MthSc 985: Topics in Discrete Mathematical Biology
Midterm
October 31, 2011

NAME: Key

Instructions

- Exam time is 50 minutes
- Open notes / book / everything.
- **Show your work.** Partial credit will be given.

Question	Points Earned	Maximum Points
1		20
2		20
3		20
4		20
5		20
Total		**100**

1. (a) In your own words, give a concise combinatorial definition of an *RNA secondary structure* and an *RNA pseudoknot structure* of length n.

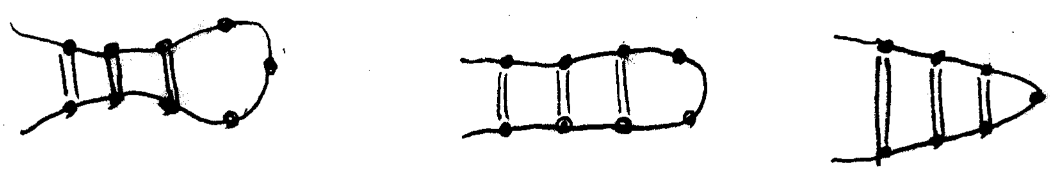
An <u>RNA secondary structure</u> is a sequence of vertices (nucleotides) $1, \ldots, n$ and arcs (i,j) such that:

1. $|i<j| \geq 2$ for each arc (i,j).
2. "Non-crossing condition:" If (i_1, j_1) and (i_2, j_2) are arcs with $i_1 < i_2$, then $i_1 < i_2 < j_2 < j_1$.

If we drop condition 2, we get an <u>RNA pseudoknot structure</u>.

(b) For both secondary and pseudoknot structures, there are additional restrictions on e.g., arc length and stack size, that lead to the notion of a *canonical* structure. Give these additional requirements, as well as biophysical justifications for why we impose them.

- <u>Minimum arc length $\lambda = 4$</u>: Small hairpin loops require energy, because the RNA strand must be folded sharply. Such a configuration is not very stable:



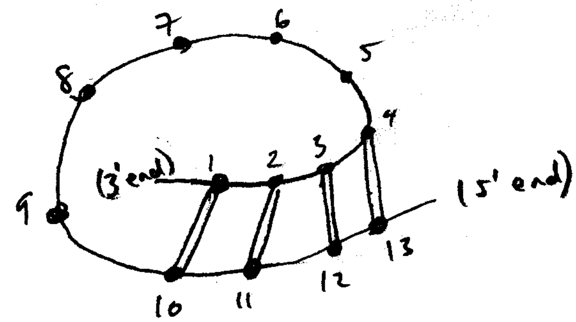
stable: $\lambda = 4$ less stable: $\lambda = 3$ very unstable: $\lambda = 2$

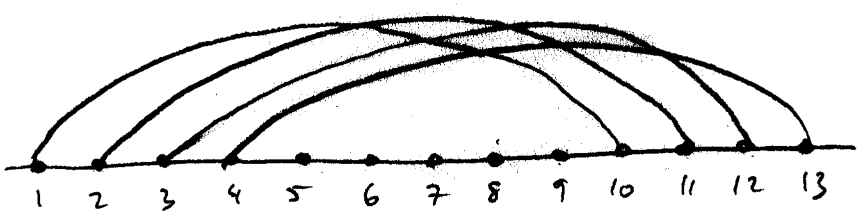
- <u>Minimum stack size $\sigma = 2$</u>: Equivalently, there are no isolated basepairs, because these are more likely to break apart than if there are several in parallel

fragile stable

2. (a) Draw an RNA strand that exhibits a pseudoknot structure, as a 5-noncrossing (but *not* 4-noncrossing) arc diagram. Draw the diagram and an actual RNA strand that it represents.

One solution:





(b) Recall the canonical bijection ψ from $*$-tableaux to arc diagrams. Compute $\psi^{-1}(G_n)$, where G_n is your diagram from Part (a).

The arcs of G_n are: $(1,10), (2,11), (3,12), (4,13)$.

RSK insert from right to left: For each arc (i,j), insert $+i$ at step j.

$$\underset{T_{13}}{\emptyset} \xrightarrow{+4} \underset{T_{12}}{\begin{array}{|c|}\hline 4\\\hline\end{array}} \xrightarrow{+3} \underset{T_{11}}{\begin{array}{|c|}\hline 3\\\hline 4\\\hline\end{array}} \xrightarrow{+2} \underset{T_{10}}{\begin{array}{|c|}\hline 2\\\hline 3\\\hline 4\\\hline\end{array}} \xrightarrow{+1} \underset{T_{9}}{\begin{array}{|c|}\hline 1\\\hline 2\\\hline 3\\\hline 4\\\hline\end{array}} \xrightarrow{\emptyset} \cdots \xrightarrow{\emptyset} \underset{T_{4}}{\begin{array}{|c|}\hline 1\\\hline 2\\\hline 3\\\hline 4\\\hline\end{array}}$$

$$\xrightarrow{-4} \underset{T_{3}}{\begin{array}{|c|}\hline 1\\\hline 2\\\hline 3\\\hline\end{array}} \xrightarrow{-3} \underset{T_{2}}{\begin{array}{|c|}\hline 1\\\hline 2\\\hline\end{array}} \xrightarrow{-2} \underset{T_{1}}{\begin{array}{|c|}\hline 1\\\hline\end{array}} \xrightarrow{-1} \underset{T_{0}}{\emptyset}$$

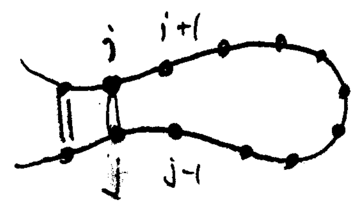
The corresponding sequence of shapes $\psi^{-1}(G_n) = (\mu_i)_{i=0}^{14}$ is

$$\left(\emptyset, \begin{array}{|c|}\hline \;\\\hline\end{array}, \begin{array}{|c|}\hline \;\\\hline \;\\\hline\end{array}, \begin{array}{|c|}\hline \;\\\hline \;\\\hline \;\\\hline\end{array}, \begin{array}{|c|}\hline \;\\\hline \;\\\hline \;\\\hline \;\\\hline\end{array}, \begin{array}{|c|}\hline \;\\\hline \;\\\hline \;\\\hline \;\\\hline\end{array}, \begin{array}{|c|}\hline \;\\\hline \;\\\hline \;\\\hline \;\\\hline\end{array}, \begin{array}{|c|}\hline \;\\\hline \;\\\hline \;\\\hline \;\\\hline\end{array}, \begin{array}{|c|}\hline \;\\\hline \;\\\hline \;\\\hline \;\\\hline\end{array}, \begin{array}{|c|}\hline \;\\\hline \;\\\hline \;\\\hline \;\\\hline\end{array}, \begin{array}{|c|}\hline \;\\\hline \;\\\hline \;\\\hline\end{array}, \begin{array}{|c|}\hline \;\\\hline \;\\\hline\end{array}, \begin{array}{|c|}\hline \;\\\hline\end{array}, \emptyset\right)$$

3. Let $\delta \in \mathcal{T}_{k,\sigma}(n)$ be a k-noncrossing, σ-canonical RNA structure.

   (a) Define a *hairpin loop* of δ purely combinatorially. Assume that the endpoints of the loop are i and j, with $i < j$.

A <u>hairpin loop</u> is an arc (i,j) with a sequence $[i+1, j-1]$ of isolated vertices.



   (b) Suppose that δ has shape $\gamma \in \mathcal{I}_k(s,m)$ (that is, s arcs, of which m are 1-arcs). What can you say about the number of hairpin loops that δ has? Justify your answer.

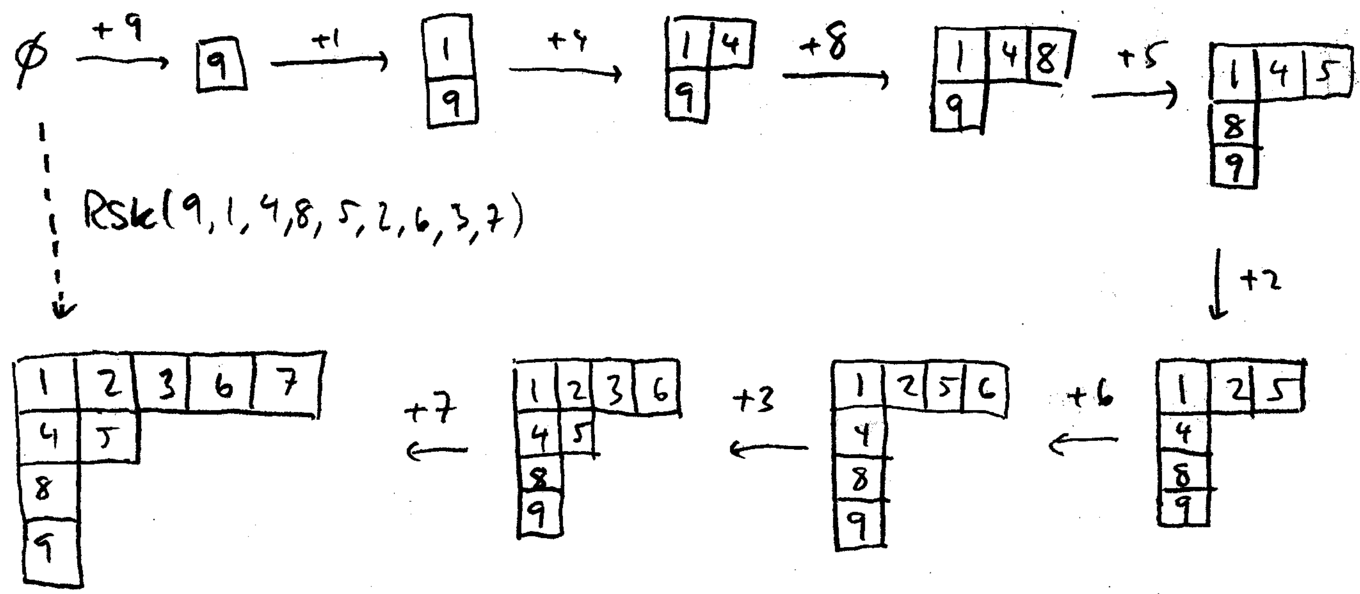
δ has exactly m hairpin loops. Consider the 2-step process of inflating δ to its shape γ:

Step I: Inflate arcs to stems

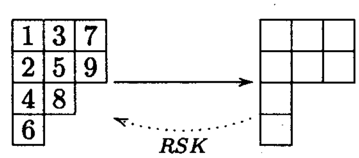
Step II: Add isolated vertices.

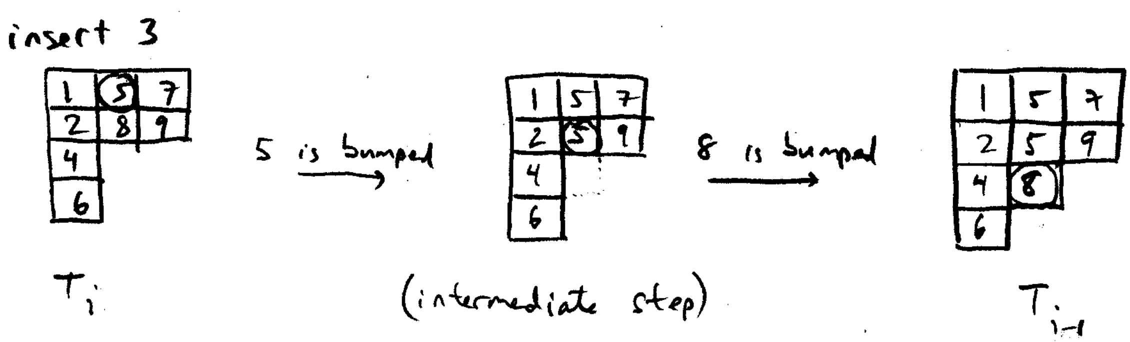
It is clear that each 1-arc gets mapped to a hairpin loop (in Step II), and moreover, that this is the only way a hairpin loop can be created.

4. (a) Starting with the empty Young tableau, apply the RSK-algorithm to the following sequence: $9, 1, 4, 8, 5, 2, 6, 3, 7$. Draw the resulting Young tableau.



(b) Find the unique Young tableau T_i (at right) with the given shape such that inserting a number into T_i via RSK gives T_{i-1} (at left).





5. In class, we constructed the combinatorial class $\mathcal{T}_\gamma$ of all k-noncrossing, σ-canonical RNA structures (that is, minimum arc length $\lambda = 2$) that have shape $\gamma \in \mathcal{I}_k(s, m)$:

$$\mathcal{T}_\gamma = [\mathcal{K}_\sigma \times \mathrm{Seq}(\mathcal{N}_\sigma)] \times [\mathcal{L}^{2s+1-m} \times (\mathcal{Z} \times \mathcal{L})^m].$$

Here, $\mathcal{Z}$ is the class of vertices, $\mathcal{L}$ is the class of vertex sequences, $\mathcal{K}_\sigma$ is the class of stacks, and $\mathcal{N}_\sigma$ the class of induced stacks.

Modify this construction to create the combinatorial class $\mathcal{F}_\gamma$ of all k-noncrossing, σ-canonical diagrams that have shape $\gamma \in \mathcal{I}_k(s, m)$. (That is, drop the minimum arc length $\lambda = 2$ requirement.)

Recall the 2-step process of inflating shapes to structures:

Step 1: Inflate arcs to stems: $\mathcal{K}_\sigma \times \mathrm{Seq}(\mathcal{N}_\sigma)$

Step 2: Add isolated vertices: $\mathcal{L}^{2s+1-m} \times (\mathcal{Z} \times \mathcal{L})^m$.

add a (possibly empty) sequence of isolated vertices in the $2s+1-m$ positions that are <u>not</u> 1-arcs.

add a <u>nonempty</u> sequence of isolated vertices under the m 1-arcs

Dropping the $\lambda = 2$ requirement amounts to allowing the isolated vertex sequences under 1-arcs to be empty.

Thus, instead of $(\mathcal{Z} \times \mathcal{L})^m$, we have $\mathcal{L}^m$.

The resulting combinatorial class is therefore

$$\mathcal{T} = [\mathcal{K}_\sigma \times \mathrm{Seq}(\mathcal{N}_\sigma)] \times \mathcal{L}^{2s+1}.$$