Stephanie Vance
Problem 4.57

We have a *discrete memoryless channel* between two nodes X and Y. For time $t = 1, 2, \ldots$ (in seconds, say), the input at node X is given by $X(t) \in \{1, \ldots, n\}$ and the output at note Y is given by $Y(t) \in \{1, \ldots, m\}$. We are given a *channel transition matrix* , $P = (p_{i,j})_{1\leq i\leq m, 1\leq j\leq n}$ with $p_{i,j} = \text{Prob}(Y(t) = i|X(t) = j)$.

Let X have probability distribution $x \in \mathbb{R}^n$; i.e., $x_j = \text{Prob}(X = j)$. The *mutual information* between X and Y is given by

$$I(X;Y) = \sum_{i=1}^{m}\sum_{j=1}^{n} x_j p_{i,j} \log_2\left(\frac{p_{i,j}}{\sum_{k=1}^{n} x_k p_{i,j}}\right)$$

and the Channel Capacity is given by

$$C = \sup_{x\succeq 0, \mathbf{1}^T x = 1} I(X;Y).$$

For this problem we need to show how C can be computed using convex optimization. To do this we introduce a new variable vector $y \in \mathbb{R}^m$ with constraint $y = Px$ and we let $c \in \mathbb{R}^n$ be the vector whose j^{th} component is $c_j = \sum_{i=1}^{m} p_{i,j}\log_2(p_{i,j})$.

Claim:

$$I(X,Y) = c^T x - \frac{1}{\log(2)}\sum_{i=1}^{m} y_i \log(y_i)$$

Proof:

$$\begin{aligned} I(X;Y) &= \sum_{i=1}^{m}\sum_{j=1}^{n} x_j p_{i,j}\log_2(p_{i,j}) - \sum_{i=1}^{m}\sum_{j=1}^{n} x_j p_{i,j}\log_2\left(\sum_{k=1}^{n} x_k p_{i,j}\right) \\ &= \sum_{j=1}^{n} x_j c_j - \sum_{i=1}^{m}\sum_{j=1}^{n} x_j p_{i,j}\log_2(y_i) \\ &= x^T c - \sum_{i=1}^{m} y_i \log_2(y_i) \\ &= x^T c - \frac{1}{\log(2)}\sum_{i=1}^{m} y_i\log(y_i) \end{aligned}$$

The Convex Optimization Problem:

$$C = \min_{(x,y)\in\mathbb{R}^{n\times m}} \left(\frac{1}{\log(2)}\sum_{i=1}^{m} y_i\log(y_i) - c^T x\right)$$

$$[-I_n|0_{n\times m}]\begin{bmatrix} x \\ y\end{bmatrix} \preceq \mathbf{0}$$

$$\begin{bmatrix} P & -I_m \\ 1_{1\times n} & 0_{1\times m}\end{bmatrix}\begin{bmatrix} x \\ y\end{bmatrix} = \begin{bmatrix} 0_{m\times 1} \\ 1_{n\times 1}\end{bmatrix}$$

Note that the objective function is a sum of convex functions and hence convex.

The Dual Optimization Problem:

$$\max_{(\lambda,v)\in\mathbb{R}^{n+(m+1)}} g(\lambda,v)$$
$$\lambda \succeq \mathbf{0}$$

Let $f_0(x,y) = \frac{1}{\log(2)}\sum_{i=1}^m y_i\log(y_i) - c^T x$, the objective function of the original minimization problem. We now compute the dual objective function g using equation (5.11) in the textbook.

$$\begin{aligned} g(\lambda,v) &= -\mathbf{0}^T\lambda - \begin{bmatrix} 0_{m\times 1} \\ 1 \end{bmatrix} v - {f_0}^*\left(\begin{bmatrix} I_n \\ 0_{n\times m}\end{bmatrix}\lambda - \begin{bmatrix} P^T & 1_{n\times 1} \\ -I_m & 0_{m\times 1}\end{bmatrix} v\right) \\ &= -v_{m+1} - {f_0}^*\left(\lambda - \begin{bmatrix} P^T \\ 1_{n\times 1}\end{bmatrix} v,\ v_1,\ldots,v_m\right) \end{aligned}$$

To finish the computation of g we need to compute ${f_0}^*(s,t)$ for $(s,t)\in\mathbb{R}^{n\times m}$ and determine $\mathrm{Dom}{f_0}^*$.

Since the function f_0 is the sum of m+n independent single variable functions,

$$\begin{aligned} {f_0}^*(s,t) &= \sum_{i=1}^m \left(\frac{1}{\log(2)} y_i\log(y_i)\right)^*(t_i) + \sum_{j=1}^n (c_j x_j)^*(s_j) \\ &= \sum_{i=1}^m \left(\frac{1}{\log(2)} y_i\log(y_i)\right)^*(t_i) + \tilde{I}_{\{c\}\times\mathbb{R}^m} \\ &= \sum_{i=1}^m \frac{2^{t_i}}{e} \end{aligned}$$

The domain of ${f_0}^*$ is the set $\{c\}\times\mathbb{R}^m \subseteq \mathbb{R}^{n\times m}$. To see the last equality directly above, let $h(u) = \alpha u\log(u)$ on $\mathbb{R}_+$ with $\alpha \geq 1$. For each $w\in\mathbb{R}$,

$$h^*(w) = \sup_{u\in\mathbb{R}_+} uw - \alpha u\log(u)$$

where the right hand side achieves its maximum when $w - \alpha - \alpha\log(u) = 0$, or equivalently when $u = e^{\frac{1}{\alpha}w-1}$. Substituting this value of u into the equation, $uw - h(u)$, yields $h^*(w) = e^{\frac{1}{\alpha}w-1}$ which is equal to $2^w/e$ when $\alpha = 1/\log(2)$.

We can now use the explicit formula for ${f_0}^*$ on $\mathrm{Dom}({f_0}^*)$ in the last expression for $g(\lambda,v)$ to obtain,

$$g(\lambda,v) = -v_{m+1} - \sum_{i=1}^m \frac{2^{v_i}}{e}$$

$$\mathrm{Dom}(g) = \left\{(\lambda,v)\ :\ \begin{bmatrix} P^T \\ -1\end{bmatrix} v = \lambda \text{ and } \lambda \succeq \mathbf{0}\right\}.$$

The Dual Optimization Problem becomes:

$$\max_{v\in\mathbb{R}^{m+1}} \left(-v_{m+1} - \sum_{i=1}^m \frac{2^{v_i}}{e}\right)$$
$$\begin{bmatrix} P^T \\ -1\end{bmatrix} v \succeq \mathbf{0}.$$