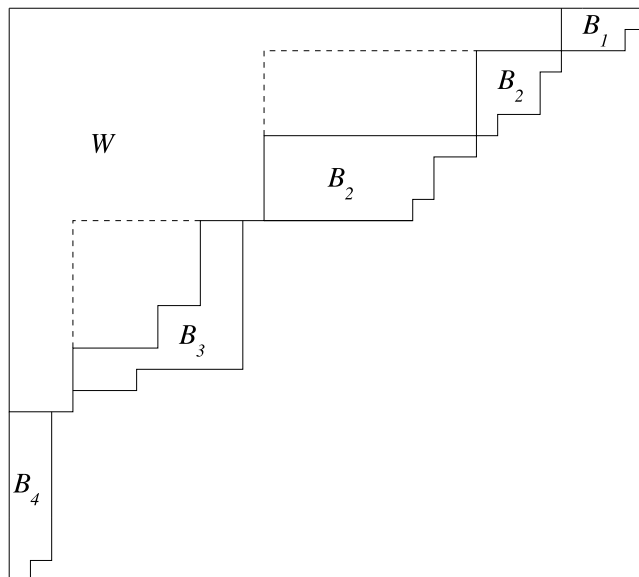**Fig. 3.** Rows and columns of W that may be removed.

Fig. 4. The top left corner of the block.

last one have the same length. The result of such a transformation on the blocks B_2 and B_3 is shown in Fig. 5.

Reduction 3. We may assume that the topmost block B_1 has a single row. Otherwise, we can shift all the cells of B_1 to the first row without changing any $A(B_i)$ or $A(W)$. We can then delete any rows of W which are of the same length in λ and μ. By similar reasoning, we may assume that the bottommost block has a single column.

Thus, we may assume that the diagrams are as shown in Fig. 6 and that the sizes of the blocks are bounded by t. As in the figure, let a_i and b_i denote the lengths of the vertical and horizontal sides of the rectangle which bounds the block B_i. Thus the area $A(W)$ can be written as a sum of areas of rectangles $a_i b_j$ whose sides are determined by the side lengths of pairs of blocks. Also note that by our construction of the blocks, $a_i b_i \leqslant 2t$.

To obtain the formulation of the lemma, suppose that we add cells to the last row of each block to complete it to a rectangle. Denote the modified blocks by B'. Then for each block, $A(B') \leqslant 2A(B)$.
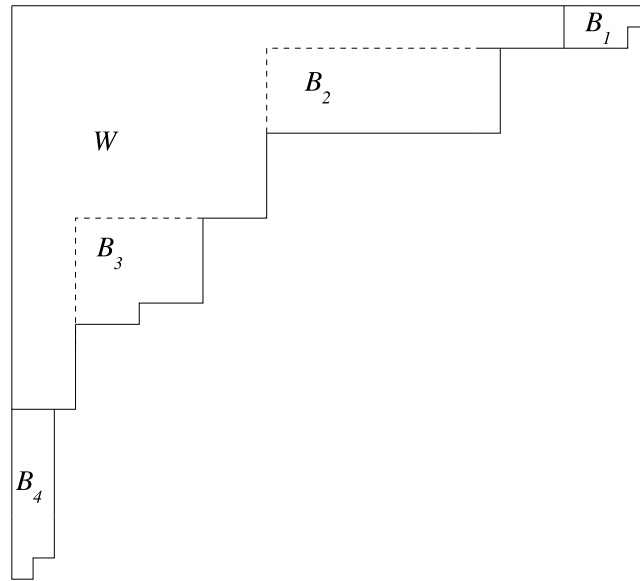
Fig. 5. All blocks are Young diagrams.
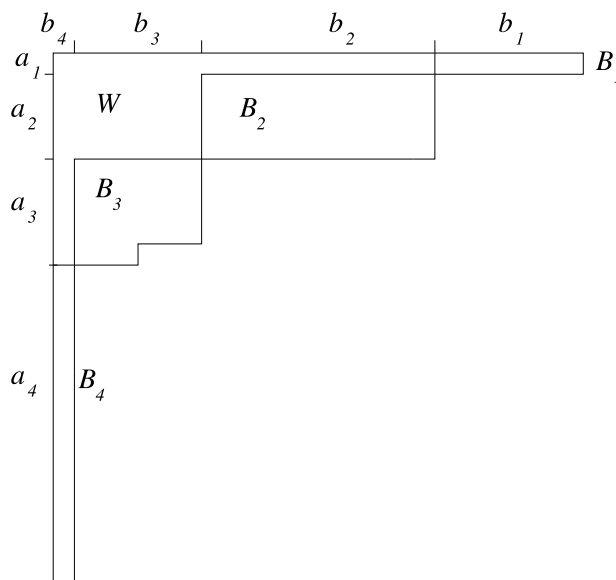
Fig. 6. A box of side lengths a_i and b_i bounds B_i.

If we show the bound for these modified diagrams with a bound of $2t$ for each block, then the bound is implied for the original diagrams since the constants can be absorbed by the $O(\cdot)$. Formally, this follows from the following inequalities.

1. $\frac{1}{2}\sum_i A(B_i) \leqslant \frac{1}{2}\sum_i A(B'_i)$.
2. $A(W) + \frac{1}{2}\sum_i A(B'_i) \leqslant 2(A(W) + \frac{1}{2}\sum_i A(B_i))$.

Thus Lemma 3.6 implies the bound (3.2) for a pair of diagrams as above and we have verified that to prove Theorem 3.3 it is sufficient to prove the lemma.

Proof of Lemma 3.6. We will minimize N/Δ^2. For $k = 2$, the lemma can be easily verified by calculation once we use the fact that $a_1 = b_2 = 1$. Thus we will assume that $k \geqslant 3$. Consider the following relaxation of the minimization problem where the a_i, b_j are not necessarily integral.

$$\min \frac{N}{\Delta^2} = \frac{\sum_{1 \leqslant i \leqslant j \leqslant k} a_i b_j}{(\sum_{i=1}^{k} a_i b_i)^2},$$

$$\text{s.t. } a_1 = b_k = 1,$$

$$a_i b_i \leqslant T, \quad 1 \leqslant i \leqslant k,$$

$$a_i \geqslant 1, \quad 2 \leqslant i \leqslant k,$$

$$b_i \geqslant 1, \quad 1 \leqslant i \leqslant k - 1.$$

We will use the method of Lagrange multipliers (see Appendix A for a brief introduction) to obtain a lower bound on the value of the objective above at any local optimum. Since the problem is a relaxation of the discrete minimization problem, this also lower bounds the objective of the discrete problem. We obtain the following Lagrangian for the relaxation above.

$$\min \mathcal{L} = \frac{N}{\Delta^2} - \sum_{i=1}^{k} \lambda_i (a_i b_i - T) - \sum_{i=1}^{k} \mu_i (a_i - 1) - \sum_{i=1}^{k} \nu_i (b_i - 1).$$

The Karush–Kuhn–Tucker conditions yield the following necessary conditions for minimality:

$$\frac{\partial}{\partial a_i} \mathcal{L} = \frac{\partial}{\partial a_i} \frac{N}{\Delta^2} - \lambda_i b_i - \mu_i = 0, \quad 1 \leqslant i \leqslant k, \tag{3.3}$$

$$\frac{\partial}{\partial b_i} \mathcal{L} = \frac{\partial}{\partial b_i} \frac{N}{\Delta^2} - \lambda_i a_i - \nu_i = 0, \quad 1 \leqslant i \leqslant k, \tag{3.4}$$

$$\lambda_i \geqslant 0, \quad \lambda_i (T - a_i b_i) = 0, \quad 1 \leqslant i \leqslant k,$$

$$\mu_i \geqslant 0, \quad \mu_i (a_i - 1) = 0, \quad 2 \leqslant i \leqslant k,$$

$$\nu_i \geqslant 0, \quad \nu_i (b_i - 1) = 0, \quad 1 \leqslant i \leqslant k - 1. \tag{3.5}$$

From these conditions, we can show that at optimality either $a_i b_i = T$ or 1. Suppose that for some i, $a_i b_i < T$. Note that by the conditions above, this implies $\lambda_i = 0$. Now, if $a_i b_i \neq 1$, at least one of a_i or b_i is > 1. Assume without loss of generality that $b_i > 1$ (the argument in the other case is exactly the same). In this case $\nu_i = 0$ by (3.5). Hence from (3.4) above, we have

$$\frac{\partial}{\partial b_i} \frac{N}{\Delta^2} = 0$$

and therefore, since $\Delta > 0$

$$\Delta^2 \frac{\partial N}{\partial b_i} = 2 \Delta N \frac{\partial \Delta}{\partial b_i}$$

$$\Rightarrow \quad \frac{\sum_{j=1}^{i} a_j}{2 a_i} = \frac{N}{\Delta}. \tag{3.6}$$

Now we show that it is possible to increase b_i by a factor $(1 + \varepsilon)$ for $\varepsilon > 0$ so that N/Δ^2 decreases and we can conclude that the solution is not optimal. This is allowed, at least for $\varepsilon > 0$ small enough, since, by assumption $a_i b_i < T$. Let N' and Δ' be the summations as defined before for the sequences where we replace b_i by $b_i(1 + \varepsilon)$.

$$\frac{N'}{(\Delta')^2} = \frac{N + b_i \varepsilon \sum_{j=1}^{i} a_j}{(\Delta + a_i b_i \varepsilon)^2}.$$

To prove the claim $N'/(\Delta')^2 < N/\Delta^2$, using the right-hand side above, it is enough to show that

$$\Delta^2 b_i \varepsilon \sum_{j=1}^{i} a_j < N \left(2 \Delta a_i b_i \varepsilon + (a_i b_i \varepsilon)^2 \right).$$

Or equivalently, dividing throughout by $2a_i b_i \varepsilon \Delta^2$, that

$$\frac{\sum_{j=1}^{i} a_j}{2a_i} < \frac{N}{\Delta} + \frac{N a_i b_i \varepsilon}{2\Delta^2}.$$

This inequality follows by (3.6). The left-hand term equals the first term on the right and $\varepsilon > 0$.

The next step is to argue that it is enough to show the claimed bound assuming that the blocks are arranged in a specific manner (i.e., the sequences are of a certain form). In particular, the blocks of area T are arranged such that a_i is increasing and b_i is decreasing. Secondly, the blocks of area 1 occur after all blocks such that $a_i \leqslant b_i$ and before all blocks such that $a_i > b_i$. This can be argued by noticing that such an arrangement can be achieved by exchanging blocks which are out of order since Δ remains unchanged and N does not increase. Thus a lower bound on N/Δ^2 for the modified sequence is a lower bound on the corresponding quantity for the original sequence.

We next argue that, in fact, w.l.o.g. no block has area 1. Recall that we wish to show

$$32 T \log T N - \Delta^2 \geqslant 0.$$

We will show that if we add a single block of area 1 then

$$32 T \log T N' - \Delta'^2 \geqslant 32 T \log T N - \Delta^2 \tag{3.7}$$

where N' and Δ' are the modified values of N and Δ. This inequality above allows us to reduce the argument to the case when there are no blocks of area 1. Let the shorter sequence have k terms. Note that $\Delta' = \Delta + 1$ and the change in the number of cells $N' - N$ is at least $\sum_{i=1}^{k} \min(a_i, b_i)$.

Recall that $k, T \geqslant 3$ and for $1 \leqslant i \leqslant k$, $a_i, b_i \geqslant 1$. Thus,

$$\begin{aligned}
(\Delta')^2 - \Delta^2 &= 2\Delta + 1 \\
&\leqslant 2kT + 1 \\
&\leqslant 32kT \log T \\
&\leqslant 32 T \log T \sum_{i=1}^{k} \min(a_i, b_i) \\
&= 32 T \log T (N' - N)
\end{aligned}$$

which implies the required inequality (3.7).

In the next step, we will make a further simplification to the picture. To summarize, we now know that we may optimize over sequences such that each block has size T, $a_1 = b_k = 1$, the sequence a_i is non-decreasing and b_i is non-increasing. The claim is that the optimal solution is of the form where there is some i such that $a_i > 1$ and $b_i > 1$. If not, then it can be checked that $\Delta = \sqrt{8NT}$ and the claimed bound holds.

We relabel the sequences $a_{-\ell_1}, \ldots, a_{-1}, a_1, \ldots, a_k, a_{k+1}, \ldots, a_{k+\ell_2}$ and $b_{-\ell_1}, \ldots, b_{-1}, b_1, \ldots, b_k,$ $b_{k+1}, \ldots, b_{k+\ell_2}$ where $\ell_1, \ell_2 \geqslant 1$ and $k \geqslant 0$ so that $a_i, b_i > 1$ for $1 \leqslant i \leqslant k$. Let N and Δ be the corresponding summations as defined before. We can reformulate the minimization problem as follows.

$$\begin{aligned}
&\min \frac{N}{\Delta^2}, \\
&\text{s.t. } a_i = 1 = b_{k+j}, \quad a_{k+j} = T = b_i, \quad -\ell_1 \leqslant i \leqslant -1, \ 1 \leqslant j \leqslant \ell_2, \\
&a_i b_i = T, \quad \forall i, \\
&a_i, b_i \geqslant 1, \quad 1 \leqslant i \leqslant k. \tag{3.8}
\end{aligned}$$

Solving this optimization problem gives the following conditions for the solutions (see Proposition B.1 in Appendix B for the detailed calculations). The sequences $b_{1 \leqslant i \leqslant k}$ and (hence $a_{1 \leqslant i \leqslant k}$) are a geometric series with

$$c = \frac{b_1}{b_2} = \cdots = \frac{b_{k-1}}{b_k}$$

and

$$c = \frac{a_2}{a_1} = \cdots = \frac{a_k}{a_{k-1}}$$

and the ratio between successive terms $c > 1$. Also,

$$b_k = (c - 1)\ell_2$$

and

$$a_1 = (c - 1)\ell_1.$$

Substituting, we also have that

$$b_1 = c^{k-1}b_k = c^{k-1}(c - 1)\ell_2,$$
$$T = a_1 b_1 = c^{k-1}(c - 1)^2 \ell_1 \ell_2. \tag{3.9}$$

Since $b_k, a_1 > 1$ we have

$$c > 1 + \frac{1}{\max\{\ell_1, \ell_2\}}.$$

Furthermore, $c^{k-1} < T$ and therefore

$$k - 1 < \frac{\log T}{\log c} \leqslant \frac{\log T}{\log(1 + 1/\max\{\ell_1, \ell_2\})}. \tag{3.10}$$

In the next step we will show the w.l.o.g. we may assume $\ell_1 = \ell_2 = 1$. Let $N_0 = \sum_{1 \leqslant i \leqslant j \leqslant k} a_i b_j + \sum_i b_i + \sum_i a_i + 1 + 2T$ and $\Delta_0 = \sum_{i=1}^{k} a_i b_i + 2T$. These are the values of the summations with the first $\ell_1 - 1$ and last $\ell_2 - 1$ members of the sequences removed. Then

$$N = N_0 + T\frac{\ell_1(\ell_1 - 1)}{2} + T\frac{\ell_2(\ell_2 - 1)}{2} + (\ell_1 - 1)\sum_{i=1}^{k} b_i + (\ell_2 - 1)\sum_{i=1}^{k} a_i + (\ell_1 - 1)(\ell_2 - 1),$$

$$\Delta = \Delta_0 + T(\ell_1 + \ell_2 - 2).$$

We have above that $\ell_1, \ell_2 \geqslant 1$. We will show that the optimal of $32T \log T N - \Delta^2$ when N_0 and Δ_0 are fixed is at $\ell_1 = \ell_2 = 1$ by showing

$$32T \log T N - \Delta^2 \geqslant 32T \log T N_0 - \Delta_0^2.$$

Without loss of generality, suppose that $\ell_1 \geqslant \ell_2$ so that $\ell_1 \geqslant 2$. Therefore by (3.10) and using the fact that for $x \leqslant 1/2$, $\log(1 + x) \geqslant x/2$, we have

$$k - 1 < \frac{\log T}{\log(1 + 1/\max\{\ell_1, \ell_2\})} \leqslant 2\log T \max\{\ell_1, \ell_2\}. \tag{3.11}$$

Now, we have

$$32T \log T (N - N_0) \geqslant 32T \log T \left(T\frac{\ell_1(\ell_1 - 1)}{2}\right).$$

On the other hand, by the bound from (3.11) on k,

$$\begin{aligned}
\Delta^2 - \Delta_0^2 &= 2\Delta_0 T(\ell_1 + \ell_2 - 2) + T^2(\ell_1 + \ell_2 - 2)^2 \\
&= 2kT^2(\ell_1 + \ell_2 - 2) + T^2(\ell_1 + \ell_2 - 2)^2 \\
&\leqslant 4kT^2(\ell_1 - 1) + 4T^2(\ell_1 - 1)^2 \\
&\leqslant 12T^2 \log T \ell_1(\ell_1 - 1) + 4T^2(\ell_1 - 1)^2
\end{aligned}$$

$$\leqslant 16 T^2 \log T \ell_1 (\ell_1 - 1)$$

$$\leqslant 32 T \log T \left(T \frac{\ell_1 (\ell_1 - 1)}{2} \right)$$

$$\leqslant 32 T \log T (N - N_0)$$

as required. Finally, if $\ell_1 = \ell_2 = 1$, then using the fact that $b_1 \leqslant T$, from (3.9), we obtain that $c > 2$. Hence by (3.11) $k < 2 \log T + 1$. Note that if $\ell_1 = \ell_2 = 1$, $\Delta = (k + 2)T$. We can then use the following straightforward bound:

$$\frac{N}{\Delta^2} \geqslant \frac{1}{\Delta} = \frac{1}{(k + 2)T} \geqslant \frac{1}{4 T \log T}.$$

Theorem 3.3 now follows. □

As we show next, the upper bound of Lemma 3.6 is tight. To construct a pair of diagrams where the a_i and b_i are integers and $\Delta = \Omega(\sqrt{NT \log T})$ we argue as follows. Let $T = 2^k$, $c = 2$, $b_1 = 1$ and $b_k = T$. Then Δ is at least $\Omega(\sqrt{NT \log T})$. However, the tightness of Theorem 3.3 does not follow from this since it is not clear that there exist corresponding permutations.

4. Conclusions

A number of interesting directions remain for further research.

Characterize extremal permutations   The permutations constructed in Section 2 achieve the maximum difference in the shapes for one transposition. There it was possible to construct the examples by carefully arranging the increasing and decreasing sequences. On the other hand, with the help of a computer, we observed several other examples whose structure we do not completely understand. We know that for Δ to achieve the upper bound, by Greene's Theorem, the permutations must be decomposable into unions of increasing sequences whose sizes are given by the required shape of the diagram. An example of such a pair of permutations from simulation for $n = 18$ is

13 14 10 15 6 1 18 2 16 <u>9 11</u> 12 3 7 17 8 4 5,

13 14 10 15 6 1 18 2 16 <u>11 9</u> 12 3 7 17 8 4 5.

Notice that in this example the permutations cannot be decomposed into conjugate increasing and decreasing sequences as done in our construction. In our view the class of such permutations is an intriguing mathematical object. We would like to know how many such permutations exist, what their structural properties are etc. This seems like a good subject for further work in this area.

Constructions for $t > 1$ transpositions   As mentioned, we do not know whether there exists a pair of permutations corresponding to the diagrams which are tight for Lemma 3.6. We do not see how to extend our construction for one transposition to this case.

Secondly, our constructions achieve $\Omega(\sqrt{nt/2})$ differences when $t \leqslant n/2$. The behavior for larger t is still unclear. For example, the maximum possible value of Δ is $n - 1$, and this is uniquely achieved with $t = \binom{n}{2}$ transpositions. We also know from Theorem 3.1 that we can make $\Delta \geqslant \Omega(n)$ with $t \leqslant O(n)$. We still do not know how large t should be to make $\Delta \geqslant \alpha n$ with α close to 1.

Dependence on transpositions   It would be interesting to obtain more detailed information about the change in Δ as a result of left-multiplication with a transposition. Knuth and Knuth-dual equivalence classes characterize transpositions which keep Δ fixed. What is the expected change in Δ for a transposition in a random permutation? How do the position of the transposition or properties of the permutation affect the change?

Other metrics   In this work we studied the adjacent transposition metric on permutations but there are a number of natural measures for the distance between two permutations which may be worth studying in this setting.

For example it can be verified that up to constants, the same bounds on the Lipschitz constant hold for the distance d' on permutations with respect to general (not necessarily adjacent) transpositions. The lower bounds from Theorems 2.5 and 3.1 hold since the constructions give permutations π and τ which differ by 1 and t transpositions respectively. On the other hand, the upper bounds on the Lipschitz constant follow (and are within a constant factor of the bounds for adjacent transpositions) since by Greene's Theorem the bounds in Proposition 2.1 change only by a small additive constant and in Lemma 3.4 this translates to the absolute value of the difference between the sums being bounded by $2t$ if the permutations differ by the multiplication of t transpositions.

Appendix A. The method of Lagrange multipliers

The method of Lagrange multipliers is used to solve for the maxima or minima of a real-valued multivariate function subject to equality constraints. In particular, the method gives *necessary conditions* for optimality which are the analog of the conditions on the gradient for unconstrained problems. The Karush–Kuhn–Tucker (KKT) conditions for optimality generalize these to the case when some of the constraints may be inequalities. Consider the following optimization problem, where $\mathbf{x} = (x_1, \ldots, x_n)$ and $\alpha_j, \beta_k \in \mathbb{R}$:

$$\min f(\mathbf{x}),$$
$$\text{s.t. } g_j(\mathbf{x}) \leqslant \alpha_j, \quad j = 1, \ldots, \ell,$$
$$h_k(\mathbf{x}) = \beta_k, \quad k = 1, \ldots, m.$$

The *Lagrangian* for this problem is defined to be the function:

$$\mathcal{L} := \mathcal{L}(\mathbf{x}, \lambda_j, \mu_k) = f(\mathbf{x}) - \sum_{j=1}^{\ell} \lambda_j \big(g_j(\mathbf{x}) - \alpha_j\big) - \sum_{k=1}^{m} \mu_k \big(h_k(\mathbf{x}) - \beta_k\big).$$

The KKT conditions say that if a local optimizer $\mathbf{x}^*$ satisfies certain technical "constraint qualifications" (explained below) then there are constants λ_j^* ($j = 1, \ldots, \ell$) and μ_k^* ($k = 1, \ldots, m$) satisfying

$$\nabla \mathcal{L} = \nabla f(\mathbf{x}^*) - \sum_{j=1}^{\ell} \lambda_j^* \nabla g_j(\mathbf{x}^*) - \sum_{k=1}^{m} \mu_k^* \nabla h_k(\mathbf{x}^*) = 0,$$
$$\lambda_i^* \geqslant 0, \quad j = 1, \ldots, \ell,$$
$$\lambda_i^* g_i(\mathbf{x}^*) = 0, \quad j = 1, \ldots, \ell.$$

A number of constraint qualifications are known to be sufficient for the result and in our case, the so-called *Mangasarian–Fromovitz* constraint qualification holds. This condition requires that at $\mathbf{x}^*$, the gradients of any active inequality constraints and the gradients of the equality constraints are *positively-linearly independent*. A collection of vectors $(v_1, \ldots, v_d)$ is positively-linearly dependent if there are $a_1 \geqslant 0, \ldots, a_d \geqslant 0$, not all 0 such that $\sum_i a_i v_i = 0$ for the optimization problems we consider, the constraint qualification can be verified without much difficulty for any possible set of active constraints, so we leave this to the reader and assume that the KKT conditions are satisfied. For more details regarding the method of Lagrange multipliers and extensions, the reader may refer to [2].

Appendix B. Solution to the minimization problem

We show below the calculations that solve the minimization problem in (3.8) which is reproduced below.

$$\min \frac{N}{\Delta^2},$$
$$\text{s.t. } a_i = 1 = b_{k+j}, \qquad a_{k+j} = T = b_i, \quad -\ell_1 \leqslant i \leqslant -1, \ 1 \leqslant j \leqslant \ell_2,$$

$$a_i b_i = T, \quad \forall i,$$

$$a_i, b_i \geqslant 1, \quad 1 \leqslant i \leqslant k.$$

Proposition B.1. *At a minimum of the optimization, the sequences $b_{1 \leqslant i \leqslant k}$ and (hence $a_{1 \leqslant i \leqslant k}$) are a geometric series with*

$$c = \frac{b_1}{b_2} = \cdots = \frac{b_{k-1}}{b_k}$$

and

$$c = \frac{a_2}{a_1} = \cdots = \frac{a_k}{a_{k-1}}$$

and the ratio between successive terms $c > 1$. Also,

$$b_k = (c - 1)\ell_2$$

and

$$a_1 = (c - 1)\ell_1.$$

Proof. We obtain the following Lagrangian.

$$\min \mathcal{L} = \frac{N}{\Delta^2} - \sum_i \lambda_i (a_i b_i - T) - \sum_{i=1}^{k} \mu_i (a_i - 1) - \sum_{i=1}^{k} \nu_i (b_i - 1).$$

From the KKT conditions for optimality, we obtain

$$\frac{\partial}{\partial a_i} \mathcal{L} = \frac{\partial}{\partial a_i} \frac{N}{\Delta^2} - \lambda_i b_i - \mu_i = 0, \quad 1 \leqslant i \leqslant k, \tag{B.1}$$

$$\frac{\partial}{\partial b_i} \mathcal{L} = \frac{\partial}{\partial b_i} \frac{N}{\Delta^2} - \lambda_i a_i - \nu_i = 0, \quad 1 \leqslant i \leqslant k, \tag{B.2}$$

$$\mu_i \geqslant 0, \quad \mu_i (a_i - 1) = 0, \quad 1 \leqslant i \leqslant k,$$

$$\nu_i \geqslant 0, \quad \nu_i (b_i - 1) = 0, \quad 1 \leqslant i \leqslant k.$$

As outlined before, we may assume that the optimal solution is such that $a_i, b_i > 1$ for $1 \leqslant i \leqslant k$. Hence by the conditions above, $\mu_i = \nu_i = 0$. Performing the differentiations in (B.1) (w.r.t. a_i) and (B.2) (w.r.t. b_i) and multiplying them by a_i and b_i respectively we obtain the following relations:

$$\frac{a_i b_i N}{\Delta^4} - \lambda_i a_i b_i - \frac{a_i (b_i + \cdots + b_k + b_{k+1} + \cdots + b_{k+\ell_2})}{\Delta^4} = 0, \quad 1 \leqslant i \leqslant k, \tag{B.3}$$

$$\frac{a_i b_i N}{\Delta^4} - \lambda_i a_i b_i - \frac{b_i (a_{-\ell_1} + \cdots + a_{-1} + a_1 + \cdots + a_i)}{\Delta^4} = 0, \quad 1 \leqslant i \leqslant k. \tag{B.4}$$

Equating (B.3) and (B.4) and canceling terms, we conclude that since $\Delta \neq 0$,

$$a_i (b_i + \cdots + b_k + \ell_2) = b_i (\ell_1 + a_1 + \cdots + a_i), \quad 1 \leqslant i \leqslant k. \tag{B.5}$$

We can solve the above set of relations as follows. Dividing (B.5) by a_i and using the equations corresponding to i and $i+1$ there, and that $a_i b_i = T$, after rearranging terms we obtain the following relations:

$$b_i + b_{i+1} + \cdots + b_k + \ell_2 = \frac{b_i}{a_i}(\ell_1 + a_1 + \cdots + a_i) = b_i^2 \left(\ell_1 + \frac{1}{b_1} + \cdots \frac{1}{b_i} \right),$$

$$b_{i+1} + \cdots + b_k + \ell_2 = \frac{b_{i+1}}{a_{i+1}}(\ell_1 + a_1 + \cdots + a_{i+1}) = b_{i+1}^2 \left(\ell_1 + \frac{1}{b_1} + \cdots + \frac{1}{b_{i+1}} \right).$$

Subtracting we obtain

$$b_i = b_i^2\left(\ell_1 + \frac{1}{b_1} + \cdots \frac{1}{b_i}\right) - b_{i+1}^2\left(\ell_1 + \frac{1}{b_1} + \cdots + \frac{1}{b_{i+1}}\right)$$

$$= (b_i^2 - b_{i+1}^2)\left(\ell_1 + \frac{1}{b_1} + \cdots \frac{1}{b_i}\right) - b_{i+1}, \quad 1 \leqslant i \leqslant k.$$

Rearranging,

$$b_i\left(\ell_1 + \frac{1}{b_1} + \cdots \frac{1}{b_i}\right) = b_{i+1}\left(\ell_1 + \frac{1}{b_1} + \cdots + \frac{1}{b_{i+1}}\right), \quad 1 \leqslant i \leqslant k. \tag{B.6}$$

Let

$$H_i = \ell_1 + \frac{1}{b_1} + \cdots \frac{1}{b_i}$$

so that

$$\frac{1}{b_i} = H_i - H_{i-1}.$$

Rearranging (B.6) and manipulating both sides, we have

$$\frac{H_i}{b_{i+1}} = \frac{H_{i+1}}{b_i}$$

$$\Rightarrow \quad H_i(H_{i+1} - H_i) = H_{i+1}(H_i - H_{i-1})$$

$$\Rightarrow \quad \frac{H_i}{H_{i-1}} = \frac{H_{i+1}}{H_i}$$

$$\Rightarrow \quad \frac{b_{i-1}}{b_i} = \frac{b_i}{b_{i+1}}, \quad 1 < i < k.$$

In other words, we can conclude that $b_{1 \leqslant i \leqslant k}$ (and hence $a_{1 \leqslant i \leqslant k}$) is a geometric series. Let

$$c = \frac{b_1}{b_2} = \cdots = \frac{b_{k-1}}{b_k}$$

and

$$c = \frac{a_2}{a_1} = \cdots = \frac{a_k}{a_{k-1}}.$$

It can be checked that $c > 1$ since for $c = 1$ (B.6) is not satisfied.

Next, suppose we multiply Eq. (B.5) by $c - 1$, we obtain

$$(c-1)a_i(b_i + \cdots + b_k + \ell_2) = (c-1)b_i(\ell_1 + a_1 + \cdots + a_i)$$

$$\Rightarrow \quad a_i(cb_i - b_k + (c-1)\ell_2) = b_i(ca_i - a_1 + (c-1)\ell_1)$$

$$\Rightarrow \quad \frac{a_i}{b_i}(-b_k + (c-1)\ell_2) = -a_1 + (c-1)\ell_1.$$

Since the right-hand side of the last equality is the same for all $1 \leqslant i \leqslant k$, from $i = 1$ and $i = k$, we obtain

$$\left(\frac{a_1}{b_1} - \frac{a_k}{b_k}\right)(-b_k + (c-1)\ell_2) = 0.$$

Now since $c > 1$, $\frac{a_1}{b_1} \neq \frac{a_k}{b_k}$ and therefore

$$b_k = (c-1)\ell_2.$$

By similar arguments,

$$a_1 = (c-1)\ell_1. \quad \square$$

References

[1] G.E. Andrews, The Theory of Partitions, Encyclopedia Math. Appl., vol. 2, Addison–Wesley Publishing Co., Reading, MA, London, Amsterdam, 1976.

[2] D.P. Bertsekas, Nonlinear Programming, Athena Scientific, Belmont, MA, 1999.

[3] P. Diaconis, Group Representations in Probability and Statistics, IMS Lecture Notes Monogr. Ser., vol. 11, Institute of Mathematical Statistics, Hayward, CA, 1988.

[4] W. Fulton, Young Tableaux, London Math. Soc. Stud. Texts, vol. 35, Cambridge University Press, 1997.

[5] C. Greene, An extension of Schensted's theorem, Adv. Math. 14 (1974) 254–265.

[6] G.D. James, The Representation Theory of the Symmetric Groups, Lecture Notes in Math., vol. 682, Springer-Verlag, Berlin, 1978.

[7] D.E. Knuth, Permutations, matrices and generalized Young tableaux, Pacific J. Math. 34 (3) (1970) 709–727.

[8] D.E. Knuth, The Art of Computer Programming, vol. 3: Sorting and Searching, Addison–Wesley, London, 1973, pp. 54–58.

[9] I.G. Macdonald, Symmetric Functions and Hall Polynomials, second ed., Oxford University Press, 1995.

[10] G. de B. Robinson, On representations of the symmetric group, Amer. J. Math. 60 (1938) 745–760.

[11] C. Schensted, Longest increasing and decreasing subsequences, Canad. J. Math. 13 (1961) 179–191.

[12] R.P. Stanley, Enumerative Combinatorics, vol. 2, Cambridge University Press, 1999.