Information Theory I

Fall Semester 2017

Prof. Dr. A. Lapidoth

Signal and Information Processing Laboratory
Institut für Signal- und Informationsverarbeitung

Karush–Kuhn–Tucker Conditions for Channel Capacity

http://www.isi.ee.ethz.ch/teaching/courses/it1.html

Theorem 1. *Denote the transition probabilities of a DMC by $W(y|x)$ and its capacity by C. For an input distribution Q, denote the output distribution by (QW), i.e.,*

$$(QW)(y) \triangleq \sum_{x \in \mathcal{X}} Q(x)W(y|x) \quad \forall y \in \mathcal{Y}.$$

a) Let Q be an input distribution, and let $\lambda \in \mathbb{R}$. If Q and λ satisfy

$$D\big(W(\cdot|x)\|(QW)(\cdot)\big) = \lambda \qquad \forall x \in \mathcal{X} : Q(x) > 0 \quad \text{and} \tag{1}$$

$$D\big(W(\cdot|x)\|(QW)(\cdot)\big) \leq \lambda \qquad \forall x \in \mathcal{X} : Q(x) = 0, \tag{2}$$

then Q achieves capacity and the capacity of the DMC is equal to λ, i.e.,

$$\mathsf{C} = I(Q, W) = \lambda.$$

b) Let Q be a capacity-achieving input distribution. Then,

$$D\big(W(\cdot|x)\|(QW)(\cdot)\big) = \mathsf{C} \qquad \forall x \in \mathcal{X} : Q(x) > 0 \quad \text{and}$$

$$D\big(W(\cdot|x)\|(QW)(\cdot)\big) \leq \mathsf{C} \qquad \forall x \in \mathcal{X} : Q(x) = 0.$$

We use the following theorem to prove Theorem 1:

Theorem 2. *Let $f(\boldsymbol{\alpha})$ be a concave function of $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_n)$ over the probability simplex $\mathcal{R} \triangleq \big\{\boldsymbol{\alpha} \colon \alpha_i \geq 0 \; \forall i, \; \sum_{i=1}^{n} \alpha_i = 1\big\}$. Assume that the partial derivatives, $\frac{\partial f(\boldsymbol{\alpha})}{\partial \alpha_k}$ are defined and continuous over the simplex $\mathcal{R}$ with the possible exception that $\lim_{\alpha_k \downarrow 0} \frac{\partial f(\boldsymbol{\alpha})}{\partial \alpha_k}$ may be $+\infty$. Then, for $\lambda' \in \mathbb{R}$,*

$$\left.\frac{\partial f(\boldsymbol{\alpha})}{\partial \alpha_k}\right|_{\boldsymbol{\alpha}=\boldsymbol{\alpha}^\star} = \lambda' \qquad \forall\, k \text{ such that } \alpha_k^\star > 0,$$

$$\left.\frac{\partial f(\boldsymbol{\alpha})}{\partial \alpha_k}\right|_{\boldsymbol{\alpha}=\boldsymbol{\alpha}^\star} \leq \lambda' \qquad \forall\, k \text{ such that } \alpha_k^\star = 0$$

are necessary and sufficient conditions on a probability vector $\boldsymbol{\alpha}^\star$ to maximize $f(\boldsymbol{\alpha})$ over $\mathcal{R}$.

Proof. See Robert G. Gallager, *Information Theory and Reliable Communication*, John Wiley & Sons, 1968, Theorem 4.4.1. ■

Proof of Theorem 1. We know that $Q \mapsto I(Q,W)$ is concave, so we apply Theorem 2 with $\boldsymbol{\alpha}$ corresponding to Q and α_k corresponding to $Q_k = Q(x_k)$. To compute the partial derivatives, we express the mutual information between the input and the output of a DMC as

$$\begin{aligned} I(Q,W) &= \sum_x \sum_y Q(x)W(y|x) \ln \frac{Q(x)W(y|x)}{Q(x)(QW)(y)} \\ &= \sum_x \sum_y Q(x)W(y|x) \ln \frac{W(y|x)}{\sum_{x'} Q(x')W(y|x')}. \end{aligned}$$

(Without loss of generality we use natural logarithms.) By the product rule and by the chain rule (note that every Q_k appears in two positions), we have

$$\begin{aligned} \frac{\partial I(Q,W)}{\partial Q_k} &= \sum_x \sum_y \mathrm{I}\{x = x_k\} W(y|x) \ln \frac{W(y|x)}{\sum_{x'} Q(x')W(y|x')} \\ &\quad + \sum_x \sum_y Q(x)W(y|x) \cdot \frac{\sum_{x'} Q(x')W(y|x')}{W(y|x)} \cdot \frac{-W(y|x)\cdot W(y|x_k)}{\left(\sum_{x'} Q(x')W(y|x')\right)^2} \\ &= \sum_y W(y|x_k) \ln \frac{W(y|x_k)}{\sum_{x'} Q(x')W(y|x')} - \sum_y \frac{W(y|x_k)}{\sum_{x'} Q(x')W(y|x')} \sum_x Q(x)W(y|x) \\ &= \sum_y W(y|x_k) \ln \frac{W(y|x_k)}{\sum_{x'} Q(x')W(y|x')} - \sum_y W(y|x_k) \\ &= \sum_y W(y|x_k) \ln \frac{W(y|x_k)}{\sum_{x'} Q(x')W(y|x')} - 1 \\ &= D\big(W(\cdot|x_k) \| (QW)(\cdot)\big) - 1. \end{aligned}$$

(This also allows to check that the partial derivatives fulfill the conditions of Theorem 2.)

We are now ready to prove Part a). Since (1) and (2) are satisfied, we can invoke Theorem 2 (with $\lambda' = \lambda - 1$) to conclude that Q maximizes $I(\cdot, W)$ over all input distributions, i.e., that Q achieves capacity. Then,

$$\begin{aligned} \mathsf{C} \overset{\text{(i)}}{=} I(Q,W) &= \sum_x \sum_y Q(x)W(y|x) \ln \frac{W(y|x)}{\sum_{x'} Q(x')W(y|x')} \\ &= \sum_x Q(x) D\big(W(\cdot|x) \| (QW)(\cdot)\big) \\ &\overset{\text{(ii)}}{=} \lambda, \end{aligned} \tag{3}$$

where (i) holds because Q achieves capacity; and (ii) follows from (1).

We finish by proving Part b). Because Q maximizes $I(\cdot, W)$ over all input distributions, we know by Theorem 2 that there exists a λ' such that

$$\begin{aligned} D\big(W(\cdot|x) \| (QW)(\cdot)\big) &= \lambda' + 1 \qquad \forall x \in \mathcal{X} : Q(x) > 0 \quad \text{and} \\ D\big(W(\cdot|x) \| (QW)(\cdot)\big) &\le \lambda' + 1 \qquad \forall x \in \mathcal{X} : Q(x) = 0. \end{aligned}$$

From the same computation as in (3) we obtain $I(Q,W) = \lambda' + 1$. Because we know that $I(Q,W) = \mathsf{C}$, it follows that $\lambda' + 1 = \mathsf{C}$. ■