TUTORIAL VIII

1 Homework 4

The main objective here is to take an algorithmic approach for the channel coding problem. The input to our algorithmic problem is the specification of a noisy channel $W_{Y|X}$ from an input set $\mathcal{X}$ to an output set $\mathcal{Y}$. We are going to use the channel only *once*. We would like to send k messages and we ask what is the minimum error probability that we can achieve.

This will be a good opportunity to introduce *submodular* functions which is an interesting property to keep in mind and a rich area of study in optimisation and approximation algorithms.

1. **Maximization of submodular functions**
   A function $f : 2^{\mathcal{X}} \to \mathbb{R}_+$ taking as input a subset $S \subseteq X$ that has the following property.

$$f(S \cup T) + f(S \cap T) \leq f(S) + f(T) \ . \tag{1}$$

   It is said to be monotone if $f(S) \leq f(T)$ whenever $S \subseteq T$.

   (a) Show that an equivalent definition for submodular function is that $f(T \cup \{j\}) - f(T) \leq f(S \cup \{j\}) - f(S)$ for any $S \subseteq T$ and any $j \in \mathcal{X}$. This can be interpreted as a "diminishing returns" property.

   (b) (Remark: this question is independent of the following questions) Let $Z_1, \dots, Z_n$ be a family of random variables. For a subset $S \subseteq \{1, \dots, n\}$, let Z_S be the collection of random variables $\{Z_i\}_{i \in S}$. Show that $f(S) = H(Z_S)$ is a submodular and monotone function.

   (c) Let f be a submodular, monotone and nonnegative function and consider the following optimization problem $\max_{S \subseteq \mathcal{X}, |S|=k} f(S)$. Let S^* of size k be such that $f(S^*) = \max_{S \subseteq \mathcal{X}, |S|=k} f(S)$. Computing such an S^* is computationally hard in general (you are even asked to show this for a special f in a later question). But there is a natural greedy algorithm for this problem: start with $S_0 = \emptyset$, then choose $S_{i+1} = S_i \cup \arg\max\{f(S_i \cup \{j\}) : j \in \mathcal{X} - S_i\}$. Show that

$$f(S^*) \leq f(S_i) + k(f(S_{i+1}) - f(S_i)) \ .$$

   (d) Prove that $f(S^*) - f(S_{i+1}) \leq (1 - \frac{1}{k})(f(S^*) - f(S_i))$.

   (e) Conclude that the greedy algorithm gives a constant factor approximation for this problem (and say what the constant is).

2. **Channel coding as a submodular optimization problem**
   Let $\mathrm{S}(W, k)$ be the largest average success probability of a code for k messages.

$$\mathrm{S}(W,k) = \max_{e,d} \frac{1}{k} \sum_{i=1}^{k} \sum_{y \in \mathcal{Y}: d(y)=i} W_{Y|X}(y|e(i)) \ , \tag{6}$$

   where the maximization is over functions $e : \{1, \dots, k\} \to \mathcal{X}$ and $d : \mathcal{Y} \to \{1, \dots, k\}$.

   (a) Show that $\mathrm{S}(W, k)$ can be written as maximizing some function f over all subsets of $\mathcal{X}$ of size k. Then show that f is submodular and monotone.

   (b) Conclude that it is possible to efficiently (here efficiently means polynomial in the description of the channel $W_{Y|X}$ and of k) find a code that achieves a success probability that is at least $(1 - 1/e) \cdot \mathrm{S}(W, k)$.

(c) Show that the following problem is **NP**-complete. You may use the **NP**-completeness of well-known problems such as 3-SAT, MAX-INDEPENDENT-SET or 3-COLORING.
Input: $W_{Y|X}$, k and a number $t \in [0, 1]$ (given in binary representation)
Output: NO if $\mathrm{S}(W, k) < t$ and YES if $\mathrm{S}(W, k) \geq t$

2 Midterm

2.1 Problem 1

For each one of these statements, say whether it is true or false and provide a brief justification.

1. Define the distribution $P_X = (1/5, 1/5, 1/5, 2/5)$. We have $H(X) = \log_2 5$.

2. For any random variable $X \in \mathcal{X}$ and any $x \in \mathcal{X}$, we have $P_X(x) \leq 2^{-H(X)}$.

3. Define the channel W with binary input and output given by $W(0|0) = 1/3, W(1|0) = 2/3, W(0|1) = 1/3, W(1|1) = 2/3$. The capacity of this channel is 0.

4. Define the tripartite mutual information $I(X : Y : Z) = I(X : Y) - I(X : Y|Z)$. For any random variables X, Y, Z, we have $I(X : Y : Z) \geq 0$.

5. For any random variables X_1, X_2, we have $H(X_1 X_2) = H(X_1) + H(X_2)$.

6. Consider the distribution $P_X = (1/2, 1/4, 1/8, 1/16, 1/16)$. The code with the shortest expected length for this source has expected length exactly $H(X)$.

7. Consider a set of points $P \subset \mathbb{R}^2$ of size m. Suppose that the projections of the set P on the x-axis and the y-axis both have at most n distinct points. Then $m \leq n^2$.

8. Let $X_1, \ldots, X_n$ be iid random variables each living in the finite set $\mathcal{X}$. Recall that a sequence $x^n = (x_1, \ldots, x_n) \in \mathcal{X}^n$ is said to be ϵ-typical if $2^{-n(H(X_1)+\epsilon)} \leq P_{X_1 \ldots X_n}(x_1 \ldots x_n) \leq 2^{-n(H(X_1)-\epsilon)}$. Now a sequence $x^n = (x_1, \ldots, x_n)$ is said to be ϵ-*strongly* typical if $(1-\epsilon)P_{X_1}(a) \leq \frac{N(a|x^n)}{n} \leq (1+\epsilon)P_{X_1}(a)$ for all $a \in \mathcal{X}$. Here $N(a|x^n)$ denotes the number of times the symbol a occurs in the sequence x^n.

   The statement is that if x^n is ϵ-strongly typical, then x^n is $c \cdot \epsilon$-typical where c is a constant that is independent of n but can depend on the distribution P_{X_1}.

9. If x^n is ϵ-typical, then it is also $c \cdot \epsilon$-strongly typical for a constant c that is independent of n but can depend on the distribution P_{X_1}.

2.2 Problem 2: Tighter analysis of the binary symmetric channel

The capacity of a channel is defined as a limit of the rate when the channel is used n times with $n \to \infty$. The objective of this problem is to obtain finite n bounds on the maximum rate of communication. We focus in this problem on the binary symmetric channel defined by

$$\mathrm{BSC}_f(b|b) = 1 - f \qquad \text{and} \qquad \mathrm{BSC}_f(1-b|b) = f \qquad \text{for any } b \in \{0, 1\} .$$

As in the homework, let us denote by $\mathrm{S}(W, k)$ the maximum over all encoding and decoding maps of the average success probability for transmitting k distinct messages over the channel W, which maps inputs $\mathcal{X}$ to outputs $\mathcal{Y}$. We can write

$$\mathrm{S}(W, k) = \max_{e,d} \frac{1}{k} \sum_{j=1}^{k} \sum_{y \in \mathcal{Y}: d(y)=j} W(y|e(j)) ,$$

where $e : \{1, \ldots, k\} \to \mathcal{X}$ and $d : \mathcal{Y} \to \{1, \ldots, k\} \cup \{\texttt{fail}\}$. In this notation, our objective is to give bounds on $\mathrm{S}(\mathrm{BSC}_f^{\otimes n}, 2^{\alpha n})$ for various values of α. Here $\mathrm{BSC}_f^{\otimes n}$ denotes n independent copies of the channel BSC_f.

1. Compute the capacity of the channel BSC_f. Draw a sketch of the graph of the capacity as a function of $f \in [0, 1]$.
2. Using the last question, what can be said on $\lim_{n\to\infty} \mathrm{S}(\mathrm{BSC}_f^{\otimes n}, 2^{\alpha n})$ as a function of α?
3. Show that for any $n \geq 1$, $f \in [0, 1/2]$ and $\alpha \in \mathbb{R}$, we have $\mathrm{S}(\mathrm{BSC}_f^{\otimes n}, 2^{\alpha n}) = \mathrm{S}(\mathrm{BSC}_{1-f}^{\otimes n}, 2^{\alpha n})$.

   Thus, in what follows, we assume that $f \in [0, 1/2]$.
4. (Achievability) We first consider the setting when α is below the capacity. Here we would like to show a lower bound on $\mathrm{S}(\mathrm{BSC}_f^{\otimes n}, 2^{\alpha n})$.
   (a) Show that for any encoding and decoding function $e : \{1, \ldots, 2^{\alpha n}\} \to \{0, 1\}^n$, $d : \{0, 1\}^n \to \{1, \ldots, 2^{\alpha n}\} \cup \{\texttt{fail}\}$, the average probability of error when transmitting a message over $\mathrm{BSC}_f^{\otimes n}$ is given by

   $$\frac{1}{2^{\alpha n}} \sum_{j=1}^{2^{\alpha n}} \mathop{\mathbf{P}}_{z\sim\mu_f} \{d(e(j) \oplus z) \neq j\} \ . \tag{8}$$

   Here μ_f denotes the distribution on $\{0, 1\}^n$ where the bits are independent and equal to 1 with probability f and $\oplus$ refers to the bitwise xor.

   (b) Now, we choose e and d for which the expression (**??**) can be upper bounded. As usual, we choose the code at random: the encoding function $e : \{1, \ldots, 2^{\alpha n}\} \to \{0, 1\}^n$ is chosen uniformly at random among all functions. For the decoder let us fix a parameter $\delta \in [0, 1]$ and define d by $d(y) = j$ if $j \in \{1, \ldots, 2^{\alpha n}\}$ is the unique j such that $\Delta(e(j), y) \leq (f + \delta)n$, otherwise, we set $d(y) = \texttt{fail}$. Here $\Delta(x, y) = |\{i \in \{1, \ldots, n\} : x_i \neq y_i\}|$ is the Hamming distance. Show that, taking the expectation (over the choice of e and d) of the probability of error (**??**) can be upper bounded by

   $$\mathop{\mathbf{P}}_{z\sim\mu_f} \{|z| > (f + \delta)n\} + (2^{\alpha n} - 1) \mathop{\mathbf{P}}_{\substack{z\sim\mu_f \\ e(1)\sim\mu_{1/2} \\ e(2)\sim\mu_{1/2}}} \{\Delta(e(1) \oplus z, e(2)) \leq (f + \delta)n\} \ . \tag{9}$$

   (c) Show that $e(1) \oplus z$ is uniformly distributed on $\{0, 1\}^n$ and conclude that $\Delta(e(1) \oplus z, e(2))$ has a binomial distribution with parameters n and $1/2$, which we denote by $\mathrm{Bin}(n, 1/2)$.

   (d) Using Chernoff's bound

   $$\mathop{\mathbf{P}}_{w\sim\mathrm{Bin}(n,f)} \{w \geq (1 + \eta)\mathbf{E}\{w\}\} \leq e^{-\frac{\eta^2 \mathbf{E}\{w\}}{3}} \qquad \text{for } \eta \in [0, 1]$$

   as well as the following inequality for $f \leq 1/2$

   $$\sum_{i=0}^{\lfloor fn \rfloor} \binom{n}{i} \leq 2^{H_2(f)n} \ ,$$

   show that for $\alpha = 1 - H_2(f) - \gamma$ with $\gamma > 0$, there is a constant $c_{\gamma,f}$ (that can depend on γ and f but not on n) such that $\mathrm{S}(\mathrm{BSC}_f^{\otimes n}, 2^{\alpha n}) \geq 1 - 2^{-c_{\gamma,f} n}$ for all $n \geq 1$.

   (e) Let again $\alpha = 1 - H_2(f) - \gamma$, how large should I take n as a function of γ to guarantee an success probability of say 0.99? Your answer can take the form $n \geq \Omega(g(\gamma))$.
5. (Strong converse) The objective of this part is to show that if $\alpha = 1 - H_2(f) + \gamma$ with $\gamma > 0$, then $\mathrm{S}(\mathrm{BSC}_f^{\otimes n}, 2^{\alpha n}) \leq 2^{-c_{\gamma,f} n}$ for $c_{\gamma,f} > 0$ and independent of n.
   (a) Let us start with a simple channel: the identity channel, i.e., BSC_0. Show that for any n, $\mathrm{S}(\mathrm{BSC}_0^{\otimes n}, 2^{\alpha n}) \leq 2^{(1-\alpha)n}$.

   (b) Show how to reduce the general case to the identity (Hint: you can see the noise z as part of the message being sent over the identity channel).