ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE
School of Computer and Communication Sciences

Handout 22 Information Theory and Coding
Solutions to Homework 9 November 26, 2013

PROBLEM 1.

(a) For all $x, y \in \mathbb{R}$, choosing $\alpha \in [0,1]$, we use the convexity of each f_i, $1 \le i \le n$, to get

$$\begin{aligned} f(\alpha x + (1-\alpha)y) &= \sum_{i=1}^{n} c_i f_i(\alpha x + (1-\alpha)y) \\ &\le \sum_{i=1}^{n} c_i \left(\alpha f_i(x) + (1-\alpha) f_i(y)\right) \\ &= \alpha \sum_{i=1}^{n} c_i f_i(x) + (1-\alpha) \sum_{i=1}^{n} c_i f_i(y) \\ &= \alpha f(x) + (1-\alpha) f(y). \end{aligned}$$

(b) For all $x = (x_1, x_2, \dots, x_n) \in \mathbb{R}^n$ and $y = (y_1, y_2, \dots, y_n) \in \mathbb{R}^n$, choosing $\alpha \in [0,1]$, observe first that $\alpha x + (1-\alpha)y = (\alpha x_1 + (1-\alpha)y_1, \alpha x_2 + (1-\alpha)y_2, \dots, \alpha x_n + (1-\alpha)y_n)$. We then use the convexity of each f_i, $1 \le i \le n$, to get

$$\begin{aligned} g(\alpha x + (1-\alpha)y) &= \sum_{i=1}^{n} c_i f_i(\alpha x_i + (1-\alpha)y_i) \\ &\le \sum_{i=1}^{n} c_i \left(\alpha f_i(x_i) + (1-\alpha) f_i(y_i)\right) \\ &= \alpha \sum_{i=1}^{n} c_i f_i(x_i) + (1-\alpha) \sum_{i=1}^{n} c_i f_i(y_i) \\ &= \alpha g(x) + (1-\alpha) g(y). \end{aligned}$$

PROBLEM 2. For all $\tilde{x} \in D$, $f(\tilde{x}) = \sup_{i \in I} f_i(x)$ iff (i) $f(\tilde{x}) \ge f_i(\tilde{x})$ for all $i \in I$ and (ii) any $s \in \mathbb{R}$ satisfying $s < f(\tilde{x})$ is such that there exists $i \in I$ satisfying $s < f_i(\tilde{x})$.

Choose $x, y \in D$ and $\alpha \in [0,1]$.

First, pick $i \in I$. With the definition of f (point (i)) and the convexity of each f_i, $i \in I$, we get

$$f_i(\alpha x + (1-\alpha)y) \le \alpha f_i(x) + (1-\alpha) f_i(y) \le \alpha f(x) + (1-\alpha) f(y).$$

Second, since the inequality $f_i(\alpha x + (1-\alpha)y) \le \alpha f(x) + (1-\alpha) f(y)$ holds for all $i \in I$, we use the definition of f (point (ii)) to claim

$$f(\alpha x + (1-\alpha)y) \le \alpha f(x) + (1-\alpha) f(y).$$

To see this, observe that, if it was not the case, then $s = \alpha f(x) + (1-\alpha) f(y) < f(\alpha x + (1-\alpha)y)$ would give the contradiction $s < f_i(\tilde{x})$ with $\tilde{x} = \alpha x + (1-\alpha)y$.

Problem 3. Choose $x, y \in U$ and $\alpha \in [0,1]$. The convexity if f associated to the fact that h is an increasing function over $[a,b]$ shows

$$g(\alpha x + (1-\alpha)y) = h\big(f\,(\alpha x + (1-\alpha)y)\,\big) \leq h\big(\alpha f(x) + (1-\alpha) f(y)\big).$$

The convexity of h gives finally

$$g(\alpha x + (1-\alpha)y) \leq \alpha h(f(x)) + (1-\alpha)h(f(y)) = \alpha g(x) + (1-\alpha)g(y).$$

Problem 4. Let us show that the function $g : \lambda \mapsto f(\lambda v_1 + (1-\lambda)v_2)$ is convex (in λ). Choosing $\lambda_x, \lambda_y \in [0,1]$ and $\alpha \in [0,1]$, we use the convexity of f in v to write

$$\begin{aligned}
g(\alpha\lambda_x + (1-\alpha)\lambda_y) &= f\big((\alpha\lambda_x + (1-\alpha)\lambda_y)v_1 + (1-(\alpha\lambda_x + (1-\alpha)\lambda_y))v_2\big)\\
&= f\big(\alpha\lambda_x v_1 + (1-\alpha)\lambda_y v_1 + v_2 - \alpha\lambda_x v_2 - (1-\alpha)\lambda_y v_2\big)\\
&= f\big(\alpha\lambda_x v_1 + (1-\alpha)\lambda_y v_1 + \big(\alpha + (1-\alpha)\big)v_2 - \alpha\lambda_x v_2 - (1-\alpha)\lambda_y v_2\big)\\
&= f\big(\alpha(\lambda_x v_1 + (1-\lambda_x)v_2) + (1-\alpha)(\lambda_y v_1 + (1-\lambda_y)v_2)\big)\\
&\leq \alpha f(\lambda_x v_1 + (1-\lambda_x)v_2) + (1-\alpha) f(\lambda_y v_1 + (1-\lambda_y)v_2)\\
&= \alpha g(\lambda_x) + (1-\alpha) g(\lambda_y).
\end{aligned}$$

Problem 5. Let $\mathcal{X} = \{x_i\}_{1\leq i\leq n}$ and $\mathcal{Y} = \{y_j\}_{1\leq j\leq m}$ be the input alphabet and output alphabet. Let $p = (p_1, p_2, \ldots, p_n) = (P(x_1), P(x_2), \ldots, P(x_n))$ denote the input probability vector. The channel is given by the probability law $\{W_{i,j} = P(y_j|x_i)\}_{i,j}$.

(a) Let us denote $q_j = P(y_j) = \sum_i W_{i,j} p_i$ and observe that the output vector $q = (q_1, q_2, \ldots, q_m)$ is a linear (convex and concave) function of p. Let $w : p \mapsto q$ denote this linear mapping. Observe also that $H(Y) = -\sum_j q_j \log q_j$ is of function of q, i.e., $H(Y) = g(q)$. To keep this in mind, let us write $H(Y) = h(q)$. We see that $H(Y) = h(q) = (h \circ w)(p)$, i.e., $g = h \circ w$. We want to show the concavity of g.

The concavity of $t \mapsto -t \log t$ and Problem 1 (b) show $h(q)$ is concave in q. Choose now two input probability vectors $p = (p_1, p_2, \ldots, p_n)$, $\tilde{p} = (\tilde{p}_1, \tilde{p}_2, \ldots, \tilde{p}_n)$ and $\alpha \in [0,1]$. We get

$$\begin{aligned}
g(\alpha p + (1-\alpha)\tilde{p}) &= h(w((\alpha p + (1-\alpha)\tilde{p}))\\
&= h((\alpha w(p) + (1-\alpha) w(\tilde{p})) && \text{since } w \text{ is linear}\\
&\geq \alpha h(w(p)) + (1-\alpha) h(w(\tilde{p})) && \text{since } h \text{ is concave}\\
&= \alpha g(p) + (1-\alpha) g(\tilde{p})
\end{aligned}$$

proving the concavity of g.

(b) Many examples might be found. A trivial one is the case when $\mathcal{X} = \{0,1\}$ and $\mathcal{Y} = \{0,1\}$ such that $P_{Y|X}(0|0) = P_{Y|X}(0|1) = 1$ for which $H(Y) = 0$ for all input distributions.

(c) Observe that $-H(Y|X) = \sum_{i,j} W_{i,j} p_i \log W_{i,j}$ is a function, call it $\xi(p)$. Clearly, $\xi(\alpha p + \tilde{p}) - \alpha\xi(p) + \xi(\tilde{p})$ showing the linearity of the function.

(d) By definition $I(X;Y) = H(Y) - H(Y|X)$. The quantity $I(X;Y)$ is therefore a function of p, call it $\iota(p) = g(p) - \xi(p)$. By (c), ξ is linear therefore $-\xi$ is concave. By (a) g is concave. By linear combination of concave function (Problem 2 (a)), we claim that $I(X;Y)$ is a concave function of the input probability vector.

Problem 6. Define

$$Q_i = \frac{a_i^{1/\lambda}}{\sum_{i=1}^n a_i^{1/\lambda}}$$

and

$$P_i = \frac{b_i^{1/(1-\lambda)}}{\sum_{i=1}^n b_i^{1/(1-\lambda)}}$$

and observe that they are non-negative numbers which sum to one.

Observe that $\phi : \lambda \mapsto Q_i^\lambda P_i^{1-\lambda}$ is a convex function for all i. To see this, note that $\phi''(\lambda) = P_i \left[\log\left(\frac{Q_i}{P_i}\right)\right]^2 \exp\left[\lambda \log\left(\frac{Q_i}{P_i}\right)\right] \geq 0$ with equality iff, for all $i \in \{1, 2, \ldots, n\}$, $P_i = Q_i$. Therefore so is $\lambda \mapsto \sum_{i=1}^n Q_i^\lambda P_i^{1-\lambda}$ by Problem 2 (a). The maximum of this function can therefore only be near the boundary $\lambda = 0$ or $\lambda = 1$. Therefore $\sum_{i=1}^n Q_i^\lambda P_i^{1-\lambda} < 1$ because of $\lambda \in (0, 1)$ and the strict convexity when $P_i \neq Q_i$. Moreover $\sum_{i=1}^n Q_i^\lambda P_i^{1-\lambda} = 1$ iff, for all $i \in \{1, 2, \ldots, n\}$, $P_i = Q_i$.

By replacing the Q_is and P_is in the inequality $\sum_{i=1}^n Q_i^\lambda P_i^{1-\lambda} \leq 1$, we get Holder's inequality, i.e.,

$$\sum_{i=1}^n a_i b_i \leq \left(\sum_{i=1}^n a_i^{1/\lambda}\right)^\lambda \left(\sum_{i=1}^n b_i^{1/(1-\lambda)}\right)^{1-\lambda}$$

with equality iff there exists some c that satisfies $a_i^{1-\lambda} = b_i^\lambda c$ for all $i \in \{1, 2, \ldots, n\}$. For the special case $\lambda = \frac{1}{2}$, this inequality is also known as the Cauchy-Schwarz inequality.