CS628 DUE BY 10/15/04

Assignment #2

---

1. (i) Let T be the number of substitutions and I the number of indels (insertion or deletion). Show that $2T + I = m + n$.

   (ii) Let S be a scoring matrix on $\Sigma \times \Sigma$ and let $\gamma(k) = g + (k-1)e$ be an affine gap penalty function. For $\alpha, \beta \in \mathbb{R}^+$ define $S'(a,b) := \alpha S(a,b) + 2\beta$ and $\gamma'(k) := \alpha\gamma(k) + \beta$. Show that a global alignment is optimal with respect to (S, γ) if and only if it is optimal with respect to (S', γ').

   (iii) Is this true for local alignments as well? If it is, prove it, if not can you modify the statement so that it will be true?

2. Say a set of k subalignments has "degree 2 pair usage" if each aligned pair of letters is not used more than twice. The non-intesecting notion of Waterman-Eggert is "degree 1 pair usage". Design an efficient algorithm for finding the *scores* of k local alignments that are an optimal degree 2 pair usage set. Assume a linear gap penalty and $O(kmn)$ is good enough.

3. Prove that the complexity of the Myers-Miller linear space alignment is $O(nm)$.

4. Would the Huang-Miller algorithm work if First(v) would be defined is some arbitrary way? If not, what would break down?

5. Show in the context of the Galil-Giancarlo algorithm for concave gap penalty function that finding $g^*_{m'+1}$ takes $O(1)$ for an affine γ and deduce that this yields a $O(mn)$ algorithm for finding the score of the optimal global alignment for such γ.

6. Set up the right scoring system so that you can find the longest common subsequence between x and y using the Needleman-Wunsch global alignment algorithm.

7. An end-space free (global) alignment is one that does not penalize indels at the end and at the beginning of the alignment. Design an efficient algorithm for finding the *score* of an optimal such global alignment.