

Notes of Suffix-Prefix Matching.

In Xparal Suffix-Prefix Matching is called the end-gaps free case.

We want to use the dynamic programming approach to find the best suffix-prefix match between two strings  $S_1$  and  $S_2$ . One motivation for this problem is in shotgun sequence assembly as discussed in class, but there are other motivations as well.

We start with a simple version of the problem, to find the longest suffix of  $S_1$  and an equal length prefix of  $S_2$  which match exactly (without mismatches or spaces).

For example, if  $S_1$  is aabcdaac and  $S_2$  is cdaacgtd then the longest exact suffix-prefix match of  $S_1$  to  $S_2$  is cdaac. The longest suffix-prefix match of  $S_2$  to  $S_1$  has length zero.

To find an efficient method to solve such problems, we will use essentially the same recurrence relations we use for global alignment, but make a few modifications. Assume that  $S_1$  is arranged on the vertical axis of the DP table, and that  $S_2$  is arranged on the horizontal axis.

1) First, we set the base conditions to be

$V(i, 0) = 0$  and  $V(0, j) = -\infty$  for any  $i$  and  $j$ . What this says is that we can consume, or align with a space, any initial part (prefix) of  $S_1$  without paying any penalty, since we only want to use a suffix of  $S_1$ , but if we align a prefix of  $S_2$  with a space, then we have to pay infinity (or “get”  $-\infty$ ). That means we will not align a prefix of  $S_2$  with a space. In the DP table, this means that we fill the zero column with all 0's, and the 0 row with all infinite values (other than the 0,0 cell).

2) Second, we set the match value to be 1 (or any other positive value) and the mismatch and space values (after the zero column) to be infinite.

3) Third, we choose the optimal value not in cell  $(n, m)$ , but as the largest value in row  $n$ .

Problems:

1) Explain what this approach looks like in terms of the alignment graph, i.e. what are the edge weights.

2) Explain as best you can, why this works.

3) Work out by hand the DP table for the example strings given above.