THE STATE UNIVERSITY OF NEW JERSEY

RUTGERS

College of Engineering
Department of Electrical and Computer Engineering

330:542 **Information Theory and Coding** **Spring 1997**

Takehome Exam Solutions

1. *(50 points)* The thing to notice is that when you define the connecting line with λ you're also parametrizing the function in λ. So looking along that line, if the parametrized function is convex then $d^2 f(g(\lambda))/d\lambda^2 > 2$ implies convexity along that line! Thus, the function is completely convex if it's convex along any line defined by choices of $\mathbf{x}_1$ and $\mathbf{x}_2$.

   So we start with $f(\lambda\mathbf{x}_1 + (1-\lambda)\mathbf{x}_2)$ and take derivatives. to obtain

   $$\frac{df(\lambda\mathbf{x}_1 + (1-\lambda)\mathbf{x}_2)}{d\lambda} = (\mathbf{x}_1 - \mathbf{x}_2)^T \begin{matrix} \partial f(\lambda\mathbf{x}_1 + (1-\lambda)\mathbf{x}_2)\partial x_1 \\ \partial f(\lambda\mathbf{x}_1 + (1-\lambda)\mathbf{x}_2)\partial x_2 \\ \vdots \partial f(\lambda\mathbf{x}_1 + (1-\lambda)\mathbf{x}_2)\partial x_n \end{matrix}$$

   and

   $$\frac{d^2 f(\lambda\mathbf{x}_1 + (1-\lambda)\mathbf{x}_2)}{d\lambda^2} = (\mathbf{x}_1 - \mathbf{x}_2)^T \begin{bmatrix} \partial f\partial x_1\partial x_1 & \partial\partial x_1\partial x_2 & \cdots\partial f\partial x_1\partial x_n & \\ \partial f\partial x_2\partial x_1 & \partial\partial x_2\partial x_2 & \cdots\partial f\partial x_2\partial x_n & \\ \vdots & \vdots & \vdots & \vdots \\ \partial f\partial x_n\partial x_1 & \partial\partial x_n\partial x_2 & \cdots\partial f\partial x_n\partial x_n & \end{bmatrix} (\mathbf{x}_1 - \mathbf{x}_2)$$

   and we require that this quantity be greater than zero for all possible $\mathbf{x}_1$ and $\mathbf{x}_2$. MY GOODNESS! That looks like the matrix defined by the partials must be positive definite everywhere!!!!!!

   So that's the condition. The multivariate function is convex if the matrix of its cross partials is positive definite for λ between zero and one and for all $\mathbf{x} - 1$ and $\mathbf{x}_2$ in the domain under consideration.

2. *(50 points)*

   (a) *(25 points)* The basic idea is that however you partition the original states the outgoing transition probability from each state in a given partition must be the same to each of the other partitions. This guarantees that where the process goes next is dependent only upon which partition it's in now... that's the markov property.

   Let i and j be partition indices. Let $s_k(i)$ be states of the hidden Markov system that are contained in partition i, $1 \leq k \leq k_i$. We must have $\sum_l P_{s_l(j)|s_k(i)} = \sum_l P_{s_l(j)|s'_k(i)}$ for any two k and k' and any two i and j.

   (b) *(25 points)* A uniform Markov model where each state can transition to any other with equal probability will satisfy the condition in the previous part for any possible partition of the original states.

3. *(50 points)*

   (a) *(25 points)* The stationary probability distribution on the states is geometric with paramter 1/2. That is, $p_i^* = 0.5^{i+1}$. For all states except the first state, the conditional entropy on the next state is $1/4 \log 4 + 1/2 \log 2 + 1/4 log 4 = 1.5$bits. For the first state it's $1/4 \log 4 + 3/4 \log 4/3 = (2 - 3/4 \log 3) = 0.811$bits.

   The entropy rate is then

   $$H(X_2|X_1) = 1/2(0.811) + 1/2(1.5) = 1.156\text{bits}$$

(b) *(25 points)* We can't define entropy rate here as for a markov process because the aggregated process is not markov. However the process is based on a process with a simple markov description. So let's see where that will take us.

The entropy rate is defined as

$$H(\mathcal{Y}) = \lim_{n \to \infty} H(Y_n | Y_{n-1} \ldots Y_1)$$

4. *(50 points)*

No one solution for this problem. I was looking for how you approached the problem. However, if you're really interested look at Robert Lucky's book. He has a very interesting chapter in there on the entropy of U.S. English.