Lecture 18

Speeding Up Internal Loop Computations

March 7, 2000
Notes: Kellie Plow

Recall from Section 17.4 that the running time for determining the internal loop free energy calculation is $O(n^4)$: each of the $O(n^2)$ exterior pairs $i \cdot j$ requires a search through the $O(n^2)$ interior pairs $i' \cdot j'$ for one that minimizes the resulting free energy. We will address this $O(n^4)$ running time computation of the free energy of bulges and internal loops, and show that it can be decreased to $O(n^3)$. This is the main result of Lyngsø *et al.* [1] and, combined with the remaining analysis in Section 17.4, shows that the entire RNA secondary structure prediction problem can be solved in time $O(n^3)$. $O(n^3)$ running time is not practical for long RNA sequences, but it does allow for secondary structure prediction for RNA sequences that are hundreds of bases in length, which would be prohibitive with an n^4 time algorithm.

18.1. Assumptions About Internal Loop Free Energy

To speed up the running time it is necessary to make some assumptions about the form of the internal loop free energy function $\mathrm{eL}(i, j, i', j')$ (see Definition 16.5). The authors cite thermodynamics studies that support the fact that these assumptions are realistic.

The authors first assume that eL is the sum of 3 contributions:

1. a term "$\mathrm{size}(i' - i + j - j' - 2)$" that is a function of the size of the loop, plus
2. stacking energies "$\mathrm{stacking}(i, j) + \mathrm{stacking}(i', j')$" for the unpaired bases adjacent on the loop to the two base pairs, plus
3. an asymmetry penalty "$\mathrm{asymmetry}(i' - i - 1, j - j' - 1)$", where $i' - i - 1$ is the number of unpaired bases between the two base pairs on one side of the loop, and $j - j' - 1$ the number on the other side.

The free energy function for bulges and internal loops is thus given by

$$\mathrm{eL}(i, j, i', j') = \mathrm{size}(i' - i + j - j' - 2) + \mathrm{stacking}(i, j) + \mathrm{stacking}(i', j') + \mathrm{asymmetry}(i' - i - 1, j - j' - 1). \tag{18.1}$$

18.2. Asymmetry Penalty

The currently used asymmetry functions are of the form

$$\mathrm{asymmetry}(n_1, n_2) = \min(E_{\max}, |n_1 - n_2| f(m)), \tag{18.2}$$

where $E_{\max}$ is the maximum asymmetry penalty assessed, f is a function whose details need not concern us, $m = \min(n_1, n_2, c)$, and c is a small constant (equal to 5 and 1, respectively, in two cited thermodynamics studies).

What is important for our purposes is that this asymmetry penalty grows linearly in $|n_1 - n_2|$, provided that $n_1 \geq c$ and $n_2 \geq c$. In particular, the only assumption we will need to make about the penalty is that

$$\text{asymmetry}(n_1, n_2) = \text{asymmetry}(n_1 + 1, n_2 + 1) \tag{18.3}$$

for all $n_1 \geq c$ and $n_2 \geq c$. This is certainly true for the particular form given in Equation (18.2).

18.3. Comparing Interior Pairs

Recall from Section 17.1.3 the recurrence

$$VBI(i, j) = \min_{\substack{i', j' \\ i < i' < j' < j}} (V(i', j') + \text{eL}(i, j, i', j')).$$

We are going to save time by not searching through all the interior pairs $i' \cdot j'$. Suppose that, for exterior pair $i \cdot j$, the interior pair $i' \cdot j'$ is better than $i'' \cdot j''$, and that both of these loops have the same size. Then, under the assumptions from Sections 18.1 and 18.2, Theorem 18.1 below demonstrates that $i' \cdot j'$ is also better for exterior pair $(i - 1) \cdot (j + 1)$. The intuition behind this theorem is that the asymmetry penalty for $i' \cdot j'$ is the same for the two different exterior pairs by Equation (18.3), as is the asymmetry penalty for $i'' \cdot j''$, and neither interior pair gains an advantage in loop size or stacking energies when you change from the smaller to the bigger loop.

Theorem 18.1: Let

$$j' - i' = j'' - i'' \tag{18.4}$$

(so as to compare internal loops of identical size). Let $i' - i - 1$, $j - j' - 1$, $i'' - i - 1$, and $j - j'' - 1$ each be at least c (so that Equation (18.3) applies to both loops). Suppose that

$$V(i', j') + \text{eL}(i, j, i', j') \leq V(i'', j'') + \text{eL}(i, j, i'', j''). \tag{18.5}$$

Then

$$V(i', j') + \text{eL}(i - 1, j + 1, i', j') \leq V(i'', j'') + \text{eL}(i - 1, j + 1, i'', j'').$$

Proof:

$$
\begin{aligned}
& V(i',j') + \mathrm{eL}(i-1,j+1,i',j') \\
&= V(i',j') + \mathrm{size}(i'-i+j-j') + \mathrm{stacking}(i-1,j+1) + \mathrm{stacking}(i',j') \\
&\quad + \mathrm{asymmetry}(i'-i,j-j') && \text{Equation (18.1)} \\
&= V(i',j') + \mathrm{eL}(i,j,i',j') \\
&\quad - \mathrm{size}(i'-i+j-j'-2) + \mathrm{size}(i'-i+j-j') \\
&\quad - \mathrm{stacking}(i,j) + \mathrm{stacking}(i-1,j+1) \\
&\quad - \mathrm{asymmetry}(i'-i-1,j-j'-1) + \mathrm{asymmetry}(i'-i,j-j') && \text{Equation (18.1)} \\
&= V(i',j') + \mathrm{eL}(i,j,i',j') \\
&\quad - \mathrm{size}(i'-i+j-j'-2) + \mathrm{size}(i'-i+j-j') \\
&\quad - \mathrm{stacking}(i,j) + \mathrm{stacking}(i-1,j+1) && \text{Equation (18.3)} \\
&\leq V(i'',j'') + \mathrm{eL}(i,j,i'',j'') \\
&\quad - \mathrm{size}(i''-i+j-j''-2) + \mathrm{size}(i''-i+j-j'') \\
&\quad - \mathrm{stacking}(i,j) + \mathrm{stacking}(i-1,j+1) && \text{Equations (18.4) \& (18.5)} \\
&= V(i'',j'') + \mathrm{eL}(i,j,i'',j'') \\
&\quad - \mathrm{size}(i''-i+j-j''-2) + \mathrm{size}(i''-i+j-j'') \\
&\quad - \mathrm{stacking}(i,j) + \mathrm{stacking}(i-1,j+1) \\
&\quad - \mathrm{asymmetry}(i''-i-1,j-j''-1) + \mathrm{asymmetry}(i''-i,j-j'') && \text{Equation (18.3)} \\
&= V(i'',j'') + \mathrm{size}(i''-i+j-j'') + \mathrm{stacking}(i-1,j+1) + \mathrm{stacking}(i'',j'') \\
&\quad + \mathrm{asymmetry}(i''-i,j-j'') && \text{Equation (18.1)} \\
&= V(i'',j'') + \mathrm{eL}(i-1,j+1,i'',j'') && \text{Equation (18.1)}
\end{aligned}
$$

□

Instead of using a two-dimensional array $VBI(i,j)$, use a three-dimensional array $VBI(i,j,l)$, where l is the loop size. This array will be filled in using dynamic programming. The entry $VBI(i,j,l)$ will store not only the free energy, but also the best interior pair $i' \cdot j'$ (subject to $i'-i-1 \geq c$ and $j-j'-1 \geq c$) that gives this energy.

Now suppose that the entry $VBI(i,j,l)$ has been calculated, and we want to calculate the entry $VBI(i-1,j+1,l+2)$. By Theorem 18.1, the interior pair $i' \cdot j'$ stored in $VBI(i,j,l)$ is the best interior pair for $VBI(i-1,j+1,l+2)$, with only two possible exceptions. These possible exceptions are the loops with exterior pair $(i-1)\cdot(j+1)$, length $l+2$, and having one or the other loop side of length exactly c.

Thus, for each of $O(n^3)$ entries in VBI, it is necessary to compare 3 loop energies and store the minimum, which takes constant time. It is also necessary to compare those loops with exterior pair $(i-1)\cdot(j+1)$

and length $l + 2$ having one or the other loop side of length less than c, but there are only a constant number of these.

References

[1] R. B. Lyngsø, M. Zuker, and C. N. S. Pedersen. Internal loops in RNA secondary structure prediction. In *RECOMB99: Proceedings of the Third Annual International Conference on Computational Molecular Biology*, pages 260–267, Lyon, France, Apr. 1999.