Supplementary Material for Lifting Prediction to Alignment of RNA Pseudoknots

Mathias Möhl[*], Sebastian Will[*], and Rolf Backofen[†]

Proof (of Lem. 1) Each instance of a fragment of degree k is uniquely determined by its $2k$ boundaries. Each boundary has one of the values $1 \dots m$. If the other boundaries are fixed, each constrained boundary can take at most two different values. Hence $\#_P^m(T) \in O(m^{2k_p - c'} 2^{c'})$, which equals $O(m^{2k_p - c'})$ since c' is considered as constant.

Each split is determined by the $2(k_p + k_1 + k_2)$ boundaries of the parent and the two children. Every two of them depend on each other: each parent boundary must coincide with some child boundary and from the remaining boundaries of the children, always two are directly adjacent. Hence, $k_p + k_1 + k_2$ values can be chosen to determine each instance. Due to the same argument concerning the constraints as before, it holds $\#_C^m(T) \in O(m^{k_p + k_1 + k_2 - c} 2^c) = O(m^{k_p + k_1 + k_2 - c})$.

Proof (of Lem. 2)

$$C(F_a, F_b) = \min_{A.\,\mathrm{align}_A(F_a,F_b)} \{C_A(F_a, F_b)\} \tag{1}$$

$$= \min_{A.\exists\ T\text{-split } (F_b^1,F_b^2) \text{ of } F_b.\,\mathrm{align}_A(F_a^1,F_b^1)\wedge\mathrm{align}_A(F_a^2,F_b^2)} \{C_A(F_a, F_b)\} \tag{2}$$

$$= \min_{A.\exists\ T\text{-split } (F_b^1,F_b^2) \text{ of } F_b.\,\mathrm{align}_A(F_a^1,F_b^1)\wedge\mathrm{align}_A(F_a^2,F_b^2)} \{C_A(F_a^1, F_b^1) + C_A(F_a^2, F_b^2)\} \tag{3}$$

$$= \min_{T\text{-split } (F_b^1,F_b^2) \text{ of } F_b} \left\{ \min_{A.\,\mathrm{align}_A(F_a^1,F_b^1)\wedge\mathrm{align}_A(F_a^2,F_b^2)} \{C_A(F_a^1, F_b^1) + C_A(F_a^2, F_b^2)\} \right\} \tag{4}$$

$$= \min_{T\text{-split } (F_b^1,F_b^2) \text{ of } F_b} \left\{ \min_{A.\,\mathrm{align}_A(F_a^1,F_b^1)} \{C_A(F_a^1, F_b^1)\} + \min_{A.\,\mathrm{align}_A(F_a^2,F_b^2)} \{C_A(F_a^2, F_b^2)\} \right\} \tag{5}$$

$$= \min_{T\text{-split } (F_b^1,F_b^2) \text{ of } F_b} \{C(F_a^1, F_b^1) + C(F_a^2, F_b^2)\} \tag{6}$$

Line (2) = (3) holds since for aligned fragments F_a, F_b and a split (F_a^1, F_a^2) of F_a, always a split (F_b^1, F_b^2) of F_b can be constructed such that F_a^1 is aligned to F_b^1 and F_a^2 is aligned to F_b^2. Line (3) = (4) relies on the fact that the split F_a^1, F_a^2 is arc-preserving, which allows to split up the computation of $C(F_a, F_b)$ into two independent parts. Line (6) is equivalent to line (5) because optimal alignments for these two parts always correspond to an optimal alignment of the fragments covering them both.

Proof (of Lem. 3) W.l.o.g. assume that the split type T contains a constraint on the first fragment. Let A be the optimal alignment of F_a and F_b and let T_b be the basetype obtained by removing all constraints from T. Assume that there is no T-split (F_b^1, F_b^2) of F_b such that A

[*]These authors contributed equally.
[†]Corresponding author

aligns F_a^1 to F_b^1 and F_a^2 to F_b^2. Lem. 2 implies that there exists at least one T_b split $(F_b^{1'}, F_b^{2'})$ with this property. Since $(F_b^{1'}, F_b^{2'})$ is not of type T, there must exist some interval $F_a^1[i] = \langle x \rangle$ such that (a) the interval has a size constraint in T (b) x is a boundary of the parent F_a (c) A aligns x to some y that is not the boundary y' of F_b corresponding to the boundary x of F_a. The last property ensures that the intereval $F_b^{1'}[i]$ has size greater than one (since it must contain both y and y') and hence the size constraint of T is not satisfiable.

Since A aligns the interval $\langle x \rangle$ to the interval containing y and y' and x is aligned to y, the boundary y' is aligned to a gap. Hence, there is a T-split that separates exactly y' from F_b, A aligns F_a with the remaining fragment of F_b.

Proof (of Lem. 4) We proof the two claims by simultaneous induction over the parse tree. That is, when we proof one of the claims for a single node, we can assume that both claims hold for the children nodes by induction hypothesis.

Clearly, the claims hold for atomic F_a (base case of induction). Otherwise F_a is split by a split of A&U split type into fragments F_a^1 and F_a^2 with n_1 and n_2 descendants, respectively.

For the case that F_a has degree 1, only split type 121 needs special attention. In this case, for computing all $C\big(F_a, F_b\big)$ we group the fragments F_b by their leftmost boundary. For all fragments in one group we need only $O(n'm^3)$ space, because we only recurse to costs $C\big(F_a^1, F_b^1\big)$ with a fix leftmost boundary of F_b^1; by claim 2, all these costs are computed in $O(n_1m^3)$ space, which is reused for each group. For the second fragment, we only need $O(n_2m^3)$ by claim 1. Finally, we store the costs in $O(m^2)$. Clearly, $O(n_1m^3 + n_2m^3 + m^2) = O(n'm^3)$.

For the case that F_a has degree 2, we only compute costs for F_b with fix leftmost boundary. For this aim, we only need costs $C\big(F_a^1, F_b^1\big)$ for F_b^1 with the same fix boundary, which are computed in $O(n_1m^3)$ due to claim 1.