Algorithms in the Real World (15-853), Fall 10
Assignment #3 Due: October 12

---

Complete all problems.
You are not permitted to look at solutions of previous year assignments. You can work together in groups, but all solutions have to be written up individually. Please submit a hard copy during class.

**Problem 1: Solving a Recurrence (5pt)**
For the space-efficient Edit Distance problem (Lecture 3), we used the recurrence:

$$\begin{aligned} T(n,m) &= T(n/2,k) + T(n/2,m-k) + mn \\ T(1,m) &= m \\ T(n,1) &= n \end{aligned}$$

Give a formal proof that $T(m,n) = O(mn)$.

**Problem 2: Weaving (10pt)**
Given two strings $S_1$ and $S_2$ and a text $T$, you want to find whether there is an occurrence of $S_1$ and $S_2$ interwoven in $T$. An interweaving is an assignment of each character in $T$ to $S_1$, $S_2$, or neither string, such that the $k$th character assigned to $S_i$ matches the $k$th character in $S_i$. The same character *may not* be assigned to both $S_1$ and $S_2$. For example, the strings `abac` and `bbc` occur interwoven in `cabcbabcca`:

```
cabcbabcca
 ab   a c
    b b c
```

Give an efficient algorithm to determine if an interweaving exists (i.e. one that is polynomial in the size of the inputs).

**Problem 3: Truncated gap model (10pt)**
Consider the following gap model – each insertion or deletion costs a unit. However, if there are more than $k$ consecutive insertions, or $k$ consecutive deletions, they cost only $k$ units. Give an algorithm that finds the minimum edit distance under this cost model in time $O(nm)$. Note that the time should not depend on $k$. (Do not worry about space efficiency).

**Problem 4: Viterbi decoding example (10pt)**
Fill-in the table for the Viterbi decoding of the background/promoter HMM given in class (Lecture 3, slide 9) for the sequence $X = GCAAATGC$, with a 0.8 probability of starting in the promoter (P) state. Also, calculate the highest-probability (Viterbi) path. (Handling of start states varies — for the purposes of this problem, there is no emission at the starting state (time 0); the first emission occurs after the first state transition.)

| | output position (time) | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| state | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
| Background | 0.2 | | | | | | | | |
| Promoter | 0.8 | | | | | | | | |

**Problem 5: Dimensionality (10pt)**

In class on Thursday (Oct 7) we discussed (will discuss) the expansion and doubling constant. The definitions also appear in the paper on cover trees.

- Can you prove any bound better than $O(n)$ on the depth of an octtree for a pointset of size $n$ in 3 dimensions with expansion constant $c$? If not give an example that is bad.
- Can you prove any bound better than $O(n)$ on the depth of an octtree for a pointset of size $n$ in 3 dimensions with doubling constant $c$? If not give an example that is arbitrarily bad.