Nussinov Algorithm

- Given an RNA sequence of length L.
  - Find the structure with the most base pairs
- There are four ways of getting the best structure between position i and j from the best structures of the smaller subsequences:
  1) Add unpaired position i onto best structure for subsequence i+1, j
  2) Add unpaired position j onto best structure for subsequence i, j-1
  3) Add i,j pair onto best structure found for subsequence i+1, j-1
  4) Combine two optimal structures i,k and k+1, j

Nussinov Algorithm: Four Cases

The numbers above each case corresponds to the enumeration on the previous slide.

Biological Sequence Analysis by Durbin et al., 2001, page 269
Nussinov Algorithm: The Table

- Compares a sequence against itself in a dynamic programming matrix.
- Since structure folds upon itself, it is only necessary to calculate half the matrix.
- The value of $M[i,j]$ is the number of base pairs in the maximally base-paired structure.
- Four rules for scoring the structure at a particular point.

Nussinov Algorithm: Initialization

- **Initialization**: fill the main diagonal and the diagonal just below it with zeros

- Formally, the scoring matrix, $M$, is initialized:
  - $M[i,i] = 0$ for $i = 1$ to $L$ (main diagonal)
  - $M[i,i-1] = 0$ for $i = 2$ to $L$ (diagonal below main diagonal)

Biological Sequence Analysis by Durbin et al., 2001, page 270
Nussinov Algorithm: Recursion

Matrix Fill: Starting with all subsequences of length 2, to length L

\[
\text{do}
\]

\[
M[i,j] = \max \text{ of the following : }
\]

\[
M[i+1,j] \quad (\text{base } i \text{ is hanging off by itself})
\]

\[
M[i,j-1] \quad (\text{base } j \text{ is hanging off by itself})
\]

\[
M[i+1,j-1] + S(x_i, x_j) \quad (\text{bases } i \text{ and } j \text{ are paired; if } x_i = \text{complement of } x_j \text{ then } S(x_i, x_j) = 1; \text{ otherwise it is } 0)
\]

\[
M[i,j] = \max_{i<k<j} (M[i,k] + M[k+1,j]) \quad (\text{merging two substructures})
\]

Biological Sequence Analysis by Durbin et al., 2001, page 270
Adapted from leah.haifa.ac.il/~zfrenkel/Sasha/CB/

Nussinov: Traceback Stage

Push (1,L) onto stack \ (Start with this coordinate: i=1, j=L)

Recursion: Repeat until stack is empty
- pop (i,j);
  if \(i > j\) continue;
else if \(M[i+1,j] = M[i,j]\) then push \((i+1,j)\);
else if \(M[i,j-1] = M[i,j]\) then push \((i,j-1)\);
else if \(M[i+1,j-1] = M[i,j] + d[i,j]\) then
  record base \(i,j\);
  push \((i+1,j-1)\);
else for \(k = i+1\) to \(j-1\):
  if \(M[i,k] + M[k+1,j] = M[i,j]\) then
    push \((k+1,j)\);
    push \((i,k)\);
    break;

Biological Sequence Analysis by Durbin et al., 2001, page 271