Algorithms in the Real World (15-853), Fall 09
Assignment #2 Due: October 1

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.

Problem 1: PPM, LPZ and BW (20pt)

The string `bcabccabc` is encoded using PPM. The (partial) dictionary constructed during encoding is given below. For the following questions, assume that escape count is given by the number of different characters for each context, and exclusion is *not* used, unless specified. Assume that the alphabet has 26 characters, and use $k = 2$.

(a) Fill in the empty spaces in the dictionary below.

Context	Counts	Context	Counts	Context	Counts
empty	$a = 2$	a	$b = 2$	ab	$c = 2$
	$b = 3$				
	$c = 4$	b			
		c			

Figure 1: Dictionary

(b) Suppose the next letter in the string is `b`. Compute the number of bits required to encode `b`, and also list the changes made to the dictionary. (You do not have to compute the exact number of bits, simply write it as an expression containing logs). The answer need not be an integer.

(c) Now assume that exclusion is used. Recompute the number of bits required to encode the character `b`.

(d) Encode the above string `bcabccabc` using LZ77, with an unbounded lookahead buffer, and a window of size 4.

(e) Encode the above string using Burrows-Wheeler. Just show the sequence of characters after the BW transform (don't bother compressing using move to front).

Problem 2: Linear Transform Encoding (10pt)

Consider the following functions defined on $x \in [1, n]$. State which transform (Cosine or Wavelets) gives a better compression in each case and justify in a sentence or two. You can say that they give about the same compression.

(A) $f_1(x) = \begin{cases} 1 & x \geq \frac{n}{4} \\ 0 & \text{otherwise} \end{cases}$

(B) $f_2(x) = x$

(C) $f_3(x) = \begin{cases} 1 & 2i - 1 \leq x \leq 2i \text{ where } i \text{ is an integer} \\ 0 & \text{otherwise} \end{cases}$

(D) $f_4(x) = \begin{cases} f_3(x) & x \leq \frac{n}{2} \\ 0 & \text{otherwise} \end{cases}$

Problem 3: Wavelets (10pt)

Describe how to invert the 5-tap/3-tap wavelet transform used by the lossless version of JPEG2000. In particular describe how given the outputs of the "high pass" filter and the "low pass" filter you can reconstruct the original sequence. You can ignore the boundary conditions.