A NOTE ON SOME CONJECTURES ABOUT COMBINATORIAL MODELS FOR RNA SECONDARY STRUCTURES

ADAM ZSOLT WAGNER

Abstract. We resolve two conjectures of Black-Drellich-Tymoczko about the numbers of valid plane trees for given primary sequences.

1. Introduction

Recall some definitions from [1, 4]: a set \( A \) is a \textit{complementary alphabet of size} \( m \) if it consists of \( m \) pairs of complementary letters \( \{A, \bar{A}, B, \bar{B}, \ldots\} \). We refer to \( \bar{A} \) as the \textit{complement} of \( A \), and vice versa \( A \) is the complement of \( \bar{A} \). An example of such an alphabet is the set of the four nucleotides \( A, C, G, U \) with the complementary pairs being the Watson–Crick complements \( A - U \) and \( C - G \).

Let \( T \) be a rooted plane tree with \( n \) edges and let \( P = p_1 p_2 \ldots p_{2n} \) be a word in a complementary alphabet \( A \). Label the edges of \( T \) by \( P \) according to a counter-clockwise walk around the boundary of \( T \), starting at the root, so that every edge receives two letters (one letter going ‘down’ and one letter going ‘up’). We say \( T \) is \( P \)-valid if every edge in \( T \) receives complementary letters by this method. As an example, consider the word \( P = AA\bar{A}BB\bar{A}BB \).

The tree on the left in \( P \)-valid, as every edge received complementary pairs of letters. The tree on the right is not \( P \)-valid, as three out of its four edges do not receive complementary letters.

Let \( \mathcal{P}(n, m) \) be the set of words \( P \) of length \( 2n \) in a complementary alphabet \( A \) of size \( m \), with at least one \( P \)-valid plane tree. Let \( V(P) \) be the set of \( P \)-valid plane...
trees for a fixed word $P$. Let $N(n, m, k)$ be the set of words $P \in \mathcal{P}(n, m)$ such that $|V(P)| = k$. Finally, let $R(n, m)$ be the set of $k \in \mathbb{N}^+$ such that $N(n, m, k)$ is non-empty.

The following two conjectures were stated in [1]:

**Conjecture 1.1.** $R(n, m) = R(n, 1)$ for all $m$. That is, the set $R(n, m)$ depends only on $m$.

We will show that Conjecture 1.1 is false, by finding a counterexample for $n = 7$.

**Conjecture 1.2.** For all $k \in \mathbb{N}^+$ there exists $n, m$ such that $k \in R(n, m)$.

We will show that slightly more is true. Our main result is the following:

**Theorem 1.3.** For all $k \in \mathbb{N}^+$ there exists $n$ such that $k \in R(n, 1)$.

Finally, we give a short application of a classical combinatorial puzzle about gas stations on a circular race track, to show that $R(n, m) \subseteq R(n + 1, m)$ for all $n, m$.

2. **Proof of main results**

We will find it easier to work with noncrossing matchings instead of rooted plane trees. Everything that follows can be translated into the language of rooted plane trees if necessary. Rooted plane trees and noncrossing perfect matchings are two of many counting problems in combinatorics whose solution is given by Catalan numbers. Given $2n$ labelled points on a circle, a noncrossing perfect matching is a perfect matching of the points, such that when drawing the edges of the matching as straight line segments inside the circle no two edges cross each other.

The standard bijection between rooted labelled plane trees and noncrossing perfect matching goes as follows. First, write the numbers $1, 2, \ldots, 2n$ on equidistant points on the unit circle $x^2 + y^2 = 1$, starting with the number 1 at position $(0, 1)$ and traversing the circle counterclockwise. Given a rooted plane tree $T$ with $n$ edges, write the numbers $1, 2, \ldots, 2n$ on the edges according to a counter-clockwise walk around the boundary of $T$, starting at the root vertex, each edge receiving two numbers (of different parity). The pairs of integers corresponding to the same edge of $T$ will be the edges of our matching $M$.

Recall that given a word $P \in \mathcal{P}(n, m)$, a $P$-valid plane tree $T$ is such that, when wrapping $P$ around $T$, each edge receives complementary letters. Hence, when translated into the language of matchings, a $P$-valid noncrossing perfect matching is such that, if we write the letters of $P$ in order along a circle, each edge in the matching receives complementary letters as its endpoints.
First we present our counterexample to Conjecture 1.1. Let $A := \{A, \bar{A}, B, \bar{B}\}$, where the pairs $(A, \bar{A})$ and $(B, \bar{B})$ represent the complementary letters. We claim that the word $P = \bar{B} \bar{A} A A A \bar{A} A B \bar{B} A A A A A$ has $|V(P)| = 11$. Indeed, there are two ways to match up the $B$s with the $\bar{B}$s. If we match the adjacent $B\bar{B}$ pairs we get 7 possible matchings, if we match the further away $B\bar{B}$ pairs we get $2 \cdot 2 = 4$ possible matchings. But according to Figure 9 in [1], we have $11 \notin R(7, 1)$.

Now we turn our attention to proving Theorem 1.3. Define the word $P_{k,\ell}$ for each $k, \ell \in \mathbb{N}^+$ as follows.

$$P_{k,\ell} = \underbrace{A A A \ldots A}_{k \text{ times}} \underbrace{A A A \ldots A}_{\ell \text{ times}} = p_1 p_2 \ldots p_{2k+2\ell}.$$ 

Note that $|V(P_{1,1})| = 2$.

**Lemma 2.1.** For $k, l \geq 1$ we have

$$|V(P_{k,l})| = \begin{cases} |V(P_{k-1,l})| & \text{if } k > l \\ |V(P_{k,l-1})| & \text{if } l > k \\ 1 + |V(P_{k,l-1})| & \text{if } k = l \end{cases}$$

**Proof.** Suppose first that $k > l$. Suppose we match $p_k$ with $p_i$. What is the value of $i$? We cannot have $i < k$, as then $p_i$ and $p_k$ are not complementary letters. But we also cannot have $i > k + 1$, because we need to have the same number of $A$s as $\bar{A}$s among $\{p_{k+1} \ldots p_i - 1\}$ to get a noncrossing matching. So $i = k + 1$. Hence $(p_k p_{k+1})$ is an edge of the matching and we may remove these two adjacent vertices to conclude the claim. The case $k < l$ is analogous to the above.
Suppose now \( k = l \). By the same logic as above, we either have \((p_k p_{k+1})\) as an edge (leaving us with \(|V(P_{k,k-1})|\) possible matchings) or we have \((p_k p_{3k+1})\) as an edge, whence both sides of our edge have a unique noncrossing matching, as required. \(\Box\)

**Proof of Theorem 1.3.** Let \( k \in \mathbb{N}^+ \). Note that \(|V(P_{k,k})| = 1 + |V(P_{k,k-1})| = 1 + |V(P_{k-1,k-1})|\), hence by induction we have \(|V(P_{k,k})| = k + 1\). So \( k + 1 \in R(2k, 1) \) for all \( k \). \(\Box\)

We finish with an application of the following classical puzzle.

**Theorem 2.2 (3, 5, 6).** Along a speed track there are some gas stations. The total amount of gasoline available in them is equal to what our car (which has a very large tank) needs for going around the track. Then there is a gas station such that if we start there with empty tank, we shall be able to go around the track without running out of gasoline.

**Proposition 2.3.** For every \( n, m \in \mathbb{N}^+ \) we have \( R(n, m) \subseteq R(n + 1, m) \).

**Proof.** Pick some \( k \in R(n, m) \) and let \( P \in \mathcal{P}(n, m) \) be such that \(|V(P)| = k\). Write \( P = p_1 p_2 \ldots p_{2n} \). Pick any letter - say \( A \), with complement \( \bar{A} \) - from the complementary alphabet. We will construct \( P' \in \mathcal{P}(n + 1, m) \) from \( P \) by inserting the string \( AA \) somewhere in \( P \), so that \(|V(P')| = k\). To achieve this, we want to make sure our two new letters are adjacent and have to get matched with each other in every noncrossing
matching. We note that in any noncrossing matching, if \((p_ip_j)\) is an edge with \(i < j\), then the number of \(A\)s and number of \(\bar{A}\)s among \(\{p_{i+1} \ldots p_{j-1}\}\) is equal.

Denote the number of \(A\)s (and hence also the number of \(\bar{A}\)s) in \(P\) by \(a\). Consider a circular speed track which is \(a\) miles long, and put \(a\) gas stations around it in clockwise order, so that the distance between the \(i\)-th and \(i+1\)-th gas station (in miles) is equal to the number of \(\bar{A}\)s between the \(i\)-th and \(i+1\)-th \(A\) in \(P\). Note that the distance between the \(a\)-th and first gas station (in miles) is equal to the number of \(\bar{A}\)s occurring after the last \(A\) or before the first \(A\) of \(P\). Assume each gas station has one liter of gasoline, and our car uses one liter per mile. According to Theorem 2.2 there exists a gas station where we can start with our car with an empty tank and go around the track clockwise without ever running out of gas.

Say this gas station corresponded to \(p_i\). Set \(P' := p_1p_2 \ldots p_{i-1}\bar{A}Ap_ip_{i+1} \ldots p_{2n}\). By the choice of \(i\), in any perfect noncrossing matching our two new added letters have to be connected to each other, hence \(|V(P')| = |V(P)| = k\). \(\square\)

Note: This manuscript was originally written, but not published, in 2015. The paper [1] it is referencing has since been updated and published, see [2].

References
[1] F. Black, E. Drellich and J. Tymoczko. “Valid plane trees: Combinatorial models for RNA secondary structures with Watson-Crick base pairs”, arXiv preprint, https://arxiv.org/abs/1501.03238v1 (2015).
[2] F. Black, E. Drellich and J. Tymoczko, “Valid Plane Trees: Combinatorial Models for RNA Secondary Structures with Watson–Crick Base Pairs”, SIAM Journal on Discrete Mathematics 31.4 (2017): 2586–2602.
[3] B. Bollobás, The Art of Mathematics: Coffee Time in Memphis, Cambridge University Press, 2006, p. 56–57.
[4] C.E. Heitsch, “Combinatorics on plane trees, motivated by RNA secondary structure configurations”, preprint.
[5] L. Lovász, Combinatorial problems and exercises, Akadémiai Kiadó, Budapest (1979) and North-Holland, Amsterdam, p. 31
[6] P. Winkler, Mathematical Puzzles: A Connoisseur’s Collection, A K Peters, 2004, p. 5