18.310A Final exam practice questions

This is a collection of practice questions, gathered randomly from previous exams and quizzes. They *may not* be representative of what will be on the final. In particular, topics not represented in this list are still likely to be on the exam.

1. The phrase "peck_pickled_peppers" has the following letter frequencies:

_	c	d	e	i	k	l	p	r	s
2	2	1	4	1	2	1	5	1	1

   where the underscore _ represents a space.

   Construct an optimal Huffman tree for this set of frequencies. How does "peppers" get encoded by this code?

   Solution: View these frequencies as probabilities. The answer is not unique; just make sure you follow the algorithm described in the notes.

2. Consider a linear code with generator matrix:

$$G = \begin{pmatrix} 1 & 1 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 0 & 1 & 1 \\ 1 & 1 & 1 & 0 & 0 & 0 \end{pmatrix}.$$

   (a) How would you encode (1 0 1)?

   Solution: 001101.

   (b) Is this a 1-error correcting code? Explain.

   Solution: Yes. After Gaussian elimination, the resulting (equivalent in terms of the set of codewords) generating matrix is

$$G' = \begin{pmatrix} 1 & 0 & 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 & 0 & 1 \end{pmatrix}.$$

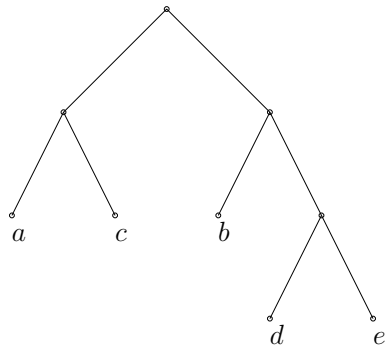
   This matrix satisfies the conditions for being 1-error correcting: considering just the 3 columns, all rows are distinct and have ≥ 2 1's.

3. Let S_1 be a source which outputs independent letters from the alphabet $A = \{a, b, c, d, e\}$ with probabilities $p_a = 1/8$, $p_b = 1/16$ $p_c = 1/2$ $p_d = 1/4$ and $p_e = 1/16$. Let S_2 be a source which outputs independent letters from the alphabet $Z = \{w, x, y, z\}$ with probabilities $p_w = 1/4$, $p_x = 1/4$ $p_y = 1/4$ $p_z = 1/4$. Suppose one wants to compress the random sequence coming out of these sources. Which source can be compressed to shorter code in average?

Solution: You just need to compute the entropies of both sources; the one with lower entropy is the more compressible.

4. Consider a source which outputs independent random letters from the alphabet $A = \{a, b, c, d, e\}$ with probabilities $p_a = 1/4$, $p_b = 1/4$, $p_c = 1/6$, $p_d = 1/6$ and $p_e = 1/6$.
   (a) Give a Huffman code for this source.

Solution: The tree is shown below. So, code(a)=00 code(c)=01 etc.



   (b) Let L_n be the random length of the Huffman code for a random sequence of n letters from that source. Compute the expectation $\mathbb{E}(L_n)$.

Solution: The random length of the code for one letter is

$$\sum_{x \in A} p_x l_x = 1/3 \times 3 + 2/3 \times 2 = 7/3.$$

So $\mathbb{E}(L_n) = 7n/3$.

   (c) Let L'_n be the (random) length of the Lempel-Ziv code for a random sequence of n letters from that source. Say if $\lim_{n\to\infty} \frac{\mathbb{E}(L_n)}{n}$ and $\lim_{n\to\infty} \frac{\mathbb{E}(L'_n)}{n}$ are equal, and if not which one is larger.

Solution:

$$\lim_{n\to\infty} \frac{\mathbb{E}(L'_n)}{n} = H < \lim_{n\to\infty} \frac{\mathbb{E}(L_n)}{n}$$

5. For the sequence

$$01011010111101101110111$$

construct the Lempel-Ziv parsing of it (i.e., its decomposition into phrases). How would it be encoded by the Lempel-Ziv code?

Solution: (Assume that the input alphabet is $\{0,1\}$.) The decomposition into phrases is

$$|0|1|01|10|101|11|011|0111|0111.$$

This yields the following list of reference-alphabet pairs:

$$0,0 \;\; 0,1 \;\; 1,1 \;\; 2,0 \;\; 4,1 \;\; 2,1 \;\; 3,1 \;\; 7,1 \;\; 7.$$

(The last phrase has no alphabet part, since it's a repeated phrase.) Encoding each dictionary reference with the correct number of bits gives

$$0 \; 01 \; 011 \; 100 \; 1001 \; 0101 \; 0111 \; 1111 \; 0111.$$

6. Consider the simple encoding scheme which simply repeats every bit of the input 3 times, and decoded by taking the majority vote.

   (a) What is the rate of this encoding scheme?

   Solution: $1/3$.

   (b) Suppose a message of length k is encoded and sent through a binary symmetric channel which flips each bit with probability p. What is the probability that a message sent using this encoding scheme is decoded correctly?

   Solution: Consider any block of 3 bits corresponding to a single bit of the input. The probability that it gets decoded correctly is the probability that at most 1 of the bits is flipped, i.e., $(1-p)^3 + 3p(1-p)^2$. The probability that each of the blocks corresponding to each bit of the input is decoded correctly is thus $((1-p)^3 + 3p(1-p)^2)^k$.

   (c) Suppose $p = 0.125$. Is it true that for k (the message length) large enough, there exists an encoding scheme with the same rate as the above simple encoding scheme and which decodes correctly with a probability of at least 99%? Explain why or why not.

   Solution: The entropy of the source is $H(p) = -\frac{1}{8}\log_2(\frac{1}{8}) - \frac{7}{8}\log_2(\frac{7}{8}) \approx 0.377 < 0.5$. Thus the channel capacacity is at least $1 - H(p) \geq 0.5$. Thus $1/3$ is above

the capacity, and by Shannon's theorem there exist good codes with the required property.

7. Let z be a primitive 32nd root of unity modulo p.
   (a) Give two possible values for p.
   (b) Write the formula for the discrete Fourier transform $a_0, a_1, \cdots, a_{31}$ of a sequence $y_0, y_1, \cdots, y_{31}$ modulo p and using z as 32nd root of unity.
   (c) Let $y_i = 1$ for $i = 0, \cdots, 31$. What is its discrete Fourier transform?

8. Find the inverse of 31 in $\mathbb{Z}_{72}^*$; show your working.

Solution: Use the Euclidean algorithm; you should get 7.

9. Let p be an odd prime. Consider the group $(\mathbb{Z}_p^*, \times)$, the group operation being multiplication modulo p. Let S be a subset of size $(p-1)/2$. Show that for some element $a \in \mathbb{Z}_p^*$, $a \in S$ and $a^{-1} \in S$.

10. Alice chooses primes $p = 5$ and $q = 11$. The public key she generates is the pair $(55, 7)$ and the private key is $(55, 23)$. Bob wants to encrypt the message $m = 53$ and send it to Alice. What is the encrypted message?

Solution: The message to be sent is

$$\begin{aligned} s &:= m^z \mod (N) = 53^7 \mod (55) \\ &= (-2)^7 \mod (55) \\ &= -128 \mod (55) \\ &= -18 \mod (55) \\ &= 37. \end{aligned}$$

11. Here is a modified RSA protocol. Instead of using a composite number $N = pq$ in the RSA cryptosystem, we just use a prime p as the main modulus. We further choose an encryption exponent e which is relatively prime with $p - 1$,yielding a public key (p, e). A message m is then encoded as $s = m^e \bmod p$. Show that this system is not secure by giving a simple (and efficient) algorithm that, given p, e and s recovers m (where $s = m^e \bmod p$).

Solution: Compute the inverse d of e modulo $p-1$ using the Euclidean algorithm (the inverse exists since $\gcd(e, p-1) = 1$). So $de = k(p-1)+1$ for some integer k. Then compute $s^d \bmod p$:

$$s^d \equiv m^{ed} = m^{k(p-1)+1} = (m^{p-1})^k \cdot m \equiv m \pmod{p}.$$

12. In implementing Pollard's rho algorithm, a colleague suggests that the following alternative algorithm might be faster:

    Modified tortoise and hare algorithm:

    1. Let $y_0 = x_0$.
    2. For $i = 1, 2, \ldots$
       (a) $x_i = f(x_{i-1})$.
       (b) $y_i = f(f(f(y_{i-1})))$ $\quad \leftarrow$ the difference is here
       (c) If $\gcd(x_i - y_i, n) \neq 1$, return this discovered factor.

    (Recall $f(x) = x^2 + 1 \bmod n$, and x_0 is some random starting element in $\mathbb{Z}_n$.) Does this work? Why, or why not? If it does work, is it faster?

Solution: This does not necessarily work. Not only might the hare jump over the tortoise, but they might never meet! Consider a cycle of length 6, label the vertices 1 through 6 in order. Suppose that at time 1, the hare is at node 3, and the tortoise at node 2. Then the hare will be at position 3 at every odd time, and 6 at every even time. But the tortoise will always be on an even position at odd times and an odd position at even times, so they will never be at the same position at the same time. Good news for the tortoise.

13. Consider the generating function

$$F(x) = \sum_{n=0}^{\infty} a_n x^n.$$

    Suppose

$$F(x) = \frac{2+2x}{1-2x-x^2}.$$

    Give an expression for a_n. Can you give a recursion for a_n and initial conditions that would give rise to this generating function?

Solution: Expanding as partial fractions,

$$F(x) = \frac{1}{\lambda_- - x} - \frac{1}{\lambda_+ + x},$$

where $\lambda_\pm = \sqrt{2} \pm 1$. Hence

$$F(x) = \frac{1}{\lambda_-} \sum_{n=0}^{\infty} (x/\lambda_-)^n - \frac{1}{\lambda_+} \sum_{n=0}^{\infty} (x/\lambda_+)^n.$$

So

$$a_n = \lambda_-^{-n-1} + \lambda_+^{-n-1}.$$

To get a recursion, rewrite the formula defining F as

$$F(x) - 2xF(x) - x^2F(x) = 2 + 2x.$$

Hence

$$\sum_{n=0}^{\infty} a_n x^n - 2\sum_{n=1}^{\infty} a_{n-1} x^n - \sum_{n=2}^{\infty} a_{n-2} x^n = 2 + 2x.$$

This gives

$$a_n - 2a_{n-1} - a_{n-2} = 0 \qquad \text{for } n \geq 2; \quad a_0 = 2, \quad a_1 - 2a_0 = 2.$$

14. Suppose we have two polynomials $p(x) = \sum_{i=0}^{d} p_i x^i$ and $q(x) = \sum_{i=0}^{d} q_i x^i$, each of degree d and with *integer* coefficients satisfying $|p_i| \leq 2$ and $|q_i| \leq 2$ for all i. Describe how one can use fast discrete fourier transforms to compute the coefficients of $p(x)q(x)$. Specify how you would select the required parameters.

Solution: *(Note: this question is quite vague.)*

We will work with the discrete Fourier transform over a finite field. The largest possible absolute value of a coefficient in $p(x)q(x)$ is $4 \cdot d$. In order to unambiguously describe integers between $-4d$ and $4d$, we should choose our field $\mathbb{Z}_p$ with $p \geq 8d+1$ (but not too much bigger).

To compute $p(x)q(x)$, first create two vectors $\vec{p}$ and $\vec{q}$ from the coefficients, but padding with zeros so that each vector has at least $2d+1$ elements. Then compute the convolution $\vec{p} \cdot \vec{q}$, by computing the DFT $\vec{c}$ of $\vec{p}$, the DFT $\vec{d}$ of $\vec{q}$, multiplying $\vec{c}$ and $\vec{d}$ pointwise, and then taking the inverse DFT of the result.

15. Let n be a positive integer. Give a concise formula for the number of sequences $a_1 < a_2 < \cdots < a_n$ of integers that satisfy $1 \leq a_i \leq 2i$ for each $1 \leq i \leq n$.

Solution: This one is tricky! First, let's write $b_i = a_i - i$ for each i. Then $b_{i+1} - b_i = a_{i+1} - a_i - 1$, so $b_1 \leq b_2 \leq \cdots \leq b_n$ (but note that equality is now possible). Also, $a_i \geq i$ (since $a_1 \geq 1$, so $a_2 > a_1 \geq 1$, etc.). Hence $0 \leq b_i \leq i$.

Now the trick is to realize that these are counted by the Catalan numbers; the answer is C_{n+1}; you might discover this by computing the first few values. So the goal is to find a bijection to some objects counted by the $(n+1)$'st Catalan number. Here is one way. Plot the points $(1, b_1), (2, b_2), \ldots, (n, b_n)$ on the plane. Note that all the points lie below the line $y = x$. Now draw a path with only horizontal and vertical steps, starting from $(0, 0)$ and ending at $(n+1, n+1)$, that stays below the line $y = x$. We begin with a step to the right, and then move up if necessary to touch the point $(1, b_1)$. Then take another step to the right, and again move up if necessary to reach $(2, b_2)$, etc., with a final step to the right and then vertical segment after (n, b_n) to reach $(n_1, n + 1)$. You can prove that this map is a bijection. Since the number of such paths is counted by the Catalan numbers (as you saw in class), you're done.

16. Let $a_0, a_1, \ldots, a_{n-1}$ be real numbers. Let $c_0, c_1, \ldots, c_{n-1}$ be the coefficients associated to the sequence $a_0, a_1, \ldots, a_{n-1}$ by the Discrete Fourier Transform. Let $c'_k = c_k$ for all $k = \{0, \ldots, n-1\}$ except $c'_1 = c_1 + 1$. Let $a'_0, a'_1, \ldots, a'_{n-1}$ be the numbers obtained from $c'_0, c'_1, \ldots, c'_{n-1}$ by the inverse Discrete Fourier Transform.

    Express $a'_0, a'_1, \ldots, a'_{n-1}$ in terms of $a_0, a_1, \ldots, a_{n-1}$.

17. Let $\vec{y} = (y_0, y_1, \ldots, y_{n-1})$ be a given vector of n complex numbers, and let $\vec{c} = (c_0, c_1, \ldots, c_{n-1})$ be the discrete Fourier transform of $\vec{y}$. Prove that

$$\sum_{j=0}^{n-1} |y_j|^2 = \sum_{k=0}^{n-1} |c_k|^2.$$

Solution: Let $z = e^{-2\pi\imath/n}$. Since $|c_k|^2 = c_k\bar{c_k}$, we have

$$\begin{aligned}
\sum_{k=0}^{n-1} |c_k|^2 &= \sum_{k=0}^{n-1}\sum_{j=0}^{n-1} y_j z^{jk} \sum_{\ell=0}^{n-1} \bar{y}_\ell z^{-\ell k} \\
&= \sum_{j=0}^{n-1}\sum_{\ell=0}^{n-1} y_j \bar{y}_\ell \sum_{k=0}^{n-1} z^{jk} z^{-\ell k} \\
&= \sum_{j=0}^{n-1}\sum_{\ell=0}^{n-1} y_j \bar{y}_\ell n\delta_{j,\ell} \\
&= \sum_{j=0}^{n-1} y_j \bar{y}_j \\
&= \sum_{j=0}^{n-1} |y_j|^2.
\end{aligned}$$

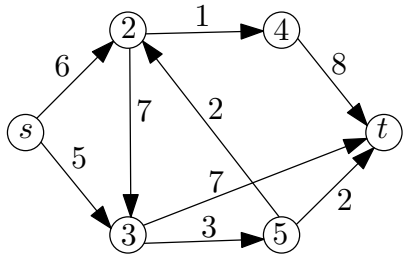
18. Determine the dual of the following LP:

$$\begin{aligned}
\min \quad & 6x - 3y - z \\
\text{s.t.} \quad & 4x - 2y + z = 4 \\
& x + 3y - z \geq 2 \\
& x, y, z \geq 0.
\end{aligned}$$

Solution:

$$\begin{aligned}
\max \quad & 4\alpha + 2\beta \\
\text{s.t.} \quad & 4\alpha + \beta \leq 6 \\
& -2\alpha + 3\beta \leq -3 \\
& \alpha - \beta \leq -1 \\
& \alpha \in \mathbb{R} \\
& \beta \geq 0.
\end{aligned}$$

19. Consider the following flow problem instance, with source s and sink t:



Find a maximum flow, and give an argument that there is no flow of any larger value.

Solution: The maximum flow has value 10 (you should be able to find such a flow); the set $S = \{s, 2, 3, 5\}$ has capacity 10, so it cannot be larger.

20. Solve the following LP using the simplex method, showing your tableau and choice of pivot clearly at each step, and writing the final answer and objective value clearly.

$$\begin{array}{lll} \text{maximize} & 3x_1 + 2x_2 + 4x_3 & \\ \text{subject to} & x_1 + x_2 + 2x_3 & \leq 4 \\ & 2x_1 \qquad + 3x_3 & \leq 5 \\ & 2x_1 + x_2 + 3x_3 & \leq 7 \\ & x_1, x_2, x_3 & \geq 0. \end{array}$$

Solution: After adding slack variables x_4, x_5 and x_6, the starting tableau is:

$-z$	x_1	x_2	x_3	x_4	x_5	x_6	
1	3	2	4				0
	1	1	2	1			4
	2	0	3		1		5
	2	1	3			1	7

After one pivot (in column x_3, 2nd row)

$-z$	x_1	x_2	x_3	x_4	x_5	x_6	
1	1/3	2			−4/3		−20/3
	−1/3	1		1	−2/3		2/3
	2/3	0	1		1/3		5/3
	0	1			−1	1	2

Next pivot is column x_2, row 1:

$-z$	x_1	x_2	x_3	x_4	x_5	x_6	
1	1			−2	0		−8
	−1/3	1		1	−2/3		2/3
	2/3		1	0	1/3		5/3
	1/3			−1	−1/3	1	4/3

Finally, pivot on column x_1, row 2:

$-z$	x_1	x_2	x_3	x_4	x_5	x_6	
1			-1.5	-2	-0.5		-10.5
		1	0.5	1	-0.5		1.5
	1		1.5	0	0.5		2.5
			-0.5	-1	-0.5	1	0.5

At this point, all coefficients in the objective row are nonpositive, so we are finished. The objective value is 10.5, and the optimal solution is

$$x_1 = 2.5, \quad x_2 = 1.5, \quad x_6 = 0.5, \quad x_3 = x_4 = x_5 = 0.$$

(In terms of the original problem without slack variables, we only need to specify $x_1 = 2.5$, $x_2 = 1.5$ and $x_3 = 0$.)

21. Consider the matrix game based on the following payoff matrix:

$$A = \begin{pmatrix} 0 & -2 & 1 \\ 2 & 0 & 3 \\ -1 & -3 & 0 \end{pmatrix}.$$

Notice that A is antisymmetric, i.e. $A = -A^T$.

(a) Write the linear programs associated with both players. Show that these linear programs are equivalent in the sense that if (x, l) is feasible for player II's linear program then $(y, g) = (x, -l)$ is feasible for player I's linear program and vice versa. Prove that $g^* = l^* = 0$.

Solution: For player I, you should get

$$\begin{aligned} & \max g \\ \text{s.t.} \quad & -2y_2 + y_3 \geq g \\ & -2y_1 - 3y_2 \geq g \\ & y_1 + 3y_2 \geq g \\ & y_1 + y_2 + y_3 = 1 \\ & y_1, y_2, y_3 \geq 0 \\ & g \in \mathbb{R} \end{aligned}$$

For player II, you should get

$$\begin{aligned} & \min l \\ \text{s.t.} \quad & -2x_2 + x_3 \leq l \\ & -2x_1 - 3x_2 \leq l \\ & x_1 + 3x_2 \leq l \\ & x_1 + x_2 + x_3 = 1 \\ & x_1, x_2, x_3 \geq 0 \\ & g \in \mathbb{R} \end{aligned}$$

Indeed, substituting $g = -l$ and $x_i = y_i$ recovers one of these LPs from the other; note how switching the sign turns a minimization into a mazimization and vice versa.

Suppose y^* is an optimal solution for player I, with guaranteed expected profit g^*. Then y^* is a feasible strategy for player II with expected loss $l^* = -g^*$, by the above. But $l^* = g^*$ by strong LP duality, so $g^* = l^* = 0$.

(b) Using part 1 and using complementary slackness, find the optimal strategies for both players.

Solution: Player I can play row 2 only (i.e., with probability 1) and symmetrically, player II can play column 2 only. This gives player I a guaranteed expected payoff of 0, and similarly player 1 a guaranteed expected loss of 0. Since these are equal, both strategies are optimal.

(Complementary slackness is not needed in the above argument.)

22. Consider the problem of tiling a rectangle of dimension $n \times 1$ with tiles of three types: white tiles of dimension 1×1, gray tiles of dimension 2×1 and black tiles of dimension 2×1. An example is given in Figure 1. Let a_n be the number of ways of tiling a rectangle of dimension $n \times 1$ (by convention $a_0 = 1$), and let $A(x) = \sum_{n=0}^{\infty} a_n x^n$.

Figure 1: A tiling of a rectangle of dimension 14×1.

(a) Give the value of a_1, a_2, a_3 (no proof needed).

Solution: $a_1 = 1$, $a_2 = 3$, $a_3 = 5$.

(b) Find an expression for the generating function $A(x)$. *(You do not need to find a formula for the coefficients a_n.)*

Solution: Tilings = Seq({white tile, gray tile, black tile}). Thus

$$A(x) = \frac{1}{1 - C(x)},$$

where $C(x)$ is the generating function of the finite set $\mathcal{C} = \{\text{white tile, gray tile, black tile}\}$. Since $\mathcal{C}$ contains 1 element of size 1 and 2 elements of size 2 one gets $C(x) = x + 2x^2$ and

$$A(x) = \frac{1}{1 - x - 2x^2}.$$