WEEK TWO NOTES

DANIELLE DEES

Problem. *You are given $2n$ words, where a word is defined as a sequence of DNA proteins of a specified length. These words will be labeled as $w_1, \overline{w_1}, w_2, \overline{w_2}, \ldots w_n, \overline{w_n}$. The concatenation of these words will occur in the following manner. The resultant strand will the $\sum_{i=1}^{n} w_i or \overline{w_i}$. The goal is to find an algorithm that will take in these $2n$ words and output a "yes" if one of the 2^n strands that result from the concatenation of these words will fold into a secondary structure or a "no" if each possible strand will not form a secondary structure.*

1. Background

There is an algorithm which will take in a single strand and return a negative number if this strand folds into a secondary structure and a zero if this strand does not fold. This algorithm is characterized by the following equations.
These are interdependant recurance relations. This is the first formula which will return the overall answer by calculating the energy of the system. Since negative energies are more stable, we will most always be taking the minimum of all the possiblities for the strand.

1.1. $W(j)$.

$$W(0) = 0 \tag{1}$$

$$W(j) = \min(W(j-1), \min_{1 \le i < j}(V(i,j) + W(i-1))), \text{for } i > 0 \tag{2}$$

The two possiblities in this equation are:

- The $W(j-1)$ part of this equation represents the option where the base s_j is not paired with anything, and thus s_j has an energy of 0 and the energy of the strand is equal to the energy of string from $s_{j-1} \text{to} s_1$ which is equal to $W(j-1)$.
- The seond possibility is that the base s_j does pair with another base in the strand. Therefore, the total energy would be the energy of structure which is closed by the pair s_j, s_i, assuming that $i < j$, which is expressed as $V(i,j)$. This energy would be added with the energy of the strand which goes form s_i to s_1, which is expressed as $W(i-1)$.

1.2. $V(i,j)$.

The second equation characterizes the equation for all non-linear strands.

$$V(i,j) = \begin{cases} +\infty \text{for} i \ge j \\ \min(eH(i,j), eS(i,j) + V(i+1, j-1), VBI(i,j), VM(i,j)), \text{for} i < j \end{cases} \tag{3}$$

The possibilities for this equation are:

- $eH(i,j)$ is free energy when i,j is the exterior pair which closes a hairpin loop.
- $eS(i,j)$ is the energy of a stacked pair, plus the enery of the structure which is enclosed by the second pair of the stack. This energy of this inner structure is given by $V(i+1, j-1)$.
- $VBI(i,j)$ is the energy of a bulge or internal loop, where i,j closes the bulge or loop.
- $VM(i,j)$ is the energy when i,j closes a multiloop.

1.3. $VBI(i,j)$. This equation will determine the energy for a bulge or an interior loop which is closed by the bases s_i and s_j.

$$VBI(i,j) = \begin{cases} +\infty \text{for} j < (i+4) \\ \min\limits_{\substack{i',j' \\ i<i'<j'<j}} (eL(i,j,i',j') + V(i',j')) \end{cases} \tag{4}$$

This equation is infinity if the value of j is less than $(i+4)$ because a bulge or internal loop would not be possible between i and j. There would only be three bases, so whereas the pair i,j could close a hairpin loop, it could not close a bulge or an interior loop. The other formula is adding the free enery of the bulge with the free enery of the structure between i' and j', which will be closed by the stacked pair of i' and j'.

1.4. $VM(i,j)$. This is the formula for the multibranched computation. The accual formula here has a rediculious running time. Here is the "infeasible slow" formula:

$$VM(i,j) = \min_{\substack{k,i_1,j_1,i_2,j_2,\ldots,i_k,j_k \\ i<i_1<j_1<i_2<j_2<\ldots<i_k<j_k<j \\ k\geq 2}} (eM(i,j,i_1,j_1,i_2,j_2,\ldots,i_k,j_k) + \sum_{h=1}^{k} V(i_h,j_h)) \tag{5}$$

However, if we use the approximation

$$eM(i,j,i_1,j_1,\ldots,i_k,j_k) = a + bk + c((i_1 - i - 1) + j - j_k - 1) + \sum_{h=0}^{k-1}(i_(h+1) - j_h - 1)), \tag{6}$$

where a, b, and c are constants, then combining these two formulas our runtime becomes $O(n^3)$. This new formula is

$$VM(i,j) = \min_{i+1<h\leq j-1}(WM(i+1,h-1) + WM(h,j-1) + a) \tag{7}$$

However, this formula requires the formula WM(i, j), which is defined as

$$WM(i,i) = c \tag{8}$$

$$WM(i,j) = \min(V(i,j)+b, \min_{i<h\leq j}(WM(i,h-1) + WM(h,j))), for i < j \tag{9}$$

This formula corresponds to the two options of:

- The pair s_i, s_j forms a pair and closes one of the k branches.
- The other option is that s_i, s_j does not form a pair. In this case, we again partition the loop into at least two pieces, and repeat the recurance.

These formulae compose the algorithm for a single stand. We are planning on using a dynamic programming technique to modify these formulae so that they work for all 2^n strands in polynomial time.

2. Beginnings of a New Algorithm

2.1. $W'_s(j)$.

$$W'_s(j) = \min_{s\in S} W_s(j) \tag{10}$$

$$W'_s(j) = \min\{w'_s(j-1), \min_{\substack{1\leq i\leq j-1 \\ b\in\{T,B\} \\ i \bmod (l+1)\neq 0}} (v'_s(b,i,j) + W'_s(b,i-1)), \min_{\substack{1\leq i\leq j-1 \\ i \bmod (i+1)=0}} (V'_s(i,j) + W'_s(i-1))\} \tag{11}$$

$$W'_s(b,j) = \min_{s\in S_{j,b}} W_s(j) \tag{12}$$

$$W'_s(b,j) = \min\{W'_s(b,j-1), \min_{1\leq i\leq j-1}(V'_s(b,i,j) + W'_s(b,i-1))\} \tag{13}$$

(14)

2.2. V'_s.

$$V'_s(i,j) = \min_{s \in S} V_s(i,j) \tag{15}$$

$$V's(i,j) = \min\{\min_{s \in S} eH_s(i,j), \tag{16}$$

$$\min_{\substack{b_i in \{B,T\} \\ b_j in \{B,T\} \\ wordnum(i)=wordnum(i+1) \\ wordnum(j)=wordnum(j-1)}} eS'(b_i,b_j,i,j) + V'_s(b_i,b_j,i+1,j-1), \tag{17}$$

$$\min_{\substack{b_i in \{B,T\} \\ c and \overline{c} in \{B,T\} \\ wordnum(i) \neq wordnum(i+1) \\ wordnum(j)=wordnum(j-1)}} eS'(b_i,c,i,j) + V'_s(b_i,\overline{c},i+1,j-1), \tag{18}$$

$$\min_{\substack{b_j in \{B,T\} \\ c and \overline{c} in \{B,T\} \\ wordnum(i)=wordnum(i+1) \\ wordnum(j) \neq wordnum(j-1)}} eS'(c,b_j,i,j) + V'_s(\overline{c},b_j,i+1,j-1), \tag{19}$$

$$\min_{\substack{wordnum(i) \neq wordnum(i+1) \\ wordnum(j) \neq wordnum(j-1)}} eS'(i,j) + V'_s(i+1,j-1), VBI, VM\} \tag{20}$$

$$V_s(i,j) = \begin{cases} +\infty \text{for} i \geq j \\ \min(eH(i,j), \min_{\substack{b \in \{B,T\} \\ i \bmod (l+1) \neq 0}} eS(b,i,j) + V(b,i+1,j-1), \\ \min_{i \bmod (l+1)=0} eS(i,j) + V(i+1,j-1), VBI(i,j), VM(i, \end{cases} \tag{21}$$

$$V'_s(b,i,j) = \min_{s \in S_{(b,j)}} V_s(i,j) \tag{22}$$

$$V'_s(b,i,j) = \begin{cases} +\infty \text{for} i \geq j \\ \min(eH(i,j), eS(b,i,j) + V(b,i+1,j-1), VBI(i,j), V \end{cases} \tag{23}$$