香港中文大學
The Chinese University of Hong Kong

9.2 The Algorithms

9.2.1 Channel Capacity

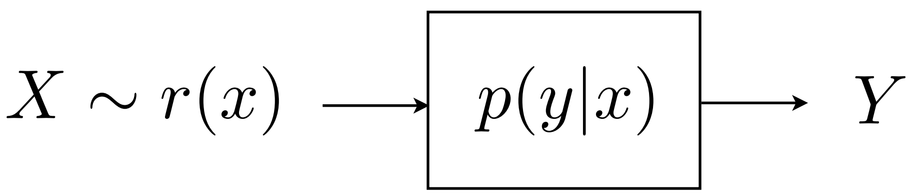
Lemma 9.1 Let $r(x)p(y|x)$ be a given joint distribution on $\mathcal{X} \times \mathcal{Y}$ such that $\mathbf{r} > 0$, and let $\mathbf{q}$ be a transition matrix from $\mathcal{Y}$ to $\mathcal{X}$. Then

Lemma 9.1 Let $r(x)p(y|x)$ be a given joint distribution on $\mathcal{X} \times \mathcal{Y}$ such that $\mathbf{r} > 0$, and let $\mathbf{q}$ be a transition matrix from $\mathcal{Y}$ to $\mathcal{X}$. Then

$$X \sim r(x) \longrightarrow \boxed{p(y|x)} \longrightarrow Y$$

Lemma 9.1 Let $r(x)p(y|x)$ be a given joint distribution on $\mathcal{X} \times \mathcal{Y}$ such that $\mathbf{r} > 0$, and let $\mathbf{q}$ be a transition matrix from $\mathcal{Y}$ to $\mathcal{X}$. Then

$$\max_{\mathbf{q}} \sum_x \sum_y r(x)p(y|x) \log \frac{q(x|y)}{r(x)}$$

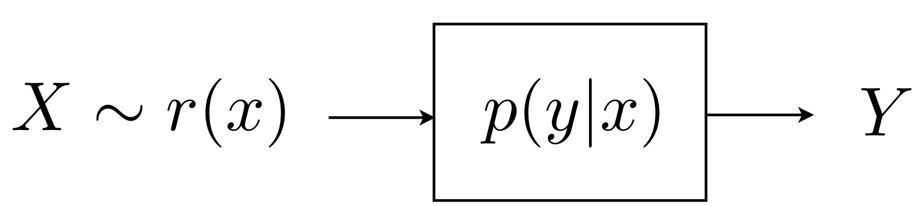


Lemma 9.1 Let $r(x)p(y|x)$ be a given joint distribution on $\mathcal{X} \times \mathcal{Y}$ such that $\mathbf{r} > 0$, and let $\mathbf{q}$ be a transition matrix from $\mathcal{Y}$ to $\mathcal{X}$. Then

$$\max_{\mathbf{q}} \sum_x \sum_y r(x)p(y|x) \log \frac{q(x|y)}{r(x)}$$

where the maximization is taken over all $\mathbf{q}$ such that

$$q(x|y) = 0 \quad \text{if and only if} \quad p(y|x) = 0,$$

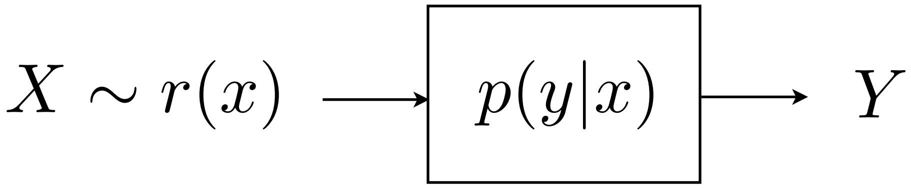


Lemma 9.1 Let $r(x)p(y|x)$ be a given joint distribution on $\mathcal{X} \times \mathcal{Y}$ such that $\mathbf{r} > 0$, and let $\mathbf{q}$ be a transition matrix from $\mathcal{Y}$ to $\mathcal{X}$. Then

$$\max_{\mathbf{q}} \sum_x \sum_y r(x)p(y|x) \log \frac{q(x|y)}{r(x)} = \sum_x \sum_y r(x)p(y|x) \log \frac{q^*(x|y)}{r(x)},$$

where the maximization is taken over all $\mathbf{q}$ such that

$$q(x|y) = 0 \quad \text{if and only if} \quad p(y|x) = 0,$$



Lemma 9.1 Let $r(x)p(y|x)$ be a given joint distribution on $\mathcal{X} \times \mathcal{Y}$ such that $\mathbf{r} > 0$, and let $\mathbf{q}$ be a transition matrix from $\mathcal{Y}$ to $\mathcal{X}$. Then

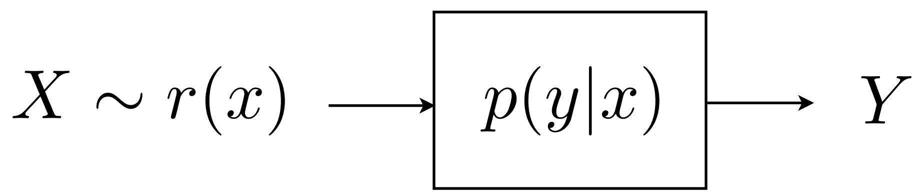
$$\max_{\mathbf{q}} \sum_x \sum_y r(x)p(y|x) \log \frac{q(x|y)}{r(x)} = \sum_x \sum_y r(x)p(y|x) \log \frac{q^*(x|y)}{r(x)},$$

where the maximization is taken over all $\mathbf{q}$ such that

$$q(x|y) = 0 \quad \text{if and only if} \quad p(y|x) = 0,$$

and

$$q^*(x|y) = \frac{r(x)p(y|x)}{\sum_{x'} r(x')p(y|x')}$$



Lemma 9.1 Let $r(x)p(y|x)$ be a given joint distribution on $\mathcal{X} \times \mathcal{Y}$ such that $\mathbf{r} > 0$, and let $\mathbf{q}$ be a transition matrix from $\mathcal{Y}$ to $\mathcal{X}$. Then

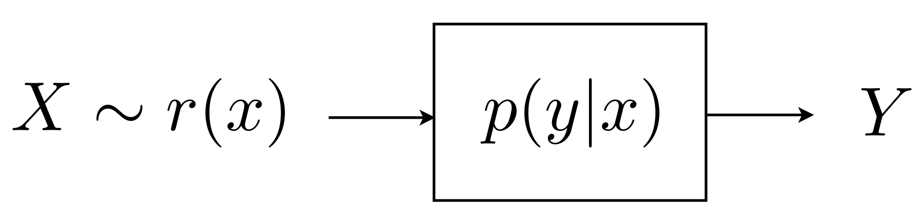
$$\max_{\mathbf{q}} \sum_x \sum_y r(x)p(y|x) \log \frac{q(x|y)}{r(x)} = \sum_x \sum_y r(x)p(y|x) \log \frac{q^*(x|y)}{r(x)},$$

where the maximization is taken over all $\mathbf{q}$ such that

$$q(x|y) = 0 \quad \text{if and only if} \quad p(y|x) = 0,$$

and

$$q^*(x|y) = \frac{r(x)p(y|x)}{\sum_{x'} r(x')p(y|x')} = \frac{q_{XY}(x,y)}{q_Y(y)}$$



Lemma 9.1 Let $r(x)p(y|x)$ be a given joint distribution on $\mathcal{X} \times \mathcal{Y}$ such that $\mathbf{r} > 0$, and let $\mathbf{q}$ be a transition matrix from $\mathcal{Y}$ to $\mathcal{X}$. Then

$$\max_{\mathbf{q}} \sum_x \sum_y r(x)p(y|x) \log \frac{q(x|y)}{r(x)} = \sum_x \sum_y r(x)p(y|x) \log \frac{q^*(x|y)}{r(x)},$$

where the maximization is taken over all $\mathbf{q}$ such that

$$q(x|y) = 0 \quad \text{if and only if} \quad p(y|x) = 0,$$

and

$$q^*(x|y) = \frac{r(x)p(y|x)}{\sum_{x'} r(x')p(y|x')} = \frac{q_{XY}(x,y)}{q_Y(y)} = q_{X|Y}(x|y)$$

$$X \sim r(x) \longrightarrow \boxed{p(y|x)} \longrightarrow Y$$

Lemma 9.1 Let $r(x)p(y|x)$ be a given joint distribution on $\mathcal{X} \times \mathcal{Y}$ such that $\mathbf{r} > 0$, and let $\mathbf{q}$ be a transition matrix from $\mathcal{Y}$ to $\mathcal{X}$. Then

$$\max_{\mathbf{q}} \sum_x \sum_y r(x)p(y|x) \log \frac{q(x|y)}{r(x)} = \sum_x \sum_y r(x)p(y|x) \log \frac{q^*(x|y)}{r(x)},$$

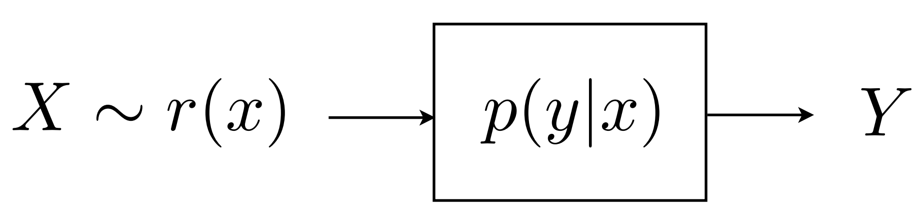
where the maximization is taken over all $\mathbf{q}$ such that

$$q(x|y) = 0 \quad \text{if and only if} \quad p(y|x) = 0,$$

and

$$q^*(x|y) = \frac{r(x)p(y|x)}{\sum_{x'} r(x')p(y|x')} = \frac{q_{XY}(x,y)}{q_Y(y)} = q_{X|Y}(x|y)$$

i.e., the maximizing $\mathbf{q}$ is the one which corresponds to the input distribution $\mathbf{r}$ and the transition matrix $p(y|x)$.



Lemma 9.1 Let $r(x)p(y|x)$ be a given joint distribution on $\mathcal{X} \times \mathcal{Y}$ such that $\mathbf{r} > 0$, and let $\mathbf{q}$ be a transition matrix from $\mathcal{Y}$ to $\mathcal{X}$. Then

$$\max_{\mathbf{q}} \sum_x \sum_y r(x)p(y|x) \log \frac{q(x|y)}{r(x)} = \sum_x \sum_y \underbrace{r(x)p(y|x)}_{q_{XY}(x,y)} \log \frac{q^*(x|y)}{r(x)},$$

where the maximization is taken over all $\mathbf{q}$ such that

$$q(x|y) = 0 \quad \text{if and only if} \quad p(y|x) = 0,$$

and

$$q^*(x|y) = \frac{r(x)p(y|x)}{\sum_{x'} r(x')p(y|x')} = \frac{q_{XY}(x,y)}{q_Y(y)} = q_{X|Y}(x|y)$$

i.e., the maximizing $\mathbf{q}$ is the one which corresponds to the input distribution $\mathbf{r}$ and the transition matrix $p(y|x)$.

$$X \sim r(x) \longrightarrow \boxed{p(y|x)} \longrightarrow Y$$

Lemma 9.1 Let $r(x)p(y|x)$ be a given joint distribution on $\mathcal{X} \times \mathcal{Y}$ such that $\mathbf{r} > 0$, and let $\mathbf{q}$ be a transition matrix from $\mathcal{Y}$ to $\mathcal{X}$. Then

$$\max_{\mathbf{q}} \sum_x \sum_y r(x)p(y|x) \log \frac{q(x|y)}{r(x)} = \sum_x \sum_y \underbrace{r(x)p(y|x)}_{q_{XY}(x,y)} \log \frac{q^*(x|y)}{r(x)},$$

$$\frac{\frac{q_{XY}(x,y)}{q_Y(y)}}{q_X(x)} = \frac{q_{XY}(x,y)}{q_X(x)q_Y(y)}$$

where the maximization is taken over all $\mathbf{q}$ such that

$$q(x|y) = 0 \quad \text{if and only if} \quad p(y|x) = 0,$$

and

$$q^*(x|y) = \frac{r(x)p(y|x)}{\sum_{x'} r(x')p(y|x')} = \frac{q_{XY}(x,y)}{q_Y(y)} = q_{X|Y}(x|y)$$

i.e., the maximizing $\mathbf{q}$ is the one which corresponds to the input distribution $\mathbf{r}$ and the transition matrix $p(y|x)$.

$$X \sim r(x) \longrightarrow \boxed{p(y|x)} \longrightarrow Y$$

Lemma 9.1 Let $r(x)p(y|x)$ be a given joint distribution on $\mathcal{X} \times \mathcal{Y}$ such that $\mathbf{r} > 0$, and let $\mathbf{q}$ be a transition matrix from $\mathcal{Y}$ to $\mathcal{X}$. Then

$$\max_{\mathbf{q}} \sum_x \sum_y r(x)p(y|x) \log \frac{q(x|y)}{r(x)} = \boxed{\sum_x \sum_y r(x)p(y|x) \log \frac{q^*(x|y)}{r(x)}}, \quad I(X;Y)$$

where the maximization is taken over all $\mathbf{q}$ such that

$$q(x|y) = 0 \quad \text{if and only if} \quad p(y|x) = 0,$$

and

$$q^*(x|y) = \frac{r(x)p(y|x)}{\sum_{x'} r(x')p(y|x')} = \frac{q_{XY}(x,y)}{q_Y(y)} = q_{X|Y}(x|y)$$

i.e., the maximizing $\mathbf{q}$ is the one which corresponds to the input distribution $\mathbf{r}$ and the transition matrix $p(y|x)$.

$$X \sim r(x) \longrightarrow \boxed{p(y|x)} \longrightarrow Y$$

Lemma 9.1 Let $r(x)p(y|x)$ be a given joint distribution on $\mathcal{X} \times \mathcal{Y}$ such that $\mathbf{r} > 0$, and let $\mathbf{q}$ be a transition matrix from $\mathcal{Y}$ to $\mathcal{X}$. Then

$$\max_{\mathbf{q}} \sum_x \sum_y r(x)p(y|x) \log \frac{q(x|y)}{r(x)} = \sum_x \sum_y r(x)p(y|x) \log \frac{q^*(x|y)}{r(x)},$$

where the maximization is taken over all $\mathbf{q}$ such that

$$q(x|y) = 0 \quad \text{if and only if} \quad p(y|x) = 0, \tag{1}$$

and

$$q^*(x|y) = \frac{r(x)p(y|x)}{\sum_{x'} r(x')p(y|x')}. \tag{2}$$

Proof

Lemma 9.1 Let $r(x)p(y|x)$ be a given joint distribution on $\mathcal{X} \times \mathcal{Y}$ such that $\mathbf{r} > 0$, and let $\mathbf{q}$ be a transition matrix from $\mathcal{Y}$ to $\mathcal{X}$. Then

$$\max_{\mathbf{q}} \sum_x \sum_y r(x)p(y|x) \log \frac{q(x|y)}{r(x)} = \sum_x \sum_y r(x)p(y|x) \log \frac{q^*(x|y)}{r(x)},$$

where the maximization is taken over all $\mathbf{q}$ such that

$$q(x|y) = 0 \quad \text{if and only if} \quad p(y|x) = 0, \tag{1}$$

and

$$q^*(x|y) = \frac{r(x)p(y|x)}{\sum_{x'} r(x')p(y|x')}. \tag{2}$$

Proof

Lemma 9.1 Let $r(x)p(y|x)$ be a given joint distribution on $\mathcal{X} \times \mathcal{Y}$ such that $\mathbf{r} > 0$, and let $\mathbf{q}$ be a transition matrix from $\mathcal{Y}$ to $\mathcal{X}$. Then

$$\max_{\mathbf{q}} \sum_x \sum_y r(x)p(y|x) \log \frac{q(x|y)}{r(x)}$$
$$= \sum_x \sum_y r(x)p(y|x) \log \frac{q^*(x|y)}{r(x)},$$

where the maximization is taken over all $\mathbf{q}$ such that

$$q(x|y) = 0 \quad \text{if and only if} \quad p(y|x) = 0, \qquad (1)$$

and

$$q^*(x|y) = \frac{r(x)p(y|x)}{\sum_{x'} r(x')p(y|x')}. \qquad (2)$$

Proof

1. In (2), let

$$w(y) = \sum_{x'} r(x')p(y|x').$$

Lemma 9.1 Let $r(x)p(y|x)$ be a given joint distribution on $\mathcal{X} \times \mathcal{Y}$ such that $\mathbf{r} > 0$, and let $\mathbf{q}$ be a transition matrix from $\mathcal{Y}$ to $\mathcal{X}$. Then

$$\max_{\mathbf{q}} \sum_x \sum_y r(x)p(y|x) \log \frac{q(x|y)}{r(x)} = \sum_x \sum_y r(x)p(y|x) \log \frac{q^*(x|y)}{r(x)},$$

where the maximization is taken over all $\mathbf{q}$ such that

$$q(x|y) = 0 \quad \text{if and only if} \quad p(y|x) = 0, \qquad (1)$$

and

$$q^*(x|y) = \frac{r(x)p(y|x)}{\cancel{\sum_{x'} r(x')p(y|x')}\ w(y)}. \qquad (2)$$

Proof

1. In (2), let

$$w(y) = \sum_{x'} r(x')p(y|x').$$

Lemma 9.1 Let $r(x)p(y|x)$ be a given joint distribution on $\mathcal{X} \times \mathcal{Y}$ such that $\mathbf{r} > 0$, and let $\mathbf{q}$ be a transition matrix from $\mathcal{Y}$ to $\mathcal{X}$. Then

$$\max_{\mathbf{q}} \sum_x \sum_y r(x)p(y|x) \log \frac{q(x|y)}{r(x)}$$
$$= \quad \sum_x \sum_y r(x)p(y|x) \log \frac{q^*(x|y)}{r(x)},$$

where the maximization is taken over all $\mathbf{q}$ such that

$$q(x|y) = 0 \quad \text{if and only if} \quad p(y|x) = 0, \qquad (1)$$

and

$$q^*(x|y) = \frac{r(x)p(y|x)}{\cancel{\sum_{x'} r(x')p(y|x')}}. \; w(y) \qquad (2)$$

Proof

1. In (2), let

$$w(y) = \sum_{x'} r(x')p(y|x').$$

2. Assume w.l.o.g. that for all $y \in \mathcal{Y}$, $p(y|x) > 0$ for some $x \in \mathcal{X}$.

Lemma 9.1 Let $r(x)p(y|x)$ be a given joint distribution on $\mathcal{X} \times \mathcal{Y}$ such that $\mathbf{r} > 0$, and let $\mathbf{q}$ be a transition matrix from $\mathcal{Y}$ to $\mathcal{X}$. Then

$$\max_{\mathbf{q}} \sum_x \sum_y r(x)p(y|x) \log \frac{q(x|y)}{r(x)} = \sum_x \sum_y r(x)p(y|x) \log \frac{q^*(x|y)}{r(x)},$$

where the maximization is taken over all $\mathbf{q}$ such that

$$q(x|y) = 0 \quad \text{if and only if} \quad p(y|x) = 0, \tag{1}$$

and

$$q^*(x|y) = \frac{r(x)p(y|x)}{\cancel{\sum_{x'} r(x')p(y|x')}\; w(y)}. \tag{2}$$

Proof

1. In (2), let

$$w(y) = \sum_{x'} r(x')p(y|x').$$

2. Assume w.l.o.g. that for all $y \in \mathcal{Y}$, $p(y|x) > 0$ for some $x \in \mathcal{X}$.

3. Since $\mathbf{r} > 0$, $w(y) > 0$ for all y, and hence $q^*(x|y)$ is well-defined.

Lemma 9.1 Let $r(x)p(y|x)$ be a given joint distribution on $\mathcal{X} \times \mathcal{Y}$ such that $\mathbf{r} > 0$, and let $\mathbf{q}$ be a transition matrix from $\mathcal{Y}$ to $\mathcal{X}$. Then

$$\begin{aligned} &\max_{\mathbf{q}} \sum_x \sum_y r(x)p(y|x) \log \frac{q(x|y)}{r(x)} \\ = \quad &\sum_x \sum_y r(x)p(y|x) \log \frac{q^*(x|y)}{r(x)}, \end{aligned}$$

where the maximization is taken over all $\mathbf{q}$ such that

$$q(x|y) = 0 \quad \text{if and only if} \quad p(y|x) = 0, \tag{1}$$

and

$$q^*(x|y) = \frac{r(x)p(y|x)}{\cancel{\sum_{x'} r(x')p(y|x')}}. \; w(y) \tag{2}$$

Proof

1. In (2), let

$$w(y) = \sum_{x'} r(x')p(y|x').$$

2. Assume w.l.o.g. that for all $y \in \mathcal{Y}$, $p(y|x) > 0$ for some $x \in \mathcal{X}$.

3. Since $\mathbf{r} > 0$, $w(y) > 0$ for all y, and hence $q^*(x|y)$ is well-defined.

4. Rearranging (2), we have

$$r(x)p(y|x) = w(y)q^*(x|y).$$

Lemma 9.1 Let $r(x)p(y|x)$ be a given joint distribution on $\mathcal{X} \times \mathcal{Y}$ such that $\mathbf{r} > 0$, and let $\mathbf{q}$ be a transition matrix from $\mathcal{Y}$ to $\mathcal{X}$. Then

$$\begin{aligned}\max_{\mathbf{q}} \sum_x \sum_y r(x)p(y|x) \log \frac{q(x|y)}{r(x)} \\ = \quad \sum_x \sum_y r(x)p(y|x) \log \frac{q^*(x|y)}{r(x)},\end{aligned}$$

where the maximization is taken over all $\mathbf{q}$ such that

$$q(x|y) = 0 \quad \text{if and only if} \quad p(y|x) = 0, \qquad (1)$$

and

$$q^*(x|y) = \frac{r(x)p(y|x)}{\sum_{x'} r(x')p(y|x')}. \qquad (2)$$

Proof

1. In (2), let

$$w(y) = \sum_{x'} r(x')p(y|x').$$

2. Assume w.l.o.g. that for all $y \in \mathcal{Y}$, $p(y|x) > 0$ for some $x \in \mathcal{X}$.

3. Since $\mathbf{r} > 0$, $w(y) > 0$ for all y, and hence $q^*(x|y)$ is well-defined.

4. Rearranging (2), we have

$$r(x)p(y|x) = w(y)q^*(x|y).$$

Lemma 9.1 Let $r(x)p(y|x)$ be a given joint distribution on $\mathcal{X} \times \mathcal{Y}$ such that $\mathbf{r} > 0$, and let $\mathbf{q}$ be a transition matrix from $\mathcal{Y}$ to $\mathcal{X}$. Then

$$\max_{\mathbf{q}} \sum_x \sum_y r(x)p(y|x) \log \frac{q(x|y)}{r(x)} = \sum_x \sum_y r(x)p(y|x) \log \frac{q^*(x|y)}{r(x)},$$

where the maximization is taken over all $\mathbf{q}$ such that

$$q(x|y) = 0 \quad \text{if and only if} \quad p(y|x) = 0, \tag{1}$$

and

$$q^*(x|y) = \frac{r(x)p(y|x)}{\sum_{x'} r(x')p(y|x')}. \tag{2}$$

Proof

1. In (2), let

$$w(y) = \sum_{x'} r(x')p(y|x').$$

2. Assume w.l.o.g. that for all $y \in \mathcal{Y}$, $p(y|x) > 0$ for some $x \in \mathcal{X}$.

3. Since $\mathbf{r} > 0$, $w(y) > 0$ for all y, and hence $q^*(x|y)$ is well-defined.

4. Rearranging (2), we have

$$r(x)p(y|x) = w(y)q^*(x|y).$$

5. For any $\mathbf{q}$ satisfying (1), consider

Lemma 9.1 Let $r(x)p(y|x)$ be a given joint distribution on $\mathcal{X} \times \mathcal{Y}$ such that $\mathbf{r} > 0$, and let $\mathbf{q}$ be a transition matrix from $\mathcal{Y}$ to $\mathcal{X}$. Then

$$\max_{\mathbf{q}} \sum_x \sum_y r(x)p(y|x) \log \frac{q(x|y)}{r(x)} = \sum_x \sum_y r(x)p(y|x) \log \frac{q^*(x|y)}{r(x)},$$

where the maximization is taken over all $\mathbf{q}$ such that

$$q(x|y) = 0 \quad \text{if and only if} \quad p(y|x) = 0, \tag{1}$$

and

$$q^*(x|y) = \frac{r(x)p(y|x)}{\sum_{x'} r(x')p(y|x')}. \tag{2}$$

Proof

1. In (2), let

$$w(y) = \sum_{x'} r(x')p(y|x').$$

2. Assume w.l.o.g. that for all $y \in \mathcal{Y}$, $p(y|x) > 0$ for some $x \in \mathcal{X}$.

3. Since $\mathbf{r} > 0$, $w(y) > 0$ for all y, and hence $q^*(x|y)$ is well-defined.

4. Rearranging (2), we have

$$r(x)p(y|x) = w(y)q^*(x|y).$$

5. For any $\mathbf{q}$ satisfying (1), consider

$$\sum_x \sum_y r(x)p(y|x) \log \frac{q^*(x|y)}{r(x)} - \sum_x \sum_y r(x)p(y|x) \log \frac{q(x|y)}{r(x)}$$

Lemma 9.1 Let $r(x)p(y|x)$ be a given joint distribution on $\mathcal{X} \times \mathcal{Y}$ such that $\mathbf{r} > 0$, and let $\mathbf{q}$ be a transition matrix from $\mathcal{Y}$ to $\mathcal{X}$. Then

$$\max_{\mathbf{q}} \sum_x \sum_y r(x)p(y|x) \log \frac{q(x|y)}{r(x)} = \sum_x \sum_y r(x)p(y|x) \log \frac{q^*(x|y)}{r(x)},$$

where the maximization is taken over all $\mathbf{q}$ such that

$$q(x|y) = 0 \quad \text{if and only if} \quad p(y|x) = 0, \qquad (1)$$

and

$$q^*(x|y) = \frac{r(x)p(y|x)}{\sum_{x'} r(x')p(y|x')}. \qquad (2)$$

Proof

1. In (2), let

$$w(y) = \sum_{x'} r(x')p(y|x').$$

2. Assume w.l.o.g. that for all $y \in \mathcal{Y}$, $p(y|x) > 0$ for some $x \in \mathcal{X}$.

3. Since $\mathbf{r} > 0$, $w(y) > 0$ for all y, and hence $q^*(x|y)$ is well-defined.

4. Rearranging (2), we have

$$r(x)p(y|x) = w(y)q^*(x|y).$$

5. For any $\mathbf{q}$ satisfying (1), consider

$$\sum_x \sum_y r(x)p(y|x) \log \frac{q^*(x|y)}{r(x)} - \sum_x \sum_y r(x)p(y|x) \log \frac{q(x|y)}{r(x)}$$

Lemma 9.1 Let $r(x)p(y|x)$ be a given joint distribution on $\mathcal{X} \times \mathcal{Y}$ such that $\mathbf{r} > 0$, and let $\mathbf{q}$ be a transition matrix from $\mathcal{Y}$ to $\mathcal{X}$. Then

$$\max_{\mathbf{q}} \sum_x \sum_y r(x)p(y|x) \log \frac{q(x|y)}{r(x)} = \sum_x \sum_y r(x)p(y|x) \log \frac{q^*(x|y)}{r(x)},$$

where the maximization is taken over all $\mathbf{q}$ such that

$$q(x|y) = 0 \quad \text{if and only if} \quad p(y|x) = 0, \tag{1}$$

and

$$q^*(x|y) = \frac{r(x)p(y|x)}{\sum_{x'} r(x')p(y|x')}. \tag{2}$$

Proof

1. In (2), let

$$w(y) = \sum_{x'} r(x')p(y|x').$$

2. Assume w.l.o.g. that for all $y \in \mathcal{Y}$, $p(y|x) > 0$ for some $x \in \mathcal{X}$.

3. Since $\mathbf{r} > 0$, $w(y) > 0$ for all y, and hence $q^*(x|y)$ is well-defined.

4. Rearranging (2), we have

$$r(x)p(y|x) = w(y)q^*(x|y).$$

5. For any $\mathbf{q}$ satisfying (1), consider

$$\sum_x \sum_y r(x)p(y|x) \log \frac{q^*(x|y)}{\cancel{r(x)}} - \sum_x \sum_y r(x)p(y|x) \log \frac{q(x|y)}{r(x)}$$

Lemma 9.1 Let $r(x)p(y|x)$ be a given joint distribution on $\mathcal{X} \times \mathcal{Y}$ such that $\mathbf{r} > 0$, and let $\mathbf{q}$ be a transition matrix from $\mathcal{Y}$ to $\mathcal{X}$. Then

$$\max_{\mathbf{q}} \sum_x \sum_y r(x)p(y|x) \log \frac{q(x|y)}{r(x)} = \sum_x \sum_y r(x)p(y|x) \log \frac{q^*(x|y)}{r(x)},$$

where the maximization is taken over all $\mathbf{q}$ such that

$$q(x|y) = 0 \quad \text{if and only if} \quad p(y|x) = 0, \qquad (1)$$

and

$$q^*(x|y) = \frac{r(x)p(y|x)}{\sum_{x'} r(x')p(y|x')}. \qquad (2)$$

Proof

1. In (2), let

$$w(y) = \sum_{x'} r(x')p(y|x').$$

2. Assume w.l.o.g. that for all $y \in \mathcal{Y}$, $p(y|x) > 0$ for some $x \in \mathcal{X}$.

3. Since $\mathbf{r} > 0$, $w(y) > 0$ for all y, and hence $q^*(x|y)$ is well-defined.

4. Rearranging (2), we have

$$r(x)p(y|x) = w(y)q^*(x|y).$$

5. For any $\mathbf{q}$ satisfying (1), consider

$$\sum_x \sum_y r(x)p(y|x) \log \frac{q^*(x|y)}{\cancel{r(x)}} - \sum_x \sum_y r(x)p(y|x) \log \frac{q(x|y)}{\cancel{r(x)}}$$

Lemma 9.1 Let $r(x)p(y|x)$ be a given joint distribution on $\mathcal{X} \times \mathcal{Y}$ such that $\mathbf{r} > 0$, and let $\mathbf{q}$ be a transition matrix from $\mathcal{Y}$ to $\mathcal{X}$. Then

$$\max_{\mathbf{q}} \sum_x \sum_y r(x)p(y|x) \log \frac{q(x|y)}{r(x)} = \sum_x \sum_y r(x)p(y|x) \log \frac{q^*(x|y)}{r(x)},$$

where the maximization is taken over all $\mathbf{q}$ such that

$$q(x|y) = 0 \quad \text{if and only if} \quad p(y|x) = 0, \tag{1}$$

and

$$q^*(x|y) = \frac{r(x)p(y|x)}{\sum_{x'} r(x')p(y|x')}. \tag{2}$$

Proof

1. In (2), let

$$w(y) = \sum_{x'} r(x')p(y|x').$$

2. Assume w.l.o.g. that for all $y \in \mathcal{Y}$, $p(y|x) > 0$ for some $x \in \mathcal{X}$.

3. Since $\mathbf{r} > 0$, $w(y) > 0$ for all y, and hence $q^*(x|y)$ is well-defined.

4. Rearranging (2), we have

$$r(x)p(y|x) = w(y)q^*(x|y).$$

5. For any $\mathbf{q}$ satisfying (1), consider

$$\sum_x \sum_y r(x)p(y|x) \log \frac{q^*(x|y)}{\cancel{r(x)}} - \sum_x \sum_y r(x)p(y|x) \log \frac{q(x|y)}{\cancel{r(x)}}$$
$$= \sum_x \sum_y r(x)p(y|x) \log \frac{q^*(x|y)}{q(x|y)}$$

Lemma 9.1 Let $r(x)p(y|x)$ be a given joint distribution on $\mathcal{X} \times \mathcal{Y}$ such that $\mathbf{r} > 0$, and let $\mathbf{q}$ be a transition matrix from $\mathcal{Y}$ to $\mathcal{X}$. Then

$$\begin{aligned}
\max_{\mathbf{q}} \sum_x \sum_y r(x)p(y|x) \log \frac{q(x|y)}{r(x)} \\
= \quad \sum_x \sum_y r(x)p(y|x) \log \frac{q^*(x|y)}{r(x)},
\end{aligned}$$

where the maximization is taken over all $\mathbf{q}$ such that

$$q(x|y) = 0 \quad \text{if and only if} \quad p(y|x) = 0, \tag{1}$$

and

$$q^*(x|y) = \frac{r(x)p(y|x)}{\sum_{x'} r(x')p(y|x')}. \tag{2}$$

Proof

1. In (2), let

$$w(y) = \sum_{x'} r(x')p(y|x').$$

2. Assume w.l.o.g. that for all $y \in \mathcal{Y}$, $p(y|x) > 0$ for some $x \in \mathcal{X}$.

3. Since $\mathbf{r} > 0$, $w(y) > 0$ for all y, and hence $q^*(x|y)$ is well-defined.

4. Rearranging (2), we have

$$r(x)p(y|x) = w(y)q^*(x|y).$$

5. For any $\mathbf{q}$ satisfying (1), consider

$$\begin{aligned}
\sum_x \sum_y r(x)p(y|x) \log \frac{q^*(x|y)}{\cancel{r(x)}} - \sum_x \sum_y r(x)p(y|x) \log \frac{q(x|y)}{\cancel{r(x)}} \\
= \quad \sum_x \sum_y r(x)p(y|x) \log \frac{q^*(x|y)}{q(x|y)}
\end{aligned}$$

Lemma 9.1 Let $r(x)p(y|x)$ be a given joint distribution on $\mathcal{X} \times \mathcal{Y}$ such that $\mathbf{r} > 0$, and let $\mathbf{q}$ be a transition matrix from $\mathcal{Y}$ to $\mathcal{X}$. Then

$$\max_{\mathbf{q}} \sum_x \sum_y r(x)p(y|x) \log \frac{q(x|y)}{r(x)} = \sum_x \sum_y r(x)p(y|x) \log \frac{q^*(x|y)}{r(x)},$$

where the maximization is taken over all $\mathbf{q}$ such that

$$q(x|y) = 0 \quad \text{if and only if} \quad p(y|x) = 0, \tag{1}$$

and

$$q^*(x|y) = \frac{r(x)p(y|x)}{\sum_{x'} r(x')p(y|x')}. \tag{2}$$

Proof

1. In (2), let

$$w(y) = \sum_{x'} r(x')p(y|x').$$

2. Assume w.l.o.g. that for all $y \in \mathcal{Y}$, $p(y|x) > 0$ for some $x \in \mathcal{X}$.

3. Since $\mathbf{r} > 0$, $w(y) > 0$ for all y, and hence $q^*(x|y)$ is well-defined.

4. Rearranging (2), we have

$$r(x)p(y|x) = w(y)q^*(x|y).$$

5. For any $\mathbf{q}$ satisfying (1), consider

$$\begin{aligned} &\sum_x \sum_y r(x)p(y|x) \log \frac{q^*(x|y)}{\cancel{r(x)}} - \sum_x \sum_y r(x)p(y|x) \log \frac{q(x|y)}{\cancel{r(x)}} \\ = &\sum_x \sum_y \underline{r(x)p(y|x)} \log \frac{q^*(x|y)}{q(x|y)} \end{aligned}$$

Lemma 9.1 Let $r(x)p(y|x)$ be a given joint distribution on $\mathcal{X} \times \mathcal{Y}$ such that $\mathbf{r} > 0$, and let $\mathbf{q}$ be a transition matrix from $\mathcal{Y}$ to $\mathcal{X}$. Then

$$\max_{\mathbf{q}} \sum_x \sum_y r(x)p(y|x) \log \frac{q(x|y)}{r(x)} = \sum_x \sum_y r(x)p(y|x) \log \frac{q^*(x|y)}{r(x)},$$

where the maximization is taken over all $\mathbf{q}$ such that

$$q(x|y) = 0 \quad \text{if and only if} \quad p(y|x) = 0, \tag{1}$$

and

$$q^*(x|y) = \frac{r(x)p(y|x)}{\sum_{x'} r(x')p(y|x')}. \tag{2}$$

Proof

1. In (2), let

$$w(y) = \sum_{x'} r(x')p(y|x').$$

2. Assume w.l.o.g. that for all $y \in \mathcal{Y}$, $p(y|x) > 0$ for some $x \in \mathcal{X}$.

3. Since $\mathbf{r} > 0$, $w(y) > 0$ for all y, and hence $q^*(x|y)$ is well-defined.

4. Rearranging (2), we have

$$r(x)p(y|x) = w(y)q^*(x|y).$$

5. For any $\mathbf{q}$ satisfying (1), consider

$$\begin{aligned}
&\sum_x \sum_y r(x)p(y|x) \log \frac{q^*(x|y)}{\cancel{r(x)}} - \sum_x \sum_y r(x)p(y|x) \log \frac{q(x|y)}{\cancel{r(x)}} \\
&= \sum_x \sum_y \underline{r(x)p(y|x)} \log \frac{q^*(x|y)}{q(x|y)} \\
&= \sum_y \sum_x \underline{w(y)q^*(x|y)} \log \frac{q^*(x|y)}{q(x|y)}
\end{aligned}$$

Lemma 9.1 Let $r(x)p(y|x)$ be a given joint distribution on $\mathcal{X} \times \mathcal{Y}$ such that $\mathbf{r} > 0$, and let $\mathbf{q}$ be a transition matrix from $\mathcal{Y}$ to $\mathcal{X}$. Then

$$\max_{\mathbf{q}} \sum_x \sum_y r(x)p(y|x) \log \frac{q(x|y)}{r(x)} = \sum_x \sum_y r(x)p(y|x) \log \frac{q^*(x|y)}{r(x)},$$

where the maximization is taken over all $\mathbf{q}$ such that

$$q(x|y) = 0 \quad \text{if and only if} \quad p(y|x) = 0, \qquad (1)$$

and

$$q^*(x|y) = \frac{r(x)p(y|x)}{\sum_{x'} r(x')p(y|x')}. \qquad (2)$$

Proof

1. In (2), let

$$w(y) = \sum_{x'} r(x')p(y|x').$$

2. Assume w.l.o.g. that for all $y \in \mathcal{Y}$, $p(y|x) > 0$ for some $x \in \mathcal{X}$.

3. Since $\mathbf{r} > 0$, $w(y) > 0$ for all y, and hence $q^*(x|y)$ is well-defined.

4. Rearranging (2), we have

$$r(x)p(y|x) = w(y)q^*(x|y).$$

5. For any $\mathbf{q}$ satisfying (1), consider

$$\begin{aligned}
&\sum_x \sum_y r(x)p(y|x) \log \frac{q^*(x|y)}{r(x)} - \sum_x \sum_y r(x)p(y|x) \log \frac{q(x|y)}{r(x)} \\
&= \sum_x \sum_y r(x)p(y|x) \log \frac{q^*(x|y)}{q(x|y)} \\
&= \sum_y \sum_x w(y)q^*(x|y) \log \frac{q^*(x|y)}{q(x|y)}
\end{aligned}$$

Lemma 9.1 Let $r(x)p(y|x)$ be a given joint distribution on $\mathcal{X} \times \mathcal{Y}$ such that $\mathbf{r} > 0$, and let $\mathbf{q}$ be a transition matrix from $\mathcal{Y}$ to $\mathcal{X}$. Then

$$\begin{aligned}\max_{\mathbf{q}} \sum_x \sum_y r(x)p(y|x) \log \frac{q(x|y)}{r(x)} \\ = \quad \sum_x \sum_y r(x)p(y|x) \log \frac{q^*(x|y)}{r(x)},\end{aligned}$$

where the maximization is taken over all $\mathbf{q}$ such that

$$q(x|y) = 0 \quad \text{if and only if} \quad p(y|x) = 0, \qquad (1)$$

and

$$q^*(x|y) = \frac{r(x)p(y|x)}{\sum_{x'} r(x')p(y|x')}. \qquad (2)$$

Proof

1. In (2), let

$$w(y) = \sum_{x'} r(x')p(y|x').$$

2. Assume w.l.o.g. that for all $y \in \mathcal{Y}$, $p(y|x) > 0$ for some $x \in \mathcal{X}$.

3. Since $\mathbf{r} > 0$, $w(y) > 0$ for all y, and hence $q^*(x|y)$ is well-defined.

4. Rearranging (2), we have

$$r(x)p(y|x) = w(y)q^*(x|y).$$

5. For any $\mathbf{q}$ satisfying (1), consider

$$\begin{aligned}&\sum_x \sum_y r(x)p(y|x) \log \frac{q^*(x|y)}{r(x)} - \sum_x \sum_y r(x)p(y|x) \log \frac{q(x|y)}{r(x)} \\ &= \sum_x \sum_y r(x)p(y|x) \log \frac{q^*(x|y)}{q(x|y)} \\ &= \sum_y \sum_x \underline{w(y)}q^*(x|y) \log \frac{q^*(x|y)}{q(x|y)} \\ &= \sum_y \underline{w(y)} \sum_x q^*(x|y) \log \frac{q^*(x|y)}{q(x|y)}\end{aligned}$$

Lemma 9.1 Let $r(x)p(y|x)$ be a given joint distribution on $\mathcal{X} \times \mathcal{Y}$ such that $\mathbf{r} > 0$, and let $\mathbf{q}$ be a transition matrix from $\mathcal{Y}$ to $\mathcal{X}$. Then

$$\max_{\mathbf{q}} \sum_x \sum_y r(x)p(y|x) \log \frac{q(x|y)}{r(x)} = \sum_x \sum_y r(x)p(y|x) \log \frac{q^*(x|y)}{r(x)},$$

where the maximization is taken over all $\mathbf{q}$ such that

$$q(x|y) = 0 \quad \text{if and only if} \quad p(y|x) = 0, \tag{1}$$

and

$$q^*(x|y) = \frac{r(x)p(y|x)}{\sum_{x'} r(x')p(y|x')}. \tag{2}$$

Proof

1. In (2), let

$$w(y) = \sum_{x'} r(x')p(y|x').$$

2. Assume w.l.o.g. that for all $y \in \mathcal{Y}$, $p(y|x) > 0$ for some $x \in \mathcal{X}$.

3. Since $\mathbf{r} > 0$, $w(y) > 0$ for all y, and hence $q^*(x|y)$ is well-defined.

4. Rearranging (2), we have

$$r(x)p(y|x) = w(y)q^*(x|y).$$

5. For any $\mathbf{q}$ satisfying (1), consider

$$\begin{aligned}
&\sum_x \sum_y r(x)p(y|x) \log \frac{q^*(x|y)}{r(x)} - \sum_x \sum_y r(x)p(y|x) \log \frac{q(x|y)}{r(x)} \\
&= \sum_x \sum_y r(x)p(y|x) \log \frac{q^*(x|y)}{q(x|y)} \\
&= \sum_y \sum_x w(y)q^*(x|y) \log \frac{q^*(x|y)}{q(x|y)} \\
&= \sum_y w(y) \underline{\sum_x q^*(x|y) \log \frac{q^*(x|y)}{q(x|y)}}
\end{aligned}$$

Lemma 9.1 Let $r(x)p(y|x)$ be a given joint distribution on $\mathcal{X} \times \mathcal{Y}$ such that $\mathbf{r} > 0$, and let $\mathbf{q}$ be a transition matrix from $\mathcal{Y}$ to $\mathcal{X}$. Then

$$\max_{\mathbf{q}} \sum_x \sum_y r(x)p(y|x) \log \frac{q(x|y)}{r(x)} = \sum_x \sum_y r(x)p(y|x) \log \frac{q^*(x|y)}{r(x)},$$

where the maximization is taken over all $\mathbf{q}$ such that

$$q(x|y) = 0 \quad \text{if and only if} \quad p(y|x) = 0, \qquad (1)$$

and

$$q^*(x|y) = \frac{r(x)p(y|x)}{\sum_{x'} r(x')p(y|x')}. \qquad (2)$$

Proof

1. In (2), let

$$w(y) = \sum_{x'} r(x')p(y|x').$$

2. Assume w.l.o.g. that for all $y \in \mathcal{Y}$, $p(y|x) > 0$ for some $x \in \mathcal{X}$.

3. Since $\mathbf{r} > 0$, $w(y) > 0$ for all y, and hence $q^*(x|y)$ is well-defined.

4. Rearranging (2), we have

$$r(x)p(y|x) = w(y)q^*(x|y).$$

5. For any $\mathbf{q}$ satisfying (1), consider

$$\begin{aligned}
&\sum_x \sum_y r(x)p(y|x) \log \frac{q^*(x|y)}{r(x)} - \sum_x \sum_y r(x)p(y|x) \log \frac{q(x|y)}{r(x)} \\
&= \sum_x \sum_y r(x)p(y|x) \log \frac{q^*(x|y)}{q(x|y)} \\
&= \sum_y \sum_x w(y)q^*(x|y) \log \frac{q^*(x|y)}{q(x|y)} \\
&= \sum_y w(y) \underline{\sum_x q^*(x|y) \log \frac{q^*(x|y)}{q(x|y)}} \\
&= \sum_y w(y) \underline{D(q^*(x|y) \| q(x|y))}
\end{aligned}$$

Lemma 9.1 Let $r(x)p(y|x)$ be a given joint distribution on $\mathcal{X} \times \mathcal{Y}$ such that $\mathbf{r} > 0$, and let $\mathbf{q}$ be a transition matrix from $\mathcal{Y}$ to $\mathcal{X}$. Then

$$\begin{aligned}
&\max_{\mathbf{q}} \sum_x \sum_y r(x)p(y|x) \log \frac{q(x|y)}{r(x)} \\
&= \sum_x \sum_y r(x)p(y|x) \log \frac{q^*(x|y)}{r(x)},
\end{aligned}$$

where the maximization is taken over all $\mathbf{q}$ such that

$$q(x|y) = 0 \quad \text{if and only if} \quad p(y|x) = 0, \qquad (1)$$

and

$$q^*(x|y) = \frac{r(x)p(y|x)}{\sum_{x'} r(x')p(y|x')}. \qquad (2)$$

Proof

1. In (2), let

$$w(y) = \sum_{x'} r(x')p(y|x').$$

2. Assume w.l.o.g. that for all $y \in \mathcal{Y}$, $p(y|x) > 0$ for some $x \in \mathcal{X}$.

3. Since $\mathbf{r} > 0$, $w(y) > 0$ for all y, and hence $q^*(x|y)$ is well-defined.

4. Rearranging (2), we have

$$r(x)p(y|x) = w(y)q^*(x|y).$$

5. For any $\mathbf{q}$ satisfying (1), consider

$$\begin{aligned}
&\sum_x \sum_y r(x)p(y|x) \log \frac{q^*(x|y)}{r(x)} - \sum_x \sum_y r(x)p(y|x) \log \frac{q(x|y)}{r(x)} \\
&= \sum_x \sum_y r(x)p(y|x) \log \frac{q^*(x|y)}{q(x|y)} \\
&= \sum_y \sum_x w(y)q^*(x|y) \log \frac{q^*(x|y)}{q(x|y)} \\
&= \sum_y w(y) \sum_x q^*(x|y) \log \frac{q^*(x|y)}{q(x|y)} \\
&= \sum_y w(y) \underline{D(q^*(x|y) \| q(x|y))} \\
&\geq 0.
\end{aligned}$$

Lemma 9.1 Let $r(x)p(y|x)$ be a given joint distribution on $\mathcal{X} \times \mathcal{Y}$ such that $\mathbf{r} > 0$, and let $\mathbf{q}$ be a transition matrix from $\mathcal{Y}$ to $\mathcal{X}$. Then

$$\begin{aligned}\max_{\mathbf{q}} \sum_x \sum_y r(x)p(y|x) \log \frac{q(x|y)}{r(x)} \\ = \quad \sum_x \sum_y r(x)p(y|x) \log \frac{q^*(x|y)}{r(x)},\end{aligned}$$

where the maximization is taken over all $\mathbf{q}$ such that

$$q(x|y) = 0 \quad \text{if and only if} \quad p(y|x) = 0, \qquad (1)$$

and

$$q^*(x|y) = \frac{r(x)p(y|x)}{\sum_{x'} r(x')p(y|x')}. \qquad (2)$$

Proof

1. In (2), let

$$w(y) = \sum_{x'} r(x')p(y|x').$$

2. Assume w.l.o.g. that for all $y \in \mathcal{Y}$, $p(y|x) > 0$ for some $x \in \mathcal{X}$.

3. Since $\mathbf{r} > 0$, $w(y) > 0$ for all y, and hence $q^*(x|y)$ is well-defined.

4. Rearranging (2), we have

$$r(x)p(y|x) = w(y)q^*(x|y).$$

5. For any $\mathbf{q}$ satisfying (1), consider

$$\begin{aligned}&\sum_x \sum_y r(x)p(y|x) \log \frac{q^*(x|y)}{r(x)} - \sum_x \sum_y r(x)p(y|x) \log \frac{q(x|y)}{r(x)} \\
&= \quad \sum_x \sum_y r(x)p(y|x) \log \frac{q^*(x|y)}{q(x|y)} \\
&= \quad \sum_y \sum_x w(y)q^*(x|y) \log \frac{q^*(x|y)}{q(x|y)} \\
&= \quad \sum_y w(y) \sum_x q^*(x|y) \log \frac{q^*(x|y)}{q(x|y)} \\
&= \quad \sum_y w(y) D(q^*(x|y) \| q(x|y)) \\
&\geq \quad 0.\end{aligned}$$

Lemma 9.1 Let $r(x)p(y|x)$ be a given joint distribution on $\mathcal{X} \times \mathcal{Y}$ such that $\mathbf{r} > 0$, and let $\mathbf{q}$ be a transition matrix from $\mathcal{Y}$ to $\mathcal{X}$. Then

$$\begin{aligned}\max_{\mathbf{q}} \sum_x \sum_y r(x)p(y|x) \log \frac{q(x|y)}{r(x)} \\ = \quad \sum_x \sum_y r(x)p(y|x) \log \frac{q^*(x|y)}{r(x)},\end{aligned}$$

where the maximization is taken over all $\mathbf{q}$ such that

$$q(x|y) = 0 \quad \text{if and only if} \quad p(y|x) = 0, \tag{1}$$

and

$$q^*(x|y) = \frac{r(x)p(y|x)}{\sum_{x'} r(x')p(y|x')}. \tag{2}$$

Proof

1. In (2), let

$$w(y) = \sum_{x'} r(x')p(y|x').$$

2. Assume w.l.o.g. that for all $y \in \mathcal{Y}$, $p(y|x) > 0$ for some $x \in \mathcal{X}$.

3. Since $\mathbf{r} > 0$, $w(y) > 0$ for all y, and hence $q^*(x|y)$ is well-defined.

4. Rearranging (2), we have

$$r(x)p(y|x) = w(y)q^*(x|y).$$

5. For any $\mathbf{q}$ satisfying (1), consider

$$\begin{aligned}\sum_x \sum_y r(x)p(y|x) \log \frac{q^*(x|y)}{r(x)} - \sum_x \sum_y r(x)p(y|x) \log \frac{q(x|y)}{r(x)} \\ = \quad & \sum_x \sum_y r(x)p(y|x) \log \frac{q^*(x|y)}{q(x|y)} \\ = \quad & \sum_y \sum_x w(y)q^*(x|y) \log \frac{q^*(x|y)}{q(x|y)} \\ = \quad & \sum_y w(y) \sum_x q^*(x|y) \log \frac{q^*(x|y)}{q(x|y)} \\ = \quad & \sum_y w(y) D(q^*(x|y) \| q(x|y)) \\ \geq \quad & 0.\end{aligned}$$

6. The proof is completed by noting in (2) that $\mathbf{q}^*$ satisfies (1) because $\mathbf{r} > 0$.

Lemma 9.1 Let $r(x)p(y|x)$ be a given joint distribution on $\mathcal{X} \times \mathcal{Y}$ such that $\mathbf{r} > 0$, and let $\mathbf{q}$ be a transition matrix from $\mathcal{Y}$ to $\mathcal{X}$. Then

$$\max_{\mathbf{q}} \sum_x \sum_y r(x)p(y|x) \log \frac{q(x|y)}{r(x)} = \sum_x \sum_y r(x)p(y|x) \log \frac{q^*(x|y)}{r(x)},$$

where the maximization is taken over all $\mathbf{q}$ such that

$$q(x|y) = 0 \quad \text{if and only if} \quad p(y|x) = 0, \tag{1}$$

and

$$q^*(x|y) = \frac{r(x)p(y|x)}{\cancel{\sum_{x'} r(x')p(y|x')}}. \tag{2}$$

$w(y) > 0$

Proof

1. In (2), let

$$w(y) = \sum_{x'} r(x')p(y|x').$$

2. Assume w.l.o.g. that for all $y \in \mathcal{Y}$, $p(y|x) > 0$ for some $x \in \mathcal{X}$.

3. Since $\mathbf{r} > 0$, $w(y) > 0$ for all y, and hence $q^*(x|y)$ is well-defined.

4. Rearranging (2), we have

$$r(x)p(y|x) = w(y)q^*(x|y).$$

5. For any $\mathbf{q}$ satisfying (1), consider

$$\begin{aligned}
&\sum_x \sum_y r(x)p(y|x) \log \frac{q^*(x|y)}{r(x)} - \sum_x \sum_y r(x)p(y|x) \log \frac{q(x|y)}{r(x)} \\
&= \sum_x \sum_y r(x)p(y|x) \log \frac{q^*(x|y)}{q(x|y)} \\
&= \sum_y \sum_x w(y)q^*(x|y) \log \frac{q^*(x|y)}{q(x|y)} \\
&= \sum_y w(y) \sum_x q^*(x|y) \log \frac{q^*(x|y)}{q(x|y)} \\
&= \sum_y w(y) D(q^*(x|y) \| q(x|y)) \\
&\geq 0.
\end{aligned}$$

6. The proof is completed by noting in (2) that $\mathbf{q}^*$ satisfies (1) because $\mathbf{r} > 0$.

Lemma 9.1 Let $r(x)p(y|x)$ be a given joint distribution on $\mathcal{X} \times \mathcal{Y}$ such that $\mathbf{r} > 0$, and let $\mathbf{q}$ be a transition matrix from $\mathcal{Y}$ to $\mathcal{X}$. Then

$$
\begin{aligned}
\max_{\mathbf{q}} \sum_x \sum_y r(x)p(y|x) \log \frac{q(x|y)}{r(x)} \\
= \quad \sum_x \sum_y r(x)p(y|x) \log \frac{q^*(x|y)}{r(x)},
\end{aligned}
$$

where the maximization is taken over all $\mathbf{q}$ such that

$$
q(x|y) = 0 \quad \text{if and only if} \quad p(y|x) = 0, \tag{1}
$$

and

> 0

$$
q^*(x|y) = \frac{r(x)p(y|x)}{\cancel{\sum_{x'} r(x')p(y|x')}}. \tag{2}
$$

$w(y) > 0$

Proof

1. In (2), let

$$
w(y) = \sum_{x'} r(x')p(y|x').
$$

2. Assume w.l.o.g. that for all $y \in \mathcal{Y}$, $p(y|x) > 0$ for some $x \in \mathcal{X}$.

3. Since $\mathbf{r} > 0$, $w(y) > 0$ for all y, and hence $q^*(x|y)$ is well-defined.

4. Rearranging (2), we have

$$
r(x)p(y|x) = w(y)q^*(x|y).
$$

5. For any $\mathbf{q}$ satisfying (1), consider

$$
\begin{aligned}
&\sum_x \sum_y r(x)p(y|x) \log \frac{q^*(x|y)}{r(x)} - \sum_x \sum_y r(x)p(y|x) \log \frac{q(x|y)}{r(x)} \\
&= \quad \sum_x \sum_y r(x)p(y|x) \log \frac{q^*(x|y)}{q(x|y)} \\
&= \quad \sum_y \sum_x w(y)q^*(x|y) \log \frac{q^*(x|y)}{q(x|y)} \\
&= \quad \sum_y w(y) \sum_x q^*(x|y) \log \frac{q^*(x|y)}{q(x|y)} \\
&= \quad \sum_y w(y) D(q^*(x|y) \| q(x|y)) \\
&\geq \quad 0.
\end{aligned}
$$

6. The proof is completed by noting in (2) that $\mathbf{q}^*$ satisfies (1) because $\mathbf{r} > 0$.

Lemma 9.1 Let $r(x)p(y|x)$ be a given joint distribution on $\mathcal{X} \times \mathcal{Y}$ such that $\mathbf{r} > 0$, and let $\mathbf{q}$ be a transition matrix from $\mathcal{Y}$ to $\mathcal{X}$. Then

$$\begin{aligned}
&\max_{\mathbf{q}} \sum_x \sum_y r(x)p(y|x) \log \frac{q(x|y)}{r(x)} \\
&= \quad \sum_x \sum_y r(x)p(y|x) \log \frac{q^*(x|y)}{r(x)},
\end{aligned}$$

where the maximization is taken over all $\mathbf{q}$ such that

$$q(x|y) = 0 \quad \text{if and only if} \quad p(y|x) = 0, \tag{1}$$

> 0

and

$$\underline{q^*(x|y)} = \frac{r(x)p(y|x)}{\cancel{\sum_{x'} r(x')p(y|x')}}. \tag{2}$$

$w(y) > 0$

Proof

1. In (2), let

$$w(y) = \sum_{x'} r(x')p(y|x').$$

2. Assume w.l.o.g. that for all $y \in \mathcal{Y}$, $p(y|x) > 0$ for some $x \in \mathcal{X}$.

3. Since $\mathbf{r} > 0$, $w(y) > 0$ for all y, and hence $q^*(x|y)$ is well-defined.

4. Rearranging (2), we have

$$r(x)p(y|x) = w(y)q^*(x|y).$$

5. For any $\mathbf{q}$ satisfying (1), consider

$$\begin{aligned}
&\sum_x \sum_y r(x)p(y|x) \log \frac{q^*(x|y)}{r(x)} - \sum_x \sum_y r(x)p(y|x) \log \frac{q(x|y)}{r(x)} \\
&= \quad \sum_x \sum_y r(x)p(y|x) \log \frac{q^*(x|y)}{q(x|y)} \\
&= \quad \sum_y \sum_x w(y)q^*(x|y) \log \frac{q^*(x|y)}{q(x|y)} \\
&= \quad \sum_y w(y) \sum_x q^*(x|y) \log \frac{q^*(x|y)}{q(x|y)} \\
&= \quad \sum_y w(y) D(q^*(x|y) \| q(x|y)) \\
&\geq \quad 0.
\end{aligned}$$

6. The proof is completed by noting in (2) that $\mathbf{q}^*$ satisfies (1) because $\mathbf{r} > 0$.

Lemma 9.1 Let $r(x)p(y|x)$ be a given joint distribution on $\mathcal{X} \times \mathcal{Y}$ such that $\mathbf{r} > 0$, and let $\mathbf{q}$ be a transition matrix from $\mathcal{Y}$ to $\mathcal{X}$. Then

$$\max_{\mathbf{q}} \sum_x \sum_y r(x)p(y|x) \log \frac{q(x|y)}{r(x)} = \sum_x \sum_y r(x)p(y|x) \log \frac{q^*(x|y)}{r(x)},$$

where the maximization is taken over all $\mathbf{q}$ such that

$$q(x|y) = 0 \quad \text{if and only if} \quad p(y|x) = 0, \tag{1}$$

and

$$\underline{q^*(x|y)} = \frac{\overset{>0}{r(x)}\,\underline{p(y|x)}}{\underset{w(y)>0}{\cancel{\sum_{x'} r(x')p(y|x')}}}. \tag{2}$$

Proof

1. In (2), let

$$w(y) = \sum_{x'} r(x')p(y|x').$$

2. Assume w.l.o.g. that for all $y \in \mathcal{Y}$, $p(y|x) > 0$ for some $x \in \mathcal{X}$.

3. Since $\mathbf{r} > 0$, $w(y) > 0$ for all y, and hence $q^*(x|y)$ is well-defined.

4. Rearranging (2), we have

$$r(x)p(y|x) = w(y)q^*(x|y).$$

5. For any $\mathbf{q}$ satisfying (1), consider

$$\begin{aligned}
&\sum_x \sum_y r(x)p(y|x) \log \frac{q^*(x|y)}{r(x)} - \sum_x \sum_y r(x)p(y|x) \log \frac{q(x|y)}{r(x)} \\
&= \sum_x \sum_y r(x)p(y|x) \log \frac{q^*(x|y)}{q(x|y)} \\
&= \sum_y \sum_x w(y)q^*(x|y) \log \frac{q^*(x|y)}{q(x|y)} \\
&= \sum_y w(y) \sum_x q^*(x|y) \log \frac{q^*(x|y)}{q(x|y)} \\
&= \sum_y w(y) D(q^*(x|y) \| q(x|y)) \\
&\geq 0.
\end{aligned}$$

6. The proof is completed by noting in (2) that $\mathbf{q}^*$ satisfies (1) because $\mathbf{r} > 0$.

Lemma 9.1 Let $r(x)p(y|x)$ be a given joint distribution on $\mathcal{X} \times \mathcal{Y}$ such that $\mathbf{r} > 0$, and let $\mathbf{q}$ be a transition matrix from $\mathcal{Y}$ to $\mathcal{X}$. Then

$$\begin{aligned}\max_{\mathbf{q}} \sum_x \sum_y r(x)p(y|x) \log \frac{q(x|y)}{r(x)} \\ = \quad \sum_x \sum_y r(x)p(y|x) \log \frac{q^*(x|y)}{r(x)},\end{aligned}$$

where the maximization is taken over all $\mathbf{q}$ such that

$$q(x|y) = 0 \quad \text{if and only if} \quad p(y|x) = 0, \tag{1}$$

and

$$q^*(x|y) = \frac{\overset{>0}{r(x)}\,p(y|x)}{\underbrace{\sum_{x'} r(x')p(y|x')}_{w(y)\,>\,0}}. \tag{2}$$

Proof

1. In (2), let

$$w(y) = \sum_{x'} r(x')p(y|x').$$

2. Assume w.l.o.g. that for all $y \in \mathcal{Y}$, $p(y|x) > 0$ for some $x \in \mathcal{X}$.

3. Since $\mathbf{r} > 0$, $w(y) > 0$ for all y, and hence $q^*(x|y)$ is well-defined.

4. Rearranging (2), we have

$$r(x)p(y|x) = w(y)q^*(x|y).$$

5. For any $\mathbf{q}$ satisfying (1), consider

$$\begin{aligned}
&\sum_x \sum_y r(x)p(y|x) \log \frac{q^*(x|y)}{r(x)} - \sum_x \sum_y r(x)p(y|x) \log \frac{q(x|y)}{r(x)} \\
&= \sum_x \sum_y r(x)p(y|x) \log \frac{q^*(x|y)}{q(x|y)} \\
&= \sum_y \sum_x w(y)q^*(x|y) \log \frac{q^*(x|y)}{q(x|y)} \\
&= \sum_y w(y) \sum_x q^*(x|y) \log \frac{q^*(x|y)}{q(x|y)} \\
&= \sum_y w(y) D(q^*(x|y) \| q(x|y)) \\
&\geq 0.
\end{aligned}$$

6. The proof is completed by noting in (2) that $\mathbf{q}^*$ satisfies (1) because $\mathbf{r} > 0$.

Lemma 9.1 Let $r(x)p(y|x)$ be a given joint distribution on $\mathcal{X} \times \mathcal{Y}$ such that $\mathbf{r} > 0$, and let $\mathbf{q}$ be a transition matrix from $\mathcal{Y}$ to $\mathcal{X}$. Then

$$\max_{\mathbf{q}} \sum_x \sum_y r(x)p(y|x) \log \frac{q(x|y)}{r(x)} = \sum_x \sum_y r(x)p(y|x) \log \frac{q^*(x|y)}{r(x)},$$

where the maximization is taken over all $\mathbf{q}$ such that

$$q(x|y) = 0 \quad \text{if and only if} \quad p(y|x) = 0, \tag{1}$$

and

$$q^*(x|y) = \frac{r(x)p(y|x)}{\sum_{x'} r(x')p(y|x')}. \tag{2}$$

Proof

1. In (2), let

$$w(y) = \sum_{x'} r(x')p(y|x').$$

2. Assume w.l.o.g. that for all $y \in \mathcal{Y}$, $p(y|x) > 0$ for some $x \in \mathcal{X}$.

3. Since $\mathbf{r} > 0$, $w(y) > 0$ for all y, and hence $q^*(x|y)$ is well-defined.

4. Rearranging (2), we have

$$r(x)p(y|x) = w(y)q^*(x|y).$$

5. For any $\mathbf{q}$ satisfying (1), consider

$$\begin{aligned}
&\sum_x \sum_y r(x)p(y|x) \log \frac{q^*(x|y)}{r(x)} - \sum_x \sum_y r(x)p(y|x) \log \frac{q(x|y)}{r(x)} \\
&= \sum_x \sum_y r(x)p(y|x) \log \frac{q^*(x|y)}{q(x|y)} \\
&= \sum_y \sum_x w(y)q^*(x|y) \log \frac{q^*(x|y)}{q(x|y)} \\
&= \sum_y w(y) \sum_x q^*(x|y) \log \frac{q^*(x|y)}{q(x|y)} \\
&= \sum_y w(y) D(q^*(x|y) \| q(x|y)) \\
&\geq 0.
\end{aligned}$$

6. The proof is completed by noting in (2) that $\mathbf{q}^*$ satisfies (1) because $\mathbf{r} > 0$.

Lemma 9.1 Let $r(x)p(y|x)$ be a given joint distribution on $\mathcal{X} \times \mathcal{Y}$ such that $\mathbf{r} > 0$, and let $\mathbf{q}$ be a transition matrix from $\mathcal{Y}$ to $\mathcal{X}$. Then

$$\begin{aligned} \max_{\mathbf{q}} \sum_x \sum_y r(x)p(y|x) \log \frac{q(x|y)}{r(x)} \\ = \quad \sum_x \sum_y r(x)p(y|x) \log \frac{q^*(x|y)}{r(x)}, \end{aligned}$$

where the maximization is taken over all $\mathbf{q}$ such that

$$q(x|y) = 0 \quad \text{if and only if} \quad p(y|x) = 0, \tag{1}$$

and

$$q^*(x|y) = \frac{r(x)p(y|x)}{\sum_{x'} r(x')p(y|x')}. \tag{2}$$

Proof

1. In (2), let

$$w(y) = \sum_{x'} r(x')p(y|x').$$

2. Assume w.l.o.g. that for all $y \in \mathcal{Y}$, $p(y|x) > 0$ for some $x \in \mathcal{X}$.

3. Since $\mathbf{r} > 0$, $w(y) > 0$ for all y, and hence $q^*(x|y)$ is well-defined.

4. Rearranging (2), we have

$$r(x)p(y|x) = w(y)q^*(x|y).$$

5. For any $\mathbf{q}$ satisfying (1), consider

$$\begin{aligned} & \sum_x \sum_y r(x)p(y|x) \log \frac{q^*(x|y)}{r(x)} - \sum_x \sum_y r(x)p(y|x) \log \frac{q(x|y)}{r(x)} \\ = & \quad \sum_x \sum_y r(x)p(y|x) \log \frac{q^*(x|y)}{q(x|y)} \\ = & \quad \sum_y \sum_x w(y)q^*(x|y) \log \frac{q^*(x|y)}{q(x|y)} \\ = & \quad \sum_y w(y) \sum_x q^*(x|y) \log \frac{q^*(x|y)}{q(x|y)} \\ = & \quad \sum_y w(y) D(q^*(x|y) \| q(x|y)) \\ \geq & \quad 0. \end{aligned}$$

6. The proof is completed by noting in (2) that $\mathbf{q}^*$ satisfies (1) because $\mathbf{r} > 0$.

Lemma 9.1 Let $r(x)p(y|x)$ be a given joint distribution on $\mathcal{X} \times \mathcal{Y}$ such that $\mathbf{r} > 0$, and let $\mathbf{q}$ be a transition matrix from $\mathcal{Y}$ to $\mathcal{X}$. Then

$$\max_{\mathbf{q}} \sum_x \sum_y r(x)p(y|x) \log \frac{q(x|y)}{r(x)} = \sum_x \sum_y r(x)p(y|x) \log \frac{q^*(x|y)}{r(x)},$$

where the maximization is taken over all $\mathbf{q}$ such that

$$q(x|y) = 0 \quad \text{if and only if} \quad p(y|x) = 0, \qquad (1)$$

and

$$q^*(x|y) = \frac{r(x)p(y|x)}{\sum_{x'} r(x')p(y|x')}. \qquad (2)$$

Proof

1. In (2), let

$$w(y) = \sum_{x'} r(x')p(y|x').$$

2. Assume w.l.o.g. that for all $y \in \mathcal{Y}$, $p(y|x) > 0$ for some $x \in \mathcal{X}$.

3. Since $\mathbf{r} > 0$, $w(y) > 0$ for all y, and hence $q^*(x|y)$ is well-defined.

4. Rearranging (2), we have

$$r(x)p(y|x) = w(y)q^*(x|y).$$

5. For any $\mathbf{q}$ satisfying (1), consider

$$\begin{aligned}
&\sum_x \sum_y r(x)p(y|x) \log \frac{q^*(x|y)}{r(x)} - \sum_x \sum_y r(x)p(y|x) \log \frac{q(x|y)}{r(x)} \\
&= \sum_x \sum_y r(x)p(y|x) \log \frac{q^*(x|y)}{q(x|y)} \\
&= \sum_y \sum_x w(y)q^*(x|y) \log \frac{q^*(x|y)}{q(x|y)} \\
&= \sum_y w(y) \sum_x q^*(x|y) \log \frac{q^*(x|y)}{q(x|y)} \\
&= \sum_y w(y) D(q^*(x|y) \| q(x|y)) \\
&\geq 0.
\end{aligned}$$

6. The proof is completed by noting in (2) that $\mathbf{q}^*$ satisfies (1) because $\mathbf{r} > 0$.

Remark The maximizing $\mathbf{q}$ in Lemma 9.1 is unique.

Lemma 9.1 Let $r(x)p(y|x)$ be a given joint distribution on $\mathcal{X} \times \mathcal{Y}$ such that $\mathbf{r} > 0$, and let $\mathbf{q}$ be a transition matrix from $\mathcal{Y}$ to $\mathcal{X}$. Then

$$\max_{\mathbf{q}} \sum_x \sum_y r(x)p(y|x) \log \frac{q(x|y)}{r(x)} = \sum_x \sum_y r(x)p(y|x) \log \frac{q^*(x|y)}{r(x)},$$

where the maximization is taken over all $\mathbf{q}$ such that

$$q(x|y) = 0 \quad \text{if and only if} \quad p(y|x) = 0, \tag{1}$$

and

$$q^*(x|y) = \frac{r(x)p(y|x)}{\sum_{x'} r(x')p(y|x')}. \tag{2}$$

Proof

1. In (2), let

$$w(y) = \sum_{x'} r(x')p(y|x').$$

2. Assume w.l.o.g. that for all $y \in \mathcal{Y}$, $p(y|x) > 0$ for some $x \in \mathcal{X}$.

3. Since $\mathbf{r} > 0$, $w(y) > 0$ for all y, and hence $q^*(x|y)$ is well-defined.

4. Rearranging (2), we have

$$r(x)p(y|x) = w(y)q^*(x|y).$$

5. For any $\mathbf{q}$ satisfying (1), consider

$$\begin{aligned}
&\sum_x \sum_y r(x)p(y|x) \log \frac{q^*(x|y)}{r(x)} - \sum_x \sum_y r(x)p(y|x) \log \frac{q(x|y)}{r(x)} \\
&= \sum_x \sum_y r(x)p(y|x) \log \frac{q^*(x|y)}{q(x|y)} \\
&= \sum_y \sum_x w(y)q^*(x|y) \log \frac{q^*(x|y)}{q(x|y)} \\
&= \sum_y w(y) \sum_x q^*(x|y) \log \frac{q^*(x|y)}{q(x|y)} \\
&= \sum_y w(y) D(q^*(x|y) \| q(x|y)) \\
&\geq 0.
\end{aligned}$$

6. The proof is completed by noting in (2) that $\mathbf{q}^*$ satisfies (1) because $\mathbf{r} > 0$.

Remark The maximizing $\mathbf{q}$ in Lemma 9.1 is unique.

Lemma 9.1 Let $r(x)p(y|x)$ be a given joint distribution on $\mathcal{X} \times \mathcal{Y}$ such that $\mathbf{r} > 0$, and let $\mathbf{q}$ be a transition matrix from $\mathcal{Y}$ to $\mathcal{X}$. Then

$$\max_{\mathbf{q}} \sum_x \sum_y r(x)p(y|x) \log \frac{q(x|y)}{r(x)} = \sum_x \sum_y r(x)p(y|x) \log \frac{q^*(x|y)}{r(x)},$$

where the maximization is taken over all $\mathbf{q}$ such that

$$q(x|y) = 0 \quad \text{if and only if} \quad p(y|x) = 0, \tag{1}$$

and

$$q^*(x|y) = \frac{r(x)p(y|x)}{\sum_{x'} r(x')p(y|x')}. \tag{2}$$

Proof

1. In (2), let

$$w(y) = \sum_{x'} r(x')p(y|x').$$

2. Assume w.l.o.g. that for all $y \in \mathcal{Y}$, $p(y|x) > 0$ for some $x \in \mathcal{X}$.

3. Since $\mathbf{r} > 0$, $w(y) > 0$ for all y, and hence $q^*(x|y)$ is well-defined.

4. Rearranging (2), we have

$$r(x)p(y|x) = w(y)q^*(x|y).$$

5. For any $\mathbf{q}$ satisfying (1), consider

$$\begin{aligned}
&\sum_x \sum_y r(x)p(y|x) \log \frac{q^*(x|y)}{r(x)} - \sum_x \sum_y r(x)p(y|x) \log \frac{q(x|y)}{r(x)} \\
&= \sum_x \sum_y r(x)p(y|x) \log \frac{q^*(x|y)}{q(x|y)} \\
&= \sum_y \sum_x w(y)q^*(x|y) \log \frac{q^*(x|y)}{q(x|y)} \\
&= \sum_y w(y) \sum_x q^*(x|y) \log \frac{q^*(x|y)}{q(x|y)} \\
&= \sum_y w(y) D(q^*(x|y) \| q(x|y)) \\
&\geq 0.
\end{aligned}$$

6. The proof is completed by noting in (2) that $\mathbf{q}^*$ satisfies (1) because $\mathbf{r} > 0$.

Remark The maximizing $\mathbf{q}$ in Lemma 9.1 is unique.

Lemma 9.1 Let $r(x)p(y|x)$ be a given joint distribution on $\mathcal{X} \times \mathcal{Y}$ such that $\mathbf{r} > 0$, and let $\mathbf{q}$ be a transition matrix from $\mathcal{Y}$ to $\mathcal{X}$. Then

$$\max_{\mathbf{q}} \sum_x \sum_y r(x)p(y|x) \log \frac{q(x|y)}{r(x)} = \sum_x \sum_y r(x)p(y|x) \log \frac{q^*(x|y)}{r(x)},$$

where the maximization is taken over all $\mathbf{q}$ such that

$$q(x|y) = 0 \quad \text{if and only if} \quad p(y|x) = 0, \tag{1}$$

and

$$q^*(x|y) = \frac{r(x)p(y|x)}{\sum_{x'} r(x')p(y|x')}. \tag{2}$$

Proof

1. In (2), let

$$w(y) = \sum_{x'} r(x')p(y|x').$$

2. Assume w.l.o.g. that for all $y \in \mathcal{Y}$, $p(y|x) > 0$ for some $x \in \mathcal{X}$.

3. Since $\mathbf{r} > 0$, $w(y) > 0$ for all y, and hence $q^*(x|y)$ is well-defined.

4. Rearranging (2), we have

$$r(x)p(y|x) = w(y)q^*(x|y).$$

5. For any $\mathbf{q}$ satisfying (1), consider

$$\begin{aligned}
&\sum_x \sum_y r(x)p(y|x) \log \frac{q^*(x|y)}{r(x)} - \sum_x \sum_y r(x)p(y|x) \log \frac{q(x|y)}{r(x)} \\
&= \sum_x \sum_y r(x)p(y|x) \log \frac{q^*(x|y)}{q(x|y)} \\
&= \sum_y \sum_x w(y)q^*(x|y) \log \frac{q^*(x|y)}{q(x|y)} \\
&= \sum_y w(y) \sum_x q^*(x|y) \log \frac{q^*(x|y)}{q(x|y)} \\
&= \sum_y w(y) D(\underline{q^*(x|y)} \| \underline{q(x|y)}) \\
&\geq 0.
\end{aligned}$$

6. The proof is completed by noting in (2) that $\mathbf{q}^*$ satisfies (1) because $\mathbf{r} > 0$.

Remark The maximizing $\mathbf{q}$ in Lemma 9.1 is unique.

Lemma 9.1 Let $r(x)p(y|x)$ be a given joint distribution on $\mathcal{X} \times \mathcal{Y}$ such that $\mathbf{r} > 0$, and let $\mathbf{q}$ be a transition matrix from $\mathcal{Y}$ to $\mathcal{X}$. Then

$$\max_{\mathbf{q}} \sum_x \sum_y r(x)p(y|x) \log \frac{q(x|y)}{r(x)} = \sum_x \sum_y r(x)p(y|x) \log \frac{q^*(x|y)}{r(x)},$$

where the maximization is taken over all $\mathbf{q}$ such that

$$q(x|y) = 0 \quad \text{if and only if} \quad p(y|x) = 0, \tag{1}$$

and

$$q^*(x|y) = \frac{r(x)p(y|x)}{\sum_{x'} r(x')p(y|x')}. \tag{2}$$

Proof

1. In (2), let

$$w(y) = \sum_{x'} r(x')p(y|x').$$

2. Assume w.l.o.g. that for all $y \in \mathcal{Y}$, $p(y|x) > 0$ for some $x \in \mathcal{X}$.

3. Since $\mathbf{r} > 0$, $w(y) > 0$ for all y, and hence $q^*(x|y)$ is well-defined.

4. Rearranging (2), we have

$$r(x)p(y|x) = w(y)q^*(x|y).$$

5. For any $\mathbf{q}$ satisfying (1), consider

$$\begin{aligned}
&\sum_x \sum_y r(x)p(y|x) \log \frac{q^*(x|y)}{r(x)} - \sum_x \sum_y r(x)p(y|x) \log \frac{q(x|y)}{r(x)} \\
&= \sum_x \sum_y r(x)p(y|x) \log \frac{q^*(x|y)}{q(x|y)} \\
&= \sum_y \sum_x w(y)q^*(x|y) \log \frac{q^*(x|y)}{q(x|y)} \\
&= \sum_y w(y) \sum_x q^*(x|y) \log \frac{q^*(x|y)}{q(x|y)} \\
&= \sum_y w(y) D(\underline{q^*(x|y)} \| \underline{q(x|y)}) \ \forall y \\
&\geq 0.
\end{aligned}$$

6. The proof is completed by noting in (2) that $\mathbf{q}^*$ satisfies (1) because $\mathbf{r} > 0$.

Remark The maximizing $\mathbf{q}$ in Lemma 9.1 is unique.

Lemma 9.1 Let $r(x)p(y|x)$ be a given joint distribution on $\mathcal{X} \times \mathcal{Y}$ such that $\mathbf{r} > 0$, and let $\mathbf{q}$ be a transition matrix from $\mathcal{Y}$ to $\mathcal{X}$. Then

$$\max_{\mathbf{q}} \sum_x \sum_y r(x)p(y|x) \log \frac{q(x|y)}{r(x)} = \sum_x \sum_y r(x)p(y|x) \log \frac{q^*(x|y)}{r(x)},$$

where the maximization is taken over all $\mathbf{q}$ such that

$$q(x|y) = 0 \quad \text{if and only if} \quad p(y|x) = 0, \qquad (1)$$

and

$$q^*(x|y) = \frac{r(x)p(y|x)}{\sum_{x'} r(x')p(y|x')}. \qquad (2)$$

Proof

1. In (2), let

$$w(y) = \sum_{x'} r(x')p(y|x').$$

2. Assume w.l.o.g. that for all $y \in \mathcal{Y}$, $p(y|x) > 0$ for some $x \in \mathcal{X}$.

3. Since $\mathbf{r} > 0$, $w(y) > 0$ for all y, and hence $q^*(x|y)$ is well-defined.

4. Rearranging (2), we have

$$r(x)p(y|x) = w(y)q^*(x|y).$$

5. For any $\mathbf{q}$ satisfying (1), consider

$$\begin{aligned}
&\sum_x \sum_y r(x)p(y|x) \log \frac{q^*(x|y)}{r(x)} - \sum_x \sum_y r(x)p(y|x) \log \frac{q(x|y)}{r(x)} \\
&= \sum_x \sum_y r(x)p(y|x) \log \frac{q^*(x|y)}{q(x|y)} \\
&= \sum_y \sum_x w(y)q^*(x|y) \log \frac{q^*(x|y)}{q(x|y)} \\
&= \sum_y w(y) \sum_x q^*(x|y) \log \frac{q^*(x|y)}{q(x|y)} \\
&= \sum_y w(y) D(\underline{q^*(x|y)} \| \underline{q(x|y)}) \quad \forall y \Rightarrow \mathbf{q} = \mathbf{q}^* \\
&\geq 0.
\end{aligned}$$

6. The proof is completed by noting in (2) that $\mathbf{q}^*$ satisfies (1) because $\mathbf{r} > 0$.

Remark The maximizing $\mathbf{q}$ in Lemma 9.1 is unique.

Theorem 9.2 For a discrete memoryless channel $p(y|x)$,

$$C = \sup_{\mathbf{r}>0} \max_{\mathbf{q}} \sum_x \sum_y r(x) p(y|x) \log \frac{q(x|y)}{r(x)},$$

where the maximization is taken over all $\mathbf{q}$ that satisfies (1) in Lemma 9.1.

Theorem 9.2 For a discrete memoryless channel $p(y|x)$,

$$C = \sup_{\mathbf{r}>0} \boxed{\max_{\mathbf{q}} \sum_x \sum_y r(x)p(y|x) \log \frac{q(x|y)}{r(x)}},$$

where the maximization is taken over all $\mathbf{q}$ that satisfies (1) in Lemma 9.1.

Theorem 9.2 For a discrete memoryless channel $p(y|x)$,

$$C = \sup_{\mathbf{r}>0} \boxed{\max_{\mathbf{q}} \sum_x \sum_y r(x)p(y|x)\log\frac{q(x|y)}{r(x)}}^{\,I(X;Y)},$$

where the maximization is taken over all $\mathbf{q}$ that satisfies (1) in Lemma 9.1.

Theorem 9.2 For a discrete memoryless channel $p(y|x)$,

$$C = \sup_{\mathbf{r}>0} \boxed{\max_{\mathbf{q}} \sum_x \sum_y r(x) p(y|x) \log \frac{q(x|y)}{r(x)}}, \quad I(X;Y)$$

where the maximization is taken over all $\mathbf{q}$ that satisfies (1) in Lemma 9.1.

Theorem 9.2 For a discrete memoryless channel $p(y|x)$,

$$C = \sup_{\mathbf{r}>0} \max_{\mathbf{q}} \sum_x \sum_y r(x)p(y|x) \log \frac{q(x|y)}{r(x)},$$

where the maximization is taken over all $\mathbf{q}$ that satisfies (1) in Lemma 9.1.

Proof

Theorem 9.2 For a discrete memoryless channel $p(y|x)$,

$$C = \sup_{\mathbf{r}>0} \max_{\mathbf{q}} \sum_x \sum_y r(x) p(y|x) \log \frac{q(x|y)}{r(x)},$$

where the maximization is taken over all $\mathbf{q}$ that satisfies (1) in Lemma 9.1.

Proof

Theorem 9.2 For a discrete memoryless channel $p(y|x)$,

$$C = \sup_{\mathbf{r}>0} \max_{\mathbf{q}} \sum_x \sum_y r(x) p(y|x) \log \frac{q(x|y)}{r(x)},$$

where the maximization is taken over all $\mathbf{q}$ that satisfies (1) in Lemma 9.1.

Proof

1. Write $I(X;Y)$ as $I(\mathbf{r}, \mathbf{p})$ where $\mathbf{r}$ is the input distribution and $\mathbf{p}$ denotes the transition matrix of the generic channel $p(y|x)$. Then

Theorem 9.2 For a discrete memoryless channel $p(y|x)$,

$$C = \sup_{\mathbf{r}>0} \max_{\mathbf{q}} \sum_x \sum_y r(x)p(y|x) \log \frac{q(x|y)}{r(x)},$$

where the maximization is taken over all $\mathbf{q}$ that satisfies (1) in Lemma 9.1.

Proof

1. Write $I(X;Y)$ as $I(\mathbf{r},\mathbf{p})$ where $\mathbf{r}$ is the input distribution and $\mathbf{p}$ denotes the transition matrix of the generic channel $p(y|x)$. Then

$$C = \max_{\mathbf{r}\geq 0} I(\mathbf{r},\mathbf{p}).$$

Theorem 9.2 For a discrete memoryless channel $p(y|x)$,

$$C = \sup_{\mathbf{r}>0} \max_{\mathbf{q}} \sum_x \sum_y r(x)p(y|x) \log \frac{q(x|y)}{r(x)},$$

where the maximization is taken over all $\mathbf{q}$ that satisfies (1) in Lemma 9.1.

Proof

1. Write $I(X;Y)$ as $I(\mathbf{r},\mathbf{p})$ where $\mathbf{r}$ is the input distribution and $\mathbf{p}$ denotes the transition matrix of the generic channel $p(y|x)$. Then

$$C = \max_{\mathbf{r}\geq 0} I(\mathbf{r},\mathbf{p}).$$

2. By Lemma 9.1, we need to prove that

$$C = \max_{\mathbf{r}\geq 0} I(\mathbf{r},\mathbf{p}) = \sup_{\mathbf{r}>0} I(\mathbf{r},\mathbf{p}).$$

Lemma 9.1

$$\begin{aligned}&\max_{\mathbf{q}} \sum_x \sum_y r(x)p(y|x) \log \frac{q(x|y)}{r(x)} \\ = &\quad \sum_x \sum_y r(x)p(y|x) \log \frac{q^*(x|y)}{r(x)}\end{aligned}$$

Theorem 9.2 For a discrete memoryless channel $p(y|x)$,

$$C = \sup_{\mathbf{r}>0} \boxed{\max_{\mathbf{q}} \sum_x \sum_y r(x)p(y|x) \log \frac{q(x|y)}{r(x)}},$$

where the maximization is taken over all $\mathbf{q}$ that satisfies (1) in Lemma 9.1.

Proof

1. Write $I(X;Y)$ as $I(\mathbf{r},\mathbf{p})$ where $\mathbf{r}$ is the input distribution and $\mathbf{p}$ denotes the transition matrix of the generic channel $p(y|x)$. Then

$$C = \max_{\mathbf{r}\geq 0} I(\mathbf{r},\mathbf{p}).$$

2. By Lemma 9.1, we need to prove that

$$C = \max_{\mathbf{r}\geq 0} I(\mathbf{r},\mathbf{p}) = \sup_{\mathbf{r}>0} I(\mathbf{r},\mathbf{p}).$$

Lemma 9.1

$$\begin{aligned} &\max_{\mathbf{q}} \sum_x \sum_y r(x)p(y|x) \log \frac{q(x|y)}{r(x)} \\ = &\quad \sum_x \sum_y r(x)p(y|x) \log \frac{q^*(x|y)}{r(x)} \end{aligned}$$

Theorem 9.2 For a discrete memoryless channel $p(y|x)$,

$$C = \sup_{\mathbf{r}>0} \boxed{\max_{\mathbf{q}} \sum_x \sum_y r(x)p(y|x) \log \frac{q(x|y)}{r(x)}}, \quad I(\mathbf{r},\mathbf{p})$$

where the maximization is taken over all $\mathbf{q}$ that satisfies (1) in Lemma 9.1.

Proof

1. Write $I(X;Y)$ as $I(\mathbf{r},\mathbf{p})$ where $\mathbf{r}$ is the input distribution and $\mathbf{p}$ denotes the transition matrix of the generic channel $p(y|x)$. Then

$$C = \max_{\mathbf{r}\geq 0} I(\mathbf{r},\mathbf{p}).$$

2. By Lemma 9.1, we need to prove that

$$C = \max_{\mathbf{r}\geq 0} I(\mathbf{r},\mathbf{p}) = \sup_{\mathbf{r}>0} I(\mathbf{r},\mathbf{p}).$$

Lemma 9.1

$$\begin{aligned} &\max_{\mathbf{q}} \sum_x \sum_y r(x)p(y|x) \log \frac{q(x|y)}{r(x)} \\ &= \quad \sum_x \sum_y r(x)p(y|x) \log \frac{q^*(x|y)}{r(x)} \end{aligned}$$

Theorem 9.2 For a discrete memoryless channel $p(y|x)$,

$$C = \sup_{\mathbf{r}>0} \underbrace{\boxed{\max_{\mathbf{q}} \sum_x \sum_y r(x) p(y|x) \log \frac{q(x|y)}{r(x)}}}_{I(\mathbf{r},\mathbf{p})},$$

where the maximization is taken over all $\mathbf{q}$ that satisfies (1) in Lemma 9.1.

Proof

1. Write $I(X;Y)$ as $I(\mathbf{r},\mathbf{p})$ where $\mathbf{r}$ is the input distribution and $\mathbf{p}$ denotes the transition matrix of the generic channel $p(y|x)$. Then

$$C = \max_{\mathbf{r}\geq 0} I(\mathbf{r},\mathbf{p}).$$

2. By Lemma 9.1, we need to prove that

$$C = \underline{\max_{\mathbf{r}\geq 0} I(\mathbf{r},\mathbf{p}) = \sup_{\mathbf{r}>0} I(\mathbf{r},\mathbf{p})}.$$

Lemma 9.1

$$\max_{\mathbf{q}} \sum_x \sum_y r(x) p(y|x) \log \frac{q(x|y)}{r(x)} = \sum_x \sum_y r(x) p(y|x) \log \frac{q^*(x|y)}{r(x)}$$

Theorem 9.2 For a discrete memoryless channel $p(y|x)$,

$$C = \sup_{\mathbf{r}>0} \boxed{\max_{\mathbf{q}} \sum_x \sum_y r(x)p(y|x) \log \frac{q(x|y)}{r(x)}}, \quad I(\mathbf{r}, \mathbf{p})$$

where the maximization is taken over all $\mathbf{q}$ that satisfies (1) in Lemma 9.1.

Proof

1. Write $I(X;Y)$ as $I(\mathbf{r}, \mathbf{p})$ where $\mathbf{r}$ is the input distribution and $\mathbf{p}$ denotes the transition matrix of the generic channel $p(y|x)$. Then

$$C = \max_{\mathbf{r}\geq 0} I(\mathbf{r}, \mathbf{p}).$$

2. By Lemma 9.1, we need to prove that

$$C = \max_{\underline{\mathbf{r}\geq 0}} I(\mathbf{r}, \mathbf{p}) = \sup_{\mathbf{r}>0} I(\mathbf{r}, \mathbf{p}).$$

Lemma 9.1

$$\max_{\mathbf{q}} \sum_x \sum_y r(x)p(y|x) \log \frac{q(x|y)}{r(x)}$$
$$= \sum_x \sum_y r(x)p(y|x) \log \frac{q^*(x|y)}{r(x)}$$

Theorem 9.2 For a discrete memoryless channel $p(y|x)$,

$$C = \sup_{\mathbf{r}>0} \boxed{\max_{\mathbf{q}} \sum_x \sum_y r(x)p(y|x) \log \frac{q(x|y)}{r(x)}}, \quad I(\mathbf{r}, \mathbf{p})$$

where the maximization is taken over all $\mathbf{q}$ that satisfies (1) in Lemma 9.1.

Proof

1. Write $I(X;Y)$ as $I(\mathbf{r}, \mathbf{p})$ where $\mathbf{r}$ is the input distribution and $\mathbf{p}$ denotes the transition matrix of the generic channel $p(y|x)$. Then

$$C = \max_{\mathbf{r}\geq 0} I(\mathbf{r}, \mathbf{p}).$$

2. By Lemma 9.1, we need to prove that

$$C = \max_{\underline{\mathbf{r}\geq 0}} I(\mathbf{r}, \mathbf{p}) = \sup_{\underline{\mathbf{r}>0}} I(\mathbf{r}, \mathbf{p}).$$

Lemma 9.1

$$\begin{aligned} &\max_{\mathbf{q}} \sum_x \sum_y r(x)p(y|x) \log \frac{q(x|y)}{r(x)} \\ = &\quad \sum_x \sum_y r(x)p(y|x) \log \frac{q^*(x|y)}{r(x)} \end{aligned}$$

Theorem 9.2 For a discrete memoryless channel $p(y|x)$,

$$C = \sup_{\mathbf{r}>0} \max_{\mathbf{q}} \sum_x \sum_y r(x)p(y|x) \log \frac{q(x|y)}{r(x)},$$

where the maximization is taken over all $\mathbf{q}$ that satisfies (1) in Lemma 9.1.

Proof

1. Write $I(X;Y)$ as $I(\mathbf{r}, \mathbf{p})$ where $\mathbf{r}$ is the input distribution and $\mathbf{p}$ denotes the transition matrix of the generic channel $p(y|x)$. Then

$$C = \max_{\mathbf{r}\geq 0} I(\mathbf{r}, \mathbf{p}).$$

2. By Lemma 9.1, we need to prove that

$$C = \max_{\mathbf{r}\geq 0} I(\mathbf{r}, \mathbf{p}) = \sup_{\mathbf{r}>0} I(\mathbf{r}, \mathbf{p}).$$

Theorem 9.2 For a discrete memoryless channel $p(y|x)$,

$$C = \sup_{\mathbf{r}>0} \max_{\mathbf{q}} \sum_x \sum_y r(x)p(y|x) \log \frac{q(x|y)}{r(x)},$$

where the maximization is taken over all $\mathbf{q}$ that satisfies (1) in Lemma 9.1.

Proof

1. Write $I(X;Y)$ as $I(\mathbf{r},\mathbf{p})$ where $\mathbf{r}$ is the input distribution and $\mathbf{p}$ denotes the transition matrix of the generic channel $p(y|x)$. Then

$$C = \max_{\mathbf{r}\geq 0} I(\mathbf{r},\mathbf{p}).$$

2. By Lemma 9.1, we need to prove that

$$C = \max_{\mathbf{r}\geq 0} I(\mathbf{r},\mathbf{p}) = \sup_{\mathbf{r}>0} I(\mathbf{r},\mathbf{p}).$$

3. Let $\mathbf{r}^*$ achieves C.

Theorem 9.2 For a discrete memoryless channel $p(y|x)$,

$$C = \sup_{\mathbf{r}>0} \max_{\mathbf{q}} \sum_x \sum_y r(x) p(y|x) \log \frac{q(x|y)}{r(x)},$$

where the maximization is taken over all $\mathbf{q}$ that satisfies (1) in Lemma 9.1.

Proof

1. Write $I(X;Y)$ as $I(\mathbf{r}, \mathbf{p})$ where $\mathbf{r}$ is the input distribution and $\mathbf{p}$ denotes the transition matrix of the generic channel $p(y|x)$. Then

$$C = \max_{\mathbf{r}\geq 0} I(\mathbf{r}, \mathbf{p}).$$

2. By Lemma 9.1, we need to prove that

$$C = \max_{\mathbf{r}\geq 0} I(\mathbf{r}, \mathbf{p}) = \sup_{\mathbf{r}>0} I(\mathbf{r}, \mathbf{p}).$$

3. Let $\mathbf{r}^*$ achieves C.

4. If $\mathbf{r}^* > 0$, then

Theorem 9.2 For a discrete memoryless channel $p(y|x)$,

$$C = \sup_{\mathbf{r}>0} \max_{\mathbf{q}} \sum_x \sum_y r(x)p(y|x) \log \frac{q(x|y)}{r(x)},$$

where the maximization is taken over all $\mathbf{q}$ that satisfies (1) in Lemma 9.1.

Proof

1. Write $I(X;Y)$ as $I(\mathbf{r},\mathbf{p})$ where $\mathbf{r}$ is the input distribution and $\mathbf{p}$ denotes the transition matrix of the generic channel $p(y|x)$. Then

$$C = \max_{\mathbf{r}\geq 0} I(\mathbf{r},\mathbf{p}).$$

2. By Lemma 9.1, we need to prove that

$$C = \max_{\mathbf{r}\geq 0} I(\mathbf{r},\mathbf{p}) = \sup_{\mathbf{r}>0} I(\mathbf{r},\mathbf{p}).$$

3. Let $\mathbf{r}^*$ achieves C.

4. If $\mathbf{r}^* > 0$, then

$$C = \max_{\mathbf{r}\geq 0} I(\mathbf{r},\mathbf{p}) = \max_{\mathbf{r}>0} I(\mathbf{r},\mathbf{p}) = \sup_{\mathbf{r}>0} I(\mathbf{r},\mathbf{p}).$$

Theorem 9.2 For a discrete memoryless channel $p(y|x)$,

$$C = \sup_{\mathbf{r}>0} \max_{\mathbf{q}} \sum_x \sum_y r(x)p(y|x) \log \frac{q(x|y)}{r(x)},$$

where the maximization is taken over all $\mathbf{q}$ that satisfies (1) in Lemma 9.1.

Proof

1. Write $I(X;Y)$ as $I(\mathbf{r},\mathbf{p})$ where $\mathbf{r}$ is the input distribution and $\mathbf{p}$ denotes the transition matrix of the generic channel $p(y|x)$. Then

$$C = \max_{\mathbf{r}\geq 0} I(\mathbf{r},\mathbf{p}).$$

2. By Lemma 9.1, we need to prove that

$$C = \max_{\mathbf{r}\geq 0} I(\mathbf{r},\mathbf{p}) = \sup_{\mathbf{r}>0} I(\mathbf{r},\mathbf{p}).$$

3. Let $\mathbf{r}^*$ achieves C.

4. If $\mathbf{r}^* > 0$, then

$$C = \max_{\underline{\mathbf{r}\geq 0}} I(\mathbf{r},\mathbf{p}) = \max_{\mathbf{r}>0} I(\mathbf{r},\mathbf{p}) = \sup_{\mathbf{r}>0} I(\mathbf{r},\mathbf{p}).$$

Theorem 9.2 For a discrete memoryless channel $p(y|x)$,

$$C = \sup_{\mathbf{r}>0} \max_{\mathbf{q}} \sum_x \sum_y r(x)p(y|x) \log \frac{q(x|y)}{r(x)},$$

where the maximization is taken over all $\mathbf{q}$ that satisfies (1) in Lemma 9.1.

Proof

1. Write $I(X;Y)$ as $I(\mathbf{r}, \mathbf{p})$ where $\mathbf{r}$ is the input distribution and $\mathbf{p}$ denotes the transition matrix of the generic channel $p(y|x)$. Then

$$C = \max_{\mathbf{r}\geq 0} I(\mathbf{r}, \mathbf{p}).$$

2. By Lemma 9.1, we need to prove that

$$C = \max_{\mathbf{r}\geq 0} I(\mathbf{r}, \mathbf{p}) = \sup_{\mathbf{r}>0} I(\mathbf{r}, \mathbf{p}).$$

3. Let $\mathbf{r}^*$ achieves C.

4. If $\mathbf{r}^* > 0$, then

$$C = \max_{\mathbf{r}\geq 0} I(\mathbf{r}, \mathbf{p}) = \max_{\mathbf{r}>0} I(\mathbf{r}, \mathbf{p}) = \sup_{\mathbf{r}>0} I(\mathbf{r}, \mathbf{p}).$$

Theorem 9.2 For a discrete memoryless channel $p(y|x)$,

$$C = \sup_{\mathbf{r}>0} \max_{\mathbf{q}} \sum_x \sum_y r(x)p(y|x) \log \frac{q(x|y)}{r(x)},$$

where the maximization is taken over all $\mathbf{q}$ that satisfies (1) in Lemma 9.1.

Proof

1. Write $I(X;Y)$ as $I(\mathbf{r},\mathbf{p})$ where $\mathbf{r}$ is the input distribution and $\mathbf{p}$ denotes the transition matrix of the generic channel $p(y|x)$. Then

$$C = \max_{\mathbf{r}\geq 0} I(\mathbf{r},\mathbf{p}).$$

2. By Lemma 9.1, we need to prove that

$$C = \max_{\mathbf{r}\geq 0} I(\mathbf{r},\mathbf{p}) = \sup_{\mathbf{r}>0} I(\mathbf{r},\mathbf{p}).$$

3. Let $\mathbf{r}^*$ achieves C.

4. If $\mathbf{r}^* > 0$, then

$$C = \max_{\mathbf{r}\geq 0} I(\mathbf{r},\mathbf{p}) = \max_{\underline{\mathbf{r}>0}} I(\mathbf{r},\mathbf{p}) = \sup_{\underline{\mathbf{r}>0}} I(\mathbf{r},\mathbf{p}).$$

Theorem 9.2 For a discrete memoryless channel $p(y|x)$,

$$C = \sup_{\mathbf{r}>0} \max_{\mathbf{q}} \sum_x \sum_y r(x) p(y|x) \log \frac{q(x|y)}{r(x)},$$

where the maximization is taken over all $\mathbf{q}$ that satisfies (1) in Lemma 9.1.

Proof

1. Write $I(X;Y)$ as $I(\mathbf{r},\mathbf{p})$ where $\mathbf{r}$ is the input distribution and $\mathbf{p}$ denotes the transition matrix of the generic channel $p(y|x)$. Then

$$C = \max_{\mathbf{r}\geq 0} I(\mathbf{r},\mathbf{p}).$$

2. By Lemma 9.1, we need to prove that

$$C = \max_{\mathbf{r}\geq 0} I(\mathbf{r},\mathbf{p}) = \sup_{\mathbf{r}>0} I(\mathbf{r},\mathbf{p}).$$

3. Let $\mathbf{r}^*$ achieves C.

4. If $\mathbf{r}^* > 0$, then

$$C = \max_{\mathbf{r}\geq 0} I(\mathbf{r},\mathbf{p}) = \max_{\mathbf{r}>0} I(\mathbf{r},\mathbf{p}) = \sup_{\mathbf{r}>0} I(\mathbf{r},\mathbf{p}).$$

5. Next, consider $\mathbf{r}^* \geq 0$. Since $I(\mathbf{r},\mathbf{p})$ is continuous in $\mathbf{r}$, for any $\epsilon > 0$, there exists $\delta > 0$ such that if

Theorem 9.2 For a discrete memoryless channel $p(y|x)$,

$$C = \sup_{\mathbf{r}>0} \max_{\mathbf{q}} \sum_x \sum_y r(x)p(y|x) \log \frac{q(x|y)}{r(x)},$$

where the maximization is taken over all $\mathbf{q}$ that satisfies (1) in Lemma 9.1.

Proof

1. Write $I(X;Y)$ as $I(\mathbf{r},\mathbf{p})$ where $\mathbf{r}$ is the input distribution and $\mathbf{p}$ denotes the transition matrix of the generic channel $p(y|x)$. Then

$$C = \max_{\mathbf{r}\geq 0} I(\mathbf{r},\mathbf{p}).$$

2. By Lemma 9.1, we need to prove that

$$C = \max_{\mathbf{r}\geq 0} I(\mathbf{r},\mathbf{p}) = \sup_{\mathbf{r}>0} I(\mathbf{r},\mathbf{p}).$$

3. Let $\mathbf{r}^*$ achieves C.

4. If $\mathbf{r}^* > 0$, then

$$C = \max_{\mathbf{r}\geq 0} I(\mathbf{r},\mathbf{p}) = \max_{\mathbf{r}>0} I(\mathbf{r},\mathbf{p}) = \sup_{\mathbf{r}>0} I(\mathbf{r},\mathbf{p}).$$

5. Next, consider $\mathbf{r}^* \geq 0$. Since $I(\mathbf{r},\mathbf{p})$ is continuous in $\mathbf{r}$, for any $\epsilon > 0$, there exists $\delta > 0$ such that if

$$\|\mathbf{r} - \mathbf{r}^*\| < \delta,$$

Theorem 9.2 For a discrete memoryless channel $p(y|x)$,

$$C = \sup_{\mathbf{r}>0} \max_{\mathbf{q}} \sum_x \sum_y r(x)p(y|x) \log \frac{q(x|y)}{r(x)},$$

where the maximization is taken over all $\mathbf{q}$ that satisfies (1) in Lemma 9.1.

Proof

1. Write $I(X;Y)$ as $I(\mathbf{r},\mathbf{p})$ where $\mathbf{r}$ is the input distribution and $\mathbf{p}$ denotes the transition matrix of the generic channel $p(y|x)$. Then

$$C = \max_{\mathbf{r}\geq 0} I(\mathbf{r},\mathbf{p}).$$

2. By Lemma 9.1, we need to prove that

$$C = \max_{\mathbf{r}\geq 0} I(\mathbf{r},\mathbf{p}) = \sup_{\mathbf{r}>0} I(\mathbf{r},\mathbf{p}).$$

3. Let $\mathbf{r}^*$ achieves C.

4. If $\mathbf{r}^* > 0$, then

$$C = \max_{\mathbf{r}\geq 0} I(\mathbf{r},\mathbf{p}) = \max_{\mathbf{r}>0} I(\mathbf{r},\mathbf{p}) = \sup_{\mathbf{r}>0} I(\mathbf{r},\mathbf{p}).$$

5. Next, consider $\mathbf{r}^* \geq 0$. Since $I(\mathbf{r},\mathbf{p})$ is continuous in $\mathbf{r}$, for any $\epsilon > 0$, there exists $\delta > 0$ such that if

$$\|\mathbf{r} - \mathbf{r}^*\| < \delta,$$

then

$$C - I(\mathbf{r},\mathbf{p}) < \epsilon.$$

Theorem 9.2 For a discrete memoryless channel $p(y|x)$,

$$C = \sup_{\mathbf{r}>0} \max_{\mathbf{q}} \sum_x \sum_y r(x) p(y|x) \log \frac{q(x|y)}{r(x)},$$

where the maximization is taken over all $\mathbf{q}$ that satisfies (1) in Lemma 9.1.

Proof

1. Write $I(X;Y)$ as $I(\mathbf{r},\mathbf{p})$ where $\mathbf{r}$ is the input distribution and $\mathbf{p}$ denotes the transition matrix of the generic channel $p(y|x)$. Then

$$C = \max_{\mathbf{r}\geq 0} I(\mathbf{r},\mathbf{p}).$$

2. By Lemma 9.1, we need to prove that

$$C = \max_{\mathbf{r}\geq 0} I(\mathbf{r},\mathbf{p}) = \sup_{\mathbf{r}>0} I(\mathbf{r},\mathbf{p}).$$

3. Let $\mathbf{r}^*$ achieves C.

4. If $\mathbf{r}^* > 0$, then

$$C = \max_{\mathbf{r}\geq 0} I(\mathbf{r},\mathbf{p}) = \max_{\mathbf{r}>0} I(\mathbf{r},\mathbf{p}) = \sup_{\mathbf{r}>0} I(\mathbf{r},\mathbf{p}).$$

5. Next, consider $\mathbf{r}^* \geq 0$. Since $I(\mathbf{r},\mathbf{p})$ is continuous in $\mathbf{r}$, for any $\epsilon > 0$, there exists $\delta > 0$ such that if

$$\|\mathbf{r} - \mathbf{r}^*\| < \delta,$$

then

$$C - I(\mathbf{r},\mathbf{p}) < \epsilon.$$

Theorem 9.2 For a discrete memoryless channel $p(y|x)$,

$$C = \sup_{\mathbf{r}>0} \max_{\mathbf{q}} \sum_x \sum_y r(x) p(y|x) \log \frac{q(x|y)}{r(x)},$$

where the maximization is taken over all $\mathbf{q}$ that satisfies (1) in Lemma 9.1.

Proof

1. Write $I(X;Y)$ as $I(\mathbf{r},\mathbf{p})$ where $\mathbf{r}$ is the input distribution and $\mathbf{p}$ denotes the transition matrix of the generic channel $p(y|x)$. Then

$$C = \max_{\mathbf{r}\geq 0} I(\mathbf{r},\mathbf{p}).$$

2. By Lemma 9.1, we need to prove that

$$C = \max_{\mathbf{r}\geq 0} I(\mathbf{r},\mathbf{p}) = \sup_{\mathbf{r}>0} I(\mathbf{r},\mathbf{p}).$$

3. Let $\mathbf{r}^*$ achieves C.

4. If $\mathbf{r}^* > 0$, then

$$C = \max_{\mathbf{r}\geq 0} I(\mathbf{r},\mathbf{p}) = \max_{\mathbf{r}>0} I(\mathbf{r},\mathbf{p}) = \sup_{\mathbf{r}>0} I(\mathbf{r},\mathbf{p}).$$

5. Next, consider $\mathbf{r}^* \geq 0$. Since $I(\mathbf{r},\mathbf{p})$ is continuous in $\mathbf{r}$, for any $\epsilon > 0$, there exists $\delta > 0$ such that if

$$\|\mathbf{r} - \mathbf{r}^*\| < \delta,$$

then

$$C - I(\mathbf{r},\mathbf{p}) < \epsilon.$$

6. In particular, there exists $\tilde{\mathbf{r}} > 0$ such that

$$\|\tilde{\mathbf{r}} - \mathbf{r}^*\| < \delta.$$

Theorem 9.2 For a discrete memoryless channel $p(y|x)$,

$$C = \sup_{\mathbf{r}>0} \max_{\mathbf{q}} \sum_x \sum_y r(x)p(y|x) \log \frac{q(x|y)}{r(x)},$$

where the maximization is taken over all $\mathbf{q}$ that satisfies (1) in Lemma 9.1.

Proof

1. Write $I(X;Y)$ as $I(\mathbf{r},\mathbf{p})$ where $\mathbf{r}$ is the input distribution and $\mathbf{p}$ denotes the transition matrix of the generic channel $p(y|x)$. Then

$$C = \max_{\mathbf{r}\geq 0} I(\mathbf{r},\mathbf{p}).$$

2. By Lemma 9.1, we need to prove that

$$C = \max_{\mathbf{r}\geq 0} I(\mathbf{r},\mathbf{p}) = \sup_{\mathbf{r}>0} I(\mathbf{r},\mathbf{p}).$$

3. Let $\mathbf{r}^*$ achieves C.

4. If $\mathbf{r}^* > 0$, then

$$C = \max_{\mathbf{r}\geq 0} I(\mathbf{r},\mathbf{p}) = \max_{\mathbf{r}>0} I(\mathbf{r},\mathbf{p}) = \sup_{\mathbf{r}>0} I(\mathbf{r},\mathbf{p}).$$

5. Next, consider $\mathbf{r}^* \geq 0$. Since $I(\mathbf{r},\mathbf{p})$ is continuous in $\mathbf{r}$, for any $\epsilon > 0$, there exists $\delta > 0$ such that if

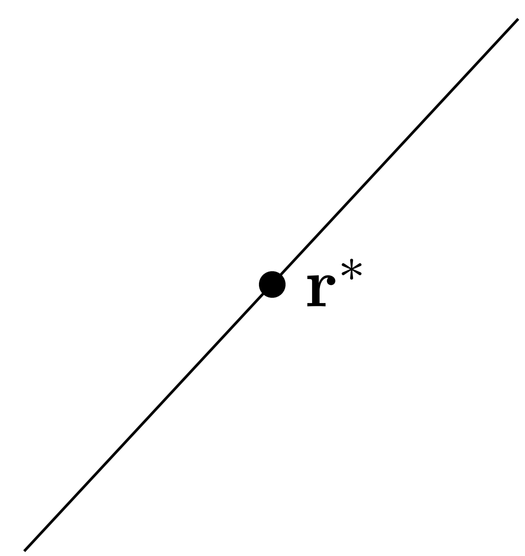
$$\|\mathbf{r} - \mathbf{r}^*\| < \delta,$$

then

$$C - I(\mathbf{r},\mathbf{p}) < \epsilon.$$

6. In particular, there exists $\tilde{\mathbf{r}} > 0$ such that

$$\|\tilde{\mathbf{r}} - \mathbf{r}^*\| < \delta.$$



Theorem 9.2 For a discrete memoryless channel $p(y|x)$,

$$C = \sup_{\mathbf{r}>0} \max_{\mathbf{q}} \sum_x \sum_y r(x)p(y|x) \log \frac{q(x|y)}{r(x)},$$

where the maximization is taken over all $\mathbf{q}$ that satisfies (1) in Lemma 9.1.

Proof

1. Write $I(X;Y)$ as $I(\mathbf{r},\mathbf{p})$ where $\mathbf{r}$ is the input distribution and $\mathbf{p}$ denotes the transition matrix of the generic channel $p(y|x)$. Then

$$C = \max_{\mathbf{r}\geq 0} I(\mathbf{r},\mathbf{p}).$$

2. By Lemma 9.1, we need to prove that

$$C = \max_{\mathbf{r}\geq 0} I(\mathbf{r},\mathbf{p}) = \sup_{\mathbf{r}>0} I(\mathbf{r},\mathbf{p}).$$

3. Let $\mathbf{r}^*$ achieves C.

4. If $\mathbf{r}^* > 0$, then

$$C = \max_{\mathbf{r}\geq 0} I(\mathbf{r},\mathbf{p}) = \max_{\mathbf{r}>0} I(\mathbf{r},\mathbf{p}) = \sup_{\mathbf{r}>0} I(\mathbf{r},\mathbf{p}).$$

5. Next, consider $\mathbf{r}^* \geq 0$. Since $I(\mathbf{r},\mathbf{p})$ is continuous in $\mathbf{r}$, for any $\epsilon > 0$, there exists $\delta > 0$ such that if

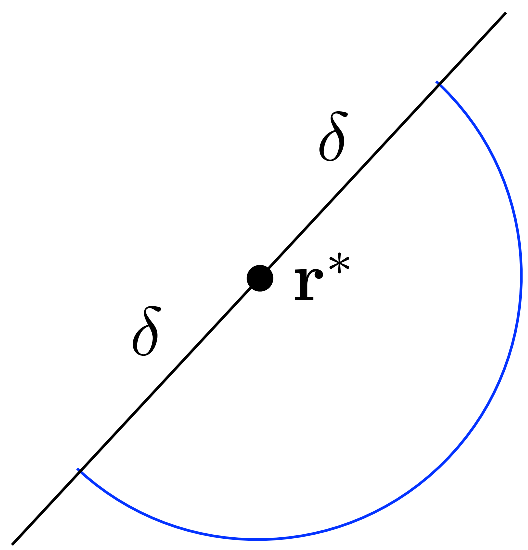
$$\|\mathbf{r} - \mathbf{r}^*\| < \delta,$$

then

$$C - I(\mathbf{r},\mathbf{p}) < \epsilon.$$

6. In particular, there exists $\tilde{\mathbf{r}} > 0$ such that

$$\|\tilde{\mathbf{r}} - \mathbf{r}^*\| < \delta.$$



Theorem 9.2 For a discrete memoryless channel $p(y|x)$,

$$C = \sup_{\mathbf{r}>0} \max_{\mathbf{q}} \sum_x \sum_y r(x) p(y|x) \log \frac{q(x|y)}{r(x)},$$

where the maximization is taken over all $\mathbf{q}$ that satisfies (1) in Lemma 9.1.

Proof

1. Write $I(X;Y)$ as $I(\mathbf{r},\mathbf{p})$ where $\mathbf{r}$ is the input distribution and $\mathbf{p}$ denotes the transition matrix of the generic channel $p(y|x)$. Then

$$C = \max_{\mathbf{r}\geq 0} I(\mathbf{r},\mathbf{p}).$$

2. By Lemma 9.1, we need to prove that

$$C = \max_{\mathbf{r}\geq 0} I(\mathbf{r},\mathbf{p}) = \sup_{\mathbf{r}>0} I(\mathbf{r},\mathbf{p}).$$

3. Let $\mathbf{r}^*$ achieves C.

4. If $\mathbf{r}^* > 0$, then

$$C = \max_{\mathbf{r}\geq 0} I(\mathbf{r},\mathbf{p}) = \max_{\mathbf{r}>0} I(\mathbf{r},\mathbf{p}) = \sup_{\mathbf{r}>0} I(\mathbf{r},\mathbf{p}).$$

5. Next, consider $\mathbf{r}^* \geq 0$. Since $I(\mathbf{r},\mathbf{p})$ is continuous in $\mathbf{r}$, for any $\epsilon > 0$, there exists $\delta > 0$ such that if

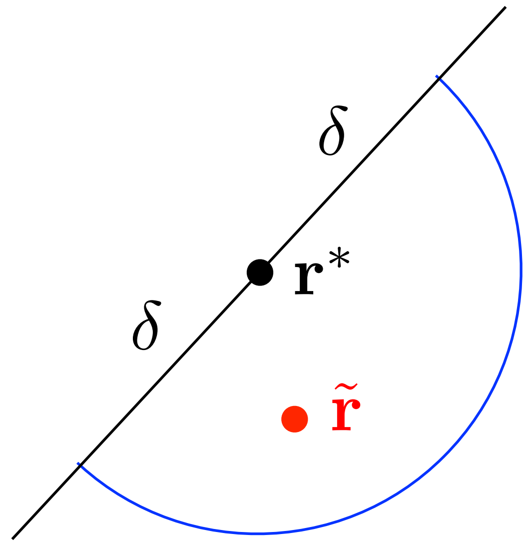
$$\|\mathbf{r} - \mathbf{r}^*\| < \delta,$$

then

$$C - I(\mathbf{r},\mathbf{p}) < \epsilon.$$

6. In particular, there exists $\tilde{\mathbf{r}} > 0$ such that

$$\|\tilde{\mathbf{r}} - \mathbf{r}^*\| < \delta.$$



Theorem 9.2 For a discrete memoryless channel $p(y|x)$,

$$C = \sup_{\mathbf{r}>0} \max_{\mathbf{q}} \sum_x \sum_y r(x)p(y|x) \log \frac{q(x|y)}{r(x)},$$

where the maximization is taken over all $\mathbf{q}$ that satisfies (1) in Lemma 9.1.

Proof

1. Write $I(X;Y)$ as $I(\mathbf{r},\mathbf{p})$ where $\mathbf{r}$ is the input distribution and $\mathbf{p}$ denotes the transition matrix of the generic channel $p(y|x)$. Then

$$\boxed{C = \max_{\mathbf{r}\geq 0} I(\mathbf{r},\mathbf{p}).}$$

2. By Lemma 9.1, we need to prove that

$$C = \max_{\mathbf{r}\geq 0} I(\mathbf{r},\mathbf{p}) = \sup_{\mathbf{r}>0} I(\mathbf{r},\mathbf{p}).$$

3. Let $\mathbf{r}^*$ achieves C.

4. If $\mathbf{r}^* > 0$, then

$$C = \max_{\mathbf{r}\geq 0} I(\mathbf{r},\mathbf{p}) = \max_{\mathbf{r}>0} I(\mathbf{r},\mathbf{p}) = \sup_{\mathbf{r}>0} I(\mathbf{r},\mathbf{p}).$$

5. Next, consider $\mathbf{r}^* \geq 0$. Since $I(\mathbf{r},\mathbf{p})$ is continuous in $\mathbf{r}$, for any $\epsilon > 0$, there exists $\delta > 0$ such that if

$$\|\mathbf{r} - \mathbf{r}^*\| < \delta,$$

then

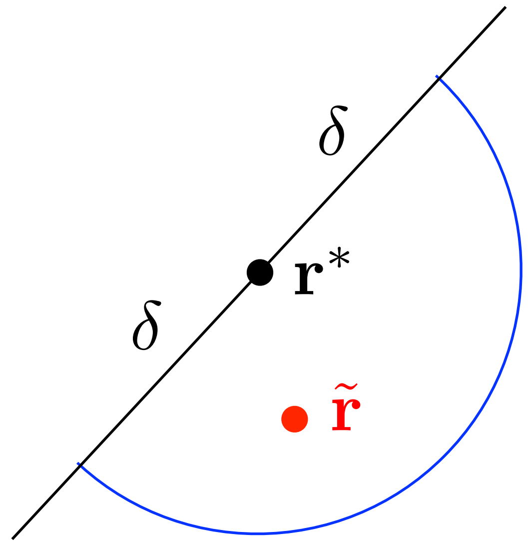
$$C - I(\mathbf{r},\mathbf{p}) < \epsilon.$$

6. In particular, there exists $\tilde{\mathbf{r}} > 0$ such that

$$\|\tilde{\mathbf{r}} - \mathbf{r}^*\| < \delta.$$

7. Then

$$C = \max_{\mathbf{r}\geq 0} I(\mathbf{r},\mathbf{p}) \geq \sup_{\mathbf{r}>0} I(\mathbf{r},\mathbf{p}) \geq I(\tilde{\mathbf{r}},\mathbf{p}) > C - \epsilon.$$



Theorem 9.2 For a discrete memoryless channel $p(y|x)$,

$$C = \sup_{\mathbf{r}>0} \max_{\mathbf{q}} \sum_x \sum_y r(x)p(y|x) \log \frac{q(x|y)}{r(x)},$$

where the maximization is taken over all $\mathbf{q}$ that satisfies (1) in Lemma 9.1.

Proof

1. Write $I(X;Y)$ as $I(\mathbf{r},\mathbf{p})$ where $\mathbf{r}$ is the input distribution and $\mathbf{p}$ denotes the transition matrix of the generic channel $p(y|x)$. Then

$$\boxed{C = \max_{\mathbf{r}\geq 0} I(\mathbf{r},\mathbf{p}).}$$

2. By Lemma 9.1, we need to prove that

$$C = \max_{\mathbf{r}\geq 0} I(\mathbf{r},\mathbf{p}) = \sup_{\mathbf{r}>0} I(\mathbf{r},\mathbf{p}).$$

3. Let $\mathbf{r}^*$ achieves C.

4. If $\mathbf{r}^* > 0$, then

$$C = \max_{\mathbf{r}\geq 0} I(\mathbf{r},\mathbf{p}) = \max_{\mathbf{r}>0} I(\mathbf{r},\mathbf{p}) = \sup_{\mathbf{r}>0} I(\mathbf{r},\mathbf{p}).$$

5. Next, consider $\mathbf{r}^* \geq 0$. Since $I(\mathbf{r},\mathbf{p})$ is continuous in $\mathbf{r}$, for any $\epsilon > 0$, there exists $\delta > 0$ such that if

$$\|\mathbf{r}-\mathbf{r}^*\| < \delta,$$

then

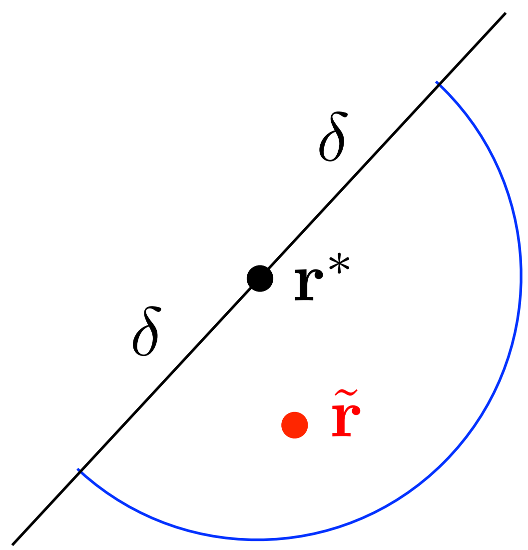
$$C - I(\mathbf{r},\mathbf{p}) < \epsilon.$$

6. In particular, there exists $\tilde{\mathbf{r}} > 0$ such that

$$\|\tilde{\mathbf{r}}-\mathbf{r}^*\| < \delta.$$

7. Then

$$C = \underline{\max_{\mathbf{r}\geq 0}} I(\mathbf{r},\mathbf{p}) \geq \sup_{\mathbf{r}>0} I(\mathbf{r},\mathbf{p}) \geq I(\tilde{\mathbf{r}},\mathbf{p}) > C - \epsilon.$$



Theorem 9.2 For a discrete memoryless channel $p(y|x)$,

$$C = \sup_{\mathbf{r}>0} \max_{\mathbf{q}} \sum_x \sum_y r(x)p(y|x) \log \frac{q(x|y)}{r(x)},$$

where the maximization is taken over all $\mathbf{q}$ that satisfies (1) in Lemma 9.1.

Proof

1. Write $I(X;Y)$ as $I(\mathbf{r},\mathbf{p})$ where $\mathbf{r}$ is the input distribution and $\mathbf{p}$ denotes the transition matrix of the generic channel $p(y|x)$. Then

$$C = \max_{\mathbf{r}\geq 0} I(\mathbf{r},\mathbf{p}).$$

2. By Lemma 9.1, we need to prove that

$$C = \max_{\mathbf{r}\geq 0} I(\mathbf{r},\mathbf{p}) = \sup_{\mathbf{r}>0} I(\mathbf{r},\mathbf{p}).$$

3. Let $\mathbf{r}^*$ achieves C.

4. If $\mathbf{r}^* > 0$, then

$$C = \max_{\mathbf{r}\geq 0} I(\mathbf{r},\mathbf{p}) = \max_{\mathbf{r}>0} I(\mathbf{r},\mathbf{p}) = \sup_{\mathbf{r}>0} I(\mathbf{r},\mathbf{p}).$$

5. Next, consider $\mathbf{r}^* \geq 0$. Since $I(\mathbf{r},\mathbf{p})$ is continuous in $\mathbf{r}$, for any $\epsilon > 0$, there exists $\delta > 0$ such that if

$$\|\mathbf{r} - \mathbf{r}^*\| < \delta,$$

then

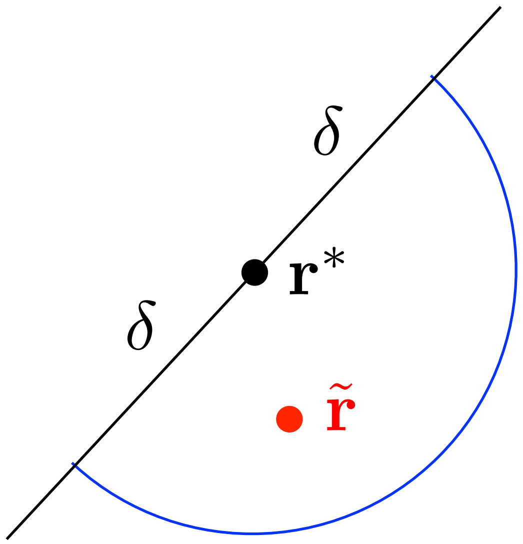
$$C - I(\mathbf{r},\mathbf{p}) < \epsilon.$$

6. In particular, there exists $\tilde{\mathbf{r}} > 0$ such that

$$\|\tilde{\mathbf{r}} - \mathbf{r}^*\| < \delta.$$

7. Then

$$C = \underline{\max_{\mathbf{r}\geq 0}} I(\mathbf{r},\mathbf{p}) \geq \underline{\sup_{\mathbf{r}>0}} I(\mathbf{r},\mathbf{p}) \geq I(\tilde{\mathbf{r}},\mathbf{p}) > C - \epsilon.$$



Theorem 9.2 For a discrete memoryless channel $p(y|x)$,

$$C = \sup_{\mathbf{r}>0} \max_{\mathbf{q}} \sum_x \sum_y r(x)p(y|x) \log \frac{q(x|y)}{r(x)},$$

where the maximization is taken over all $\mathbf{q}$ that satisfies (1) in Lemma 9.1.

Proof

1. Write $I(X;Y)$ as $I(\mathbf{r},\mathbf{p})$ where $\mathbf{r}$ is the input distribution and $\mathbf{p}$ denotes the transition matrix of the generic channel $p(y|x)$. Then

$$C = \max_{\mathbf{r}\geq 0} I(\mathbf{r},\mathbf{p}).$$

2. By Lemma 9.1, we need to prove that

$$C = \max_{\mathbf{r}\geq 0} I(\mathbf{r},\mathbf{p}) = \sup_{\mathbf{r}>0} I(\mathbf{r},\mathbf{p}).$$

3. Let $\mathbf{r}^*$ achieves C.

4. If $\mathbf{r}^* > 0$, then

$$C = \max_{\mathbf{r}\geq 0} I(\mathbf{r},\mathbf{p}) = \max_{\mathbf{r}>0} I(\mathbf{r},\mathbf{p}) = \sup_{\mathbf{r}>0} I(\mathbf{r},\mathbf{p}).$$

5. Next, consider $\mathbf{r}^* \geq 0$. Since $I(\mathbf{r},\mathbf{p})$ is continuous in $\mathbf{r}$, for any $\epsilon > 0$, there exists $\delta > 0$ such that if

$$\|\mathbf{r} - \mathbf{r}^*\| < \delta,$$

then

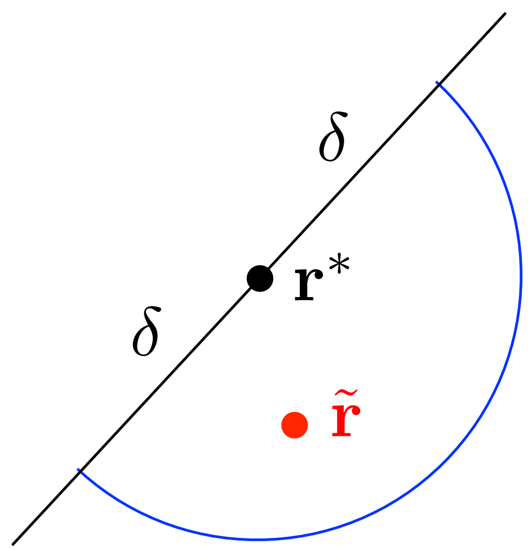
$$C - I(\mathbf{r},\mathbf{p}) < \epsilon.$$

6. In particular, there exists $\tilde{\mathbf{r}} > 0$ such that

$$\|\tilde{\mathbf{r}} - \mathbf{r}^*\| < \delta.$$

7. Then

$$C = \max_{\mathbf{r}\geq 0} I(\mathbf{r},\mathbf{p}) \geq \sup_{\mathbf{r}>0} I(\mathbf{r},\mathbf{p}) \geq I(\tilde{\mathbf{r}},\mathbf{p}) > C - \epsilon.$$



Theorem 9.2 For a discrete memoryless channel $p(y|x)$,

$$C = \sup_{\mathbf{r}>0} \max_{\mathbf{q}} \sum_x \sum_y r(x) p(y|x) \log \frac{q(x|y)}{r(x)},$$

where the maximization is taken over all $\mathbf{q}$ that satisfies (1) in Lemma 9.1.

Proof

1. Write $I(X;Y)$ as $I(\mathbf{r},\mathbf{p})$ where $\mathbf{r}$ is the input distribution and $\mathbf{p}$ denotes the transition matrix of the generic channel $p(y|x)$. Then

$$C = \max_{\mathbf{r}\geq 0} I(\mathbf{r},\mathbf{p}).$$

2. By Lemma 9.1, we need to prove that

$$C = \max_{\mathbf{r}\geq 0} I(\mathbf{r},\mathbf{p}) = \sup_{\mathbf{r}>0} I(\mathbf{r},\mathbf{p}).$$

3. Let $\mathbf{r}^*$ achieves C.

4. If $\mathbf{r}^* > 0$, then

$$C = \max_{\mathbf{r}\geq 0} I(\mathbf{r},\mathbf{p}) = \max_{\mathbf{r}>0} I(\mathbf{r},\mathbf{p}) = \sup_{\mathbf{r}>0} I(\mathbf{r},\mathbf{p}).$$

5. Next, consider $\mathbf{r}^* \geq 0$. Since $I(\mathbf{r},\mathbf{p})$ is continuous in $\mathbf{r}$, for any $\epsilon > 0$, there exists $\delta > 0$ such that if

$$\|\mathbf{r} - \mathbf{r}^*\| < \delta,$$

then

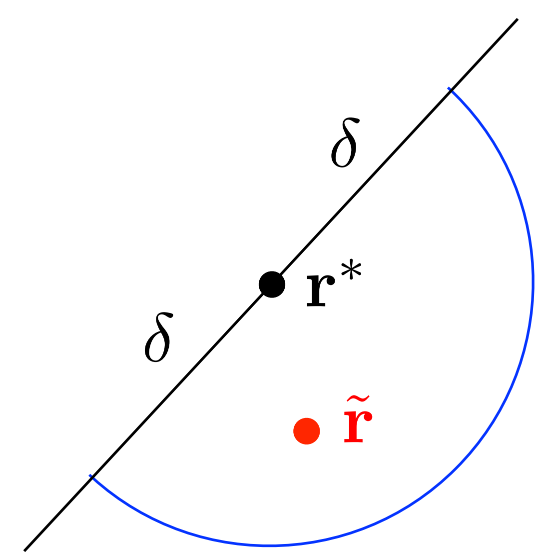
$$C - I(\mathbf{r},\mathbf{p}) < \epsilon.$$

6. In particular, there exists $\tilde{\mathbf{r}} > 0$ such that

$$\|\tilde{\mathbf{r}} - \mathbf{r}^*\| < \delta.$$

7. Then

$$C = \max_{\mathbf{r}\geq 0} I(\mathbf{r},\mathbf{p}) \geq \sup_{\underline{\mathbf{r}>0}} I(\mathbf{r},\mathbf{p}) \geq \underline{I(\tilde{\mathbf{r}},\mathbf{p})} > C - \epsilon.$$



Theorem 9.2 For a discrete memoryless channel $p(y|x)$,

$$C = \sup_{\mathbf{r}>0} \max_{\mathbf{q}} \sum_x \sum_y r(x)p(y|x) \log \frac{q(x|y)}{r(x)},$$

where the maximization is taken over all $\mathbf{q}$ that satisfies (1) in Lemma 9.1.

Proof

1. Write $I(X;Y)$ as $I(\mathbf{r},\mathbf{p})$ where $\mathbf{r}$ is the input distribution and $\mathbf{p}$ denotes the transition matrix of the generic channel $p(y|x)$. Then

$$C = \max_{\mathbf{r}\geq 0} I(\mathbf{r},\mathbf{p}).$$

2. By Lemma 9.1, we need to prove that

$$C = \max_{\mathbf{r}\geq 0} I(\mathbf{r},\mathbf{p}) = \sup_{\mathbf{r}>0} I(\mathbf{r},\mathbf{p}).$$

3. Let $\mathbf{r}^*$ achieves C.

4. If $\mathbf{r}^* > 0$, then

$$C = \max_{\mathbf{r}\geq 0} I(\mathbf{r},\mathbf{p}) = \max_{\mathbf{r}>0} I(\mathbf{r},\mathbf{p}) = \sup_{\mathbf{r}>0} I(\mathbf{r},\mathbf{p}).$$

5. Next, consider $\mathbf{r}^* \geq 0$. Since $I(\mathbf{r},\mathbf{p})$ is continuous in $\mathbf{r}$, for any $\epsilon > 0$, there exists $\delta > 0$ such that if

$$\|\mathbf{r} - \mathbf{r}^*\| < \delta,$$

then

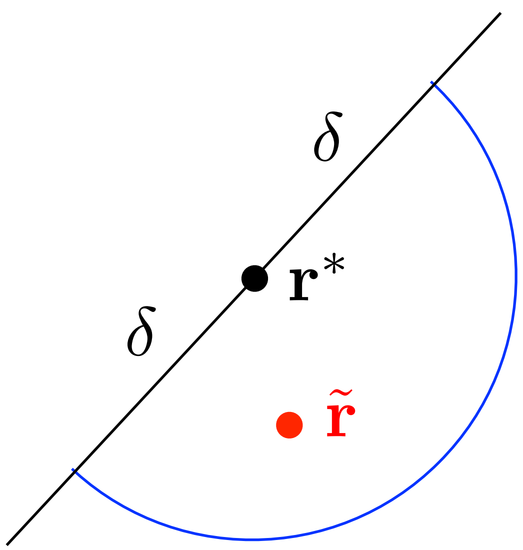
$$C - I(\mathbf{r},\mathbf{p}) < \epsilon.$$

6. In particular, there exists $\tilde{\mathbf{r}} > 0$ such that

$$\|\tilde{\mathbf{r}} - \mathbf{r}^*\| < \delta.$$

7. Then

$$C = \max_{\mathbf{r}\geq 0} I(\mathbf{r},\mathbf{p}) \geq \underline{\sup_{\mathbf{r}>0}} I(\mathbf{r},\mathbf{p}) \geq \underline{I(\tilde{\mathbf{r}},\mathbf{p})} > C - \epsilon.$$



Theorem 9.2 For a discrete memoryless channel $p(y|x)$,

$$C = \sup_{\mathbf{r}>0} \max_{\mathbf{q}} \sum_x \sum_y r(x)p(y|x) \log \frac{q(x|y)}{r(x)},$$

where the maximization is taken over all $\mathbf{q}$ that satisfies (1) in Lemma 9.1.

Proof

1. Write $I(X;Y)$ as $I(\mathbf{r},\mathbf{p})$ where $\mathbf{r}$ is the input distribution and $\mathbf{p}$ denotes the transition matrix of the generic channel $p(y|x)$. Then

$$C = \max_{\mathbf{r}\geq 0} I(\mathbf{r},\mathbf{p}).$$

2. By Lemma 9.1, we need to prove that

$$C = \max_{\mathbf{r}\geq 0} I(\mathbf{r},\mathbf{p}) = \sup_{\mathbf{r}>0} I(\mathbf{r},\mathbf{p}).$$

3. Let $\mathbf{r}^*$ achieves C.

4. If $\mathbf{r}^* > 0$, then

$$C = \max_{\mathbf{r}\geq 0} I(\mathbf{r},\mathbf{p}) = \max_{\mathbf{r}>0} I(\mathbf{r},\mathbf{p}) = \sup_{\mathbf{r}>0} I(\mathbf{r},\mathbf{p}).$$

5. Next, consider $\mathbf{r}^* \geq 0$. Since $I(\mathbf{r},\mathbf{p})$ is continuous in $\mathbf{r}$, for any $\epsilon > 0$, there exists $\delta > 0$ such that if

$$\|\mathbf{r} - \mathbf{r}^*\| < \delta,$$

then

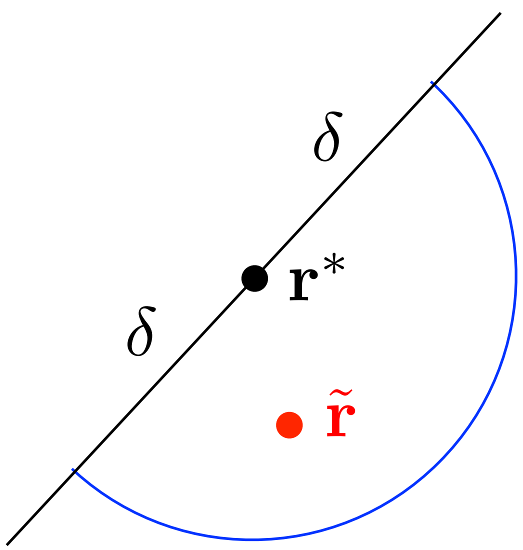
$$C - I(\mathbf{r},\mathbf{p}) < \epsilon.$$

6. In particular, there exists $\tilde{\mathbf{r}} > 0$ such that

$$\|\tilde{\mathbf{r}} - \mathbf{r}^*\| < \delta.$$

7. Then

$$C = \max_{\mathbf{r}\geq 0} I(\mathbf{r},\mathbf{p}) \geq \sup_{\underline{\mathbf{r}>0}} I(\mathbf{r},\mathbf{p}) \geq \underline{I(\tilde{\mathbf{r}},\mathbf{p})} > C - \epsilon.$$



Theorem 9.2 For a discrete memoryless channel $p(y|x)$,

$$C = \sup_{\mathbf{r}>0} \max_{\mathbf{q}} \sum_x \sum_y r(x) p(y|x) \log \frac{q(x|y)}{r(x)},$$

where the maximization is taken over all $\mathbf{q}$ that satisfies (1) in Lemma 9.1.

Proof

1. Write $I(X;Y)$ as $I(\mathbf{r}, \mathbf{p})$ where $\mathbf{r}$ is the input distribution and $\mathbf{p}$ denotes the transition matrix of the generic channel $p(y|x)$. Then

$$C = \max_{\mathbf{r}\geq 0} I(\mathbf{r}, \mathbf{p}).$$

2. By Lemma 9.1, we need to prove that

$$C = \max_{\mathbf{r}\geq 0} I(\mathbf{r}, \mathbf{p}) = \sup_{\mathbf{r}>0} I(\mathbf{r}, \mathbf{p}).$$

3. Let $\mathbf{r}^*$ achieves C.

4. If $\mathbf{r}^* > 0$, then

$$C = \max_{\mathbf{r}\geq 0} I(\mathbf{r}, \mathbf{p}) = \max_{\mathbf{r}>0} I(\mathbf{r}, \mathbf{p}) = \sup_{\mathbf{r}>0} I(\mathbf{r}, \mathbf{p}).$$

5. Next, consider $\mathbf{r}^* \geq 0$. Since $I(\mathbf{r}, \mathbf{p})$ is continuous in $\mathbf{r}$, for any $\epsilon > 0$, there exists $\delta > 0$ such that if

$$\|\mathbf{r} - \mathbf{r}^*\| < \delta,$$

then

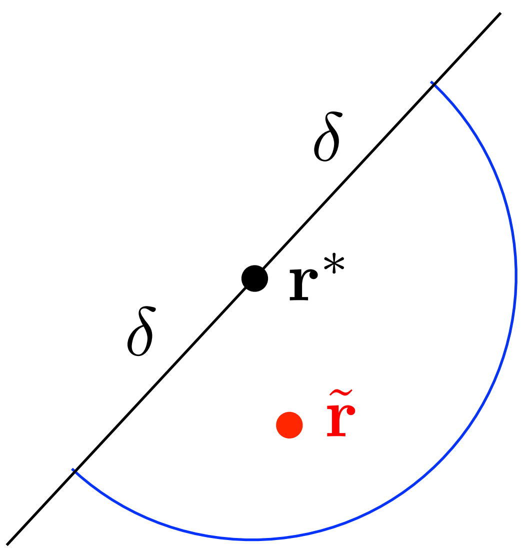
$$C - I(\mathbf{r}, \mathbf{p}) < \epsilon.$$

6. In particular, there exists $\tilde{\mathbf{r}} > 0$ such that

$$\|\tilde{\mathbf{r}} - \mathbf{r}^*\| < \delta.$$

7. Then

$$C = \max_{\mathbf{r}\geq 0} I(\mathbf{r}, \mathbf{p}) \geq \sup_{\mathbf{r}>0} I(\mathbf{r}, \mathbf{p}) \geq \underline{I(\tilde{\mathbf{r}}, \mathbf{p})} > \underline{C - \epsilon}.$$



Theorem 9.2 For a discrete memoryless channel $p(y|x)$,

$$C = \sup_{\mathbf{r}>0} \max_{\mathbf{q}} \sum_x \sum_y r(x)p(y|x) \log \frac{q(x|y)}{r(x)},$$

where the maximization is taken over all $\mathbf{q}$ that satisfies (1) in Lemma 9.1.

Proof

1. Write $I(X;Y)$ as $I(\mathbf{r},\mathbf{p})$ where $\mathbf{r}$ is the input distribution and $\mathbf{p}$ denotes the transition matrix of the generic channel $p(y|x)$. Then

$$C = \max_{\mathbf{r}\geq 0} I(\mathbf{r},\mathbf{p}).$$

2. By Lemma 9.1, we need to prove that

$$C = \max_{\mathbf{r}\geq 0} I(\mathbf{r},\mathbf{p}) = \sup_{\mathbf{r}>0} I(\mathbf{r},\mathbf{p}).$$

3. Let $\mathbf{r}^*$ achieves C.

4. If $\mathbf{r}^* > 0$, then

$$C = \max_{\mathbf{r}\geq 0} I(\mathbf{r},\mathbf{p}) = \max_{\mathbf{r}>0} I(\mathbf{r},\mathbf{p}) = \sup_{\mathbf{r}>0} I(\mathbf{r},\mathbf{p}).$$

5. Next, consider $\mathbf{r}^* \geq 0$. Since $I(\mathbf{r},\mathbf{p})$ is continuous in $\mathbf{r}$, for any $\epsilon > 0$, there exists $\delta > 0$ such that if

$$\|\mathbf{r}-\mathbf{r}^*\| < \delta,$$

then

$$C - I(\mathbf{r},\mathbf{p}) < \epsilon.$$

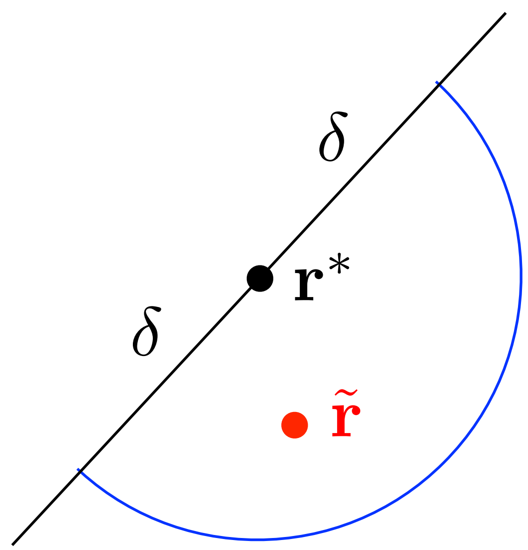
6. In particular, there exists $\tilde{\mathbf{r}} > 0$ such that

$$\|\tilde{\mathbf{r}}-\mathbf{r}^*\| < \delta.$$

7. Then

$$C = \max_{\mathbf{r}\geq 0} I(\mathbf{r},\mathbf{p}) \geq \sup_{\mathbf{r}>0} I(\mathbf{r},\mathbf{p}) \geq I(\tilde{\mathbf{r}},\mathbf{p}) > C - \epsilon.$$

8. Let $\epsilon \to 0$ to conclude that



Theorem 9.2 For a discrete memoryless channel $p(y|x)$,

$$C = \sup_{\mathbf{r}>0} \max_{\mathbf{q}} \sum_x \sum_y r(x)p(y|x) \log \frac{q(x|y)}{r(x)},$$

where the maximization is taken over all $\mathbf{q}$ that satisfies (1) in Lemma 9.1.

Proof

1. Write $I(X;Y)$ as $I(\mathbf{r},\mathbf{p})$ where $\mathbf{r}$ is the input distribution and $\mathbf{p}$ denotes the transition matrix of the generic channel $p(y|x)$. Then

$$C = \max_{\mathbf{r}\geq 0} I(\mathbf{r},\mathbf{p}).$$

2. By Lemma 9.1, we need to prove that

$$C = \max_{\mathbf{r}\geq 0} I(\mathbf{r},\mathbf{p}) = \sup_{\mathbf{r}>0} I(\mathbf{r},\mathbf{p}).$$

3. Let $\mathbf{r}^*$ achieves C.

4. If $\mathbf{r}^* > 0$, then

$$C = \max_{\mathbf{r}\geq 0} I(\mathbf{r},\mathbf{p}) = \max_{\mathbf{r}>0} I(\mathbf{r},\mathbf{p}) = \sup_{\mathbf{r}>0} I(\mathbf{r},\mathbf{p}).$$

5. Next, consider $\mathbf{r}^* \geq 0$. Since $I(\mathbf{r},\mathbf{p})$ is continuous in $\mathbf{r}$, for any $\epsilon > 0$, there exists $\delta > 0$ such that if

$$\|\mathbf{r} - \mathbf{r}^*\| < \delta,$$

then

$$C - I(\mathbf{r},\mathbf{p}) < \epsilon.$$

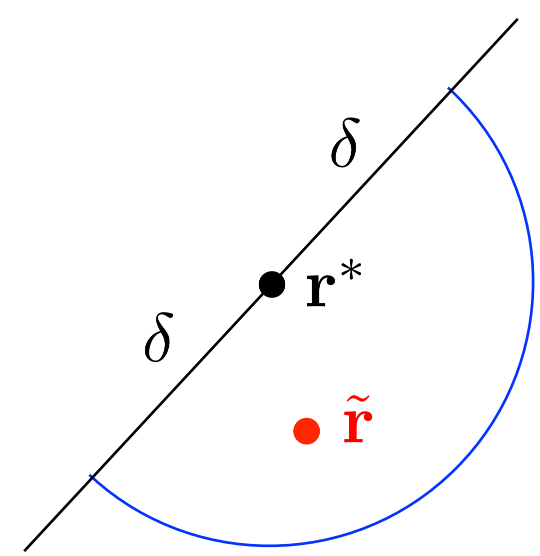
6. In particular, there exists $\tilde{\mathbf{r}} > 0$ such that

$$\|\tilde{\mathbf{r}} - \mathbf{r}^*\| < \delta.$$

7. Then

$$C = \max_{\mathbf{r}\geq 0} I(\mathbf{r},\mathbf{p}) \geq \sup_{\mathbf{r}>0} I(\mathbf{r},\mathbf{p}) \geq I(\tilde{\mathbf{r}},\mathbf{p}) > C - \cancel{\epsilon}.$$

8. Let $\epsilon \to 0$ to conclude that



Theorem 9.2 For a discrete memoryless channel $p(y|x)$,

$$C = \sup_{\mathbf{r}>0} \max_{\mathbf{q}} \sum_x \sum_y r(x) p(y|x) \log \frac{q(x|y)}{r(x)},$$

where the maximization is taken over all $\mathbf{q}$ that satisfies (1) in Lemma 9.1.

Proof

1. Write $I(X;Y)$ as $I(\mathbf{r},\mathbf{p})$ where $\mathbf{r}$ is the input distribution and $\mathbf{p}$ denotes the transition matrix of the generic channel $p(y|x)$. Then

$$C = \max_{\mathbf{r}\geq 0} I(\mathbf{r},\mathbf{p}).$$

2. By Lemma 9.1, we need to prove that

$$C = \max_{\mathbf{r}\geq 0} I(\mathbf{r},\mathbf{p}) = \sup_{\mathbf{r}>0} I(\mathbf{r},\mathbf{p}).$$

3. Let $\mathbf{r}^*$ achieves C.

4. If $\mathbf{r}^* > 0$, then

$$C = \max_{\mathbf{r}\geq 0} I(\mathbf{r},\mathbf{p}) = \max_{\mathbf{r}>0} I(\mathbf{r},\mathbf{p}) = \sup_{\mathbf{r}>0} I(\mathbf{r},\mathbf{p}).$$

5. Next, consider $\mathbf{r}^* \geq 0$. Since $I(\mathbf{r},\mathbf{p})$ is continuous in $\mathbf{r}$, for any $\epsilon > 0$, there exists $\delta > 0$ such that if

$$\|\mathbf{r} - \mathbf{r}^*\| < \delta,$$

then

$$C - I(\mathbf{r},\mathbf{p}) < \epsilon.$$

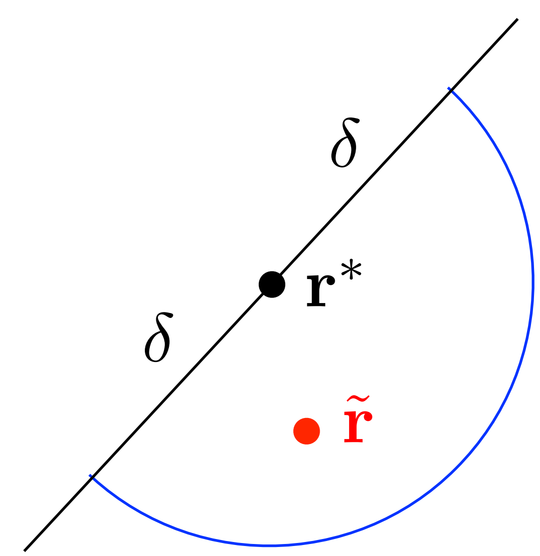
6. In particular, there exists $\tilde{\mathbf{r}} > 0$ such that

$$\|\tilde{\mathbf{r}} - \mathbf{r}^*\| < \delta.$$

7. Then

$$C = \max_{\mathbf{r}\geq 0} I(\mathbf{r},\mathbf{p}) \geq \sup_{\mathbf{r}>0} I(\mathbf{r},\mathbf{p}) \geq I(\tilde{\mathbf{r}},\mathbf{p}) > C - \cancel{\epsilon}.$$

8. Let $\epsilon \to 0$ to conclude that



Theorem 9.2 For a discrete memoryless channel $p(y|x)$,

$$C = \sup_{\mathbf{r}>0} \max_{\mathbf{q}} \sum_x \sum_y r(x)p(y|x) \log \frac{q(x|y)}{r(x)},$$

where the maximization is taken over all $\mathbf{q}$ that satisfies (1) in Lemma 9.1.

Proof

1. Write $I(X;Y)$ as $I(\mathbf{r},\mathbf{p})$ where $\mathbf{r}$ is the input distribution and $\mathbf{p}$ denotes the transition matrix of the generic channel $p(y|x)$. Then

$$C = \max_{\mathbf{r}\geq 0} I(\mathbf{r},\mathbf{p}).$$

2. By Lemma 9.1, we need to prove that

$$C = \max_{\mathbf{r}\geq 0} I(\mathbf{r},\mathbf{p}) = \sup_{\mathbf{r}>0} I(\mathbf{r},\mathbf{p}).$$

3. Let $\mathbf{r}^*$ achieves C.

4. If $\mathbf{r}^* > 0$, then

$$C = \max_{\mathbf{r}\geq 0} I(\mathbf{r},\mathbf{p}) = \max_{\mathbf{r}>0} I(\mathbf{r},\mathbf{p}) = \sup_{\mathbf{r}>0} I(\mathbf{r},\mathbf{p}).$$

5. Next, consider $\mathbf{r}^* \geq 0$. Since $I(\mathbf{r},\mathbf{p})$ is continuous in $\mathbf{r}$, for any $\epsilon > 0$, there exists $\delta > 0$ such that if

$$\|\mathbf{r} - \mathbf{r}^*\| < \delta,$$

then

$$C - I(\mathbf{r},\mathbf{p}) < \epsilon.$$

6. In particular, there exists $\tilde{\mathbf{r}} > 0$ such that

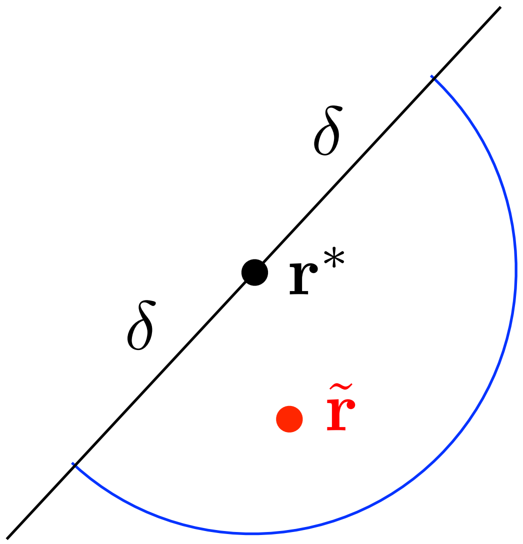
$$\|\tilde{\mathbf{r}} - \mathbf{r}^*\| < \delta.$$

7. Then

$$C = \max_{\mathbf{r}\geq 0} I(\mathbf{r},\mathbf{p}) \geq \sup_{\mathbf{r}>0} I(\mathbf{r},\mathbf{p}) \geq I(\tilde{\mathbf{r}},\mathbf{p}) > C - \cancel{\epsilon}.$$

8. Let $\epsilon \to 0$ to conclude that

$$C = \sup_{\mathbf{r}>0} I(\mathbf{r},\mathbf{p}).$$



The BA Algorithm for Computing C

The BA Algorithm for Computing C

Recall the double supremum in Section 9.1:

$$\sup_{\mathbf{u}_1 \in A_1} \sup_{\mathbf{u}_2 \in A_2} f(\mathbf{u}_1, \mathbf{u}_2),$$

The BA Algorithm for Computing C

Recall the double supremum in Section 9.1:

$$\sup_{\mathbf{u}_1 \in A_1} \sup_{\mathbf{u}_2 \in A_2} f(\mathbf{u}_1, \mathbf{u}_2),$$

where

1. A_i is a convex subset of $\Re^{n_i}$ for $i = 1, 2$.

The BA Algorithm for Computing C

Recall the double supremum in Section 9.1:

$$\sup_{\mathbf{u}_1 \in A_1} \sup_{\mathbf{u}_2 \in A_2} f(\mathbf{u}_1, \mathbf{u}_2),$$

where

1. A_i is a convex subset of $\Re^{n_i}$ for $i = 1, 2$.
2. $f : A_1 \times A_2 \to \Re$ is bounded from above and is such that

The BA Algorithm for Computing *C*

Recall the double supremum in Section 9.1:

$$\sup_{\mathbf{u}_1 \in A_1} \sup_{\mathbf{u}_2 \in A_2} f(\mathbf{u}_1, \mathbf{u}_2),$$

where

1. A_i is a convex subset of $\Re^{n_i}$ for $i = 1, 2$.

2. $f : A_1 \times A_2 \rightarrow \Re$ is bounded from above and is such that

- f is continuous and has continuous partial derivatives on $A_1 \times A_2$

The BA Algorithm for Computing C

Recall the double supremum in Section 9.1:

$$\sup_{\mathbf{u}_1 \in A_1} \sup_{\mathbf{u}_2 \in A_2} f(\mathbf{u}_1, \mathbf{u}_2),$$

where

1. A_i is a convex subset of $\Re^{n_i}$ for $i = 1, 2$.

2. $f : A_1 \times A_2 \to \Re$ is bounded from above and is such that

- f is continuous and has continuous partial derivatives on $A_1 \times A_2$
- For all $\mathbf{u}_2 \in A_2$, there exists a unique $c_1(\mathbf{u}_2) \in A_1$ such that

$$f(c_1(\mathbf{u}_2), \mathbf{u}_2) = \max_{\mathbf{u}_1' \in A_1} f(\mathbf{u}_1', \mathbf{u}_2),$$

The BA Algorithm for Computing C

Recall the double supremum in Section 9.1:

$$\sup_{\mathbf{u}_1 \in A_1} \sup_{\mathbf{u}_2 \in A_2} f(\mathbf{u}_1, \mathbf{u}_2),$$

where

1. A_i is a convex subset of $\Re^{n_i}$ for $i = 1, 2$.

2. $f : A_1 \times A_2 \to \Re$ is bounded from above and is such that

- f is continuous and has continuous partial derivatives on $A_1 \times A_2$
- For all $\mathbf{u}_2 \in A_2$, there exists a unique $c_1(\mathbf{u}_2) \in A_1$ such that

$$f(c_1(\mathbf{u}_2), \mathbf{u}_2) = \max_{\mathbf{u}_1' \in A_1} f(\mathbf{u}_1', \mathbf{u}_2),$$

and for all $\mathbf{u}_1 \in A_1$, there exists a unique $c_2(\mathbf{u}_1) \in A_2$ such that

$$f(\mathbf{u}_1, c_2(\mathbf{u}_1)) = \max_{\mathbf{u}_2' \in A_2} f(\mathbf{u}_1, \mathbf{u}_2').$$

The BA Algorithm for Computing C

Recall the double supremum in Section 9.1:

$$\sup_{\mathbf{u}_1 \in A_1} \sup_{\mathbf{u}_2 \in A_2} f(\mathbf{u}_1, \mathbf{u}_2),$$

where

1. A_i is a convex subset of $\Re^{n_i}$ for $i = 1, 2$.

2. $f : A_1 \times A_2 \rightarrow \Re$ is bounded from above and is such that

- f is continuous and has continuous partial derivatives on $A_1 \times A_2$
- For all $\mathbf{u}_2 \in A_2$, there exists a unique $c_1(\mathbf{u}_2) \in A_1$ such that

$$f(c_1(\mathbf{u}_2), \mathbf{u}_2) = \max_{\mathbf{u}'_1 \in A_1} f(\mathbf{u}'_1, \mathbf{u}_2),$$

and for all $\mathbf{u}_1 \in A_1$, there exists a unique $c_2(\mathbf{u}_1) \in A_2$ such that

$$f(\mathbf{u}_1, c_2(\mathbf{u}_1)) = \max_{\mathbf{u}'_2 \in A_2} f(\mathbf{u}_1, \mathbf{u}'_2).$$

The BA Algorithm for Computing C

Recall the double supremum in Section 9.1:

$$\sup_{\mathbf{u}_1 \in A_1} \sup_{\mathbf{u}_2 \in A_2} f(\mathbf{u}_1, \mathbf{u}_2),$$

where

1. A_i is a convex subset of $\Re^{n_i}$ for $i = 1, 2$.

2. $f : A_1 \times A_2 \to \Re$ is bounded from above and is such that

- f is continuous and has continuous partial derivatives on $A_1 \times A_2$
- For all $\mathbf{u}_2 \in A_2$, there exists a unique $c_1(\mathbf{u}_2) \in A_1$ such that

$$f(c_1(\mathbf{u}_2), \mathbf{u}_2) = \max_{\mathbf{u}'_1 \in A_1} f(\mathbf{u}'_1, \mathbf{u}_2),$$

and for all $\mathbf{u}_1 \in A_1$, there exists a unique $c_2(\mathbf{u}_1) \in A_2$ such that

$$f(\mathbf{u}_1, c_2(\mathbf{u}_1)) = \max_{\mathbf{u}'_2 \in A_2} f(\mathbf{u}_1, \mathbf{u}'_2).$$

Cast the computation of C into this optimization problem:

The BA Algorithm for Computing C

Recall the double supremum in Section 9.1:

$$\sup_{\mathbf{u}_1 \in A_1} \sup_{\mathbf{u}_2 \in A_2} f(\mathbf{u}_1, \mathbf{u}_2),$$

where

1. A_i is a convex subset of $\Re^{n_i}$ for $i = 1, 2$.

2. $f : A_1 \times A_2 \to \Re$ is bounded from above and is such that

- f is continuous and has continuous partial derivatives on $A_1 \times A_2$
- For all $\mathbf{u}_2 \in A_2$, there exists a unique $c_1(\mathbf{u}_2) \in A_1$ such that

$$f(c_1(\mathbf{u}_2), \mathbf{u}_2) = \max_{\mathbf{u}'_1 \in A_1} f(\mathbf{u}'_1, \mathbf{u}_2),$$

and for all $\mathbf{u}_1 \in A_1$, there exists a unique $c_2(\mathbf{u}_1) \in A_2$ such that

$$f(\mathbf{u}_1, c_2(\mathbf{u}_1)) = \max_{\mathbf{u}'_2 \in A_2} f(\mathbf{u}_1, \mathbf{u}'_2).$$

Cast the computation of C into this optimization problem:

1. Let

$$f(\mathbf{r}, \mathbf{q}) = \sum_x \sum_y r(x) p(y|x) \log \frac{q(x|y)}{r(x)},$$

where $\mathbf{u}_1 \leftarrow \mathbf{r}$ and $\mathbf{u}_2 \leftarrow \mathbf{q}$.

The BA Algorithm for Computing C

Recall the double supremum in Section 9.1:

$$\sup_{\mathbf{u}_1 \in A_1} \sup_{\mathbf{u}_2 \in A_2} f(\mathbf{u}_1, \mathbf{u}_2),$$

where

1. A_i is a convex subset of $\Re^{n_i}$ for $i = 1, 2$.

2. $f : A_1 \times A_2 \to \Re$ is bounded from above and is such that

• f is continuous and has continuous partial derivatives on $A_1 \times A_2$

• For all $\mathbf{u}_2 \in A_2$, there exists a unique $c_1(\mathbf{u}_2) \in A_1$ such that

$$f(c_1(\mathbf{u}_2), \mathbf{u}_2) = \max_{\mathbf{u}_1' \in A_1} f(\mathbf{u}_1', \mathbf{u}_2),$$

and for all $\mathbf{u}_1 \in A_1$, there exists a unique $c_2(\mathbf{u}_1) \in A_2$ such that

$$f(\mathbf{u}_1, c_2(\mathbf{u}_1)) = \max_{\mathbf{u}_2' \in A_2} f(\mathbf{u}_1, \mathbf{u}_2').$$

Cast the computation of C into this optimization problem:

1. Let

$$f(\mathbf{r}, \mathbf{q}) = \sum_x \sum_y r(x) p(y|x) \log \frac{q(x|y)}{r(x)},$$

where $\mathbf{u}_1 \leftarrow \mathbf{r}$ and $\mathbf{u}_2 \leftarrow \mathbf{q}$.

2. Let

$$A_1 = \left\{ (r(x), x \in \mathcal{X}) : r(x) > 0 \text{ and } \textstyle\sum_x r(x) = 1 \right\} \subset \Re^{|\mathcal{X}|}$$

The BA Algorithm for Computing C

Recall the double supremum in Section 9.1:

$$\sup_{\mathbf{u}_1 \in A_1} \sup_{\mathbf{u}_2 \in A_2} f(\mathbf{u}_1, \mathbf{u}_2),$$

where

1. A_i is a convex subset of $\Re^{n_i}$ for $i = 1, 2$.

2. $f : A_1 \times A_2 \to \Re$ is bounded from above and is such that

• f is continuous and has continuous partial derivatives on $A_1 \times A_2$

• For all $\mathbf{u}_2 \in A_2$, there exists a unique $c_1(\mathbf{u}_2) \in A_1$ such that

$$f(c_1(\mathbf{u}_2), \mathbf{u}_2) = \max_{\mathbf{u}'_1 \in A_1} f(\mathbf{u}'_1, \mathbf{u}_2),$$

and for all $\mathbf{u}_1 \in A_1$, there exists a unique $c_2(\mathbf{u}_1) \in A_2$ such that

$$f(\mathbf{u}_1, c_2(\mathbf{u}_1)) = \max_{\mathbf{u}'_2 \in A_2} f(\mathbf{u}_1, \mathbf{u}'_2).$$

Cast the computation of C into this optimization problem:

1. Let

$$f(\mathbf{r}, \mathbf{q}) = \sum_x \sum_y r(x) p(y|x) \log \frac{q(x|y)}{r(x)},$$

where $\mathbf{u}_1 \leftarrow \mathbf{r}$ and $\mathbf{u}_2 \leftarrow \mathbf{q}$.

2. Let

$$A_1 = \left\{(r(x), x \in \mathcal{X}) : r(x) > 0 \text{ and } \textstyle\sum_x r(x) = 1\right\} \subset \Re^{|\mathcal{X}|}$$

and

$$\begin{aligned} A_2 &= \{(q(x|y), (x, y) \in \mathcal{X} \times \mathcal{Y}) : q(x|y) \geq 0, \\ &\quad q(x|y) > 0 \text{ iff } p(y|x) > 0, \textstyle\sum_x q(x|y) = 1 \ \forall y \in \mathcal{Y}\} \\ &\subset \Re^{|\mathcal{X}||\mathcal{Y}|}. \end{aligned}$$

The BA Algorithm for Computing C

Recall the double supremum in Section 9.1:

$$\sup_{\mathbf{u}_1 \in A_1} \sup_{\mathbf{u}_2 \in A_2} f(\mathbf{u}_1, \mathbf{u}_2),$$

where

1. A_i is a convex subset of $\Re^{n_i}$ for $i = 1, 2$.

2. $f : A_1 \times A_2 \to \Re$ is bounded from above and is such that

- f is continuous and has continuous partial derivatives on $A_1 \times A_2$
- For all $\mathbf{u}_2 \in A_2$, there exists a unique $c_1(\mathbf{u}_2) \in A_1$ such that

$$f(c_1(\mathbf{u}_2), \mathbf{u}_2) = \max_{\mathbf{u}'_1 \in A_1} f(\mathbf{u}'_1, \mathbf{u}_2),$$

and for all $\mathbf{u}_1 \in A_1$, there exists a unique $c_2(\mathbf{u}_1) \in A_2$ such that

$$f(\mathbf{u}_1, c_2(\mathbf{u}_1)) = \max_{\mathbf{u}'_2 \in A_2} f(\mathbf{u}_1, \mathbf{u}'_2).$$

Cast the computation of C into this optimization problem:

1. Let

$$f(\mathbf{r}, \mathbf{q}) = \sum_x \sum_y r(x) p(y|x) \log \frac{q(x|y)}{r(x)},$$

where $\mathbf{u}_1 \leftarrow \mathbf{r}$ and $\mathbf{u}_2 \leftarrow \mathbf{q}$.

2. Let

$$A_1 = \left\{(r(x), x \in \mathcal{X}) : \ r(x) > 0 \text{ and } \textstyle\sum_x r(x) = 1\right\} \subset \Re^{|\mathcal{X}|}$$

and

$$\begin{aligned} A_2 &= \{(q(x|y), (x, y) \in \mathcal{X} \times \mathcal{Y}) : \ q(x|y) \ge 0, \\ &\quad q(x|y) > 0 \text{ iff } p(y|x) > 0, \ \textstyle\sum_x q(x|y) = 1 \ \forall y \in \mathcal{Y}\} \\ &\subset \Re^{|\mathcal{X}||\mathcal{Y}|}. \end{aligned}$$

The BA Algorithm for Computing C

Recall the double supremum in Section 9.1:

$$\sup_{\mathbf{u}_1 \in A_1} \sup_{\mathbf{u}_2 \in A_2} f(\mathbf{u}_1, \mathbf{u}_2),$$

where

1. A_i is a convex subset of $\Re^{n_i}$ for $i = 1, 2$.

2. $f : A_1 \times A_2 \to \Re$ is bounded from above and is such that

- f is continuous and has continuous partial derivatives on $A_1 \times A_2$
- For all $\mathbf{u}_2 \in A_2$, there exists a unique $c_1(\mathbf{u}_2) \in A_1$ such that

$$f(c_1(\mathbf{u}_2), \mathbf{u}_2) = \max_{\mathbf{u}'_1 \in A_1} f(\mathbf{u}'_1, \mathbf{u}_2),$$

and for all $\mathbf{u}_1 \in A_1$, there exists a unique $c_2(\mathbf{u}_1) \in A_2$ such that

$$f(\mathbf{u}_1, c_2(\mathbf{u}_1)) = \max_{\mathbf{u}'_2 \in A_2} f(\mathbf{u}_1, \mathbf{u}'_2).$$

Cast the computation of C into this optimization problem:

1. Let

$$f(\mathbf{r}, \mathbf{q}) = \sum_x \sum_y r(x) p(y|x) \log \frac{q(x|y)}{r(x)},$$

where $\mathbf{u}_1 \leftarrow \mathbf{r}$ and $\mathbf{u}_2 \leftarrow \mathbf{q}$.

2. Let

$$A_1 = \left\{(r(x), x \in \mathcal{X}) : \; r(x) > 0 \text{ and } \textstyle\sum_x r(x) = 1\right\} \subset \Re^{|\mathcal{X}|}$$

and

$$\begin{aligned} A_2 &= \{(q(x|y), (x, y) \in \mathcal{X} \times \mathcal{Y}) : \; q(x|y) \geq 0, \\ &\quad q(x|y) > 0 \text{ iff } p(y|x) > 0, \; \textstyle\sum_x q(x|y) = 1 \; \forall y \in \mathcal{Y}\} \\ &\subset \Re^{|\mathcal{X}||\mathcal{Y}|}. \end{aligned}$$

Remarks

The BA Algorithm for Computing C

Recall the double supremum in Section 9.1:

$$\sup_{\mathbf{u}_1 \in A_1} \sup_{\mathbf{u}_2 \in A_2} f(\mathbf{u}_1, \mathbf{u}_2),$$

where

1. A_i is a convex subset of $\Re^{n_i}$ for $i = 1, 2$.

2. $f : A_1 \times A_2 \to \Re$ is bounded from above and is such that

• f is continuous and has continuous partial derivatives on $A_1 \times A_2$

• For all $\mathbf{u}_2 \in A_2$, there exists a unique $c_1(\mathbf{u}_2) \in A_1$ such that

$$f(c_1(\mathbf{u}_2), \mathbf{u}_2) = \max_{\mathbf{u}_1' \in A_1} f(\mathbf{u}_1', \mathbf{u}_2),$$

and for all $\mathbf{u}_1 \in A_1$, there exists a unique $c_2(\mathbf{u}_1) \in A_2$ such that

$$f(\mathbf{u}_1, c_2(\mathbf{u}_1)) = \max_{\mathbf{u}_2' \in A_2} f(\mathbf{u}_1, \mathbf{u}_2').$$

Cast the computation of C into this optimization problem:

1. Let

$$f(\mathbf{r}, \mathbf{q}) = \sum_x \sum_y r(x) p(y|x) \log \frac{q(x|y)}{r(x)},$$

where $\mathbf{u}_1 \leftarrow \mathbf{r}$ and $\mathbf{u}_2 \leftarrow \mathbf{q}$.

2. Let

$$A_1 = \left\{(r(x), x \in \mathcal{X}) : r(x) > 0 \text{ and } \textstyle\sum_x r(x) = 1\right\} \subset \Re^{|\mathcal{X}|}$$

and

$$\begin{aligned} A_2 &= \{(q(x|y), (x, y) \in \mathcal{X} \times \mathcal{Y}) : q(x|y) \geq 0, \\ &\quad q(x|y) > 0 \text{ iff } p(y|x) > 0, \ \textstyle\sum_x q(x|y) = 1 \ \forall y \in \mathcal{Y}\} \\ &\subset \Re^{|\mathcal{X}||\mathcal{Y}|}. \end{aligned}$$

Remarks

1. Both A_1 and A_2 are convex.

The BA Algorithm for Computing C

Recall the double supremum in Section 9.1:

$$\sup_{\mathbf{u}_1 \in A_1} \sup_{\mathbf{u}_2 \in A_2} f(\mathbf{u}_1, \mathbf{u}_2),$$

where

1. A_i is a convex subset of $\Re^{n_i}$ for $i = 1, 2$.
2. $f : A_1 \times A_2 \to \Re$ is bounded from above and is such that

- f is continuous and has continuous partial derivatives on $A_1 \times A_2$
- For all $\mathbf{u}_2 \in A_2$, there exists a unique $c_1(\mathbf{u}_2) \in A_1$ such that

$$f(c_1(\mathbf{u}_2), \mathbf{u}_2) = \max_{\mathbf{u}'_1 \in A_1} f(\mathbf{u}'_1, \mathbf{u}_2),$$

and for all $\mathbf{u}_1 \in A_1$, there exists a unique $c_2(\mathbf{u}_1) \in A_2$ such that

$$f(\mathbf{u}_1, c_2(\mathbf{u}_1)) = \max_{\mathbf{u}'_2 \in A_2} f(\mathbf{u}_1, \mathbf{u}'_2).$$

Cast the computation of C into this optimization problem:

1. Let

$$f(\mathbf{r}, \mathbf{q}) = \sum_x \sum_y r(x) p(y|x) \log \frac{q(x|y)}{r(x)},$$

where $\mathbf{u}_1 \leftarrow \mathbf{r}$ and $\mathbf{u}_2 \leftarrow \mathbf{q}$.

2. Let

$$A_1 = \left\{(r(x), x \in \mathcal{X}) : r(x) > 0 \text{ and } \textstyle\sum_x r(x) = 1\right\} \subset \Re^{|\mathcal{X}|}$$

and

$$\begin{aligned} A_2 &= \{(q(x|y), (x,y) \in \mathcal{X} \times \mathcal{Y}) : q(x|y) \geq 0, \\ &\quad q(x|y) > 0 \text{ iff } p(y|x) > 0, \textstyle\sum_x q(x|y) = 1 \; \forall y \in \mathcal{Y}\} \\ &\subset \Re^{|\mathcal{X}||\mathcal{Y}|}. \end{aligned}$$

Remarks

1. Both A_1 and A_2 are convex.

The BA Algorithm for Computing C

Recall the double supremum in Section 9.1:

$$\sup_{\mathbf{u}_1 \in A_1} \sup_{\mathbf{u}_2 \in A_2} f(\mathbf{u}_1, \mathbf{u}_2),$$

where

1. A_i is a convex subset of $\Re^{n_i}$ for $i = 1, 2$.
2. $f : A_1 \times A_2 \to \Re$ is bounded from above and is such that

- f is continuous and has continuous partial derivatives on $A_1 \times A_2$
- For all $\mathbf{u}_2 \in A_2$, there exists a unique $c_1(\mathbf{u}_2) \in A_1$ such that

$$f(c_1(\mathbf{u}_2), \mathbf{u}_2) = \max_{\mathbf{u}_1' \in A_1} f(\mathbf{u}_1', \mathbf{u}_2),$$

and for all $\mathbf{u}_1 \in A_1$, there exists a unique $c_2(\mathbf{u}_1) \in A_2$ such that

$$f(\mathbf{u}_1, c_2(\mathbf{u}_1)) = \max_{\mathbf{u}_2' \in A_2} f(\mathbf{u}_1, \mathbf{u}_2').$$

Cast the computation of C into this optimization problem:

1. Let

$$f(\mathbf{r}, \mathbf{q}) = \sum_x \sum_y r(x) p(y|x) \log \frac{q(x|y)}{r(x)},$$

where $\mathbf{u}_1 \leftarrow \mathbf{r}$ and $\mathbf{u}_2 \leftarrow \mathbf{q}$.

2. Let

$$A_1 = \left\{(r(x), x \in \mathcal{X}) : \ r(x) > 0 \text{ and } \textstyle\sum_x r(x) = 1\right\} \subset \Re^{|\mathcal{X}|}$$

and

$$\begin{aligned} A_2 &= \{(q(x|y), (x,y) \in \mathcal{X} \times \mathcal{Y}) : \ q(x|y) \geq 0, \\ &\quad q(x|y) > 0 \text{ iff } p(y|x) > 0, \ \textstyle\sum_x q(x|y) = 1 \ \forall y \in \mathcal{Y}\} \\ &\subset \Re^{|\mathcal{X}||\mathcal{Y}|}. \end{aligned}$$

Remarks

1. Both A_1 and A_2 are convex.
2. f is bounded from above.

The BA Algorithm for Computing C

Recall the double supremum in Section 9.1:

$$\sup_{\mathbf{u}_1 \in A_1} \sup_{\mathbf{u}_2 \in A_2} f(\mathbf{u}_1, \mathbf{u}_2),$$

where

1. A_i is a convex subset of $\Re^{n_i}$ for $i = 1, 2$.
2. $f : A_1 \times A_2 \to \Re$ is bounded from above and is such that

- f is continuous and has continuous partial derivatives on $A_1 \times A_2$
- For all $\mathbf{u}_2 \in A_2$, there exists a unique $c_1(\mathbf{u}_2) \in A_1$ such that

$$f(c_1(\mathbf{u}_2), \mathbf{u}_2) = \max_{\mathbf{u}'_1 \in A_1} f(\mathbf{u}'_1, \mathbf{u}_2),$$

and for all $\mathbf{u}_1 \in A_1$, there exists a unique $c_2(\mathbf{u}_1) \in A_2$ such that

$$f(\mathbf{u}_1, c_2(\mathbf{u}_1)) = \max_{\mathbf{u}'_2 \in A_2} f(\mathbf{u}_1, \mathbf{u}'_2).$$

Cast the computation of C into this optimization problem:

1. Let

$\le I(X;Y) \le H(X) \le \log|\mathcal{X}|$

$$f(\mathbf{r}, \mathbf{q}) = \sum_x \sum_y r(x) p(y|x) \log \frac{q(x|y)}{r(x)},$$

where $\mathbf{u}_1 \leftarrow \mathbf{r}$ and $\mathbf{u}_2 \leftarrow \mathbf{q}$.

2. Let

$$A_1 = \left\{(r(x), x \in \mathcal{X}) : r(x) > 0 \text{ and } \textstyle\sum_x r(x) = 1\right\} \subset \Re^{|\mathcal{X}|}$$

and

$$\begin{aligned} A_2 &= \{(q(x|y), (x,y) \in \mathcal{X} \times \mathcal{Y}) : q(x|y) \ge 0, \\ &\quad q(x|y) > 0 \text{ iff } p(y|x) > 0, \textstyle\sum_x q(x|y) = 1 \; \forall y \in \mathcal{Y}\} \\ &\subset \Re^{|\mathcal{X}||\mathcal{Y}|}. \end{aligned}$$

Remarks

1. Both A_1 and A_2 are convex.
2. f is bounded from above.

The BA Algorithm for Computing C

Recall the double supremum in Section 9.1:

$$\sup_{\mathbf{u}_1 \in A_1} \sup_{\mathbf{u}_2 \in A_2} f(\mathbf{u}_1, \mathbf{u}_2),$$

where

1. A_i is a convex subset of $\Re^{n_i}$ for $i = 1, 2$.
2. $f : A_1 \times A_2 \to \Re$ is bounded from above and is such that

- f is continuous and has continuous partial derivatives on $A_1 \times A_2$
- For all $\mathbf{u}_2 \in A_2$, there exists a unique $c_1(\mathbf{u}_2) \in A_1$ such that

$$f(c_1(\mathbf{u}_2), \mathbf{u}_2) = \max_{\mathbf{u}_1' \in A_1} f(\mathbf{u}_1', \mathbf{u}_2),$$

and for all $\mathbf{u}_1 \in A_1$, there exists a unique $c_2(\mathbf{u}_1) \in A_2$ such that

$$f(\mathbf{u}_1, c_2(\mathbf{u}_1)) = \max_{\mathbf{u}_2' \in A_2} f(\mathbf{u}_1, \mathbf{u}_2').$$

Cast the computation of C into this optimization problem:

1. Let

$\leq I(X;Y) \leq H(X) \leq \log|\mathcal{X}|$

$$f(\mathbf{r}, \mathbf{q}) = \sum_x \sum_y r(x) p(y|x) \log \frac{q(x|y)}{r(x)},$$

where $\mathbf{u}_1 \leftarrow \mathbf{r}$ and $\mathbf{u}_2 \leftarrow \mathbf{q}$.

2. Let

$$A_1 = \left\{(r(x), x \in \mathcal{X}) : r(x) > 0 \text{ and } \textstyle\sum_x r(x) = 1\right\} \subset \Re^{|\mathcal{X}|}$$

and

$$\begin{aligned} A_2 &= \{(q(x|y), (x,y) \in \mathcal{X} \times \mathcal{Y}) : q(x|y) \geq 0, \\ &\quad q(x|y) > 0 \text{ iff } p(y|x) > 0, \textstyle\sum_x q(x|y) = 1 \ \forall y \in \mathcal{Y}\} \\ &\subset \Re^{|\mathcal{X}||\mathcal{Y}|}. \end{aligned}$$

Remarks

1. Both A_1 and A_2 are convex.
2. f is bounded from above.

The BA Algorithm for Computing C

Recall the double supremum in Section 9.1:

$$\sup_{\mathbf{u}_1 \in A_1} \sup_{\mathbf{u}_2 \in A_2} f(\mathbf{u}_1, \mathbf{u}_2),$$

where

1. A_i is a convex subset of $\Re^{n_i}$ for $i = 1, 2$.
2. $f : A_1 \times A_2 \to \Re$ is bounded from above and is such that

- f is continuous and has continuous partial derivatives on $A_1 \times A_2$
- For all $\mathbf{u}_2 \in A_2$, there exists a unique $c_1(\mathbf{u}_2) \in A_1$ such that

$$f(c_1(\mathbf{u}_2), \mathbf{u}_2) = \max_{\mathbf{u}_1' \in A_1} f(\mathbf{u}_1', \mathbf{u}_2),$$

and for all $\mathbf{u}_1 \in A_1$, there exists a unique $c_2(\mathbf{u}_1) \in A_2$ such that

$$f(\mathbf{u}_1, c_2(\mathbf{u}_1)) = \max_{\mathbf{u}_2' \in A_2} f(\mathbf{u}_1, \mathbf{u}_2').$$

Cast the computation of C into this optimization problem:

1. Let

$$f(\mathbf{r}, \mathbf{q}) = \sum_x \sum_y r(x) p(y|x) \log \frac{q(x|y)}{r(x)},$$

where $\mathbf{u}_1 \leftarrow \mathbf{r}$ and $\mathbf{u}_2 \leftarrow \mathbf{q}$.

2. Let

$$A_1 = \left\{(r(x), x \in \mathcal{X}) : \ r(x) > 0 \text{ and } \textstyle\sum_x r(x) = 1\right\} \subset \Re^{|\mathcal{X}|}$$

and

$$\begin{aligned} A_2 &= \{(q(x|y), (x, y) \in \mathcal{X} \times \mathcal{Y}) : \ q(x|y) \geq 0, \\ &\quad\ q(x|y) > 0 \text{ iff } p(y|x) > 0, \ \textstyle\sum_x q(x|y) = 1 \ \forall y \in \mathcal{Y}\} \\ &\subset \Re^{|\mathcal{X}||\mathcal{Y}|}. \end{aligned}$$

Remarks

1. Both A_1 and A_2 are convex.
2. f is bounded from above.
3. In $f(\mathbf{r}, \mathbf{q})$, the double summation by convention is over all x such that $r(x) > 0$ and all y such that $p(y|x) > 0$.

The BA Algorithm for Computing C

Recall the double supremum in Section 9.1:

$$\sup_{\mathbf{u}_1 \in A_1} \sup_{\mathbf{u}_2 \in A_2} f(\mathbf{u}_1, \mathbf{u}_2),$$

where

1. A_i is a convex subset of $\Re^{n_i}$ for $i = 1, 2$.
2. $f : A_1 \times A_2 \to \Re$ is bounded from above and is such that

- f is continuous and has continuous partial derivatives on $A_1 \times A_2$
- For all $\mathbf{u}_2 \in A_2$, there exists a unique $c_1(\mathbf{u}_2) \in A_1$ such that

$$f(c_1(\mathbf{u}_2), \mathbf{u}_2) = \max_{\mathbf{u}_1' \in A_1} f(\mathbf{u}_1', \mathbf{u}_2),$$

and for all $\mathbf{u}_1 \in A_1$, there exists a unique $c_2(\mathbf{u}_1) \in A_2$ such that

$$f(\mathbf{u}_1, c_2(\mathbf{u}_1)) = \max_{\mathbf{u}_2' \in A_2} f(\mathbf{u}_1, \mathbf{u}_2').$$

Cast the computation of C into this optimization problem:

1. Let

$$f(\mathbf{r}, \mathbf{q}) = \sum_x \sum_y r(x) p(y|x) \log \frac{q(x|y)}{r(x)},$$

where $\mathbf{u}_1 \leftarrow \mathbf{r}$ and $\mathbf{u}_2 \leftarrow \mathbf{q}$.

2. Let

$$A_1 = \{(r(x), x \in \mathcal{X}) : r(x) > 0 \text{ and } \textstyle\sum_x r(x) = 1\} \subset \Re^{|\mathcal{X}|}$$

and

$$\begin{aligned} A_2 &= \{(q(x|y), (x, y) \in \mathcal{X} \times \mathcal{Y}) : q(x|y) \geq 0, \\ &\quad q(x|y) > 0 \text{ iff } p(y|x) > 0, \textstyle\sum_x q(x|y) = 1 \ \forall y \in \mathcal{Y}\} \\ &\subset \Re^{|\mathcal{X}||\mathcal{Y}|}. \end{aligned}$$

Remarks

1. Both A_1 and A_2 are convex.
2. f is bounded from above.
3. In $f(\mathbf{r}, \mathbf{q})$, the double summation by convention is over all x such that $r(x) > 0$ and all y such that $p(y|x) > 0$.
4. Since $q(x|y) > 0$ whenever $p(y|x) > 0$, all the probabilities involved in the double summation are positive.

The BA Algorithm for Computing C

Recall the double supremum in Section 9.1:

$$\sup_{\mathbf{u}_1 \in A_1} \sup_{\mathbf{u}_2 \in A_2} f(\mathbf{u}_1, \mathbf{u}_2),$$

where

1. A_i is a convex subset of $\Re^{n_i}$ for $i = 1, 2$.
2. $f : A_1 \times A_2 \to \Re$ is bounded from above and is such that

- f is continuous and has continuous partial derivatives on $A_1 \times A_2$
- For all $\mathbf{u}_2 \in A_2$, there exists a unique $c_1(\mathbf{u}_2) \in A_1$ such that

$$f(c_1(\mathbf{u}_2), \mathbf{u}_2) = \max_{\mathbf{u}'_1 \in A_1} f(\mathbf{u}'_1, \mathbf{u}_2),$$

and for all $\mathbf{u}_1 \in A_1$, there exists a unique $c_2(\mathbf{u}_1) \in A_2$ such that

$$f(\mathbf{u}_1, c_2(\mathbf{u}_1)) = \max_{\mathbf{u}'_2 \in A_2} f(\mathbf{u}_1, \mathbf{u}'_2).$$

Cast the computation of C into this optimization problem:

1. Let

$$f(\mathbf{r}, \mathbf{q}) = \sum_x \sum_y r(x) p(y|x) \log \frac{q(x|y)}{r(x)},$$

where $\mathbf{u}_1 \leftarrow \mathbf{r}$ and $\mathbf{u}_2 \leftarrow \mathbf{q}$.

2. Let

$$A_1 = \{(r(x), x \in \mathcal{X}) : r(x) > 0 \text{ and } \textstyle\sum_x r(x) = 1\} \subset \Re^{|\mathcal{X}|}$$

and

$$\begin{aligned} A_2 &= \{(q(x|y), (x,y) \in \mathcal{X} \times \mathcal{Y}) : q(x|y) \geq 0, \\ &\quad q(x|y) > 0 \text{ iff } p(y|x) > 0, \textstyle\sum_x q(x|y) = 1 \ \forall y \in \mathcal{Y}\} \\ &\subset \Re^{|\mathcal{X}||\mathcal{Y}|}. \end{aligned}$$

Remarks

1. Both A_1 and A_2 are convex.
2. f is bounded from above.
3. In $f(\mathbf{r}, \mathbf{q})$, the double summation by convention is over all x such that $r(x) > 0$ and all y such that $p(y|x) > 0$.
4. Since $q(x|y) > 0$ whenever $p(y|x) > 0$, all the probabilities involved in the double summation are positive.

The BA Algorithm for Computing C

Recall the double supremum in Section 9.1:

$$\sup_{\mathbf{u}_1 \in A_1} \sup_{\mathbf{u}_2 \in A_2} f(\mathbf{u}_1, \mathbf{u}_2),$$

where

1. A_i is a convex subset of $\Re^{n_i}$ for $i = 1, 2$.

2. $f : A_1 \times A_2 \to \Re$ is bounded from above and is such that

- f is continuous and has continuous partial derivatives on $A_1 \times A_2$
- For all $\mathbf{u}_2 \in A_2$, there exists a unique $c_1(\mathbf{u}_2) \in A_1$ such that

$$f(c_1(\mathbf{u}_2), \mathbf{u}_2) = \max_{\mathbf{u}'_1 \in A_1} f(\mathbf{u}'_1, \mathbf{u}_2),$$

and for all $\mathbf{u}_1 \in A_1$, there exists a unique $c_2(\mathbf{u}_1) \in A_2$ such that

$$f(\mathbf{u}_1, c_2(\mathbf{u}_1)) = \max_{\mathbf{u}'_2 \in A_2} f(\mathbf{u}_1, \mathbf{u}'_2).$$

Cast the computation of C into this optimization problem:

1. Let

$$f(\mathbf{r}, \mathbf{q}) = \sum_x \sum_y r(x) p(y|x) \log \frac{q(x|y)}{r(x)},$$

where $\mathbf{u}_1 \leftarrow \mathbf{r}$ and $\mathbf{u}_2 \leftarrow \mathbf{q}$.

2. Let

$$A_1 = \{(r(x), x \in \mathcal{X}) : r(x) > 0 \text{ and } \textstyle\sum_x r(x) = 1\} \subset \Re^{|\mathcal{X}|}$$

and

$$\begin{aligned} A_2 &= \{(q(x|y), (x,y) \in \mathcal{X} \times \mathcal{Y}) : q(x|y) \geq 0, \\ &\quad q(x|y) > 0 \text{ iff } p(y|x) > 0, \textstyle\sum_x q(x|y) = 1 \ \forall y \in \mathcal{Y}\} \\ &\subset \Re^{|\mathcal{X}||\mathcal{Y}|}. \end{aligned}$$

Remarks

1. Both A_1 and A_2 are convex.
2. f is bounded from above.
3. In $f(\mathbf{r}, \mathbf{q})$, the double summation by convention is over all x such that $r(x) > 0$ and all y such that $p(y|x) > 0$.
4. Since $q(x|y) > 0$ whenever $p(y|x) > 0$, all the probabilities involved in the double summation are positive.

The BA Algorithm for Computing C

Recall the double supremum in Section 9.1:

$$\sup_{\mathbf{u}_1 \in A_1} \sup_{\mathbf{u}_2 \in A_2} f(\mathbf{u}_1, \mathbf{u}_2),$$

where

1. A_i is a convex subset of $\Re^{n_i}$ for $i = 1, 2$.
2. $f : A_1 \times A_2 \to \Re$ is bounded from above and is such that
   - f is continuous and has continuous partial derivatives on $A_1 \times A_2$
   - For all $\mathbf{u}_2 \in A_2$, there exists a unique $c_1(\mathbf{u}_2) \in A_1$ such that

$$f(c_1(\mathbf{u}_2), \mathbf{u}_2) = \max_{\mathbf{u}_1' \in A_1} f(\mathbf{u}_1', \mathbf{u}_2),$$

and for all $\mathbf{u}_1 \in A_1$, there exists a unique $c_2(\mathbf{u}_1) \in A_2$ such that

$$f(\mathbf{u}_1, c_2(\mathbf{u}_1)) = \max_{\mathbf{u}_2' \in A_2} f(\mathbf{u}_1, \mathbf{u}_2').$$

Cast the computation of C into this optimization problem:

1. Let

$$f(\mathbf{r}, \mathbf{q}) = \sum_x \sum_y r(x) p(y|x) \log \frac{q(x|y)}{r(x)},$$

where $\mathbf{u}_1 \leftarrow \mathbf{r}$ and $\mathbf{u}_2 \leftarrow \mathbf{q}$.

2. Let

$$A_1 = \{(r(x), x \in \mathcal{X}) : r(x) > 0 \text{ and } \textstyle\sum_x r(x) = 1\} \subset \Re^{|\mathcal{X}|}$$

and

$$\begin{aligned} A_2 &= \{(q(x|y), (x, y) \in \mathcal{X} \times \mathcal{Y}) : q(x|y) \ge 0, \\ &\quad q(x|y) > 0 \text{ iff } p(y|x) > 0, \textstyle\sum_x q(x|y) = 1 \ \forall y \in \mathcal{Y}\} \\ &\subset \Re^{|\mathcal{X}||\mathcal{Y}|}. \end{aligned}$$

Remarks

1. Both A_1 and A_2 are convex.
2. f is bounded from above.
3. In $f(\mathbf{r}, \mathbf{q})$, the double summation by convention is over all x such that $r(x) > 0$ and all y such that $p(y|x) > 0$.
4. Since $q(x|y) > 0$ whenever $p(y|x) > 0$, all the probabilities involved in the double summation are positive.
5. Therefore, f is continuous and has continuous partial derivatives on $A = A_1 \times A_2$.

The BA Algorithm for Computing C

Recall the double supremum in Section 9.1:

$$\sup_{\mathbf{u}_1 \in A_1} \sup_{\mathbf{u}_2 \in A_2} f(\mathbf{u}_1, \mathbf{u}_2),$$

where

1. A_i is a convex subset of $\Re^{n_i}$ for $i = 1, 2$.
2. $f : A_1 \times A_2 \to \Re$ is bounded from above and is such that

- f is continuous and has continuous partial derivatives on $A_1 \times A_2$
- For all $\mathbf{u}_2 \in A_2$, there exists a unique $c_1(\mathbf{u}_2) \in A_1$ such that

$$f(c_1(\mathbf{u}_2), \mathbf{u}_2) = \max_{\mathbf{u}'_1 \in A_1} f(\mathbf{u}'_1, \mathbf{u}_2),$$

and for all $\mathbf{u}_1 \in A_1$, there exists a unique $c_2(\mathbf{u}_1) \in A_2$ such that

$$f(\mathbf{u}_1, c_2(\mathbf{u}_1)) = \max_{\mathbf{u}'_2 \in A_2} f(\mathbf{u}_1, \mathbf{u}'_2).$$

Cast the computation of C into this optimization problem:

1. Let

$$f(\mathbf{r}, \mathbf{q}) = \sum_x \sum_y r(x) p(y|x) \log \frac{q(x|y)}{r(x)},$$

where $\mathbf{u}_1 \leftarrow \mathbf{r}$ and $\mathbf{u}_2 \leftarrow \mathbf{q}$.

2. Let

$$A_1 = \{(r(x), x \in \mathcal{X}) : r(x) > 0 \text{ and } \textstyle\sum_x r(x) = 1\} \subset \Re^{|\mathcal{X}|}$$

and

$$\begin{aligned} A_2 &= \{(q(x|y), (x, y) \in \mathcal{X} \times \mathcal{Y}) : q(x|y) \geq 0, \\ &\quad q(x|y) > 0 \text{ iff } p(y|x) > 0, \textstyle\sum_x q(x|y) = 1 \ \forall y \in \mathcal{Y}\} \\ &\subset \Re^{|\mathcal{X}||\mathcal{Y}|}. \end{aligned}$$

Remarks

1. Both A_1 and A_2 are convex.
2. f is bounded from above.
3. In $f(\mathbf{r}, \mathbf{q})$, the double summation by convention is over all x such that $r(x) > 0$ and all y such that $p(y|x) > 0$.
4. Since $q(x|y) > 0$ whenever $p(y|x) > 0$, all the probabilities involved in the double summation are positive.
5. Therefore, f is continuous and has continuous partial derivatives on $A = A_1 \times A_2$.

The BA Algorithm for Computing C

Recall the double supremum in Section 9.1:

$$\sup_{\mathbf{u}_1 \in A_1} \sup_{\mathbf{u}_2 \in A_2} f(\mathbf{u}_1, \mathbf{u}_2),$$

where

1. A_i is a convex subset of $\Re^{n_i}$ for $i = 1, 2$.
2. $f : A_1 \times A_2 \to \Re$ is bounded from above and is such that

- f is continuous and has continuous partial derivatives on $A_1 \times A_2$
- For all $\mathbf{u}_2 \in A_2$, there exists a unique $c_1(\mathbf{u}_2) \in A_1$ such that

$$f(c_1(\mathbf{u}_2), \mathbf{u}_2) = \max_{\mathbf{u}_1' \in A_1} f(\mathbf{u}_1', \mathbf{u}_2),$$

and for all $\mathbf{u}_1 \in A_1$, there exists a unique $c_2(\mathbf{u}_1) \in A_2$ such that

$$f(\mathbf{u}_1, c_2(\mathbf{u}_1)) = \max_{\mathbf{u}_2' \in A_2} f(\mathbf{u}_1, \mathbf{u}_2').$$

Cast the computation of C into this optimization problem:

1. Let

$$f(\mathbf{r}, \mathbf{q}) = \sum_x \sum_y r(x) p(y|x) \log \frac{q(x|y)}{r(x)},$$

where $\mathbf{u}_1 \leftarrow \mathbf{r}$ and $\mathbf{u}_2 \leftarrow \mathbf{q}$.

2. Let

$$A_1 = \left\{(r(x), x \in \mathcal{X}) : r(x) > 0 \text{ and } \textstyle\sum_x r(x) = 1\right\} \subset \Re^{|\mathcal{X}|}$$

and

$$\begin{aligned} A_2 &= \{(q(x|y), (x, y) \in \mathcal{X} \times \mathcal{Y}) : q(x|y) \geq 0, \\ &\quad q(x|y) > 0 \text{ iff } p(y|x) > 0, \textstyle\sum_x q(x|y) = 1 \ \forall y \in \mathcal{Y}\} \\ &\subset \Re^{|\mathcal{X}||\mathcal{Y}|}. \end{aligned}$$

The BA Algorithm for Computing C

Recall the double supremum in Section 9.1:

$$\sup_{\mathbf{u}_1 \in A_1} \sup_{\mathbf{u}_2 \in A_2} f(\mathbf{u}_1, \mathbf{u}_2),$$

where

1. A_i is a convex subset of $\Re^{n_i}$ for $i = 1, 2$.
2. $f : A_1 \times A_2 \to \Re$ is bounded from above and is such that

- f is continuous and has continuous partial derivatives on $A_1 \times A_2$
- For all $\mathbf{u}_2 \in A_2$, there exists a unique $c_1(\mathbf{u}_2) \in A_1$ such that

$$f(c_1(\mathbf{u}_2), \mathbf{u}_2) = \max_{\mathbf{u}'_1 \in A_1} f(\mathbf{u}'_1, \mathbf{u}_2),$$

and for all $\mathbf{u}_1 \in A_1$, there exists a unique $c_2(\mathbf{u}_1) \in A_2$ such that

$$f(\mathbf{u}_1, c_2(\mathbf{u}_1)) = \max_{\mathbf{u}'_2 \in A_2} f(\mathbf{u}_1, \mathbf{u}'_2).$$

Cast the computation of C into this optimization problem:

1. Let

$$f(\mathbf{r}, \mathbf{q}) = \sum_x \sum_y r(x) p(y|x) \log \frac{q(x|y)}{r(x)},$$

where $\mathbf{u}_1 \leftarrow \mathbf{r}$ and $\mathbf{u}_2 \leftarrow \mathbf{q}$.

2. Let

$$A_1 = \{(r(x), x \in \mathcal{X}) : r(x) > 0 \text{ and } \textstyle\sum_x r(x) = 1\} \subset \Re^{|\mathcal{X}|}$$

and

$$\begin{aligned} A_2 &= \{(q(x|y), (x, y) \in \mathcal{X} \times \mathcal{Y}) : q(x|y) \geq 0, \\ &\quad q(x|y) > 0 \text{ iff } p(y|x) > 0, \textstyle\sum_x q(x|y) = 1 \ \forall y \in \mathcal{Y}\} \\ &\subset \Re^{|\mathcal{X}||\mathcal{Y}|}. \end{aligned}$$

3. The double supremum now becomes

$$\sup_{\mathbf{r} \in A_1} \sup_{\mathbf{q} \in A_2} f(\mathbf{r}, \mathbf{q}) = \sup_{\mathbf{r} \in A_1} \sup_{\mathbf{q} \in A_2} \sum_x \sum_y r(x) p(y|x) \log \frac{q(x|y)}{r(x)},$$

The BA Algorithm for Computing C

Recall the double supremum in Section 9.1:

$$\sup_{\mathbf{u}_1 \in A_1} \sup_{\mathbf{u}_2 \in A_2} f(\mathbf{u}_1, \mathbf{u}_2),$$

where

1. A_i is a convex subset of $\Re^{n_i}$ for $i = 1, 2$.
2. $f : A_1 \times A_2 \to \Re$ is bounded from above and is such that

- f is continuous and has continuous partial derivatives on $A_1 \times A_2$
- For all $\mathbf{u}_2 \in A_2$, there exists a unique $c_1(\mathbf{u}_2) \in A_1$ such that

$$f(c_1(\mathbf{u}_2), \mathbf{u}_2) = \max_{\mathbf{u}_1' \in A_1} f(\mathbf{u}_1', \mathbf{u}_2),$$

and for all $\mathbf{u}_1 \in A_1$, there exists a unique $c_2(\mathbf{u}_1) \in A_2$ such that

$$f(\mathbf{u}_1, c_2(\mathbf{u}_1)) = \max_{\mathbf{u}_2' \in A_2} f(\mathbf{u}_1, \mathbf{u}_2').$$

Cast the computation of C into this optimization problem:

1. Let

$$f(\mathbf{r}, \mathbf{q}) = \sum_x \sum_y r(x) p(y|x) \log \frac{q(x|y)}{r(x)},$$

where $\mathbf{u}_1 \leftarrow \mathbf{r}$ and $\mathbf{u}_2 \leftarrow \mathbf{q}$.

2. Let

$$A_1 = \{(r(x), x \in \mathcal{X}) : r(x) > 0 \text{ and } \textstyle\sum_x r(x) = 1\} \subset \Re^{|\mathcal{X}|}$$

and

$$\begin{aligned} A_2 &= \{(q(x|y), (x, y) \in \mathcal{X} \times \mathcal{Y}) : q(x|y) \geq 0, \\ &\quad q(x|y) > 0 \text{ iff } p(y|x) > 0, \textstyle\sum_x q(x|y) = 1 \ \forall y \in \mathcal{Y}\} \\ &\subset \Re^{|\mathcal{X}||\mathcal{Y}|}. \end{aligned}$$

3. The double supremum now becomes

$$\sup_{\mathbf{r} \in A_1} \sup_{\mathbf{q} \in A_2} f(\mathbf{r}, \mathbf{q}) = \sup_{\mathbf{r} \in A_1} \underbrace{\sup_{\mathbf{q} \in A_2} \sum_x \sum_y r(x) p(y|x) \log \frac{q(x|y)}{r(x)}}_{= I(\mathbf{r}, \mathbf{p})},$$

where the supremum over all $\mathbf{q} \in A_2$ is in fact a maximum by Lemma 9.1, and by Theorem 9.2,

The BA Algorithm for Computing C

Recall the double supremum in Section 9.1:

$$\sup_{\mathbf{u}_1 \in A_1} \sup_{\mathbf{u}_2 \in A_2} f(\mathbf{u}_1, \mathbf{u}_2),$$

where

✓ 1. A_i is a convex subset of $\Re^{n_i}$ for $i = 1, 2$.

✓ 2. $f : A_1 \times A_2 \to \Re$ is bounded from above and is such that

✓ • f is continuous and has continuous partial derivatives on $A_1 \times A_2$

• For all $\mathbf{u}_2 \in A_2$, there exists a unique $c_1(\mathbf{u}_2) \in A_1$ such that

$$f(c_1(\mathbf{u}_2), \mathbf{u}_2) = \max_{\mathbf{u}_1' \in A_1} f(\mathbf{u}_1', \mathbf{u}_2),$$

and for all $\mathbf{u}_1 \in A_1$, there exists a unique $c_2(\mathbf{u}_1) \in A_2$ such that

$$f(\mathbf{u}_1, c_2(\mathbf{u}_1)) = \max_{\mathbf{u}_2' \in A_2} f(\mathbf{u}_1, \mathbf{u}_2').$$

Cast the computation of C into this optimization problem:

1. Let

$$f(\mathbf{r}, \mathbf{q}) = \sum_x \sum_y r(x) p(y|x) \log \frac{q(x|y)}{r(x)},$$

where $\mathbf{u}_1 \leftarrow \mathbf{r}$ and $\mathbf{u}_2 \leftarrow \mathbf{q}$.

2. Let

$$A_1 = \{(r(x), x \in \mathcal{X}) : r(x) > 0 \text{ and } \textstyle\sum_x r(x) = 1\} \subset \Re^{|\mathcal{X}|}$$

and

$$\begin{aligned} A_2 &= \{(q(x|y), (x, y) \in \mathcal{X} \times \mathcal{Y}) : q(x|y) \geq 0, \\ &\quad q(x|y) > 0 \text{ iff } p(y|x) > 0, \textstyle\sum_x q(x|y) = 1 \ \forall y \in \mathcal{Y}\} \\ &\subset \Re^{|\mathcal{X}||\mathcal{Y}|}. \end{aligned}$$

3. The double supremum now becomes

$$\sup_{\mathbf{r} \in A_1} \sup_{\mathbf{q} \in A_2} f(\mathbf{r}, \mathbf{q}) = \sup_{\mathbf{r} \in A_1} \sup_{\mathbf{q} \in A_2} \sum_x \sum_y r(x) p(y|x) \log \frac{q(x|y)}{r(x)},$$

where the supremum over all $\mathbf{q} \in A_2$ is in fact a maximum by Lemma 9.1, and by Theorem 9.2,

$$f^* = \sup_{\mathbf{r} \in A_1} \sup_{\mathbf{q} \in A_2} f(\mathbf{r}, \mathbf{q}) = C.$$

Algorithm Details

$$f^* = \sup_{\mathbf{r} \in A_1} \sup_{\mathbf{q} \in A_2} f(\mathbf{r}, \mathbf{q})$$

$$f(\mathbf{r}, \mathbf{q}) = \sum_x \sum_y r(x) p(y|x) \log \frac{q(x|y)}{r(x)}$$

Algorithm Details

$$f^* = \sup_{\mathbf{r} \in A_1} \sup_{\mathbf{q} \in A_2} f(\mathbf{r}, \mathbf{q})$$

$$f(\mathbf{r}, \mathbf{q}) = \sum_x \sum_y r(x) p(y|x) \log \frac{q(x|y)}{r(x)}$$

1. By Lemma 9.1, for any given $\mathbf{r} \in A_1$, the unique $\mathbf{q} \in A_2$ that maximizes f is given by

Algorithm Details

$$f^* = \sup_{\mathbf{r} \in A_1} \sup_{\mathbf{q} \in A_2} f(\mathbf{r}, \mathbf{q})$$

$$f(\mathbf{r}, \mathbf{q}) = \sum_x \sum_y r(x) p(y|x) \log \frac{q(x|y)}{r(x)}$$

1. By Lemma 9.1, for any given $\mathbf{r} \in A_1$, the unique $\mathbf{q} \in A_2$ that maximizes f is given by

$$q(x|y) = \frac{r(x)p(y|x)}{\sum_{x'} r(x')p(y|x')}. \tag{1}$$

Algorithm Details

$$f^* = \sup_{\mathbf{r} \in A_1} \sup_{\mathbf{q} \in A_2} f(\mathbf{r}, \mathbf{q})$$

$$f(\mathbf{r}, \mathbf{q}) = \sum_x \sum_y r(x) p(y|x) \log \frac{q(x|y)}{r(x)}$$

1. By Lemma 9.1, for any given $\mathbf{r} \in A_1$, the unique $\mathbf{q} \in A_2$ that maximizes f is given by

$$q(x|y) = \frac{r(x)p(y|x)}{\sum_{x'} r(x')p(y|x')}. \tag{1}$$

2. By Lagrange multipliers, it can be shown that for any given $\mathbf{q} \in A_2$, the unique input distribution $\mathbf{r}$ that maximizes f is given by

Algorithm Details

$$f^* = \sup_{\mathbf{r} \in A_1} \sup_{\mathbf{q} \in A_2} f(\mathbf{r}, \mathbf{q})$$

$$f(\mathbf{r}, \mathbf{q}) = \sum_x \sum_y r(x) p(y|x) \log \frac{q(x|y)}{r(x)}$$

1. By Lemma 9.1, for any given $\mathbf{r} \in A_1$, the unique $\mathbf{q} \in A_2$ that maximizes f is given by

$$q(x|y) = \frac{r(x) p(y|x)}{\sum_{x'} r(x') p(y|x')}. \tag{1}$$

2. By Lagrange multipliers, it can be shown that for any given $\mathbf{q} \in A_2$, the unique input distribution $\mathbf{r}$ that maximizes f is given by

$$r(x) = \frac{\prod_y q(x|y)^{p(y|x)}}{\sum_{x'} \prod_y q(x'|y)^{p(y|x')}}, \tag{2}$$

Algorithm Details

$$f^* = \sup_{\mathbf{r} \in A_1} \sup_{\mathbf{q} \in A_2} f(\mathbf{r}, \mathbf{q})$$

$$f(\mathbf{r}, \mathbf{q}) = \sum_x \sum_y r(x) p(y|x) \log \frac{q(x|y)}{r(x)}$$

1. By Lemma 9.1, for any given $\mathbf{r} \in A_1$, the unique $\mathbf{q} \in A_2$ that maximizes f is given by

$$q(x|y) = \frac{r(x) p(y|x)}{\sum_{x'} r(x') p(y|x')}. \tag{1}$$

2. By Lagrange multipliers, it can be shown that for any given $\mathbf{q} \in A_2$, the unique input distribution $\mathbf{r}$ that maximizes f is given by

$$r(x) = \frac{\prod_y q(x|y)^{p(y|x)}}{\sum_{x'} \prod_y q(x'|y)^{p(y|x')}}, \tag{2}$$

where $\prod_y$ is over all y such that $p(y|x) > 0$.

Algorithm Details

$$f^* = \sup_{\mathbf{r}\in A_1} \sup_{\mathbf{q}\in A_2} f(\mathbf{r}, \mathbf{q})$$

$$f(\mathbf{r}, \mathbf{q}) = \sum_x \sum_y r(x) p(y|x) \log \frac{q(x|y)}{r(x)}$$

1. By Lemma 9.1, for any given $\mathbf{r} \in A_1$, the unique $\mathbf{q} \in A_2$ that maximizes f is given by

$$q(x|y) = \frac{r(x)p(y|x)}{\sum_{x'} r(x')p(y|x')}. \tag{1}$$

2. By Lagrange multipliers, it can be shown that for any given $\mathbf{q} \in A_2$, the unique input distribution $\mathbf{r}$ that maximizes f is given by

$$r(x) = \frac{\prod_y q(x|y)^{p(y|x)}}{\sum_{x'} \prod_y q(x'|y)^{p(y|x')}}, \tag{2}$$

where $\prod_y$ is over all y such that $p(y|x) > 0$.

3. Let $\mathbf{r}^{(0)}$ be an arbitrarily chosen strictly positive input distribution in A_1. Then $\mathbf{q}^{(0)} \in A_2$ can be computed according to (1). This forms $(\mathbf{r}^{(0)}, \mathbf{q}^{(0)})$.

Algorithm Details

$$f^* = \sup_{\mathbf{r} \in A_1} \sup_{\mathbf{q} \in A_2} f(\mathbf{r}, \mathbf{q})$$

$$f(\mathbf{r}, \mathbf{q}) = \sum_x \sum_y r(x) p(y|x) \log \frac{q(x|y)}{r(x)}$$

1. By Lemma 9.1, for any given $\mathbf{r} \in A_1$, the unique $\mathbf{q} \in A_2$ that maximizes f is given by

$$q(x|y) = \frac{r(x)p(y|x)}{\sum_{x'} r(x')p(y|x')}. \qquad (1)$$

2. By Lagrange multipliers, it can be shown that for any given $\mathbf{q} \in A_2$, the unique input distribution $\mathbf{r}$ that maximizes f is given by

$$r(x) = \frac{\prod_y q(x|y)^{p(y|x)}}{\sum_{x'} \prod_y q(x'|y)^{p(y|x')}}, \qquad (2)$$

where $\prod_y$ is over all y such that $p(y|x) > 0$.

3. Let $\mathbf{r}^{(0)}$ be an arbitrarily chosen strictly positive input distribution in A_1. Then $\mathbf{q}^{(0)} \in A_2$ can be computed according to (1). This forms $(\mathbf{r}^{(0)}, \mathbf{q}^{(0)})$.

Algorithm Details

$$f^* = \sup_{\mathbf{r} \in A_1} \sup_{\mathbf{q} \in A_2} f(\mathbf{r}, \mathbf{q})$$

$$f(\mathbf{r}, \mathbf{q}) = \sum_x \sum_y r(x) p(y|x) \log \frac{q(x|y)}{r(x)}$$

1. By Lemma 9.1, for any given $\mathbf{r} \in A_1$, the unique $\mathbf{q} \in A_2$ that maximizes f is given by

$$q(x|y) = \frac{r(x)p(y|x)}{\sum_{x'} r(x')p(y|x')}. \qquad (1)$$

2. By Lagrange multipliers, it can be shown that for any given $\mathbf{q} \in A_2$, the unique input distribution $\mathbf{r}$ that maximizes f is given by

$$r(x) = \frac{\prod_y q(x|y)^{p(y|x)}}{\sum_{x'} \prod_y q(x'|y)^{p(y|x')}}, \qquad (2)$$

where $\prod_y$ is over all y such that $p(y|x) > 0$.

3. Let $\mathbf{r}^{(0)}$ be an arbitrarily chosen strictly positive input distribution in A_1. Then $\mathbf{q}^{(0)} \in A_2$ can be computed according to (1). This forms $(\mathbf{r}^{(0)}, \mathbf{q}^{(0)})$.

4. Compute $\mathbf{r}^{(1)}$, $\mathbf{q}^{(1)}$, $\mathbf{r}^{(2)}$, $\mathbf{q}^{(2)}$, $\cdots$ iteratively by applying (2) and (1) alternately.

Algorithm Details

$$f^* = \sup_{\mathbf{r} \in A_1} \sup_{\mathbf{q} \in A_2} f(\mathbf{r}, \mathbf{q})$$

$$f(\mathbf{r}, \mathbf{q}) = \sum_x \sum_y r(x) p(y|x) \log \frac{q(x|y)}{r(x)}$$

1. By Lemma 9.1, for any given $\mathbf{r} \in A_1$, the unique $\mathbf{q} \in A_2$ that maximizes f is given by

$$q(x|y) = \frac{r(x)p(y|x)}{\sum_{x'} r(x')p(y|x')}. \qquad (1)$$

2. By Lagrange multipliers, it can be shown that for any given $\mathbf{q} \in A_2$, the unique input distribution $\mathbf{r}$ that maximizes f is given by

$$r(x) = \frac{\prod_y q(x|y)^{p(y|x)}}{\sum_{x'} \prod_y q(x'|y)^{p(y|x')}}, \qquad (2)$$

where $\prod_y$ is over all y such that $p(y|x) > 0$.

3. Let $\mathbf{r}^{(0)}$ be an arbitrarily chosen strictly positive input distribution in A_1. Then $\mathbf{q}^{(0)} \in A_2$ can be computed according to (1). This forms $(\mathbf{r}^{(0)}, \mathbf{q}^{(0)})$.

4. Compute $\mathbf{r}^{(1)}$, $\mathbf{q}^{(1)}$, $\mathbf{r}^{(2)}$, $\mathbf{q}^{(2)}, \cdots$ iteratively by applying (2) and (1) alternately.

Algorithm Details

$$f^* = \sup_{\mathbf{r}\in A_1} \sup_{\mathbf{q}\in A_2} f(\mathbf{r},\mathbf{q})$$

$$f(\mathbf{r},\mathbf{q}) = \sum_x \sum_y r(x)p(y|x)\log\frac{q(x|y)}{r(x)}$$

1. By Lemma 9.1, for any given $\mathbf{r} \in A_1$, the unique $\mathbf{q} \in A_2$ that maximizes f is given by

$$q(x|y) = \frac{r(x)p(y|x)}{\sum_{x'} r(x')p(y|x')}. \qquad (1)$$

2. By Lagrange multipliers, it can be shown that for any given $\mathbf{q} \in A_2$, the unique input distribution $\mathbf{r}$ that maximizes f is given by

$$r(x) = \frac{\prod_y q(x|y)^{p(y|x)}}{\sum_{x'} \prod_y q(x'|y)^{p(y|x')}}, \qquad (2)$$

where $\prod_y$ is over all y such that $p(y|x) > 0$.

3. Let $\mathbf{r}^{(0)}$ be an arbitrarily chosen strictly positive input distribution in A_1. Then $\mathbf{q}^{(0)} \in A_2$ can be computed according to (1). This forms $(\mathbf{r}^{(0)}, \mathbf{q}^{(0)})$.

4. Compute $\mathbf{r}^{(1)}$, $\mathbf{q}^{(1)}$, $\mathbf{r}^{(2)}$, $\mathbf{q}^{(2)}$, $\cdots$ iteratively by applying (2) and (1) alternately.

Algorithm Details

$$f^* = \sup_{\mathbf{r} \in A_1} \sup_{\mathbf{q} \in A_2} f(\mathbf{r}, \mathbf{q})$$

$$f(\mathbf{r}, \mathbf{q}) = \sum_x \sum_y r(x) p(y|x) \log \frac{q(x|y)}{r(x)}$$

1. By Lemma 9.1, for any given $\mathbf{r} \in A_1$, the unique $\mathbf{q} \in A_2$ that maximizes f is given by

$$q(x|y) = \frac{r(x)p(y|x)}{\sum_{x'} r(x')p(y|x')}. \tag{1}$$

2. By Lagrange multipliers, it can be shown that for any given $\mathbf{q} \in A_2$, the unique input distribution $\mathbf{r}$ that maximizes f is given by

$$r(x) = \frac{\prod_y q(x|y)^{p(y|x)}}{\sum_{x'} \prod_y q(x'|y)^{p(y|x')}}, \tag{2}$$

where $\prod_y$ is over all y such that $p(y|x) > 0$.

3. Let $\mathbf{r}^{(0)}$ be an arbitrarily chosen strictly positive input distribution in A_1. Then $\mathbf{q}^{(0)} \in A_2$ can be computed according to (1). This forms $(\mathbf{r}^{(0)}, \mathbf{q}^{(0)})$.

4. Compute $\mathbf{r}^{(1)}, \mathbf{q}^{(1)}, \mathbf{r}^{(2)}, \mathbf{q}^{(2)}, \cdots$ iteratively by applying (2) and (1) alternately.

5. It can be verified from (1) that if $\mathbf{r}^{(k)} \in A_1$, i.e., $\mathbf{r}^{(k)} > 0$, then $q^{(k)}(x|y) > 0$ iff $p(y|x) > 0$, i.e., $\mathbf{q}^{(k)} \in A_2$.

Algorithm Details

$$f^* = \sup_{\mathbf{r} \in A_1} \sup_{\mathbf{q} \in A_2} f(\mathbf{r}, \mathbf{q})$$

$$f(\mathbf{r}, \mathbf{q}) = \sum_x \sum_y r(x) p(y|x) \log \frac{q(x|y)}{r(x)}$$

1. By Lemma 9.1, for any given $\mathbf{r} \in A_1$, the unique $\mathbf{q} \in A_2$ that maximizes f is given by

$$q(x|y) = \frac{r(x)p(y|x)}{\sum_{x'} r(x')p(y|x')}. \qquad (1)$$

2. By Lagrange multipliers, it can be shown that for any given $\mathbf{q} \in A_2$, the unique input distribution $\mathbf{r}$ that maximizes f is given by

$$r(x) = \frac{\prod_y q(x|y)^{p(y|x)}}{\sum_{x'} \prod_y q(x'|y)^{p(y|x')}}, \qquad (2)$$

where $\prod_y$ is over all y such that $p(y|x) > 0$.

3. Let $\mathbf{r}^{(0)}$ be an arbitrarily chosen strictly positive input distribution in A_1. Then $\mathbf{q}^{(0)} \in A_2$ can be computed according to (1). This forms $(\mathbf{r}^{(0)}, \mathbf{q}^{(0)})$.

4. Compute $\mathbf{r}^{(1)}$, $\mathbf{q}^{(1)}$, $\mathbf{r}^{(2)}$, $\mathbf{q}^{(2)}$, $\cdots$ iteratively by applying (2) and (1) alternately.

5. It can be verified from (1) that if $\mathbf{r}^{(k)} \in A_1$, i.e., $\mathbf{r}^{(k)} > 0$, then $q^{(k)}(x|y) > 0$ iff $p(y|x) > 0$, i.e., $\mathbf{q}^{(k)} \in A_2$.

Algorithm Details

$$f^* = \sup_{\mathbf{r}\in A_1} \sup_{\mathbf{q}\in A_2} f(\mathbf{r}, \mathbf{q})$$

$$f(\mathbf{r}, \mathbf{q}) = \sum_x \sum_y r(x)p(y|x) \log \frac{q(x|y)}{r(x)}$$

1. By Lemma 9.1, for any given $\mathbf{r} \in A_1$, the unique $\mathbf{q} \in A_2$ that maximizes f is given by

$$q(x|y) = \frac{r(x)p(y|x)}{\sum_{x'} r(x')p(y|x')}. \qquad (1)$$

2. By Lagrange multipliers, it can be shown that for any given $\mathbf{q} \in A_2$, the unique input distribution $\mathbf{r}$ that maximizes f is given by

$$r(x) = \frac{\prod_y q(x|y)^{p(y|x)}}{\sum_{x'} \prod_y q(x'|y)^{p(y|x')}}, \qquad (2)$$

where $\prod_y$ is over all y such that $p(y|x) > 0$.

3. Let $\mathbf{r}^{(0)}$ be an arbitrarily chosen strictly positive input distribution in A_1. Then $\mathbf{q}^{(0)} \in A_2$ can be computed according to (1). This forms $(\mathbf{r}^{(0)}, \mathbf{q}^{(0)})$.

4. Compute $\mathbf{r}^{(1)}, \mathbf{q}^{(1)}, \mathbf{r}^{(2)}, \mathbf{q}^{(2)}, \cdots$ iteratively by applying (2) and (1) alternately.

5. It can be verified from (1) that if $\mathbf{r}^{(k)} \in A_1$, i.e., $\mathbf{r}^{(k)} > 0$, then $q^{(k)}(x|y) > 0$ iff $p(y|x) > 0$, i.e., $\mathbf{q}^{(k)} \in A_2$.

Algorithm Details

$$f^* = \sup_{\mathbf{r} \in A_1} \sup_{\mathbf{q} \in A_2} f(\mathbf{r}, \mathbf{q})$$

$$f(\mathbf{r}, \mathbf{q}) = \sum_x \sum_y r(x) p(y|x) \log \frac{q(x|y)}{r(x)}$$

1. By Lemma 9.1, for any given $\mathbf{r} \in A_1$, the unique $\mathbf{q} \in A_2$ that maximizes f is given by

$$q(x|y) = \frac{r(x)p(y|x)}{\sum_{x'} r(x')p(y|x')}. \qquad (1)$$

2. By Lagrange multipliers, it can be shown that for any given $\mathbf{q} \in A_2$, the unique input distribution $\mathbf{r}$ that maximizes f is given by

$$r(x) = \frac{\prod_y q(x|y)^{p(y|x)}}{\sum_{x'} \prod_y q(x'|y)^{p(y|x')}}, \qquad (2)$$

where $\prod_y$ is over all y such that $p(y|x) > 0$.

3. Let $\mathbf{r}^{(0)}$ be an arbitrarily chosen strictly positive input distribution in A_1. Then $\mathbf{q}^{(0)} \in A_2$ can be computed according to (1). This forms $(\mathbf{r}^{(0)}, \mathbf{q}^{(0)})$.

4. Compute $\mathbf{r}^{(1)}, \mathbf{q}^{(1)}, \mathbf{r}^{(2)}, \mathbf{q}^{(2)}, \cdots$ iteratively by applying (2) and (1) alternately.

5. It can be verified from (1) that if $\mathbf{r}^{(k)} \in A_1$, i.e., $\mathbf{r}^{(k)} > 0$, then $q^{(k)}(x|y) > 0$ iff $p(y|x) > 0$, i.e., $\mathbf{q}^{(k)} \in A_2$.

Algorithm Details

$$f^* = \sup_{\mathbf{r}\in A_1} \sup_{\mathbf{q}\in A_2} f(\mathbf{r},\mathbf{q})$$

$$f(\mathbf{r},\mathbf{q}) = \sum_x \sum_y r(x)p(y|x) \log \frac{q(x|y)}{r(x)}$$

1. By Lemma 9.1, for any given $\mathbf{r} \in A_1$, the unique $\mathbf{q} \in A_2$ that maximizes f is given by

$$q(x|y) = \frac{r(x)p(y|x)}{\sum_{x'} r(x')p(y|x')}. \tag{1}$$

2. By Lagrange multipliers, it can be shown that for any given $\mathbf{q} \in A_2$, the unique input distribution $\mathbf{r}$ that maximizes f is given by

$$r(x) = \frac{\prod_y q(x|y)^{p(y|x)}}{\sum_{x'} \prod_y q(x'|y)^{p(y|x')}}, \tag{2}$$

where $\prod_y$ is over all y such that $p(y|x) > 0$.

3. Let $\mathbf{r}^{(0)}$ be an arbitrarily chosen strictly positive input distribution in A_1. Then $\mathbf{q}^{(0)} \in A_2$ can be computed according to (1). This forms $(\mathbf{r}^{(0)}, \mathbf{q}^{(0)})$.

4. Compute $\mathbf{r}^{(1)}$, $\mathbf{q}^{(1)}$, $\mathbf{r}^{(2)}$, $\mathbf{q}^{(2)}$, $\cdots$ iteratively by applying (2) and (1) alternately.

5. It can be verified from (1) that if $\mathbf{r}^{(k)} \in A_1$, i.e., $\mathbf{r}^{(k)} > 0$, then $q^{(k)}(x|y) > 0$ iff $p(y|x) > 0$, i.e., $\mathbf{q}^{(k)} \in A_2$.

6. Likewise, it can be verified from (2) that if $\mathbf{q}^{(k)} \in A_2$, then $\mathbf{r}^{(k+1)} > 0$, i.e., $\mathbf{r}^{(k+1)} \in A_1$.

Algorithm Details

$$f^* = \sup_{\mathbf{r} \in A_1} \sup_{\mathbf{q} \in A_2} f(\mathbf{r}, \mathbf{q})$$

$$f(\mathbf{r}, \mathbf{q}) = \sum_x \sum_y r(x) p(y|x) \log \frac{q(x|y)}{r(x)}$$

1. By Lemma 9.1, for any given $\mathbf{r} \in A_1$, the unique $\mathbf{q} \in A_2$ that maximizes f is given by

$$q(x|y) = \frac{r(x)p(y|x)}{\sum_{x'} r(x')p(y|x')}. \qquad (1)$$

2. By Lagrange multipliers, it can be shown that for any given $\mathbf{q} \in A_2$, the unique input distribution $\mathbf{r}$ that maximizes f is given by

$$r(x) = \frac{\prod_y q(x|y)^{p(y|x)}}{\sum_{x'} \prod_y q(x'|y)^{p(y|x')}}, \qquad (2)$$

where $\prod_y$ is over all y such that $p(y|x) > 0$.

3. Let $\mathbf{r}^{(0)}$ be an arbitrarily chosen strictly positive input distribution in A_1. Then $\mathbf{q}^{(0)} \in A_2$ can be computed according to (1). This forms $(\mathbf{r}^{(0)}, \mathbf{q}^{(0)})$.

4. Compute $\mathbf{r}^{(1)}, \mathbf{q}^{(1)}, \mathbf{r}^{(2)}, \mathbf{q}^{(2)}, \cdots$ iteratively by applying (2) and (1) alternately.

5. It can be verified from (1) that if $\mathbf{r}^{(k)} \in A_1$, i.e., $\mathbf{r}^{(k)} > 0$, then $q^{(k)}(x|y) > 0$ iff $p(y|x) > 0$, i.e., $\mathbf{q}^{(k)} \in A_2$.

6. Likewise, it can be verified from (2) that if $\mathbf{q}^{(k)} \in A_2$, then $\mathbf{r}^{(k+1)} > 0$, i.e., $\mathbf{r}^{(k+1)} \in A_1$.

Algorithm Details

$$f^* = \sup_{\mathbf{r} \in A_1} \sup_{\mathbf{q} \in A_2} f(\mathbf{r}, \mathbf{q})$$

$$f(\mathbf{r}, \mathbf{q}) = \sum_x \sum_y r(x) p(y|x) \log \frac{q(x|y)}{r(x)}$$

1. By Lemma 9.1, for any given $\mathbf{r} \in A_1$, the unique $\mathbf{q} \in A_2$ that maximizes f is given by

$$q(x|y) = \frac{r(x)p(y|x)}{\sum_{x'} r(x')p(y|x')}. \tag{1}$$

2. By Lagrange multipliers, it can be shown that for any given $\mathbf{q} \in A_2$, the unique input distribution $\mathbf{r}$ that maximizes f is given by

$$r(x) = \frac{\prod_y q(x|y)^{p(y|x)}}{\sum_{x'} \prod_y q(x'|y)^{p(y|x')}}, \tag{2}$$

where $\prod_y$ is over all y such that $p(y|x) > 0$.

3. Let $\mathbf{r}^{(0)}$ be an arbitrarily chosen strictly positive input distribution in A_1. Then $\mathbf{q}^{(0)} \in A_2$ can be computed according to (1). This forms $(\mathbf{r}^{(0)}, \mathbf{q}^{(0)})$.

4. Compute $\mathbf{r}^{(1)}$, $\mathbf{q}^{(1)}$, $\mathbf{r}^{(2)}$, $\mathbf{q}^{(2)}$, $\cdots$ iteratively by applying (2) and (1) alternately.

5. It can be verified from (1) that if $\mathbf{r}^{(k)} \in A_1$, i.e., $\mathbf{r}^{(k)} > 0$, then $q^{(k)}(x|y) > 0$ iff $p(y|x) > 0$, i.e., $\mathbf{q}^{(k)} \in A_2$.

6. Likewise, it can be verified from (2) that if $\mathbf{q}^{(k)} \in A_2$, then $\mathbf{r}^{(k+1)} > 0$, i.e., $\mathbf{r}^{(k+1)} \in A_1$.

Algorithm Details

$$f^* = \sup_{\mathbf{r} \in A_1} \sup_{\mathbf{q} \in A_2} f(\mathbf{r}, \mathbf{q})$$

$$f(\mathbf{r}, \mathbf{q}) = \sum_x \sum_y r(x) p(y|x) \log \frac{q(x|y)}{r(x)}$$

1. By Lemma 9.1, for any given $\mathbf{r} \in A_1$, the unique $\mathbf{q} \in A_2$ that maximizes f is given by

$$q(x|y) = \frac{r(x)p(y|x)}{\sum_{x'} r(x')p(y|x')}. \quad (1)$$

2. By Lagrange multipliers, it can be shown that for any given $\mathbf{q} \in A_2$, the unique input distribution $\mathbf{r}$ that maximizes f is given by

$$r(x) = \frac{\prod_y q(x|y)^{p(y|x)}}{\sum_{x'} \prod_y q(x'|y)^{p(y|x')}}, \quad (2)$$

where $\prod_y$ is over all y such that $p(y|x) > 0$.

3. Let $\mathbf{r}^{(0)}$ be an arbitrarily chosen strictly positive input distribution in A_1. Then $\mathbf{q}^{(0)} \in A_2$ can be computed according to (1). This forms $(\mathbf{r}^{(0)}, \mathbf{q}^{(0)})$.

4. Compute $\mathbf{r}^{(1)}$, $\mathbf{q}^{(1)}$, $\mathbf{r}^{(2)}$, $\mathbf{q}^{(2)}$, $\cdots$ iteratively by applying (2) and (1) alternately.

5. It can be verified from (1) that if $\mathbf{r}^{(k)} \in A_1$, i.e., $\mathbf{r}^{(k)} > 0$, then $q^{(k)}(x|y) > 0$ iff $p(y|x) > 0$, i.e., $\mathbf{q}^{(k)} \in A_2$.

6. Likewise, it can be verified from (2) that if $\mathbf{q}^{(k)} \in A_2$, then $\mathbf{r}^{(k+1)} > 0$, i.e., $\mathbf{r}^{(k+1)} \in A_1$.

Algorithm Details

$$f^* = \sup_{\mathbf{r} \in A_1} \sup_{\mathbf{q} \in A_2} f(\mathbf{r}, \mathbf{q})$$

$$f(\mathbf{r}, \mathbf{q}) = \sum_x \sum_y r(x) p(y|x) \log \frac{q(x|y)}{r(x)}$$

1. By Lemma 9.1, for any given $\mathbf{r} \in A_1$, the unique $\mathbf{q} \in A_2$ that maximizes f is given by

$$q(x|y) = \frac{r(x)p(y|x)}{\sum_{x'} r(x')p(y|x')}. \tag{1}$$

2. By Lagrange multipliers, it can be shown that for any given $\mathbf{q} \in A_2$, the unique input distribution $\mathbf{r}$ that maximizes f is given by

$$r(x) = \frac{\prod_y q(x|y)^{p(y|x)}}{\sum_{x'} \prod_y q(x'|y)^{p(y|x')}}, \tag{2}$$

where $\prod_y$ is over all y such that $p(y|x) > 0$.

3. Let $\mathbf{r}^{(0)}$ be an arbitrarily chosen strictly positive input distribution in A_1. Then $\mathbf{q}^{(0)} \in A_2$ can be computed according to (1). This forms $(\mathbf{r}^{(0)}, \mathbf{q}^{(0)})$.

4. Compute $\mathbf{r}^{(1)}, \mathbf{q}^{(1)}, \mathbf{r}^{(2)}, \mathbf{q}^{(2)}, \cdots$ iteratively by applying (2) and (1) alternately.

5. It can be verified from (1) that if $\mathbf{r}^{(k)} \in A_1$, i.e., $\mathbf{r}^{(k)} > 0$, then $q^{(k)}(x|y) > 0$ iff $p(y|x) > 0$, i.e., $\mathbf{q}^{(k)} \in A_2$.

6. Likewise, it can be verified from (2) that if $\mathbf{q}^{(k)} \in A_2$, then $\mathbf{r}^{(k+1)} > 0$, i.e., $\mathbf{r}^{(k+1)} \in A_1$.

7. Therefore, $\mathbf{r}^{(k)} \in A_1$ and $\mathbf{q}^{(k)} \in A_2$ for all $k \geq 0$.

Algorithm Details

$$f^* = \sup_{\mathbf{r}\in A_1} \sup_{\mathbf{q}\in A_2} f(\mathbf{r},\mathbf{q})$$

$$f(\mathbf{r},\mathbf{q}) = \sum_x \sum_y r(x)p(y|x) \log \frac{q(x|y)}{r(x)}$$

1. By Lemma 9.1, for any given $\mathbf{r} \in A_1$, the unique $\mathbf{q} \in A_2$ that maximizes f is given by

$$q(x|y) = \frac{r(x)p(y|x)}{\sum_{x'} r(x')p(y|x')}. \tag{1}$$

2. By Lagrange multipliers, it can be shown that for any given $\mathbf{q} \in A_2$, the unique input distribution $\mathbf{r}$ that maximizes f is given by

$$r(x) = \frac{\prod_y q(x|y)^{p(y|x)}}{\sum_{x'} \prod_y q(x'|y)^{p(y|x')}}, \tag{2}$$

where $\prod_y$ is over all y such that $p(y|x) > 0$.

3. Let $\mathbf{r}^{(0)}$ be an arbitrarily chosen strictly positive input distribution in A_1. Then $\mathbf{q}^{(0)} \in A_2$ can be computed according to (1). This forms $(\mathbf{r}^{(0)}, \mathbf{q}^{(0)})$.

4. Compute $\mathbf{r}^{(1)}$, $\mathbf{q}^{(1)}$, $\mathbf{r}^{(2)}$, $\mathbf{q}^{(2)}$, $\cdots$ iteratively by applying (2) and (1) alternately.

5. It can be verified from (1) that if $\mathbf{r}^{(k)} \in A_1$, i.e., $\mathbf{r}^{(k)} > 0$, then $q^{(k)}(x|y) > 0$ iff $p(y|x) > 0$, i.e., $\mathbf{q}^{(k)} \in A_2$.

6. Likewise, it can be verified from (2) that if $\mathbf{q}^{(k)} \in A_2$, then $\mathbf{r}^{(k+1)} > 0$, i.e., $\mathbf{r}^{(k+1)} \in A_1$.

7. Therefore, $\mathbf{r}^{(k)} \in A_1$ and $\mathbf{q}^{(k)} \in A_2$ for all $k \geq 0$.

8. It will be shown in Section 9.3 that $f^{(k)} = f(\mathbf{r}^{(k)}, \mathbf{q}^{(k)}) \to f^* = C$.

Maximizing $f(\mathbf{r},\mathbf{q})$ for a Fixed q

$$\max_{\mathbf{r}\in A_1} \sum_x \sum_y r(x)p(y|x)\log\frac{q(x|y)}{r(x)}$$

Maximizing *f*(r,q) for a Fixed q

$$\max_{\mathbf{r} \in A_1} \sum_x \sum_y r(x) p(y|x) \log \frac{q(x|y)}{r(x)}$$

1. The constraints on $\mathbf{r}$ are

$$\sum_x r(x) = 1 \tag{1}$$

Maximizing *f*(r,q) for a Fixed q

$$\boxed{\max_{\mathbf{r} \in A_1} \sum_x \sum_y r(x) p(y|x) \log \frac{q(x|y)}{r(x)}}$$

1. The constraints on $\mathbf{r}$ are

$$\sum_x r(x) = 1 \tag{1}$$

and

$$r(x) > 0 \quad \text{for all } x \in \mathcal{X}. \tag{2}$$

Maximizing f(r,q) for a Fixed q

$$\max_{\mathbf{r}\in A_1} \sum_x \sum_y r(x)p(y|x)\log\frac{q(x|y)}{r(x)}$$

1. The constraints on $\mathbf{r}$ are

$$\sum_x r(x) = 1 \tag{1}$$

and

$$r(x) > 0 \quad \text{for all } x \in \mathcal{X}. \tag{2}$$

2. Use the method of Lagrange multipliers to find the best $\mathbf{r}$ by ignoring temporarily the positivity constraints on $\mathbf{r}$ in (2). Let

Maximizing f(r,q) for a Fixed q

$$\boxed{\max_{\mathbf{r}\in A_1} \sum_x \sum_y r(x) p(y|x) \log \frac{q(x|y)}{r(x)}}$$

1. The constraints on $\mathbf{r}$ are

$$\sum_x r(x) = 1 \tag{1}$$

and

$$r(x) > 0 \quad \text{for all } x \in \mathcal{X}. \tag{2}$$

2. Use the method of Lagrange multipliers to find the best $\mathbf{r}$ by ignoring temporarily the positivity constraints on $\mathbf{r}$ in (2). Let

Maximizing f(r,q) for a Fixed q

$$\max_{\mathbf{r}\in A_1} \sum_x \sum_y r(x)p(y|x)\log\frac{q(x|y)}{r(x)}$$

1. The constraints on $\mathbf{r}$ are

$$\sum_x r(x) = 1 \tag{1}$$

and

$$r(x) > 0 \quad \text{for all } x \in \mathcal{X}. \tag{2}$$

2. Use the method of Lagrange multipliers to find the best $\mathbf{r}$ by ignoring temporarily the positivity constraints on $\mathbf{r}$ in (2). Let

$$J = \underline{\sum_x \sum_y r(x)p(y|x)\log\frac{q(x|y)}{r(x)}} - \lambda \sum_x r(x).$$

Maximizing f(r,q) for a Fixed q

$$\max_{\mathbf{r}\in A_1} \sum_x \sum_y r(x)p(y|x)\log\frac{q(x|y)}{r(x)}$$

1. The constraints on $\mathbf{r}$ are

$$\sum_x r(x) = 1 \tag{1}$$

and

$$r(x) > 0 \quad \text{for all } x \in \mathcal{X}. \tag{2}$$

2. Use the method of Lagrange multipliers to find the best $\mathbf{r}$ by ignoring temporarily the positivity constraints on $\mathbf{r}$ in (2). Let

$$J = \sum_x \sum_y r(x)p(y|x)\log\frac{q(x|y)}{r(x)} - \lambda \underline{\sum_x r(x)}.$$

Maximizing f(r,q) for a Fixed q

$$\boxed{\max_{\mathbf{r}\in A_1} \sum_x \sum_y r(x)p(y|x) \log \frac{q(x|y)}{r(x)}}$$

1. The constraints on $\mathbf{r}$ are

$$\sum_x r(x) = 1 \tag{1}$$

and

$$r(x) > 0 \quad \text{for all } x \in \mathcal{X}. \tag{2}$$

2. Use the method of Lagrange multipliers to find the best $\mathbf{r}$ by ignoring temporarily the positivity constraints on $\mathbf{r}$ in (2). Let

$$J = \sum_x \sum_y r(x)p(y|x) \log \frac{q(x|y)}{r(x)} - \lambda \sum_x r(x).$$

For convenience sake, we assume that the logarithm is the natural logarithm. Differentiating with respect to $r(x)$ gives

Maximizing f(r,q) for a Fixed q

$$\max_{\mathbf{r}\in A_1} \sum_x \sum_y r(x)p(y|x)\log\frac{q(x|y)}{r(x)}$$

1. The constraints on $\mathbf{r}$ are

$$\sum_x r(x) = 1 \tag{1}$$

and

$$r(x) > 0 \quad \text{for all } x \in \mathcal{X}. \tag{2}$$

2. Use the method of Lagrange multipliers to find the best $\mathbf{r}$ by ignoring temporarily the positivity constraints on $\mathbf{r}$ in (2). Let

$$J = \sum_x \sum_y r(x)p(y|x)\log\frac{q(x|y)}{r(x)} - \lambda \sum_x r(x).$$

For convenience sake, we assume that the logarithm is the natural logarithm. Differentiating with respect to $r(x)$ gives

$$\frac{\partial J}{\partial r(x)} = \sum_y p(y|x)\log q(x|y) - \log r(x) - 1 - \lambda.$$

Maximizing *f*(r,q) for a Fixed q

$$\boxed{\max_{\mathbf{r}\in A_1} \sum_x \sum_y r(x) p(y|x) \log \frac{q(x|y)}{r(x)}}$$

1. The constraints on $\mathbf{r}$ are

$$\sum_x r(x) = 1 \tag{1}$$

and

$$r(x) > 0 \quad \text{for all } x \in \mathcal{X}. \tag{2}$$

2. Use the method of Lagrange multipliers to find the best $\mathbf{r}$ by ignoring temporarily the positivity constraints on $\mathbf{r}$ in (2). Let

$$J = \sum_x \sum_y r(x) p(y|x) \log \frac{q(x|y)}{r(x)} - \lambda \sum_x r(x).$$

For convenience sake, we assume that the logarithm is the natural logarithm. Differentiating with respect to $r(x)$ gives

$$\frac{\partial J}{\partial r(x)} = \sum_y p(y|x) \log q(x|y) - \log r(x) - 1 - \lambda.$$

Upon setting $\frac{\partial J}{\partial r(x)} = 0$, we have

Maximizing $f(\mathbf{r},\mathbf{q})$ for a Fixed q

$$\max_{\mathbf{r}\in A_1} \sum_x \sum_y r(x)p(y|x) \log \frac{q(x|y)}{r(x)}$$

1. The constraints on $\mathbf{r}$ are

$$\sum_x r(x) = 1 \tag{1}$$

and

$$r(x) > 0 \quad \text{for all } x \in \mathcal{X}. \tag{2}$$

2. Use the method of Lagrange multipliers to find the best $\mathbf{r}$ by ignoring temporarily the positivity constraints on $\mathbf{r}$ in (2). Let

$$J = \sum_x \sum_y r(x)p(y|x) \log \frac{q(x|y)}{r(x)} - \lambda \sum_x r(x).$$

For convenience sake, we assume that the logarithm is the natural logarithm. Differentiating with respect to $r(x)$ gives

$$\cancel{\frac{\partial J}{\partial r(x)}}^{\,0} = \sum_y p(y|x) \log q(x|y) - \log r(x) - 1 - \lambda.$$

Upon setting $\frac{\partial J}{\partial r(x)} = 0$, we have

Maximizing *f*(r,q) for a Fixed q

$$\boxed{\max_{\mathbf{r}\in A_1} \sum_x \sum_y r(x)p(y|x) \log \frac{q(x|y)}{r(x)}}$$

1. The constraints on $\mathbf{r}$ are

$$\sum_x r(x) = 1 \tag{1}$$

and

$$r(x) > 0 \quad \text{for all } x \in \mathcal{X}. \tag{2}$$

2. Use the method of Lagrange multipliers to find the best $\mathbf{r}$ by ignoring temporarily the positivity constraints on $\mathbf{r}$ in (2). Let

$$J = \sum_x \sum_y r(x)p(y|x) \log \frac{q(x|y)}{r(x)} - \lambda \sum_x r(x).$$

For convenience sake, we assume that the logarithm is the natural logarithm. Differentiating with respect to $r(x)$ gives

$$\cancelto{0}{\frac{\partial J}{\partial r(x)}} = \sum_y p(y|x) \log q(x|y) - \underline{\log r(x)} - 1 - \lambda.$$

Upon setting $\frac{\partial J}{\partial r(x)} = 0$, we have

Maximizing *f*(r,q) for a Fixed q

$$\boxed{\max_{\mathbf{r}\in A_1} \sum_x \sum_y r(x)p(y|x) \log \frac{q(x|y)}{r(x)}}$$

1. The constraints on $\mathbf{r}$ are

$$\sum_x r(x) = 1 \tag{1}$$

and

$$r(x) > 0 \quad \text{for all } x \in \mathcal{X}. \tag{2}$$

2. Use the method of Lagrange multipliers to find the best $\mathbf{r}$ by ignoring temporarily the positivity constraints on $\mathbf{r}$ in (2). Let

$$J = \sum_x \sum_y r(x)p(y|x) \log \frac{q(x|y)}{r(x)} - \lambda \sum_x r(x).$$

For convenience sake, we assume that the logarithm is the natural logarithm. Differentiating with respect to $r(x)$ gives

$$\cancelto{0}{\frac{\partial J}{\partial r(x)}} = \sum_y p(y|x) \log q(x|y) - \underline{\log r(x)} - 1 - \lambda.$$

Upon setting $\frac{\partial J}{\partial r(x)} = 0$, we have

$$\underline{\log r(x)} = \sum_y p(y|x) \log q(x|y) - 1 - \lambda,$$

Maximizing f(r,q) for a Fixed q

$$\max_{\mathbf{r}\in A_1} \sum_x \sum_y r(x) p(y|x) \log \frac{q(x|y)}{r(x)}$$

1. The constraints on $\mathbf{r}$ are

$$\sum_x r(x) = 1 \tag{1}$$

and

$$r(x) > 0 \quad \text{for all } x \in \mathcal{X}. \tag{2}$$

2. Use the method of Lagrange multipliers to find the best $\mathbf{r}$ by ignoring temporarily the positivity constraints on $\mathbf{r}$ in (2). Let

$$J = \sum_x \sum_y r(x) p(y|x) \log \frac{q(x|y)}{r(x)} - \lambda \sum_x r(x).$$

For convenience sake, we assume that the logarithm is the natural logarithm. Differentiating with respect to $r(x)$ gives

$$\frac{\partial J}{\partial r(x)} = \sum_y p(y|x) \log q(x|y) - \log r(x) - 1 - \lambda.$$

Upon setting $\frac{\partial J}{\partial r(x)} = 0$, we have

$$\log r(x) = \sum_y p(y|x) \log q(x|y) - 1 - \lambda,$$

Maximizing *f*(r,q) for a Fixed q

$$\max_{\mathbf{r} \in A_1} \sum_x \sum_y r(x) p(y|x) \log \frac{q(x|y)}{r(x)}$$

1. The constraints on $\mathbf{r}$ are

$$\sum_x r(x) = 1 \tag{1}$$

and

$$r(x) > 0 \quad \text{for all } x \in \mathcal{X}. \tag{2}$$

2. Use the method of Lagrange multipliers to find the best $\mathbf{r}$ by ignoring temporarily the positivity constraints on $\mathbf{r}$ in (2). Let

$$J = \sum_x \sum_y r(x) p(y|x) \log \frac{q(x|y)}{r(x)} - \lambda \sum_x r(x).$$

For convenience sake, we assume that the logarithm is the natural logarithm. Differentiating with respect to $r(x)$ gives

$$\frac{\partial J}{\partial r(x)} = \sum_y p(y|x) \log q(x|y) - \log r(x) - 1 - \lambda.$$

Upon setting $\frac{\partial J}{\partial r(x)} = 0$, we have

$$\log r(x) = \sum_y p(y|x) \log q(x|y) - 1 - \lambda,$$

or

$$r(x) = e^{-(\lambda+1)} \prod_y q(x|y)^{p(y|x)}.$$

Maximizing $f(r,q)$ for a Fixed q

$$\max_{\mathbf{r}\in A_1} \sum_x \sum_y r(x)p(y|x)\log\frac{q(x|y)}{r(x)}$$

1. The constraints on $\mathbf{r}$ are

$$\sum_x r(x) = 1 \tag{1}$$

and

$$r(x) > 0 \quad \text{for all } x \in \mathcal{X}. \tag{2}$$

2. Use the method of Lagrange multipliers to find the best $\mathbf{r}$ by ignoring temporarily the positivity constraints on $\mathbf{r}$ in (2). Let

$$J = \sum_x \sum_y r(x)p(y|x)\log\frac{q(x|y)}{r(x)} - \lambda \sum_x r(x).$$

For convenience sake, we assume that the logarithm is the natural logarithm. Differentiating with respect to $r(x)$ gives

$$\frac{\partial J}{\partial r(x)} = \sum_y p(y|x)\log q(x|y) - \log r(x) - 1 - \lambda.$$

Upon setting $\frac{\partial J}{\partial r(x)} = 0$, we have

$$\log r(x) = \sum_y p(y|x)\log q(x|y) - 1 - \lambda,$$

or

$$r(x) = e^{-(\lambda+1)} \prod_y q(x|y)^{p(y|x)}.$$

3. By considering the normalization constraint in (1), we can eliminate λ and obtain

Maximizing *f*(r,q) for a Fixed q

$$\max_{\mathbf{r}\in A_1} \sum_x \sum_y r(x)p(y|x) \log \frac{q(x|y)}{r(x)}$$

1. The constraints on $\mathbf{r}$ are

$$\sum_x r(x) = 1 \tag{1}$$

and

$$r(x) > 0 \quad \text{for all } x \in \mathcal{X}. \tag{2}$$

2. Use the method of Lagrange multipliers to find the best $\mathbf{r}$ by ignoring temporarily the positivity constraints on $\mathbf{r}$ in (2). Let

$$J = \sum_x \sum_y r(x)p(y|x) \log \frac{q(x|y)}{r(x)} - \lambda \sum_x r(x).$$

For convenience sake, we assume that the logarithm is the natural logarithm. Differentiating with respect to $r(x)$ gives

$$\frac{\partial J}{\partial r(x)} = \sum_y p(y|x) \log q(x|y) - \log r(x) - 1 - \lambda.$$

Upon setting $\frac{\partial J}{\partial r(x)} = 0$, we have

$$\log r(x) = \sum_y p(y|x) \log q(x|y) - 1 - \lambda,$$

or

$$r(x) = e^{-(\lambda+1)} \prod_y q(x|y)^{p(y|x)}.$$

3. By considering the normalization constraint in (1), we can eliminate λ and obtain

Maximizing *f*(r,q) for a Fixed q

$$\boxed{\max_{\mathbf{r}\in A_1} \sum_x \sum_y r(x)p(y|x)\log \frac{q(x|y)}{r(x)}}$$

1. The constraints on $\mathbf{r}$ are

$$\sum_x r(x) = 1 \tag{1}$$

and

$$r(x) > 0 \quad \text{for all } x \in \mathcal{X}. \tag{2}$$

2. Use the method of Lagrange multipliers to find the best $\mathbf{r}$ by ignoring temporarily the positivity constraints on $\mathbf{r}$ in (2). Let

$$J = \sum_x \sum_y r(x)p(y|x)\log \frac{q(x|y)}{r(x)} - \lambda \sum_x r(x).$$

For convenience sake, we assume that the logarithm is the natural logarithm. Differentiating with respect to $r(x)$ gives

$$\frac{\partial J}{\partial r(x)} = \sum_y p(y|x)\log q(x|y) - \log r(x) - 1 - \lambda.$$

Upon setting $\frac{\partial J}{\partial r(x)} = 0$, we have

$$\log r(x) = \sum_y p(y|x)\log q(x|y) - 1 - \lambda,$$

or

$$r(x) = \underline{e^{-(\lambda+1)}} \prod_y q(x|y)^{p(y|x)}.$$

3. By considering the normalization constraint in (1), we can eliminate λ and obtain

Maximizing f(r,q) for a Fixed q

$$\boxed{\max_{\mathbf{r} \in A_1} \sum_x \sum_y r(x) p(y|x) \log \frac{q(x|y)}{r(x)}}$$

1. The constraints on $\mathbf{r}$ are

$$\sum_x r(x) = 1 \tag{1}$$

and

$$r(x) > 0 \quad \text{for all } x \in \mathcal{X}. \tag{2}$$

2. Use the method of Lagrange multipliers to find the best $\mathbf{r}$ by ignoring temporarily the positivity constraints on $\mathbf{r}$ in (2). Let

$$J = \sum_x \sum_y r(x) p(y|x) \log \frac{q(x|y)}{r(x)} - \lambda \sum_x r(x).$$

For convenience sake, we assume that the logarithm is the natural logarithm. Differentiating with respect to $r(x)$ gives

$$\frac{\partial J}{\partial r(x)} = \sum_y p(y|x) \log q(x|y) - \log r(x) - 1 - \lambda.$$

Upon setting $\frac{\partial J}{\partial r(x)} = 0$, we have

$$\log r(x) = \sum_y p(y|x) \log q(x|y) - 1 - \lambda,$$

or

$$r(x) = \underline{e^{-(\lambda+1)}} \prod_y q(x|y)^{p(y|x)}.$$

3. By considering the normalization constraint in (1), we can eliminate λ and obtain

$$r(x) = \frac{\prod_y q(x|y)^{p(y|x)}}{\sum_{x'} \prod_y q(x'|y)^{p(y|x')}}. \tag{3}$$

Maximizing *f*(r,q) for a Fixed q

$$\boxed{\max_{\mathbf{r}\in A_1} \sum_x \sum_y r(x)p(y|x)\log\frac{q(x|y)}{r(x)}}$$

1. The constraints on $\mathbf{r}$ are

$$\sum_x r(x) = 1 \tag{1}$$

and

$$r(x) > 0 \quad \text{for all } x \in \mathcal{X}. \tag{2}$$

2. Use the method of Lagrange multipliers to find the best $\mathbf{r}$ by ignoring temporarily the positivity constraints on $\mathbf{r}$ in (2). Let

$$J = \sum_x \sum_y r(x)p(y|x)\log\frac{q(x|y)}{r(x)} - \lambda\sum_x r(x).$$

For convenience sake, we assume that the logarithm is the natural logarithm. Differentiating with respect to $r(x)$ gives

$$\frac{\partial J}{\partial r(x)} = \sum_y p(y|x)\log q(x|y) - \log r(x) - 1 - \lambda.$$

Upon setting $\frac{\partial J}{\partial r(x)} = 0$, we have

$$\log r(x) = \sum_y p(y|x)\log q(x|y) - 1 - \lambda,$$

or

$$r(x) = \underline{e^{-(\lambda+1)}} \prod_y q(x|y)^{p(y|x)}.$$

3. By considering the normalization constraint in (1), we can eliminate λ and obtain

$$r(x) = \frac{\prod_y q(x|y)^{p(y|x)}}{\sum_{x'}\prod_y q(x'|y)^{p(y|x')}}. \tag{3}$$

Maximizing $f(\mathbf{r},\mathbf{q})$ for a Fixed q

$$\boxed{\max_{\mathbf{r} \in A_1} \sum_x \sum_y r(x) p(y|x) \log \frac{q(x|y)}{r(x)}}$$

1. The constraints on $\mathbf{r}$ are

$$\sum_x r(x) = 1 \tag{1}$$

and

$$r(x) > 0 \quad \text{for all } x \in \mathcal{X}. \tag{2}$$

2. Use the method of Lagrange multipliers to find the best $\mathbf{r}$ by ignoring temporarily the positivity constraints on $\mathbf{r}$ in (2). Let

$$J = \sum_x \sum_y r(x) p(y|x) \log \frac{q(x|y)}{r(x)} - \lambda \sum_x r(x).$$

For convenience sake, we assume that the logarithm is the natural logarithm. Differentiating with respect to $r(x)$ gives

$$\frac{\partial J}{\partial r(x)} = \sum_y p(y|x) \log q(x|y) - \log r(x) - 1 - \lambda.$$

Upon setting $\frac{\partial J}{\partial r(x)} = 0$, we have

$$\log r(x) = \sum_y p(y|x) \log q(x|y) - 1 - \lambda,$$

or

$$r(x) = \underline{e^{-(\lambda+1)}} \prod_y q(x|y)^{p(y|x)}.$$

3. By considering the normalization constraint in (1), we can eliminate λ and obtain

$$r(x) = \frac{\prod_y q(x|y)^{p(y|x)}}{\sum_{x'} \prod_y q(x'|y)^{p(y|x')}}. \tag{3}$$

Maximizing f(r,q) for a Fixed q

$$\boxed{\max_{\mathbf{r}\in A_1} \sum_x \sum_y r(x)p(y|x)\log\frac{q(x|y)}{r(x)}}$$

1. The constraints on $\mathbf{r}$ are

$$\sum_x r(x) = 1 \tag{1}$$

and

$$r(x) > 0 \quad \text{for all } x \in \mathcal{X}. \tag{2}$$

2. Use the method of Lagrange multipliers to find the best $\mathbf{r}$ by ignoring temporarily the positivity constraints on $\mathbf{r}$ in (2). Let

$$J = \sum_x \sum_y r(x)p(y|x)\log\frac{q(x|y)}{r(x)} - \lambda \sum_x r(x).$$

For convenience sake, we assume that the logarithm is the natural logarithm. Differentiating with respect to $r(x)$ gives

$$\frac{\partial J}{\partial r(x)} = \sum_y p(y|x)\log q(x|y) - \log r(x) - 1 - \lambda.$$

Upon setting $\frac{\partial J}{\partial r(x)} = 0$, we have

$$\log r(x) = \sum_y p(y|x)\log q(x|y) - 1 - \lambda,$$

or

$$r(x) = \underline{e^{-(\lambda+1)}} \prod_y q(x|y)^{p(y|x)}.$$

3. By considering the normalization constraint in (1), we can eliminate λ and obtain

$$r(x) = \frac{\prod_y q(x|y)^{p(y|x)}}{\sum_{x'} \prod_y q(x'|y)^{p(y|x')}}. \tag{3}$$

Maximizing f(r,q) for a Fixed q

$$\boxed{\max_{\mathbf{r}\in A_1} \sum_x \sum_y r(x)p(y|x) \log \frac{q(x|y)}{r(x)}}$$

1. The constraints on $\mathbf{r}$ are

$$\sum_x r(x) = 1 \tag{1}$$

and

$$r(x) > 0 \quad \text{for all } x \in \mathcal{X}. \tag{2}$$

2. Use the method of Lagrange multipliers to find the best $\mathbf{r}$ by ignoring temporarily the positivity constraints on $\mathbf{r}$ in (2). Let

$$J = \sum_x \sum_y r(x)p(y|x) \log \frac{q(x|y)}{r(x)} - \lambda \sum_x r(x).$$

For convenience sake, we assume that the logarithm is the natural logarithm. Differentiating with respect to $r(x)$ gives

$$\frac{\partial J}{\partial r(x)} = \sum_y p(y|x) \log q(x|y) - \log r(x) - 1 - \lambda.$$

Upon setting $\frac{\partial J}{\partial r(x)} = 0$, we have

$$\log r(x) = \sum_y p(y|x) \log q(x|y) - 1 - \lambda,$$

or

$$r(x) = \underline{e^{-(\lambda+1)}} \prod_y q(x|y)^{p(y|x)}.$$

3. By considering the normalization constraint in (1), we can eliminate λ and obtain

$$r(x) = \frac{\prod_y q(x|y)^{p(y|x)}}{\sum_{x'} \prod_y q(x'|y)^{p(y|x')}}. \tag{3}$$

Maximizing $f(\mathbf{r},\mathbf{q})$ for a Fixed $\mathbf{q}$

$$\max_{\mathbf{r}\in A_1} \sum_x \sum_y r(x)p(y|x)\log\frac{q(x|y)}{r(x)}$$

1. The constraints on $\mathbf{r}$ are

$$\sum_x r(x) = 1 \tag{1}$$

and

$$r(x) > 0 \quad \text{for all } x \in \mathcal{X}. \tag{2}$$

2. Use the method of Lagrange multipliers to find the best $\mathbf{r}$ by ignoring temporarily the positivity constraints on $\mathbf{r}$ in (2). Let

$$J = \sum_x \sum_y r(x)p(y|x)\log\frac{q(x|y)}{r(x)} - \lambda \sum_x r(x).$$

For convenience sake, we assume that the logarithm is the natural logarithm. Differentiating with respect to $r(x)$ gives

$$\frac{\partial J}{\partial r(x)} = \sum_y p(y|x)\log q(x|y) - \log r(x) - 1 - \lambda.$$

Upon setting $\frac{\partial J}{\partial r(x)} = 0$, we have

$$\log r(x) = \sum_y p(y|x)\log q(x|y) - 1 - \lambda,$$

or

$$r(x) = e^{-(\lambda+1)} \prod_y q(x|y)^{p(y|x)}.$$

3. By considering the normalization constraint in (1), we can eliminate λ and obtain

$$r(x) = \frac{\prod_y q(x|y)^{p(y|x)}}{\sum_{x'} \prod_y q(x'|y)^{p(y|x')}}. \tag{3}$$

4. The above product is over all y such that $p(y|x) > 0$, and $q(x|y) > 0$ for all such y. This implies that both the numerator and the denominator on the right hand side above are positive, and therefore $r(x) > 0$.

Maximizing f(r,q) for a Fixed q

$$\max_{\mathbf{r} \in A_1} \sum_x \sum_y r(x) p(y|x) \log \frac{q(x|y)}{r(x)}$$

1. The constraints on $\mathbf{r}$ are

$$\sum_x r(x) = 1 \tag{1}$$

and

$$r(x) > 0 \quad \text{for all } x \in \mathcal{X}. \tag{2}$$

2. Use the method of Lagrange multipliers to find the best $\mathbf{r}$ by ignoring temporarily the positivity constraints on $\mathbf{r}$ in (2). Let

$$J = \sum_x \sum_y r(x) p(y|x) \log \frac{q(x|y)}{r(x)} - \lambda \sum_x r(x).$$

For convenience sake, we assume that the logarithm is the natural logarithm. Differentiating with respect to $r(x)$ gives

$$\frac{\partial J}{\partial r(x)} = \sum_y p(y|x) \log q(x|y) - \log r(x) - 1 - \lambda.$$

Upon setting $\frac{\partial J}{\partial r(x)} = 0$, we have

$$\log r(x) = \sum_y p(y|x) \log q(x|y) - 1 - \lambda,$$

or

$$r(x) = e^{-(\lambda+1)} \prod_y q(x|y)^{p(y|x)}.$$

3. By considering the normalization constraint in (1), we can eliminate λ and obtain

$$r(x) = \frac{\prod_y q(x|y)^{\underline{p(y|x)}}}{\sum_{x'} \prod_y q(x'|y)^{p(y|x')}}. \tag{3}$$

4. The above product is over all y such that $p(y|x) > 0$, and $q(x|y) > 0$ for all such y. This implies that both the numerator and the denominator on the right hand side above are positive, and therefore $r(x) > 0$.

Maximizing $f(\mathbf{r},\mathbf{q})$ for a Fixed q

$$\boxed{\max_{\mathbf{r}\in A_1} \sum_x \sum_y r(x)p(y|x)\log \frac{q(x|y)}{r(x)}}$$

1. The constraints on $\mathbf{r}$ are

$$\sum_x r(x) = 1 \tag{1}$$

and

$$r(x) > 0 \quad \text{for all } x \in \mathcal{X}. \tag{2}$$

2. Use the method of Lagrange multipliers to find the best $\mathbf{r}$ by ignoring temporarily the positivity constraints on $\mathbf{r}$ in (2). Let

$$J = \sum_x \sum_y r(x)p(y|x)\log \frac{q(x|y)}{r(x)} - \lambda \sum_x r(x).$$

For convenience sake, we assume that the logarithm is the natural logarithm. Differentiating with respect to $r(x)$ gives

$$\frac{\partial J}{\partial r(x)} = \sum_y p(y|x)\log q(x|y) - \log r(x) - 1 - \lambda.$$

Upon setting $\frac{\partial J}{\partial r(x)} = 0$, we have

$$\log r(x) = \sum_y p(y|x)\log q(x|y) - 1 - \lambda,$$

or

$$r(x) = e^{-(\lambda+1)} \prod_y q(x|y)^{p(y|x)}.$$

3. By considering the normalization constraint in (1), we can eliminate λ and obtain

$$r(x) = \frac{\prod_y \underline{q(x|y)^{p(y|x)}}}{\sum_{x'} \prod_y q(x'|y)^{p(y|x')}}. \tag{3}$$

4. The above product is over all y such that $p(y|x) > 0$, and $q(x|y) > 0$ for all such y. This implies that both the numerator and the denominator on the right hand side above are positive, and therefore $r(x) > 0$.

Maximizing $f(\mathbf{r},\mathbf{q})$ for a Fixed $\mathbf{q}$

$$\max_{\mathbf{r}\in A_1} \sum_x \sum_y r(x) p(y|x) \log \frac{q(x|y)}{r(x)}$$

1. The constraints on $\mathbf{r}$ are

$$\sum_x r(x) = 1 \tag{1}$$

and

$$r(x) > 0 \quad \text{for all } x \in \mathcal{X}. \tag{2}$$

2. Use the method of Lagrange multipliers to find the best $\mathbf{r}$ by ignoring temporarily the positivity constraints on $\mathbf{r}$ in (2). Let

$$J = \sum_x \sum_y r(x) p(y|x) \log \frac{q(x|y)}{r(x)} - \lambda \sum_x r(x).$$

For convenience sake, we assume that the logarithm is the natural logarithm. Differentiating with respect to $r(x)$ gives

$$\frac{\partial J}{\partial r(x)} = \sum_y p(y|x) \log q(x|y) - \log r(x) - 1 - \lambda.$$

Upon setting $\frac{\partial J}{\partial r(x)} = 0$, we have

$$\log r(x) = \sum_y p(y|x) \log q(x|y) - 1 - \lambda,$$

or

$$r(x) = e^{-(\lambda+1)} \prod_y q(x|y)^{p(y|x)}.$$

3. By considering the normalization constraint in (1), we can eliminate λ and obtain

$$r(x) = \frac{\prod_y \underline{q(x|y)}^{\underline{p(y|x)}}}{\sum_{x'} \prod_y q(x'|y)^{p(y|x')}}. \tag{3}$$

4. The above product is over all y such that $p(y|x) > 0$, and $q(x|y) > 0$ for all such y. This implies that both the numerator and the denominator on the right hand side above are positive, and therefore $\underline{r(x) > 0}$.

Maximizing $f(\mathbf{r},\mathbf{q})$ for a Fixed $\mathbf{q}$

$$\max_{\mathbf{r}\in A_1} \sum_x \sum_y r(x)p(y|x) \log \frac{q(x|y)}{r(x)}$$

1. The constraints on $\mathbf{r}$ are

$$\sum_x r(x) = 1 \tag{1}$$

and

$$r(x) > 0 \quad \text{for all } x \in \mathcal{X}. \tag{2}$$

2. Use the method of Lagrange multipliers to find the best $\mathbf{r}$ by ignoring temporarily the positivity constraints on $\mathbf{r}$ in (2). Let

$$J = \sum_x \sum_y r(x)p(y|x) \log \frac{q(x|y)}{r(x)} - \lambda \sum_x r(x).$$

For convenience sake, we assume that the logarithm is the natural logarithm. Differentiating with respect to $r(x)$ gives

$$\frac{\partial J}{\partial r(x)} = \sum_y p(y|x) \log q(x|y) - \log r(x) - 1 - \lambda.$$

Upon setting $\frac{\partial J}{\partial r(x)} = 0$, we have

$$\log r(x) = \sum_y p(y|x) \log q(x|y) - 1 - \lambda,$$

or

$$r(x) = e^{-(\lambda+1)} \prod_y q(x|y)^{p(y|x)}.$$

3. By considering the normalization constraint in (1), we can eliminate λ and obtain

$$r(x) = \frac{\prod_y q(x|y)^{p(y|x)}}{\sum_{x'} \prod_y q(x'|y)^{p(y|x')}}. \tag{3}$$

4. The above product is over all y such that $p(y|x) > 0$, and $q(x|y) > 0$ for all such y. This implies that both the numerator and the denominator on the right hand side above are positive, and therefore $r(x) > 0$.

5. In other words, the $\mathbf{r}$ thus obtained happen to satisfy the positivity constraints in (2) although these constraints were ignored when we set up the Lagrange multipliers.

Maximizing $f(\mathbf{r},\mathbf{q})$ for a Fixed q

$$\boxed{\max_{\mathbf{r}\in A_1} \sum_x \sum_y r(x)p(y|x)\log\frac{q(x|y)}{r(x)}}$$

1. The constraints on $\mathbf{r}$ are

$$\sum_x r(x) = 1 \tag{1}$$

and

$$r(x) > 0 \quad \text{for all } x \in \mathcal{X}. \tag{2}$$

2. Use the method of Lagrange multipliers to find the best $\mathbf{r}$ by ignoring temporarily the positivity constraints on $\mathbf{r}$ in (2). Let

$$J = \sum_x \sum_y r(x)p(y|x)\log\frac{q(x|y)}{r(x)} - \lambda\sum_x r(x).$$

For convenience sake, we assume that the logarithm is the natural logarithm. Differentiating with respect to $r(x)$ gives

$$\frac{\partial J}{\partial r(x)} = \sum_y p(y|x)\log q(x|y) - \log r(x) - 1 - \lambda.$$

Upon setting $\frac{\partial J}{\partial r(x)} = 0$, we have

$$\log r(x) = \sum_y p(y|x)\log q(x|y) - 1 - \lambda,$$

or

$$r(x) = e^{-(\lambda+1)}\prod_y q(x|y)^{p(y|x)}.$$

3. By considering the normalization constraint in (1), we can eliminate λ and obtain

$$r(x) = \frac{\prod_y q(x|y)^{p(y|x)}}{\sum_{x'}\prod_y q(x'|y)^{p(y|x')}}. \tag{3}$$

4. The above product is over all y such that $p(y|x) > 0$, and $q(x|y) > 0$ for all such y. This implies that both the numerator and the denominator on the right hand side above are positive, and therefore $r(x) > 0$.

5. In other words, the $\mathbf{r}$ thus obtained happen to satisfy the positivity constraints in (2) although these constraints were ignored when we set up the Lagrange multipliers.

Maximizing *f*(r,q) for a Fixed q

$$\max_{\mathbf{r}\in A_1} \sum_x \sum_y r(x)p(y|x)\log\frac{q(x|y)}{r(x)}$$

1. The constraints on $\mathbf{r}$ are

$$\sum_x r(x) = 1 \tag{1}$$

and

$$r(x) > 0 \quad \text{for all } x \in \mathcal{X}. \tag{2}$$

2. Use the method of Lagrange multipliers to find the best $\mathbf{r}$ by ignoring temporarily the positivity constraints on $\mathbf{r}$ in (2). Let

$$J = \sum_x \sum_y r(x)p(y|x)\log\frac{q(x|y)}{r(x)} - \lambda \sum_x r(x).$$

For convenience sake, we assume that the logarithm is the natural logarithm. Differentiating with respect to $r(x)$ gives

$$\frac{\partial J}{\partial r(x)} = \sum_y p(y|x)\log q(x|y) - \log r(x) - 1 - \lambda.$$

Upon setting $\frac{\partial J}{\partial r(x)} = 0$, we have

$$\log r(x) = \sum_y p(y|x)\log q(x|y) - 1 - \lambda,$$

or

$$r(x) = e^{-(\lambda+1)} \prod_y q(x|y)^{p(y|x)}.$$

3. By considering the normalization constraint in (1), we can eliminate λ and obtain

$$r(x) = \frac{\prod_y q(x|y)^{p(y|x)}}{\sum_{x'} \prod_y q(x'|y)^{p(y|x')}}. \tag{3}$$

4. The above product is over all y such that $p(y|x) > 0$, and $q(x|y) > 0$ for all such y. This implies that both the numerator and the denominator on the right hand side above are positive, and therefore $r(x) > 0$.

5. In other words, the $\mathbf{r}$ thus obtained happen to satisfy the positivity constraints in (2) although these constraints were ignored when we set up the Lagrange multipliers.

6. We will show in Section 9.3.2 that f is concave. Then $\mathbf{r}$ as given in (3), which is unique, indeed achieves the maximum of f for a given $\mathbf{q} \in A_2$ because $\mathbf{r}$ is in the interior of A_1.

Maximizing *f*(r,q) for a Fixed q

$$\max_{\mathbf{r}\in A_1} \sum_x \sum_y r(x)p(y|x)\log\frac{q(x|y)}{r(x)}$$

1. The constraints on $\mathbf{r}$ are

$$\sum_x r(x) = 1 \tag{1}$$

and

$$r(x) > 0 \quad \text{for all } x \in \mathcal{X}. \tag{2}$$

2. Use the method of Lagrange multipliers to find the best $\mathbf{r}$ by ignoring temporarily the positivity constraints on $\mathbf{r}$ in (2). Let

$$J = \sum_x \sum_y r(x)p(y|x)\log\frac{q(x|y)}{r(x)} - \lambda \sum_x r(x).$$

For convenience sake, we assume that the logarithm is the natural logarithm. Differentiating with respect to $r(x)$ gives

$$\frac{\partial J}{\partial r(x)} = \sum_y p(y|x)\log q(x|y) - \log r(x) - 1 - \lambda.$$

Upon setting $\frac{\partial J}{\partial r(x)} = 0$, we have

$$\log r(x) = \sum_y p(y|x)\log q(x|y) - 1 - \lambda,$$

or

$$r(x) = e^{-(\lambda+1)} \prod_y q(x|y)^{p(y|x)}.$$

3. By considering the normalization constraint in (1), we can eliminate λ and obtain

$$r(x) = \frac{\prod_y q(x|y)^{p(y|x)}}{\sum_{x'} \prod_y q(x'|y)^{p(y|x')}}. \tag{3}$$

4. The above product is over all y such that $p(y|x) > 0$, and $q(x|y) > 0$ for all such y. This implies that both the numerator and the denominator on the right hand side above are positive, and therefore $r(x) > 0$.

5. In other words, the $\mathbf{r}$ thus obtained happen to satisfy the positivity constraints in (2) although these constraints were ignored when we set up the Lagrange multipliers.

6. We will show in Section 9.3.2 that f is concave. Then $\mathbf{r}$ as given in (3), which is unique, indeed achieves the maximum of f for a given $\mathbf{q} \in A_2$ because $\mathbf{r}$ is in the interior of A_1.

Maximizing *f*(r,q) for a Fixed q

$$\max_{\mathbf{r}\in A_1} \sum_x \sum_y r(x)p(y|x)\log\frac{q(x|y)}{r(x)}$$

1. The constraints on $\mathbf{r}$ are

$$\sum_x r(x) = 1 \tag{1}$$

and

$$r(x) > 0 \quad \text{for all } x \in \mathcal{X}. \tag{2}$$

2. Use the method of Lagrange multipliers to find the best $\mathbf{r}$ by ignoring temporarily the positivity constraints on $\mathbf{r}$ in (2). Let

$$J = \sum_x \sum_y r(x)p(y|x)\log\frac{q(x|y)}{r(x)} - \lambda\sum_x r(x).$$

For convenience sake, we assume that the logarithm is the natural logarithm. Differentiating with respect to $r(x)$ gives

$$\frac{\partial J}{\partial r(x)} = \sum_y p(y|x)\log q(x|y) - \log r(x) - 1 - \lambda.$$

Upon setting $\frac{\partial J}{\partial r(x)} = 0$, we have

$$\log r(x) = \sum_y p(y|x)\log q(x|y) - 1 - \lambda,$$

or

$$r(x) = e^{-(\lambda+1)}\prod_y q(x|y)^{p(y|x)}.$$

3. By considering the normalization constraint in (1), we can eliminate λ and obtain

$$r(x) = \frac{\prod_y q(x|y)^{p(y|x)}}{\sum_{x'}\prod_y q(x'|y)^{p(y|x')}}. \tag{3}$$

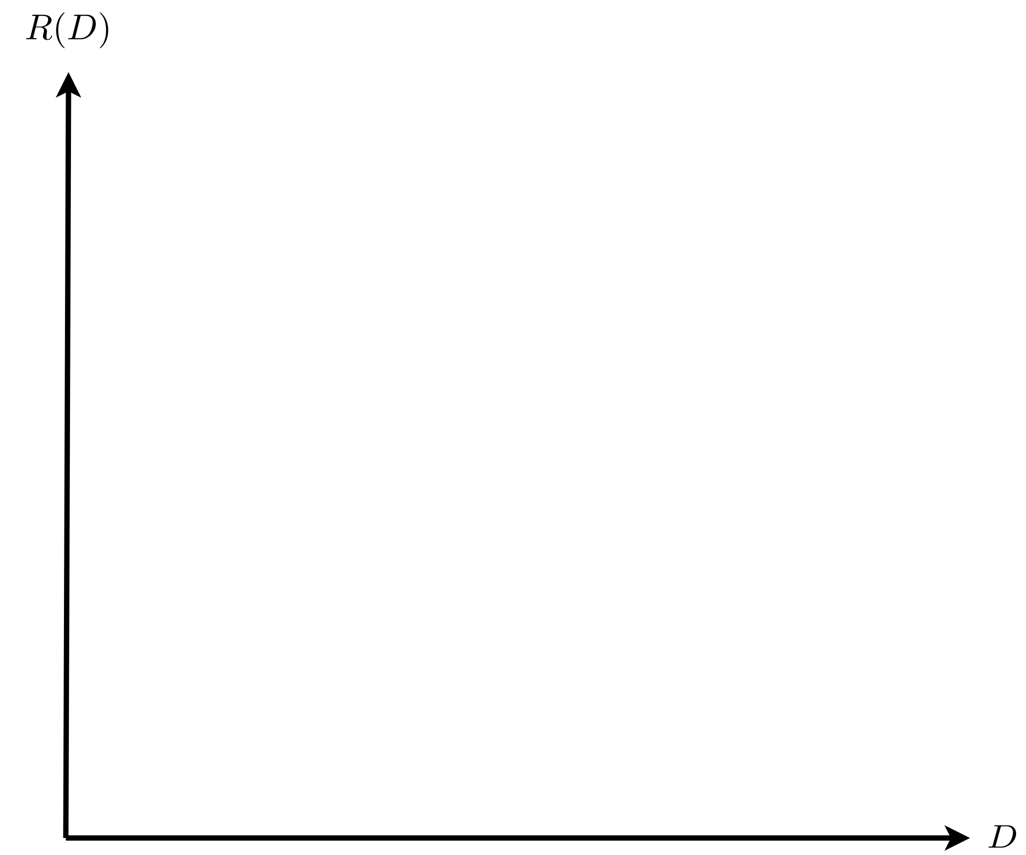
4. The above product is over all y such that $p(y|x) > 0$, and $q(x|y) > 0$ for all such y. This implies that both the numerator and the denominator on the right hand side above are positive, and therefore $r(x) > 0$.

5. In other words, the $\mathbf{r}$ thus obtained happen to satisfy the positivity constraints in (2) although these constraints were ignored when we set up the Lagrange multipliers.

6. We will show in Section 9.3.2 that f is concave. Then $\mathbf{r}$ as given in (3), which is unique, indeed achieves the maximum of f for a given $\mathbf{q} \in A_2$ because $\mathbf{r}$ is in the interior of A_1.

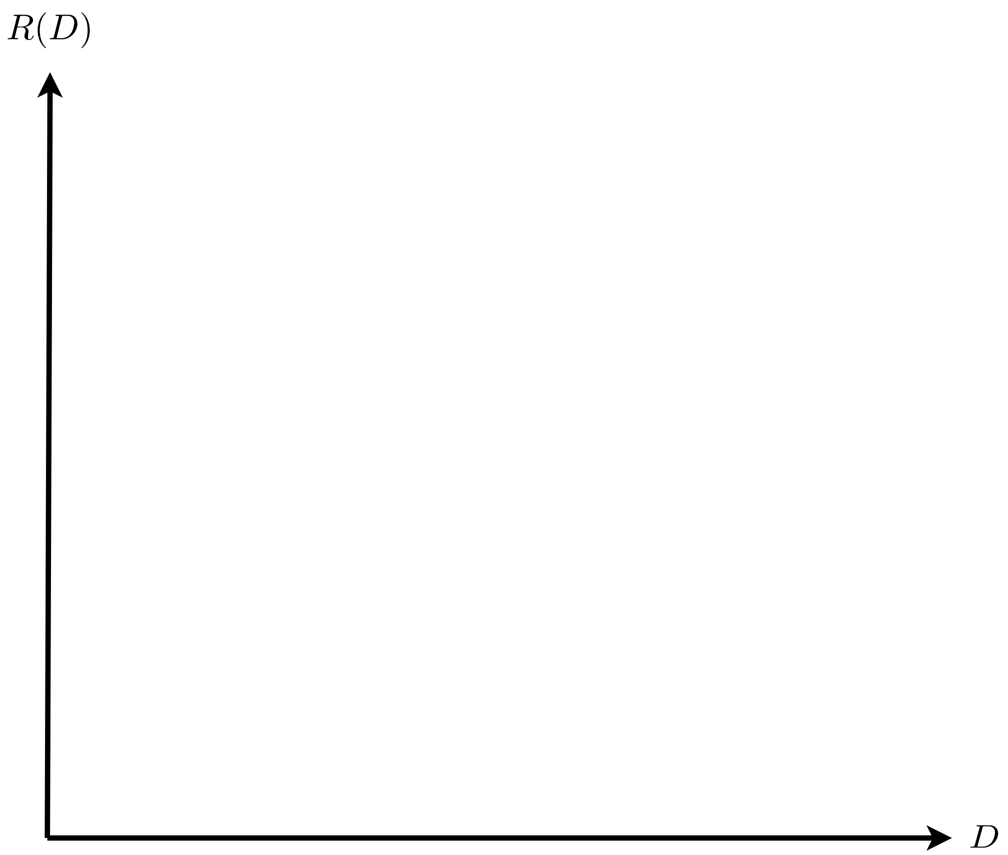
9.2.2 The Rate-Distortion Function

The $R(D)$ Curve



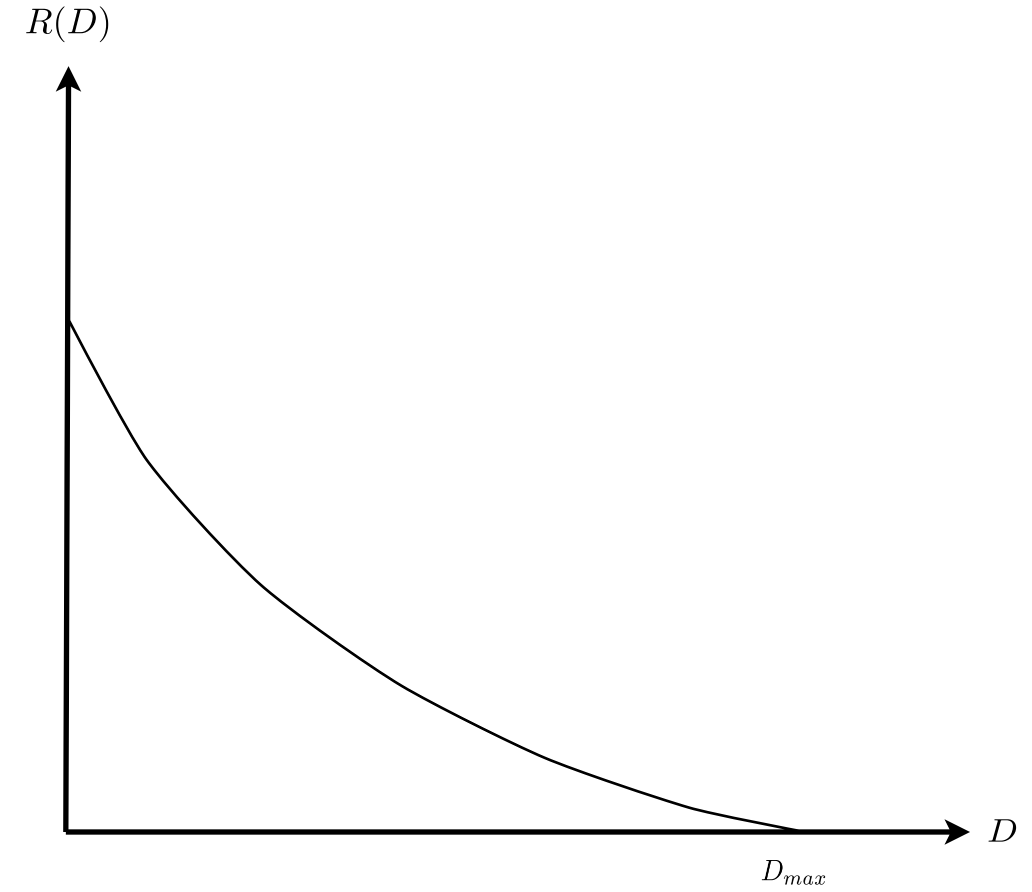
The $R(D)$ Curve

1. Assume $R(0) > 0$, so that $R(D)$ is strictly decreasing for $0 \leq D \leq D_{max}$.



The $R(D)$ Curve

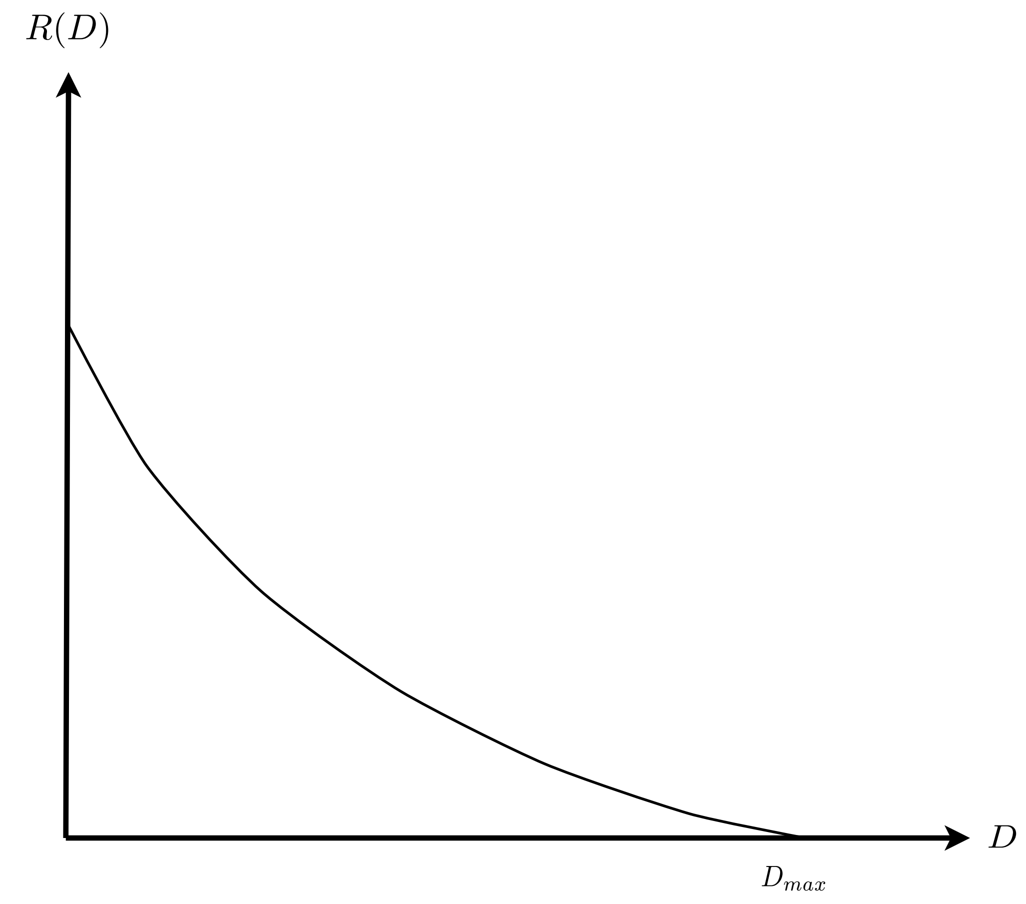
1. Assume $R(0) > 0$, so that $R(D)$ is strictly decreasing for $0 \leq D \leq D_{max}$.



The $R(D)$ Curve

1. Assume $R(0) > 0$, so that $R(D)$ is strictly decreasing for $0 \leq D \leq D_{max}$.

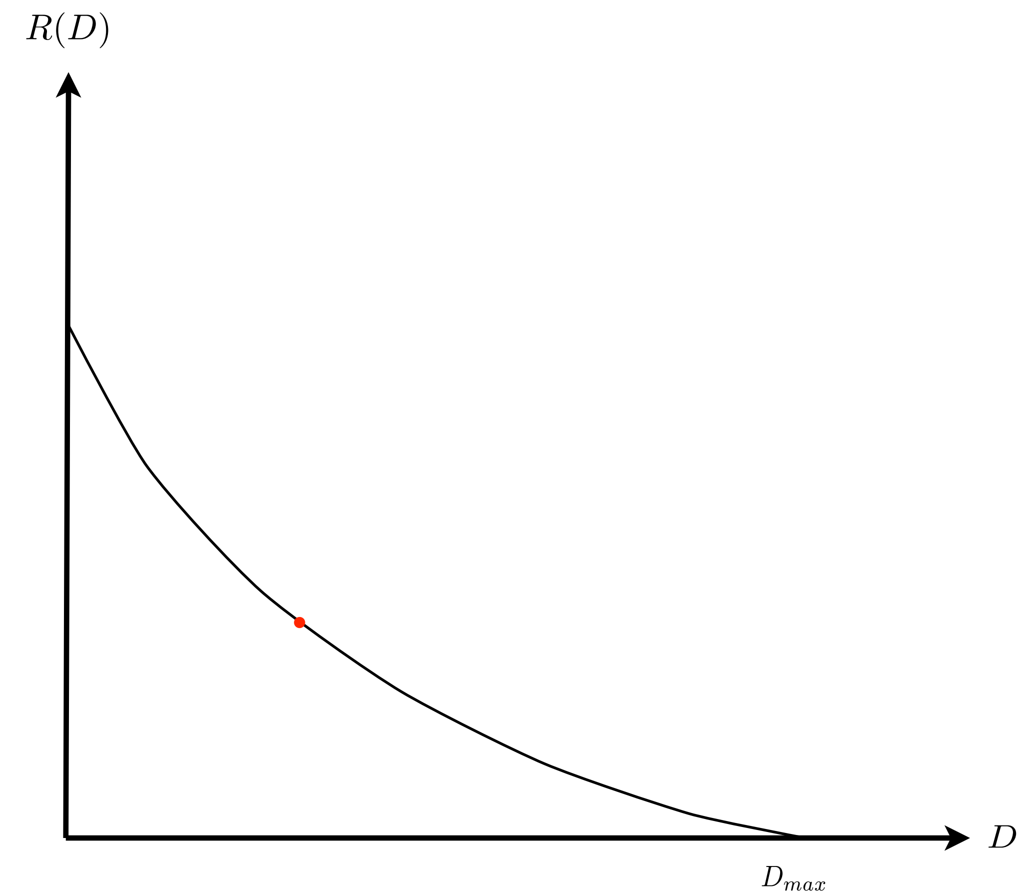
2. Since $R(D)$ is convex, for any $s \leq 0$, there exists a point on the $R(D)$ curve for $0 \leq D \leq D_{max}$ such that the slope of a tangent to the $R(D)$ curve at that point is equal to s.



The $R(D)$ Curve

1. Assume $R(0) > 0$, so that $R(D)$ is strictly decreasing for $0 \le D \le D_{max}$.

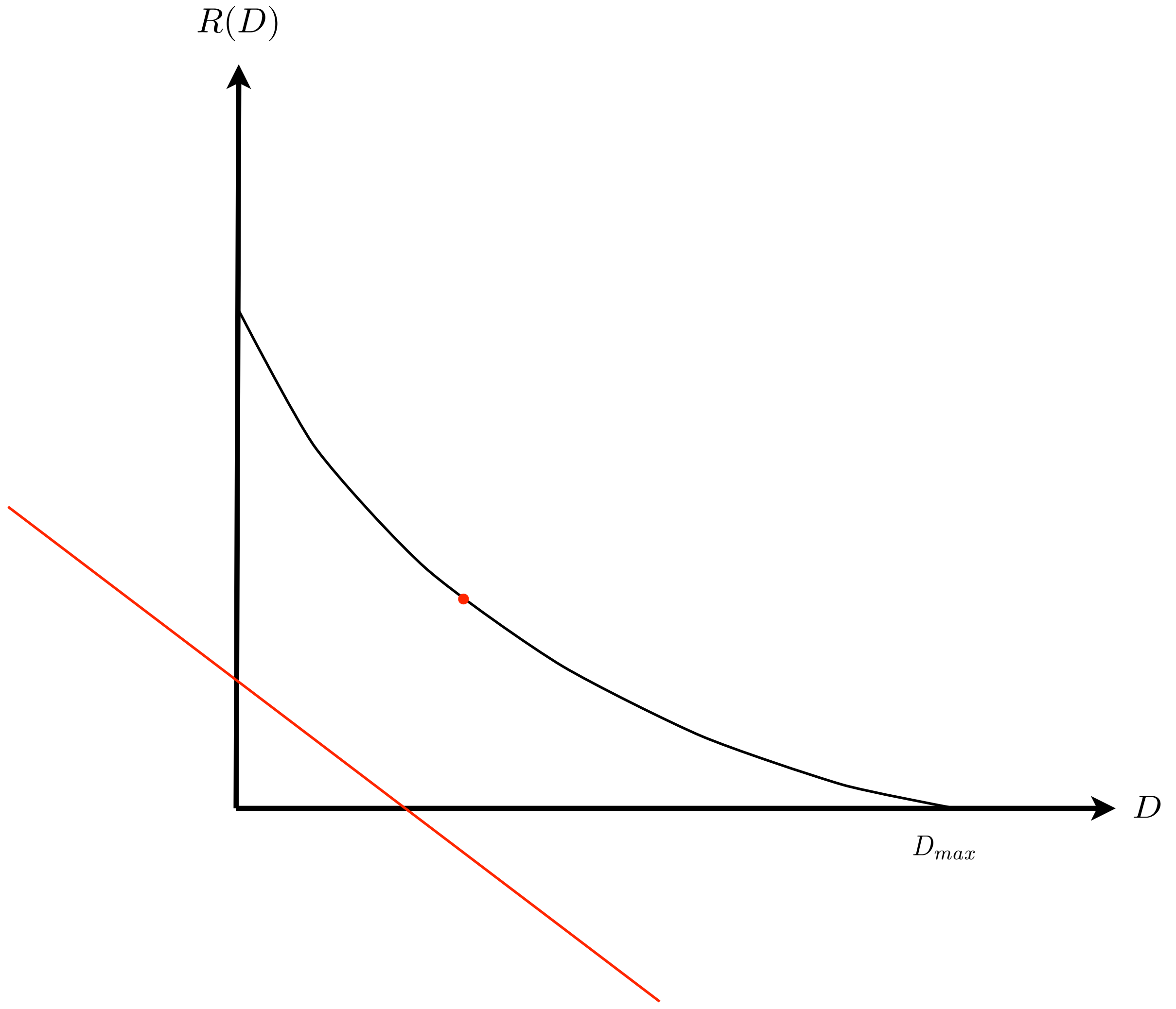
2. Since $R(D)$ is convex, for any $s \le 0$, there exists a point on the $R(D)$ curve for $0 \le D \le D_{max}$ such that the slope of a tangent to the $R(D)$ curve at that point is equal to s.



The $R(D)$ Curve

1. Assume $R(0) > 0$, so that $R(D)$ is strictly decreasing for $0 \leq D \leq D_{max}$.

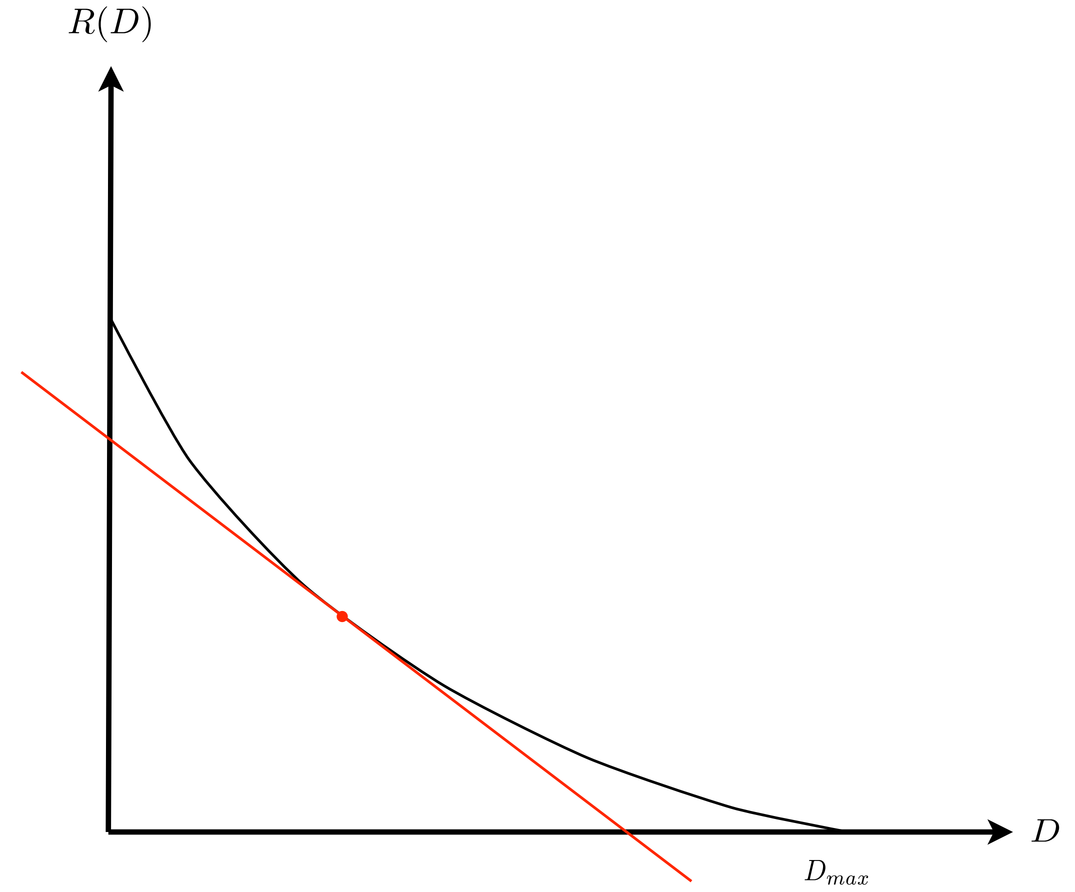
2. Since $R(D)$ is convex, for any $s \leq 0$, there exists a point on the $R(D)$ curve for $0 \leq D \leq D_{max}$ such that the slope of a tangent to the $R(D)$ curve at that point is equal to s.



The $R(D)$ Curve

1. Assume $R(0) > 0$, so that $R(D)$ is strictly decreasing for $0 \leq D \leq D_{max}$.

2. Since $R(D)$ is convex, for any $s \leq 0$, there exists a point on the $R(D)$ curve for $0 \leq D \leq D_{max}$ such that the slope of a tangent to the $R(D)$ curve at that point is equal to s.

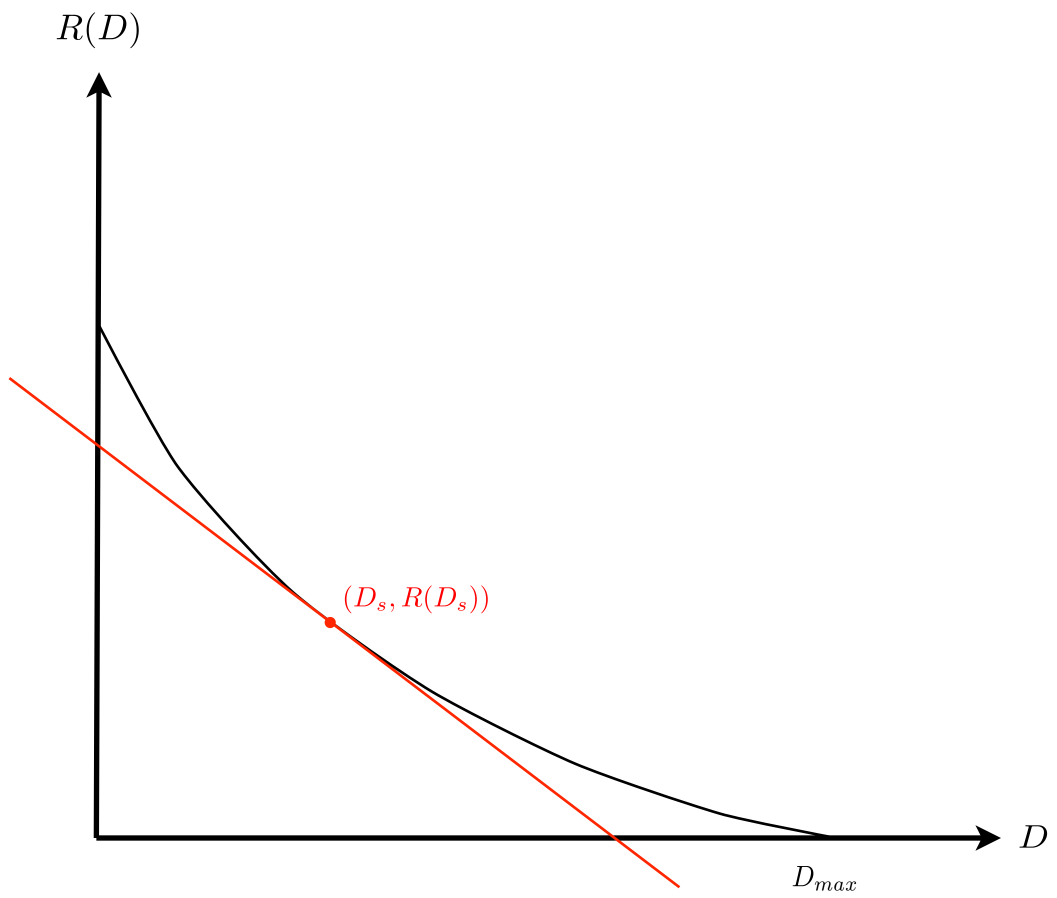


The $R(D)$ Curve

1. Assume $R(0) > 0$, so that $R(D)$ is strictly decreasing for $0 \leq D \leq D_{max}$.

2. Since $R(D)$ is convex, for any $s \leq 0$, there exists a point on the $R(D)$ curve for $0 \leq D \leq D_{max}$ such that the slope of a tangent to the $R(D)$ curve at that point is equal to s.

3. Denote such a point on the $R(D)$ curve by $(D_s, R(D_s))$, which is not necessarily unique.



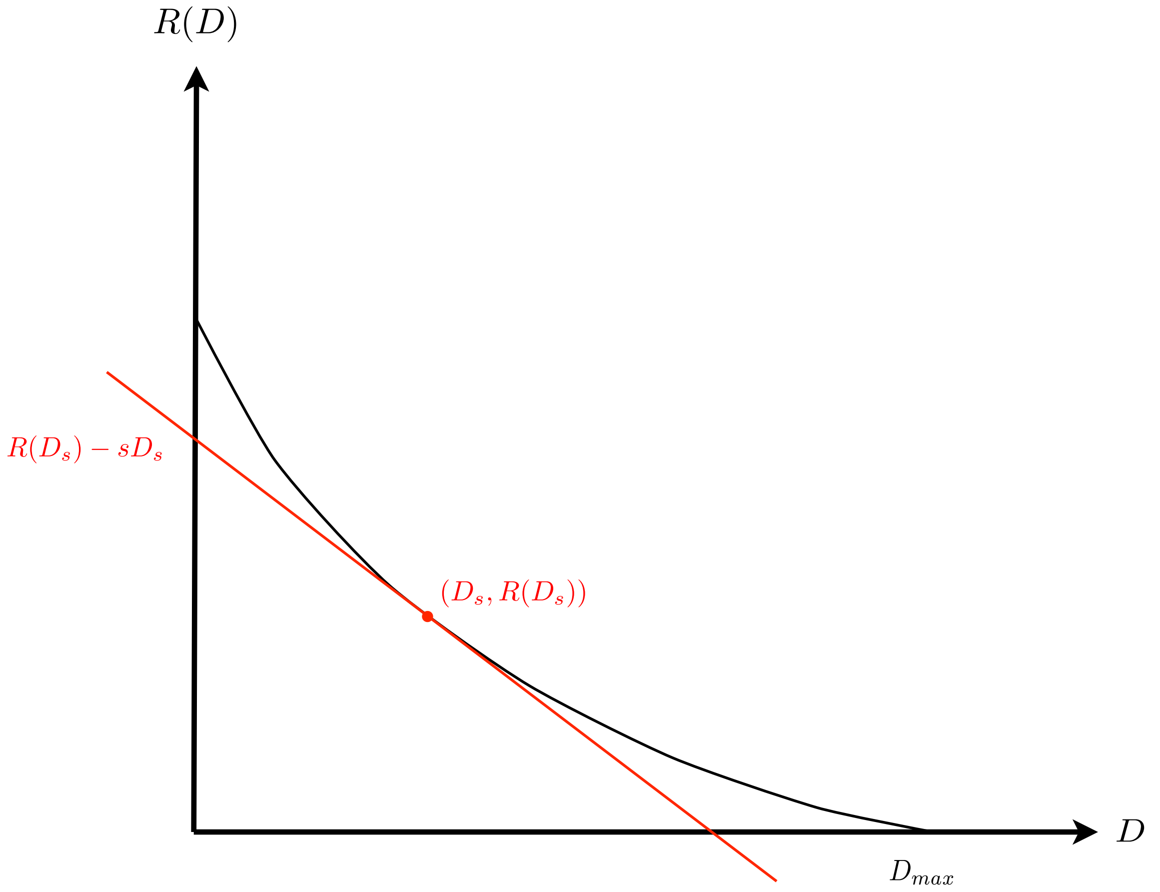
The $R(D)$ Curve

1. Assume $R(0) > 0$, so that $R(D)$ is strictly decreasing for $0 \leq D \leq D_{max}$.

2. Since $R(D)$ is convex, for any $s \leq 0$, there exists a point on the $R(D)$ curve for $0 \leq D \leq D_{max}$ such that the slope of a tangent to the $R(D)$ curve at that point is equal to s.

3. Denote such a point on the $R(D)$ curve by $(D_s, R(D_s))$, which is not necessarily unique.

4. Then this tangent intersects with the ordinate at $R(D_s) - sD_s$.



The $R(D)$ Curve

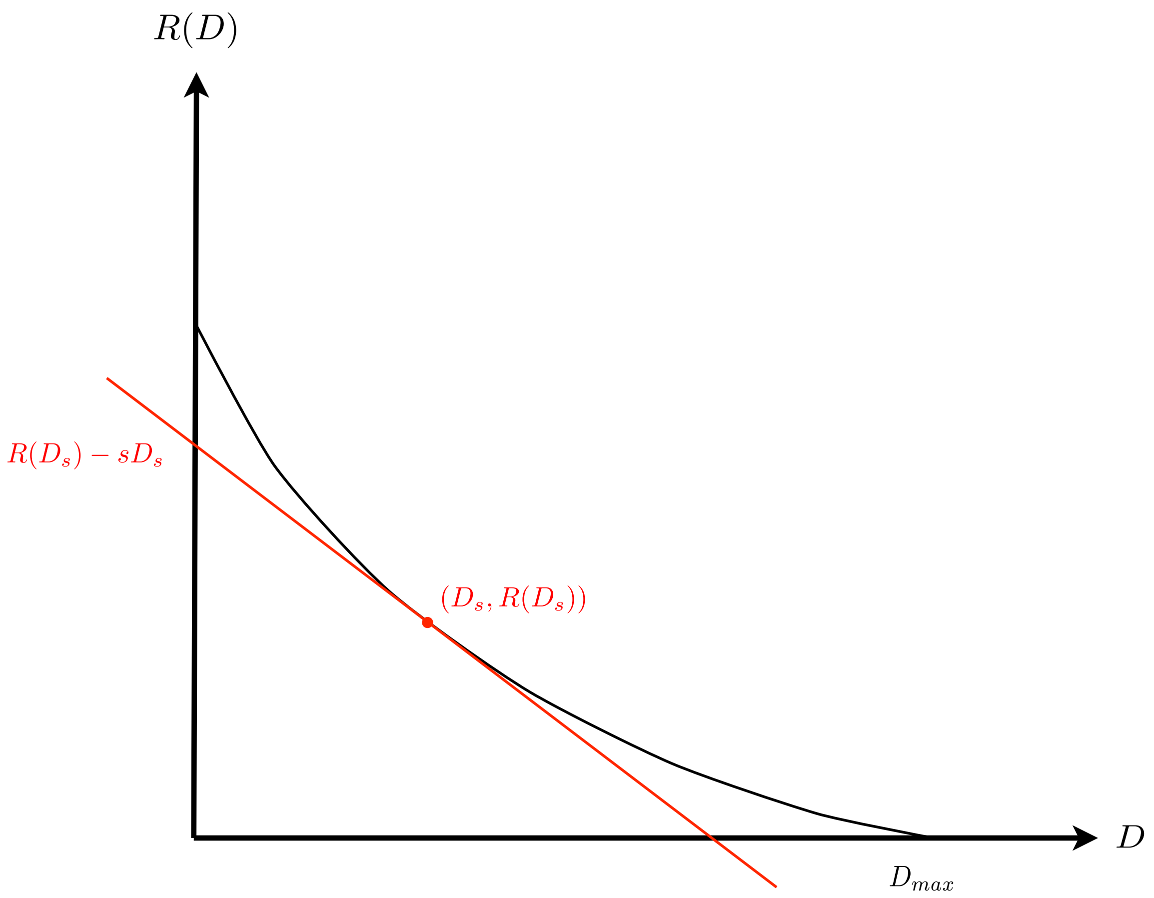
1. Assume $R(0) > 0$, so that $R(D)$ is strictly decreasing for $0 \leq D \leq D_{max}$.

2. Since $R(D)$ is convex, for any $s \leq 0$, there exists a point on the $R(D)$ curve for $0 \leq D \leq D_{max}$ such that the slope of a tangent to the $R(D)$ curve at that point is equal to s.

3. Denote such a point on the $R(D)$ curve by $(D_s, R(D_s))$, which is not necessarily unique.

4. Then this tangent intersects with the ordinate at $R(D_s) - sD_s$.

5. Write $I(X;\hat{X})$ and $Ed(X,\hat{X})$ as $I(\mathbf{p},\mathbf{Q})$ and $D(\mathbf{p},\mathbf{Q})$, respectively, where $\mathbf{p}$ is the distribution for X and $\mathbf{Q}$ is the transition matrix from $\mathcal{X}$ to $\hat{\mathcal{X}}$ defining $\hat{X}$.



The $R(D)$ Curve

1. Assume $R(0) > 0$, so that $R(D)$ is strictly decreasing for $0 \le D \le D_{max}$.

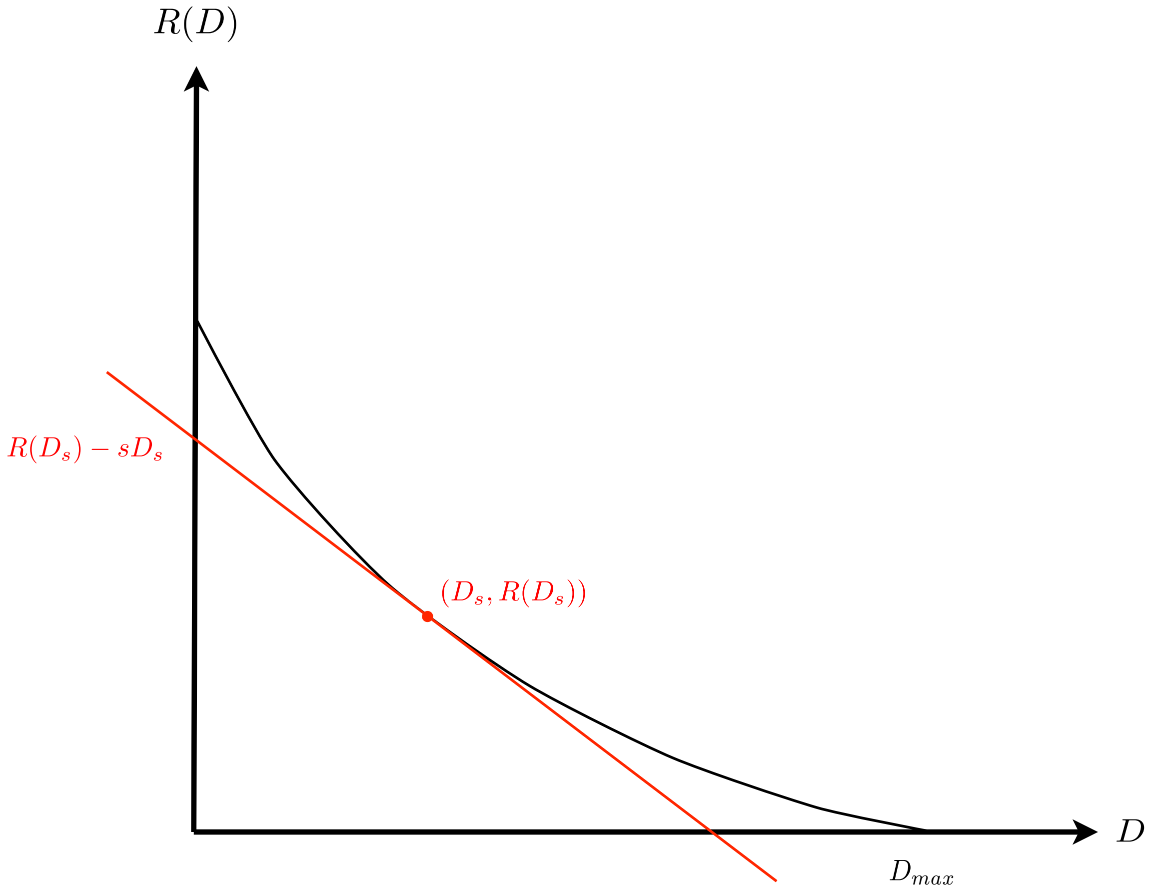
2. Since $R(D)$ is convex, for any $s \le 0$, there exists a point on the $R(D)$ curve for $0 \le D \le D_{max}$ such that the slope of a tangent to the $R(D)$ curve at that point is equal to s.

3. Denote such a point on the $R(D)$ curve by $(D_s, R(D_s))$, which is not necessarily unique.

4. Then this tangent intersects with the ordinate at $R(D_s) - sD_s$.

5. Write $I(X;\hat{X})$ and $Ed(X,\hat{X})$ as $I(\mathbf{p},\mathbf{Q})$ and $D(\mathbf{p},\mathbf{Q})$, respectively, where $\mathbf{p}$ is the distribution for X and $\mathbf{Q}$ is the transition matrix from $\mathcal{X}$ to $\hat{\mathcal{X}}$ defining $\hat{X}$.

6. For any $\mathbf{Q}$, $(D(\mathbf{p},\mathbf{Q}), I(\mathbf{p},\mathbf{Q}))$ is a point in the rate-distortion region, and the line with slope s passing through $(D(\mathbf{p},\mathbf{Q}), I(\mathbf{p},\mathbf{Q}))$ intersects the ordinate at $I(\mathbf{p},\mathbf{Q}) - sD(\mathbf{p},\mathbf{Q})$.



The $R(D)$ Curve

1. Assume $R(0) > 0$, so that $R(D)$ is strictly decreasing for $0 \leq D \leq D_{max}$.

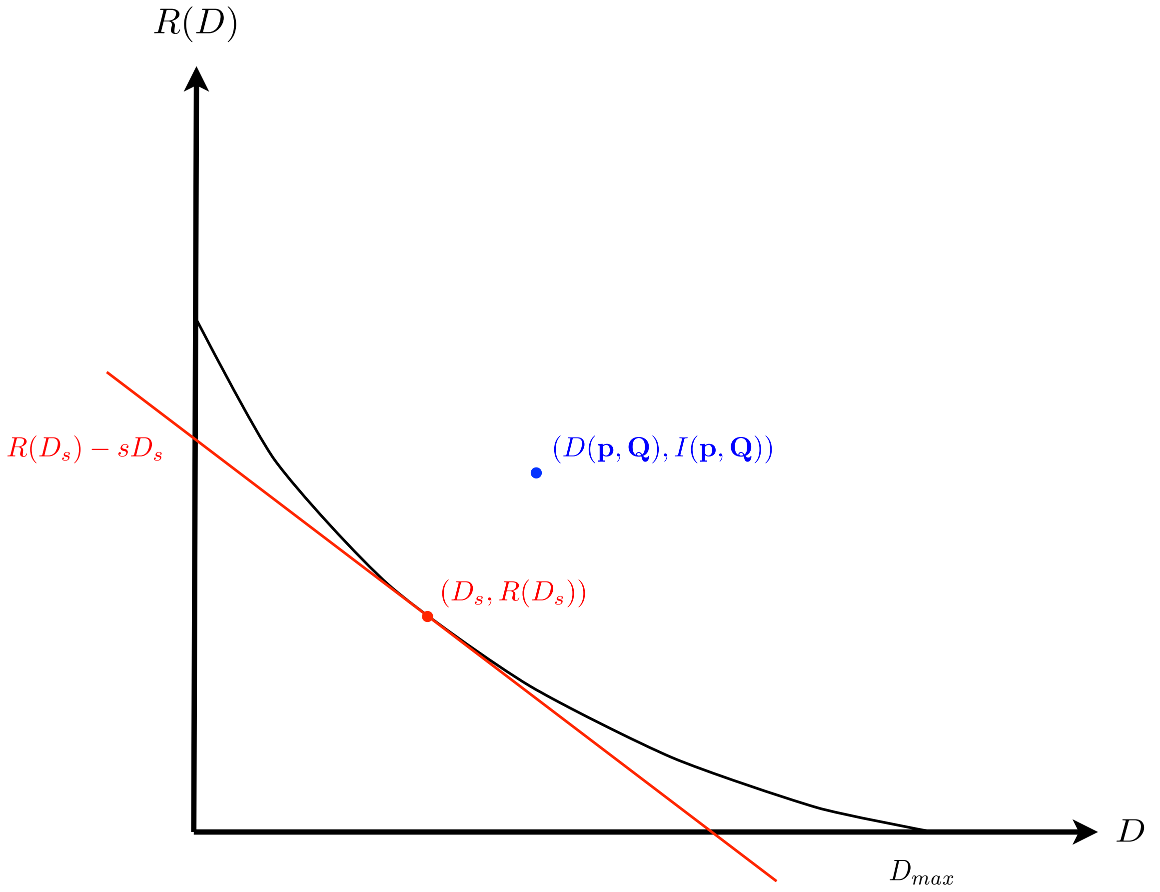
2. Since $R(D)$ is convex, for any $s \leq 0$, there exists a point on the $R(D)$ curve for $0 \leq D \leq D_{max}$ such that the slope of a tangent to the $R(D)$ curve at that point is equal to s.

3. Denote such a point on the $R(D)$ curve by $(D_s, R(D_s))$, which is not necessarily unique.

4. Then this tangent intersects with the ordinate at $R(D_s) - sD_s$.

5. Write $I(X;\hat{X})$ and $Ed(X,\hat{X})$ as $I(\mathbf{p},\mathbf{Q})$ and $D(\mathbf{p},\mathbf{Q})$, respectively, where $\mathbf{p}$ is the distribution for X and $\mathbf{Q}$ is the transition matrix from $\mathcal{X}$ to $\hat{\mathcal{X}}$ defining $\hat{X}$.

6. For any $\mathbf{Q}$, $(D(\mathbf{p},\mathbf{Q}), I(\mathbf{p},\mathbf{Q}))$ is a point in the rate-distortion region, and the line with slope s passing through $(D(\mathbf{p},\mathbf{Q}), I(\mathbf{p},\mathbf{Q}))$ intersects the ordinate at $I(\mathbf{p},\mathbf{Q}) - sD(\mathbf{p},\mathbf{Q})$.



The $R(D)$ Curve

1. Assume $R(0) > 0$, so that $R(D)$ is strictly decreasing for $0 \le D \le D_{max}$.

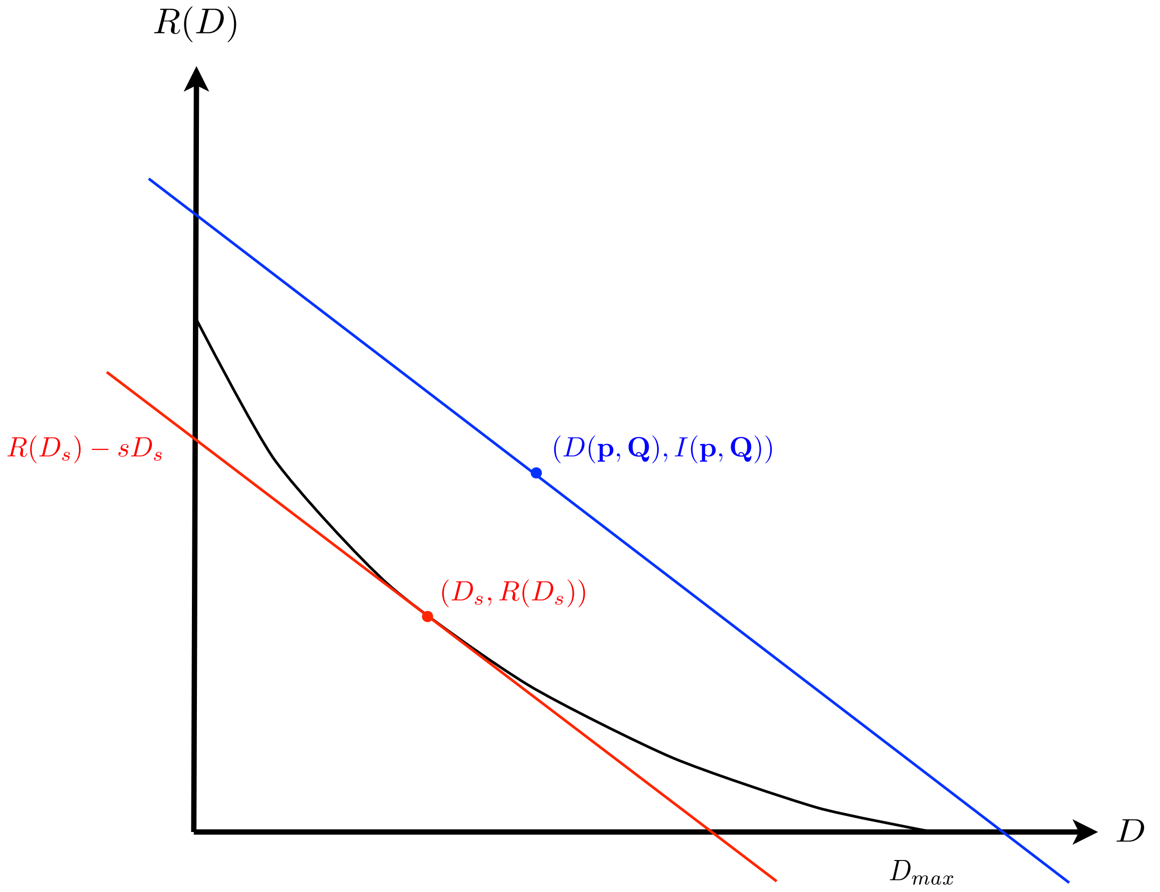
2. Since $R(D)$ is convex, for any $s \le 0$, there exists a point on the $R(D)$ curve for $0 \le D \le D_{max}$ such that the slope of a tangent to the $R(D)$ curve at that point is equal to s.

3. Denote such a point on the $R(D)$ curve by $(D_s, R(D_s))$, which is not necessarily unique.

4. Then this tangent intersects with the ordinate at $R(D_s) - sD_s$.

5. Write $I(X;\hat{X})$ and $Ed(X,\hat{X})$ as $I(\mathbf{p},\mathbf{Q})$ and $D(\mathbf{p},\mathbf{Q})$, respectively, where $\mathbf{p}$ is the distribution for X and $\mathbf{Q}$ is the transition matrix from $\mathcal{X}$ to $\hat{\mathcal{X}}$ defining $\hat{X}$.

6. For any $\mathbf{Q}$, $(D(\mathbf{p},\mathbf{Q}), I(\mathbf{p},\mathbf{Q}))$ is a point in the rate-distortion region, and the line with slope s passing through $(D(\mathbf{p},\mathbf{Q}), I(\mathbf{p},\mathbf{Q}))$ intersects the ordinate at $I(\mathbf{p},\mathbf{Q}) - sD(\mathbf{p},\mathbf{Q})$.



The $R(D)$ Curve

1. Assume $R(0) > 0$, so that $R(D)$ is strictly decreasing for $0 \leq D \leq D_{max}$.

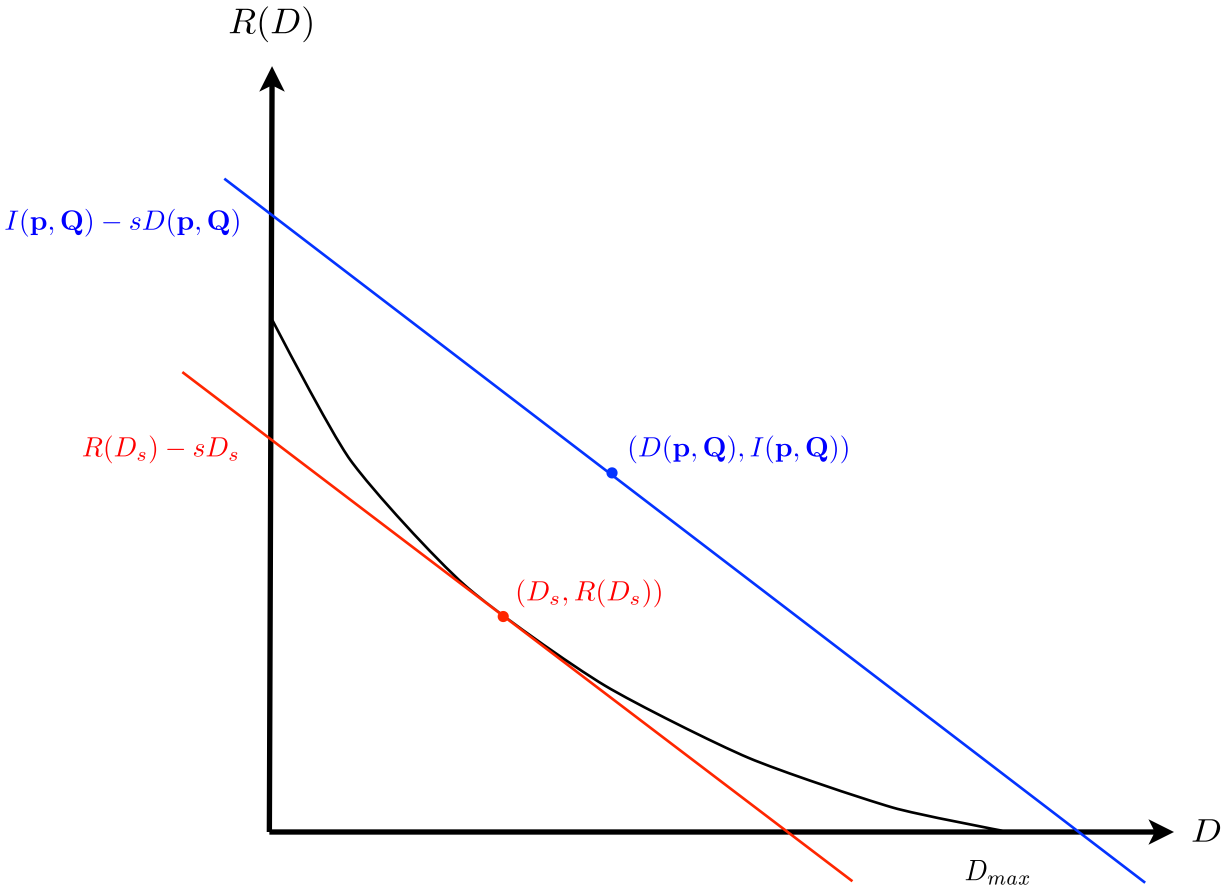
2. Since $R(D)$ is convex, for any $s \leq 0$, there exists a point on the $R(D)$ curve for $0 \leq D \leq D_{max}$ such that the slope of a tangent to the $R(D)$ curve at that point is equal to s.

3. Denote such a point on the $R(D)$ curve by $(D_s, R(D_s))$, which is not necessarily unique.

4. Then this tangent intersects with the ordinate at $R(D_s) - sD_s$.

5. Write $I(X;\hat{X})$ and $Ed(X,\hat{X})$ as $I(\mathbf{p},\mathbf{Q})$ and $D(\mathbf{p},\mathbf{Q})$, respectively, where $\mathbf{p}$ is the distribution for X and $\mathbf{Q}$ is the transition matrix from $\mathcal{X}$ to $\hat{\mathcal{X}}$ defining $\hat{X}$.

6. For any $\mathbf{Q}$, $(D(\mathbf{p},\mathbf{Q}), I(\mathbf{p},\mathbf{Q}))$ is a point in the rate-distortion region, and the line with slope s passing through $(D(\mathbf{p},\mathbf{Q}), I(\mathbf{p},\mathbf{Q}))$ intersects the ordinate at $I(\mathbf{p},\mathbf{Q}) - sD(\mathbf{p},\mathbf{Q})$.



The $R(D)$ Curve

1. Assume $R(0) > 0$, so that $R(D)$ is strictly decreasing for $0 \leq D \leq D_{max}$.

2. Since $R(D)$ is convex, for any $s \leq 0$, there exists a point on the $R(D)$ curve for $0 \leq D \leq D_{max}$ such that the slope of a tangent to the $R(D)$ curve at that point is equal to s.

3. Denote such a point on the $R(D)$ curve by $(D_s, R(D_s))$, which is not necessarily unique.

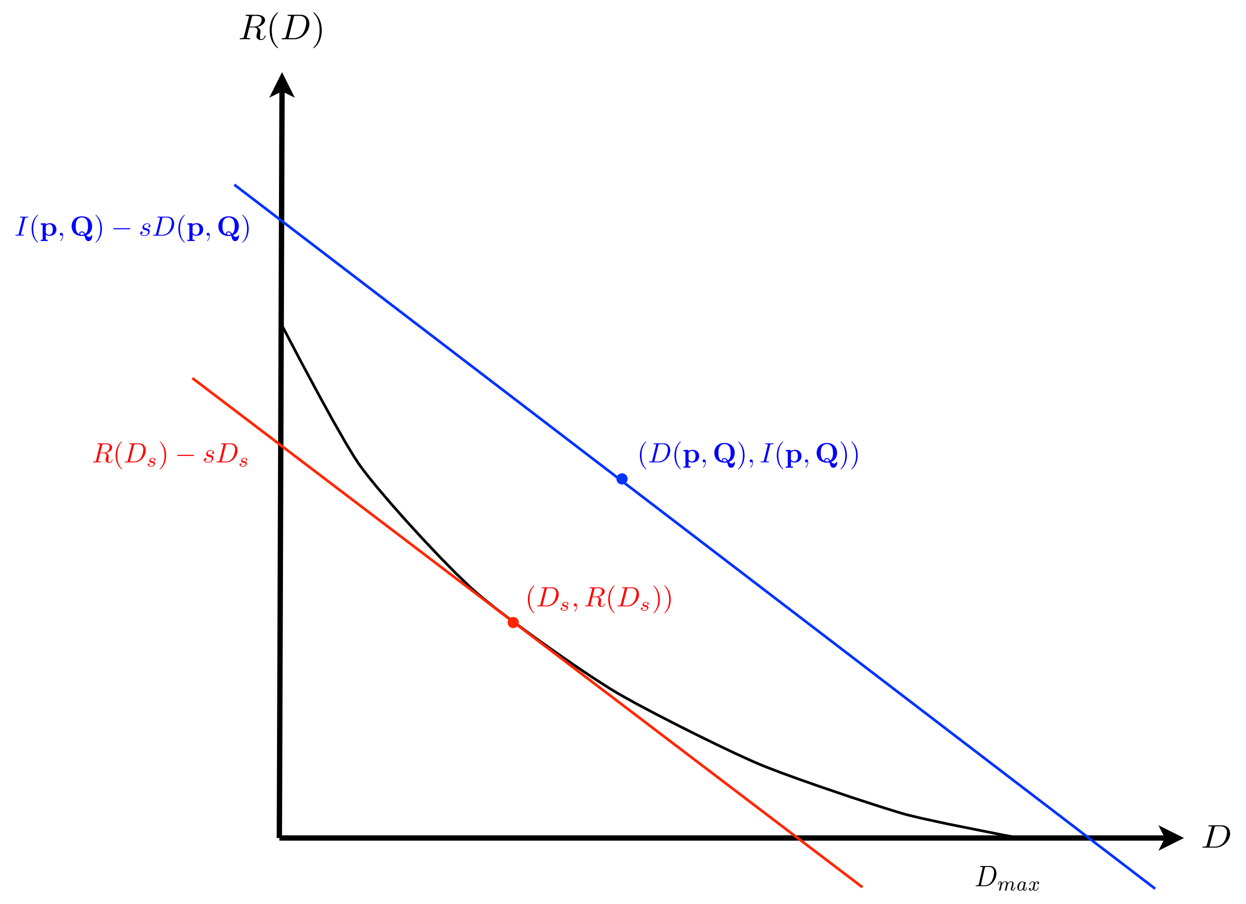
4. Then this tangent intersects with the ordinate at $R(D_s) - sD_s$.

5. Write $I(X;\hat{X})$ and $Ed(X,\hat{X})$ as $I(\mathbf{p},\mathbf{Q})$ and $D(\mathbf{p},\mathbf{Q})$, respectively, where $\mathbf{p}$ is the distribution for X and $\mathbf{Q}$ is the transition matrix from $\mathcal{X}$ to $\hat{\mathcal{X}}$ defining $\hat{X}$.

6. For any $\mathbf{Q}$, $(D(\mathbf{p},\mathbf{Q}), I(\mathbf{p},\mathbf{Q}))$ is a point in the rate-distortion region, and the line with slope s passing through $(D(\mathbf{p},\mathbf{Q}), I(\mathbf{p},\mathbf{Q}))$ intersects the ordinate at $I(\mathbf{p},\mathbf{Q}) - sD(\mathbf{p},\mathbf{Q})$.

7. Then

$$R(D_s) - sD_s = \min_{\mathbf{Q}} [I(\mathbf{p},\mathbf{Q}) - sD(\mathbf{p},\mathbf{Q})].$$



The $R(D)$ Curve

1. Assume $R(0) > 0$, so that $R(D)$ is strictly decreasing for $0 \le D \le D_{max}$.

2. Since $R(D)$ is convex, for any $s \le 0$, there exists a point on the $R(D)$ curve for $0 \le D \le D_{max}$ such that the slope of a tangent to the $R(D)$ curve at that point is equal to s.

3. Denote such a point on the $R(D)$ curve by $(D_s, R(D_s))$, which is not necessarily unique.

4. Then this tangent intersects with the ordinate at $R(D_s) - sD_s$.

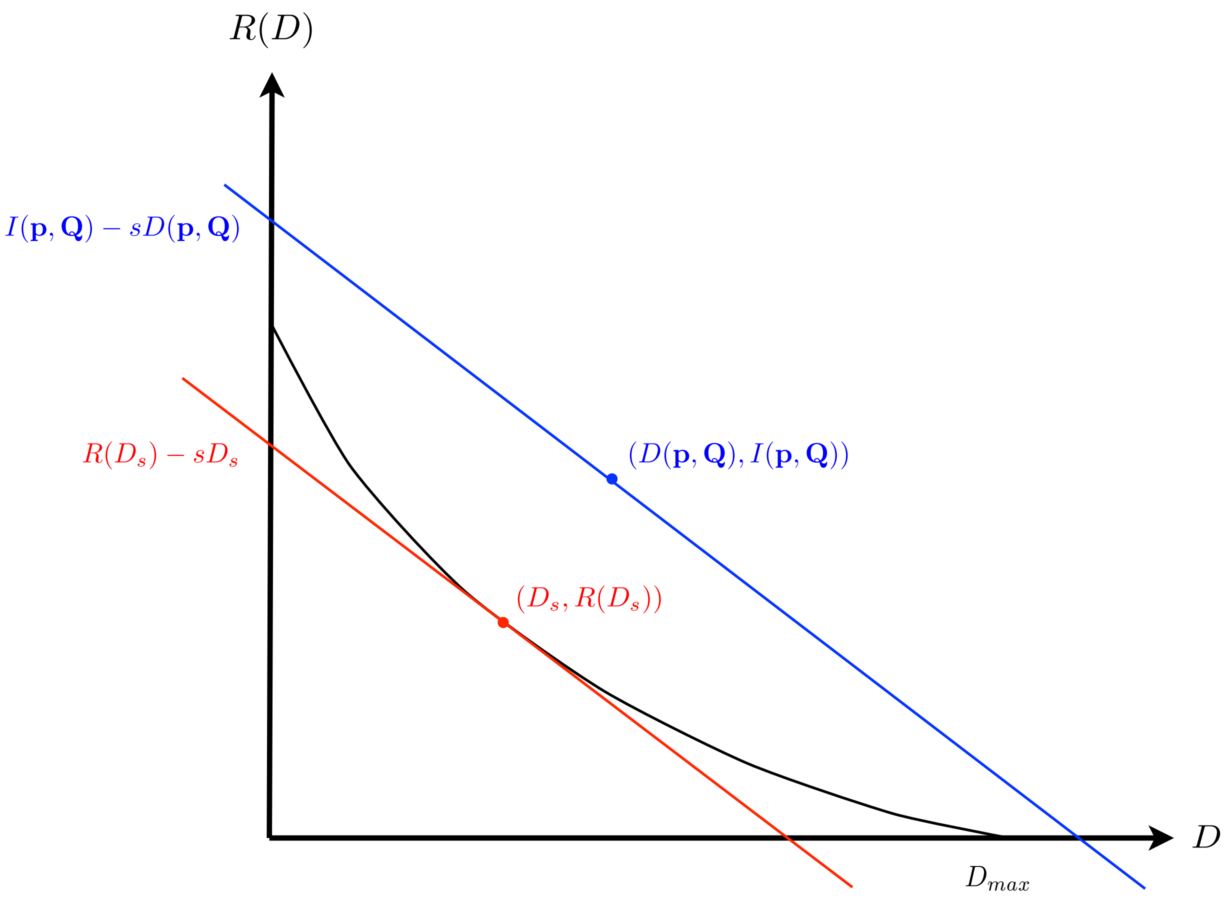
5. Write $I(X; \hat{X})$ and $Ed(X, \hat{X})$ as $I(\mathbf{p}, \mathbf{Q})$ and $D(\mathbf{p}, \mathbf{Q})$, respectively, where $\mathbf{p}$ is the distribution for X and $\mathbf{Q}$ is the transition matrix from $\mathcal{X}$ to $\hat{\mathcal{X}}$ defining $\hat{X}$.

6. For any $\mathbf{Q}$, $(D(\mathbf{p}, \mathbf{Q}), I(\mathbf{p}, \mathbf{Q}))$ is a point in the rate-distortion region, and the line with slope s passing through $(D(\mathbf{p}, \mathbf{Q}), I(\mathbf{p}, \mathbf{Q}))$ intersects the ordinate at $I(\mathbf{p}, \mathbf{Q}) - sD(\mathbf{p}, \mathbf{Q})$.

7. Then

$$R(D_s) - sD_s = \min_{\mathbf{Q}} [I(\mathbf{p}, \mathbf{Q}) - sD(\mathbf{p}, \mathbf{Q})].$$

8. By varying over all $s \le 0$, we can then trace out the whole $R(D)$ curve.



Lemma 9.3 Let $p(x)Q(\hat{x}|x)$ be a given joint distribution on $\mathcal{X} \times \hat{\mathcal{X}}$ such that $\mathbf{Q} > 0$, and let $\mathbf{t}$ be any distribution on $\hat{\mathcal{X}}$ such that $\mathbf{t} > 0$. Then

$$\min_{\mathbf{t}>0} \sum_x \sum_{\hat{x}} p(x)Q(\hat{x}|x) \log \frac{Q(\hat{x}|x)}{t(\hat{x})} = \sum_x \sum_{\hat{x}} p(x)Q(\hat{x}|x) \log \frac{Q(\hat{x}|x)}{t^*(\hat{x})},$$

where

$$t^*(\hat{x}) = \sum_x p(x)Q(\hat{x}|x),$$

i.e., the minimizing $\mathbf{t}$ is the one which corresponds to the input distribution $\mathbf{p}$ and the transition matrix $\mathbf{Q}$.

Proof

Lemma 9.3 Let $p(x)Q(\hat{x}|x)$ be a given joint distribution on $\mathcal{X} \times \hat{\mathcal{X}}$ such that $\mathbf{Q} > 0$, and let $\mathbf{t}$ be any distribution on $\hat{\mathcal{X}}$ such that $\mathbf{t} > 0$. Then

$$\min_{\mathbf{t}>0} \sum_x \sum_{\hat{x}} p(x)Q(\hat{x}|x) \log \frac{Q(\hat{x}|x)}{t(\hat{x})} = \sum_x \sum_{\hat{x}} p(x)Q(\hat{x}|x) \log \frac{Q(\hat{x}|x)}{t^*(\hat{x})},$$

where

$$t^*(\hat{x}) = \sum_x p(x)Q(\hat{x}|x),$$

i.e., the minimizing $\mathbf{t}$ is the one which corresponds to the input distribution $\mathbf{p}$ and the transition matrix $\mathbf{Q}$.

Proof

- Similar to Lemma 9.1.

Lemma 9.3 Let $p(x)Q(\hat{x}|x)$ be a given joint distribution on $\mathcal{X} \times \hat{\mathcal{X}}$ such that $\mathbf{Q} > 0$, and let $\mathbf{t}$ be any distribution on $\hat{\mathcal{X}}$ such that $\mathbf{t} > 0$. Then

$$\min_{\mathbf{t}>0} \sum_x \sum_{\hat{x}} p(x)Q(\hat{x}|x) \log \frac{Q(\hat{x}|x)}{t(\hat{x})} = \sum_x \sum_{\hat{x}} p(x)Q(\hat{x}|x) \log \frac{Q(\hat{x}|x)}{t^*(\hat{x})},$$

where

$$t^*(\hat{x}) = \sum_x p(x)Q(\hat{x}|x),$$

i.e., the minimizing $\mathbf{t}$ is the one which corresponds to the input distribution $\mathbf{p}$ and the transition matrix $\mathbf{Q}$.

Proof

- Similar to Lemma 9.1.
- Note that $\mathbf{t}^* > 0$ because $\mathbf{Q} > 0$, so that it suffices to minimize over all $\mathbf{t} > 0$ instead of $\mathbf{t} \geq 0$.

Lemma 9.3 Let $p(x)Q(\hat{x}|x)$ be a given joint distribution on $\mathcal{X} \times \hat{\mathcal{X}}$ such that $\mathbf{Q} > 0$, and let $\mathbf{t}$ be any distribution on $\hat{\mathcal{X}}$ such that $\mathbf{t} > 0$. Then

$$\min_{\mathbf{t}>0} \sum_x \sum_{\hat{x}} p(x)Q(\hat{x}|x) \log \frac{Q(\hat{x}|x)}{t(\hat{x})} = \sum_x \sum_{\hat{x}} p(x)Q(\hat{x}|x) \log \frac{Q(\hat{x}|x)}{t^*(\hat{x})},$$

where

$$t^*(\hat{x}) = \sum_x p(x)Q(\hat{x}|x),$$

i.e., the minimizing $\mathbf{t}$ is the one which corresponds to the input distribution $\mathbf{p}$ and the transition matrix $\mathbf{Q}$.

Proof

- Similar to Lemma 9.1.
- Note that $\mathbf{t}^* > 0$ because $\mathbf{Q} > 0$, so that it suffices to minimize over all $\mathbf{t} > 0$ instead of $\mathbf{t} \geq 0$.

Lemma 9.3 Let $p(x)Q(\hat{x}|x)$ be a given joint distribution on $\mathcal{X} \times \hat{\mathcal{X}}$ such that $\mathbf{Q} > 0$, and let $\mathbf{t}$ be any distribution on $\hat{\mathcal{X}}$ such that $\mathbf{t} > 0$. Then

$$\min_{\underline{\mathbf{t}>0}} \sum_x \sum_{\hat{x}} p(x)Q(\hat{x}|x) \log \frac{Q(\hat{x}|x)}{t(\hat{x})} = \sum_x \sum_{\hat{x}} p(x)Q(\hat{x}|x) \log \frac{Q(\hat{x}|x)}{t^*(\hat{x})},$$

where

$$t^*(\hat{x}) = \sum_x p(x)Q(\hat{x}|x),$$

i.e., the minimizing $\mathbf{t}$ is the one which corresponds to the input distribution $\mathbf{p}$ and the transition matrix $\mathbf{Q}$.

Proof

- Similar to Lemma 9.1.
- Note that $\mathbf{t}^* > 0$ because $\mathbf{Q} > 0$, so that it suffices to minimize over all $\mathbf{t} > 0$ instead of $\mathbf{t} \geq 0$.

Lemma 9.3 Let $p(x)Q(\hat{x}|x)$ be a given joint distribution on $\mathcal{X} \times \hat{\mathcal{X}}$ such that $\mathbf{Q} > 0$, and let $\mathbf{t}$ be any distribution on $\hat{\mathcal{X}}$ such that $\mathbf{t} > 0$. Then

$$\min_{\mathbf{t}>0} \sum_{x} \sum_{\hat{x}} p(x)Q(\hat{x}|x) \log \frac{Q(\hat{x}|x)}{t(\hat{x})} = \underbrace{\boxed{\sum_{x} \sum_{\hat{x}} p(x)Q(\hat{x}|x) \log \frac{Q(\hat{x}|x)}{t^*(\hat{x})}}}_{I(\mathbf{p},\mathbf{Q})},$$

where

$$t^*(\hat{x}) = \sum_{x} p(x)Q(\hat{x}|x),$$

i.e., the minimizing $\mathbf{t}$ is the one which corresponds to the input distribution $\mathbf{p}$ and the transition matrix $\mathbf{Q}$.

Proof

- Similar to Lemma 9.1.
- Note that $\mathbf{t}^* > 0$ because $\mathbf{Q} > 0$, so that it suffices to minimize over all $\mathbf{t} > 0$ instead of $\mathbf{t} \geq 0$.

An Expression for $R(D_s)$ - sD_s

$$\boxed{R(D_S) - sD_S = \min_{\mathbf{Q}} \left[I(\mathbf{p}, \mathbf{Q}) - sD(\mathbf{p}, \mathbf{Q}) \right].}$$

An Expression for $R(D_s)$ - sD_s

$$\boxed{R(D_s) - sD_s = \min_{\mathbf{Q}} \left[I(\mathbf{p}, \mathbf{Q}) - sD(\mathbf{p}, \mathbf{Q}) \right].}$$

1. Since $I(\mathbf{p}, \mathbf{Q})$ and $D(\mathbf{p}, \mathbf{Q})$ are continuous in $\mathbf{Q}$, the minimum over all $\mathbf{Q}$ can be replaced by the infimum over all $\mathbf{Q} > 0$ (cf. Theorem 9.2).

An Expression for $R(D_s)$ - sD_s

$$R(D_s) - sD_s = \inf_{\mathbf{Q} > 0} [I(\mathbf{p}, \mathbf{Q}) - sD(\mathbf{p}, \mathbf{Q})].$$

1. Since $I(\mathbf{p}, \mathbf{Q})$ and $D(\mathbf{p}, \mathbf{Q})$ are continuous in $\mathbf{Q}$, the minimum over all $\mathbf{Q}$ can be replaced by the infimum over all $\mathbf{Q} > 0$ (cf. Theorem 9.2).

An Expression for $R(D_s)$ - sD_s

$$R(D_s) - sD_s = \inf_{\mathbf{Q}>0} [I(\mathbf{p}, \mathbf{Q}) - sD(\mathbf{p}, \mathbf{Q})].$$

1. Since $I(\mathbf{p}, \mathbf{Q})$ and $D(\mathbf{p}, \mathbf{Q})$ are continuous in $\mathbf{Q}$, the minimum over all $\mathbf{Q}$ can be replaced by the infimum over all $\mathbf{Q} > 0$ (cf. Theorem 9.2).

2. Note that

$$D(\mathbf{p}, \mathbf{Q}) = \sum_x \sum_{\hat{x}} p(x)Q(\hat{x}|x)d(x, \hat{x}).$$

An Expression for $R(D_s)$ - sD_s

$$\boxed{R(D_s) - sD_s = \inf_{\mathbf{Q}>0} [I(\mathbf{p},\mathbf{Q}) - sD(\mathbf{p},\mathbf{Q})].}$$

1. Since $I(\mathbf{p},\mathbf{Q})$ and $D(\mathbf{p},\mathbf{Q})$ are continuous in $\mathbf{Q}$, the minimum over all $\mathbf{Q}$ can be replaced by the infimum over all $\mathbf{Q} > 0$ (cf. Theorem 9.2).
2. Note that

$$D(\mathbf{p},\mathbf{Q}) = \sum_x \sum_{\hat{x}} p(x)Q(\hat{x}|x)d(x,\hat{x}).$$

3. Then

$R(D_s) - sD_s$

An Expression for $R(D_s)$ - sD_s

$$R(D_s) - sD_s = \inf_{\mathbf{Q}>0} [I(\mathbf{p}, \mathbf{Q}) - sD(\mathbf{p}, \mathbf{Q})].$$

1. Since $I(\mathbf{p}, \mathbf{Q})$ and $D(\mathbf{p}, \mathbf{Q})$ are continuous in $\mathbf{Q}$, the minimum over all $\mathbf{Q}$ can be replaced by the infimum over all $\mathbf{Q} > 0$ (cf. Theorem 9.2).
2. Note that

$$D(\mathbf{p}, \mathbf{Q}) = \sum_x \sum_{\hat{x}} p(x)Q(\hat{x}|x)d(x, \hat{x}).$$

3. Then

$$\begin{aligned} & R(D_s) - sD_s \\ = & \inf_{\mathbf{Q}>0} [I(\mathbf{p}, \mathbf{Q}) - sD(\mathbf{p}, \mathbf{Q})] \end{aligned}$$

An Expression for $R(D_s)$ - sD_s

$$R(D_s) - sD_s = \inf_{\mathbf{Q}>0} [I(\mathbf{p}, \mathbf{Q}) - sD(\mathbf{p}, \mathbf{Q})].$$

1. Since $I(\mathbf{p}, \mathbf{Q})$ and $D(\mathbf{p}, \mathbf{Q})$ are continuous in $\mathbf{Q}$, the minimum over all $\mathbf{Q}$ can be replaced by the infimum over all $\mathbf{Q} > 0$ (cf. Theorem 9.2).
2. Note that

$$D(\mathbf{p}, \mathbf{Q}) = \sum_x \sum_{\hat{x}} p(x) Q(\hat{x}|x) d(x, \hat{x}).$$

3. Then

$$\begin{aligned} &R(D_s) - sD_s \\ &\quad = \inf_{\mathbf{Q}>0} [\underline{I(\mathbf{p}, \mathbf{Q})} - sD(\mathbf{p}, \mathbf{Q})] \end{aligned}$$

An Expression for $R(D_s)$ - sD_s

$$R(D_s) - sD_s = \inf_{\mathbf{Q}>0} [I(\mathbf{p}, \mathbf{Q}) - sD(\mathbf{p}, \mathbf{Q})].$$

1. Since $I(\mathbf{p}, \mathbf{Q})$ and $D(\mathbf{p}, \mathbf{Q})$ are continuous in $\mathbf{Q}$, the minimum over all $\mathbf{Q}$ can be replaced by the infimum over all $\mathbf{Q} > 0$ (cf. Theorem 9.2).

2. Note that

$$D(\mathbf{p}, \mathbf{Q}) = \sum_x \sum_{\hat{x}} p(x)Q(\hat{x}|x)d(x, \hat{x}).$$

3. Then

$$\begin{aligned} & R(D_s) - sD_s \\ = \quad & \inf_{\mathbf{Q}>0} [\underline{I(\mathbf{p}, \mathbf{Q})} - sD(\mathbf{p}, \mathbf{Q})] \end{aligned}$$

Lemma 9.3

$$I(\mathbf{p}, \mathbf{Q}) = \min_{\mathbf{t}>0} \sum_x \sum_{\hat{x}} p(x)Q(\hat{x}|x) \log \frac{Q(\hat{x}|x)}{t(\hat{x})}$$

An Expression for $R(D_s)$ - sD_s

$$R(D_s) - sD_s = \inf_{\mathbf{Q}>0} [I(\mathbf{p},\mathbf{Q}) - sD(\mathbf{p},\mathbf{Q})].$$

1. Since $I(\mathbf{p},\mathbf{Q})$ and $D(\mathbf{p},\mathbf{Q})$ are continuous in $\mathbf{Q}$, the minimum over all $\mathbf{Q}$ can be replaced by the infimum over all $\mathbf{Q} > 0$ (cf. Theorem 9.2).

2. Note that

$$D(\mathbf{p},\mathbf{Q}) = \sum_x \sum_{\hat{x}} p(x)Q(\hat{x}|x)d(x,\hat{x}).$$

3. Then

$$\begin{aligned}
&R(D_s) - sD_s \\
&= \inf_{\mathbf{Q}>0} [\underline{I(\mathbf{p},\mathbf{Q})} - sD(\mathbf{p},\mathbf{Q})] \\
&= \inf_{\mathbf{Q}>0} \left[\underline{\min_{\mathbf{t}>0} \sum_{x,\hat{x}} p(x)Q(\hat{x}|x)\log\frac{Q(\hat{x}|x)}{t(\hat{x})}} - s\sum_{x,\hat{x}} p(x)Q(\hat{x}|x)d(x,\hat{x}) \right]
\end{aligned}$$

Lemma 9.3

$$I(\mathbf{p},\mathbf{Q}) = \min_{\mathbf{t}>0} \sum_x \sum_{\hat{x}} p(x)Q(\hat{x}|x)\log\frac{Q(\hat{x}|x)}{t(\hat{x})}$$

An Expression for $R(D_s)$ - sD_s

$$R(D_s) - sD_s = \inf_{\mathbf{Q}>0} [I(\mathbf{p},\mathbf{Q}) - sD(\mathbf{p},\mathbf{Q})].$$

Lemma 9.3

$$I(\mathbf{p},\mathbf{Q}) = \min_{\mathbf{t}>0} \sum_x \sum_{\hat{x}} p(x)Q(\hat{x}|x) \log \frac{Q(\hat{x}|x)}{t(\hat{x})}$$

1. Since $I(\mathbf{p},\mathbf{Q})$ and $D(\mathbf{p},\mathbf{Q})$ are continuous in $\mathbf{Q}$, the minimum over all $\mathbf{Q}$ can be replaced by the infimum over all $\mathbf{Q} > 0$ (cf. Theorem 9.2).

2. Note that

$$D(\mathbf{p},\mathbf{Q}) = \sum_x \sum_{\hat{x}} p(x)Q(\hat{x}|x)d(x,\hat{x}).$$

3. Then

$$\begin{aligned}
&R(D_s) - sD_s \\
&\quad = \inf_{\mathbf{Q}>0} [I(\mathbf{p},\mathbf{Q}) - s\underline{D(\mathbf{p},\mathbf{Q})}] \\
&\quad = \inf_{\mathbf{Q}>0} \left[\min_{\mathbf{t}>0} \sum_{x,\hat{x}} p(x)Q(\hat{x}|x) \log \frac{Q(\hat{x}|x)}{t(\hat{x})} - s \sum_{x,\hat{x}} p(x)Q(\hat{x}|x)d(x,\hat{x}) \right]
\end{aligned}$$

An Expression for $R(D_s)$ - sD_s

$$R(D_s) - sD_s = \inf_{\mathbf{Q}>0} [I(\mathbf{p}, \mathbf{Q}) - sD(\mathbf{p}, \mathbf{Q})].$$

1. Since $I(\mathbf{p}, \mathbf{Q})$ and $D(\mathbf{p}, \mathbf{Q})$ are continuous in $\mathbf{Q}$, the minimum over all $\mathbf{Q}$ can be replaced by the infimum over all $\mathbf{Q} > 0$ (cf. Theorem 9.2).

2. Note that

$$D(\mathbf{p}, \mathbf{Q}) = \sum_x \sum_{\hat{x}} p(x)Q(\hat{x}|x)d(x, \hat{x}).$$

3. Then

$$\begin{aligned}
&R(D_s) - sD_s \\
&= \inf_{\mathbf{Q}>0} [I(\mathbf{p}, \mathbf{Q}) - s\underline{D(\mathbf{p}, \mathbf{Q})}] \\
&= \inf_{\mathbf{Q}>0} \left[\min_{\mathbf{t}>0} \sum_{x,\hat{x}} p(x)Q(\hat{x}|x) \log \frac{Q(\hat{x}|x)}{t(\hat{x})} - s \sum_{x,\hat{x}} p(x)Q(\hat{x}|x)d(x, \hat{x})\right]
\end{aligned}$$

Lemma 9.3

$$I(\mathbf{p}, \mathbf{Q}) = \min_{\mathbf{t}>0} \sum_x \sum_{\hat{x}} p(x)Q(\hat{x}|x) \log \frac{Q(\hat{x}|x)}{t(\hat{x})}$$

An Expression for $R(D_s)$ - sD_s

$$R(D_s) - sD_s = \inf_{\mathbf{Q}>0} [I(\mathbf{p}, \mathbf{Q}) - sD(\mathbf{p}, \mathbf{Q})].$$

1. Since $I(\mathbf{p}, \mathbf{Q})$ and $D(\mathbf{p}, \mathbf{Q})$ are continuous in $\mathbf{Q}$, the minimum over all $\mathbf{Q}$ can be replaced by the infimum over all $\mathbf{Q} > 0$ (cf. Theorem 9.2).
2. Note that

$$D(\mathbf{p}, \mathbf{Q}) = \sum_x \sum_{\hat{x}} p(x)Q(\hat{x}|x)d(x, \hat{x}).$$

3. Then

$$\begin{aligned}
&R(D_s) - sD_s \\
&= \inf_{\mathbf{Q}>0} [I(\mathbf{p}, \mathbf{Q}) - s\underline{D(\mathbf{p}, \mathbf{Q})}] \\
&= \inf_{\mathbf{Q}>0} \left[\min_{\mathbf{t}>0} \sum_{x,\hat{x}} p(x)Q(\hat{x}|x) \log \frac{Q(\hat{x}|x)}{t(\hat{x})} - s \underline{\sum_{x,\hat{x}} p(x)Q(\hat{x}|x)d(x, \hat{x})} \right]
\end{aligned}$$

Lemma 9.3

$$I(\mathbf{p}, \mathbf{Q}) = \min_{\mathbf{t}>0} \sum_x \sum_{\hat{x}} p(x)Q(\hat{x}|x) \log \frac{Q(\hat{x}|x)}{t(\hat{x})}$$

An Expression for $R(D_s)$ - sD_s

$$R(D_s) - sD_s = \inf_{\mathbf{Q}>0} [I(\mathbf{p}, \mathbf{Q}) - sD(\mathbf{p}, \mathbf{Q})].$$

Lemma 9.3

$$I(\mathbf{p}, \mathbf{Q}) = \min_{\mathbf{t}>0} \sum_x \sum_{\hat{x}} p(x)Q(\hat{x}|x) \log \frac{Q(\hat{x}|x)}{t(\hat{x})}$$

1. Since $I(\mathbf{p}, \mathbf{Q})$ and $D(\mathbf{p}, \mathbf{Q})$ are continuous in $\mathbf{Q}$, the minimum over all $\mathbf{Q}$ can be replaced by the infimum over all $\mathbf{Q} > 0$ (cf. Theorem 9.2).

2. Note that

$$D(\mathbf{p}, \mathbf{Q}) = \sum_x \sum_{\hat{x}} p(x)Q(\hat{x}|x)d(x, \hat{x}).$$

3. Then

$$\begin{aligned} & R(D_s) - sD_s \\ & \quad = \inf_{\mathbf{Q}>0} [I(\mathbf{p}, \mathbf{Q}) - sD(\mathbf{p}, \mathbf{Q})] \\ & \quad = \inf_{\mathbf{Q}>0} \left[\min_{\mathbf{t}>0} \sum_{x,\hat{x}} p(x)Q(\hat{x}|x) \log \frac{Q(\hat{x}|x)}{t(\hat{x})} - s \sum_{x,\hat{x}} p(x)Q(\hat{x}|x)d(x, \hat{x}) \right] \end{aligned}$$

An Expression for $R(D_s)$ - sD_s

$$R(D_s) - sD_s = \inf_{\mathbf{Q}>0} [I(\mathbf{p},\mathbf{Q}) - sD(\mathbf{p},\mathbf{Q})].$$

1. Since $I(\mathbf{p},\mathbf{Q})$ and $D(\mathbf{p},\mathbf{Q})$ are continuous in $\mathbf{Q}$, the minimum over all $\mathbf{Q}$ can be replaced by the infimum over all $\mathbf{Q} > 0$ (cf. Theorem 9.2).

2. Note that

$$D(\mathbf{p},\mathbf{Q}) = \sum_x \sum_{\hat{x}} p(x)Q(\hat{x}|x)d(x,\hat{x}).$$

3. Then

$$\begin{aligned}
& R(D_s) - sD_s \\
&= \inf_{\mathbf{Q}>0} [I(\mathbf{p},\mathbf{Q}) - sD(\mathbf{p},\mathbf{Q})] \\
&= \inf_{\mathbf{Q}>0} \left[\underline{\min_{\mathbf{t}>0}} \sum_{x,\hat{x}} p(x)Q(\hat{x}|x)\log\frac{Q(\hat{x}|x)}{t(\hat{x})} - s\sum_{x,\hat{x}} p(x)Q(\hat{x}|x)d(x,\hat{x}) \right] \\
&= \inf_{\mathbf{Q}>0} \underline{\min_{\mathbf{t}>0}} \left[\sum_{x,\hat{x}} p(x)Q(\hat{x}|x)\log\frac{Q(\hat{x}|x)}{t(\hat{x})} - s\sum_{x,\hat{x}} p(x)Q(\hat{x}|x)d(x,\hat{x}) \right].
\end{aligned}$$

Lemma 9.3

$$I(\mathbf{p},\mathbf{Q}) = \min_{\mathbf{t}>0} \sum_x \sum_{\hat{x}} p(x)Q(\hat{x}|x)\log\frac{Q(\hat{x}|x)}{t(\hat{x})}$$

An Expression for $R(D_s)$ - sD_s

$$R(D_s) - sD_s = \inf_{\mathbf{Q}>0} [I(\mathbf{p}, \mathbf{Q}) - sD(\mathbf{p}, \mathbf{Q})].$$

1. Since $I(\mathbf{p}, \mathbf{Q})$ and $D(\mathbf{p}, \mathbf{Q})$ are continuous in $\mathbf{Q}$, the minimum over all $\mathbf{Q}$ can be replaced by the infimum over all $\mathbf{Q} > 0$ (cf. Theorem 9.2).

2. Note that

$$D(\mathbf{p}, \mathbf{Q}) = \sum_x \sum_{\hat{x}} p(x)Q(\hat{x}|x)d(x, \hat{x}).$$

3. Then

$$\begin{aligned}
&R(D_s) - sD_s \\
&\quad = \inf_{\mathbf{Q}>0} [I(\mathbf{p}, \mathbf{Q}) - sD(\mathbf{p}, \mathbf{Q})] \\
&\quad = \inf_{\mathbf{Q}>0} \left[\min_{\mathbf{t}>0} \sum_{x,\hat{x}} p(x)Q(\hat{x}|x) \log \frac{Q(\hat{x}|x)}{t(\hat{x})} - s \sum_{x,\hat{x}} p(x)Q(\hat{x}|x)d(x, \hat{x}) \right] \\
&\quad = \inf_{\mathbf{Q}>0} \min_{\mathbf{t}>0} \left[\sum_{x,\hat{x}} p(x)Q(\hat{x}|x) \log \frac{Q(\hat{x}|x)}{t(\hat{x})} - s \sum_{x,\hat{x}} p(x)Q(\hat{x}|x)d(x, \hat{x}) \right].
\end{aligned}$$

Lemma 9.3

$$I(\mathbf{p}, \mathbf{Q}) = \min_{\mathbf{t}>0} \sum_x \sum_{\hat{x}} p(x)Q(\hat{x}|x) \log \frac{Q(\hat{x}|x)}{t(\hat{x})}$$

The BA Algorithm for Computing $R(D_s) - sD_s$

The BA Algorithm for Computing $R(D_s)$ - sD_s

Recall the double infimum in Section 9.1:

$$\inf_{\mathbf{u}_1 \in A_1} \inf_{\mathbf{u}_2 \in A_2} f(\mathbf{u}_1, \mathbf{u}_2).$$

The BA Algorithm for Computing $R(D_s)$ - sD_s

Recall the double infimum in Section 9.1:

$$\inf_{\mathbf{u}_1 \in A_1} \inf_{\mathbf{u}_2 \in A_2} f(\mathbf{u}_1, \mathbf{u}_2).$$

where

1. A_i is a convex subset of $\Re^{n_i}$ for $i = 1, 2$.

The BA Algorithm for Computing $R(D_s)$ - sD_s

Recall the double infimum in Section 9.1:

$$\inf_{\mathbf{u}_1 \in A_1} \inf_{\mathbf{u}_2 \in A_2} f(\mathbf{u}_1, \mathbf{u}_2).$$

where

1. A_i is a convex subset of $\Re^{n_i}$ for $i = 1, 2$.
2. $f : A_1 \times A_2 \to \Re$ is bounded from below and is such that

The BA Algorithm for Computing $R(D_s) - sD_s$

Recall the double infimum in Section 9.1:

$$\inf_{\mathbf{u}_1 \in A_1} \inf_{\mathbf{u}_2 \in A_2} f(\mathbf{u}_1, \mathbf{u}_2).$$

where

1. A_i is a convex subset of $\Re^{n_i}$ for $i = 1, 2$.
2. $f : A_1 \times A_2 \to \Re$ is bounded from below and is such that

- f is continuous and has continuous partial derivatives on $A_1 \times A_2$

The BA Algorithm for Computing $R(D_s)$ - sD_s

Recall the double infimum in Section 9.1:

$$\inf_{\mathbf{u}_1 \in A_1} \inf_{\mathbf{u}_2 \in A_2} f(\mathbf{u}_1, \mathbf{u}_2).$$

where

1. A_i is a convex subset of $\Re^{n_i}$ for $i = 1, 2$.

2. $f : A_1 \times A_2 \rightarrow \Re$ is bounded from below and is such that

- f is continuous and has continuous partial derivatives on $A_1 \times A_2$
- For all $\mathbf{u}_2 \in A_2$, there exists a unique $c_1(\mathbf{u}_2) \in A_1$ such that

The BA Algorithm for Computing $R(D_s) - sD_s$

Recall the double infimum in Section 9.1:

$$\inf_{\mathbf{u}_1 \in A_1} \inf_{\mathbf{u}_2 \in A_2} f(\mathbf{u}_1, \mathbf{u}_2).$$

where

1. A_i is a convex subset of $\Re^{n_i}$ for $i = 1, 2$.

2. $f : A_1 \times A_2 \to \Re$ is bounded from below and is such that

- f is continuous and has continuous partial derivatives on $A_1 \times A_2$
- For all $\mathbf{u}_2 \in A_2$, there exists a unique $c_1(\mathbf{u}_2) \in A_1$ such that

$$f(c_1(\mathbf{u}_2), \mathbf{u}_2) = \min_{\mathbf{u}_1' \in A_1} f(\mathbf{u}_1', \mathbf{u}_2),$$

The BA Algorithm for Computing $R(D_s)$ - sD_s

Recall the double infimum in Section 9.1:

$$\inf_{\mathbf{u}_1 \in A_1} \inf_{\mathbf{u}_2 \in A_2} f(\mathbf{u}_1, \mathbf{u}_2).$$

where

1. A_i is a convex subset of $\Re^{n_i}$ for $i = 1, 2$.

2. $f : A_1 \times A_2 \to \Re$ is bounded from below and is such that

- f is continuous and has continuous partial derivatives on $A_1 \times A_2$
- For all $\mathbf{u}_2 \in A_2$, there exists a unique $c_1(\mathbf{u}_2) \in A_1$ such that

$$f(c_1(\mathbf{u}_2), \mathbf{u}_2) = \min_{\mathbf{u}_1' \in A_1} f(\mathbf{u}_1', \mathbf{u}_2),$$

and for all $\mathbf{u}_1 \in A_1$, there exists a unique $c_2(\mathbf{u}_1) \in A_2$ such that

The BA Algorithm for Computing $R(D_s) - sD_s$

Recall the double infimum in Section 9.1:

$$\inf_{\mathbf{u}_1 \in A_1} \inf_{\mathbf{u}_2 \in A_2} f(\mathbf{u}_1, \mathbf{u}_2).$$

where

1. A_i is a convex subset of $\Re^{n_i}$ for $i = 1, 2$.

2. $f : A_1 \times A_2 \to \Re$ is bounded from below and is such that

- f is continuous and has continuous partial derivatives on $A_1 \times A_2$
- For all $\mathbf{u}_2 \in A_2$, there exists a unique $c_1(\mathbf{u}_2) \in A_1$ such that

$$f(c_1(\mathbf{u}_2), \mathbf{u}_2) = \min_{\mathbf{u}_1' \in A_1} f(\mathbf{u}_1', \mathbf{u}_2),$$

and for all $\mathbf{u}_1 \in A_1$, there exists a unique $c_2(\mathbf{u}_1) \in A_2$ such that

$$f(\mathbf{u}_1, c_2(\mathbf{u}_1)) = \min_{\mathbf{u}_2' \in A_2} f(\mathbf{u}_1, \mathbf{u}_2').$$

The BA Algorithm for Computing $R(D_s)$ - sD_s

Recall the double infimum in Section 9.1:

$$\inf_{\mathbf{u}_1 \in A_1} \inf_{\mathbf{u}_2 \in A_2} f(\mathbf{u}_1, \mathbf{u}_2).$$

where

1. A_i is a convex subset of $\Re^{n_i}$ for $i = 1, 2$.

2. $f : A_1 \times A_2 \to \Re$ is bounded from below and is such that

- f is continuous and has continuous partial derivatives on $A_1 \times A_2$
- For all $\mathbf{u}_2 \in A_2$, there exists a unique $c_1(\mathbf{u}_2) \in A_1$ such that

$$f(c_1(\mathbf{u}_2), \mathbf{u}_2) = \min_{\mathbf{u}'_1 \in A_1} f(\mathbf{u}'_1, \mathbf{u}_2),$$

and for all $\mathbf{u}_1 \in A_1$, there exists a unique $c_2(\mathbf{u}_1) \in A_2$ such that

$$f(\mathbf{u}_1, c_2(\mathbf{u}_1)) = \min_{\mathbf{u}'_2 \in A_2} f(\mathbf{u}_1, \mathbf{u}'_2).$$

The BA Algorithm for Computing $R(D_s)$ - sD_s

Recall the double infimum in Section 9.1:

$$\inf_{\mathbf{u}_1 \in A_1} \inf_{\mathbf{u}_2 \in A_2} f(\mathbf{u}_1, \mathbf{u}_2).$$

where

1. A_i is a convex subset of $\Re^{n_i}$ for $i = 1, 2$.

2. $f : A_1 \times A_2 \to \Re$ is bounded from below and is such that

- f is continuous and has continuous partial derivatives on $A_1 \times A_2$
- For all $\mathbf{u}_2 \in A_2$, there exists a unique $c_1(\mathbf{u}_2) \in A_1$ such that

$$f(c_1(\mathbf{u}_2), \mathbf{u}_2) = \min_{\mathbf{u}_1' \in A_1} f(\mathbf{u}_1', \mathbf{u}_2),$$

and for all $\mathbf{u}_1 \in A_1$, there exists a unique $c_2(\mathbf{u}_1) \in A_2$ such that

$$f(\mathbf{u}_1, c_2(\mathbf{u}_1)) = \min_{\mathbf{u}_2' \in A_2} f(\mathbf{u}_1, \mathbf{u}_2').$$

Cast the computation of $R(D_s) - sD_s$ into this optimization problem:

The BA Algorithm for Computing $R(D_s)$ - sD_s

Recall the double infimum in Section 9.1:

$$\inf_{\mathbf{u}_1 \in A_1} \inf_{\mathbf{u}_2 \in A_2} f(\mathbf{u}_1, \mathbf{u}_2).$$

where

1. A_i is a convex subset of $\Re^{n_i}$ for $i = 1, 2$.

2. $f : A_1 \times A_2 \rightarrow \Re$ is bounded from below and is such that

- f is continuous and has continuous partial derivatives on $A_1 \times A_2$
- For all $\mathbf{u}_2 \in A_2$, there exists a unique $c_1(\mathbf{u}_2) \in A_1$ such that

$$f(c_1(\mathbf{u}_2), \mathbf{u}_2) = \min_{\mathbf{u}_1' \in A_1} f(\mathbf{u}_1', \mathbf{u}_2),$$

and for all $\mathbf{u}_1 \in A_1$, there exists a unique $c_2(\mathbf{u}_1) \in A_2$ such that

$$f(\mathbf{u}_1, c_2(\mathbf{u}_1)) = \min_{\mathbf{u}_2' \in A_2} f(\mathbf{u}_1, \mathbf{u}_2').$$

Cast the computation of $R(D_s) - sD_s$ into this optimization problem:

1. Let

$$f(\mathbf{Q}, \mathbf{t}) = \sum_x \sum_{\hat{x}} p(x) Q(\hat{x}|x) \log \frac{Q(\hat{x}|x)}{t(\hat{x})} - s \sum_x \sum_{\hat{x}} p(x) Q(\hat{x}|x) d(x, \hat{x}),$$

where $\mathbf{u}_1 \leftarrow \mathbf{Q}$ and $\mathbf{u}_2 \leftarrow \mathbf{t}$.

The BA Algorithm for Computing $R(D_s)$ - sD_s

Recall the double infimum in Section 9.1:

$$\inf_{\mathbf{u}_1 \in A_1} \inf_{\mathbf{u}_2 \in A_2} f(\mathbf{u}_1, \mathbf{u}_2).$$

where

1. A_i is a convex subset of $\Re^{n_i}$ for $i = 1, 2$.

2. $f : A_1 \times A_2 \to \Re$ is bounded from below and is such that

- f is continuous and has continuous partial derivatives on $A_1 \times A_2$

- For all $\mathbf{u}_2 \in A_2$, there exists a unique $c_1(\mathbf{u}_2) \in A_1$ such that

$$f(c_1(\mathbf{u}_2), \mathbf{u}_2) = \min_{\mathbf{u}_1' \in A_1} f(\mathbf{u}_1', \mathbf{u}_2),$$

and for all $\mathbf{u}_1 \in A_1$, there exists a unique $c_2(\mathbf{u}_1) \in A_2$ such that

$$f(\mathbf{u}_1, c_2(\mathbf{u}_1)) = \min_{\mathbf{u}_2' \in A_2} f(\mathbf{u}_1, \mathbf{u}_2').$$

Cast the computation of $R(D_s) - sD_s$ into this optimization problem:

1. Let

$$f(\mathbf{Q}, \mathbf{t}) = \sum_x \sum_{\hat{x}} p(x) Q(\hat{x}|x) \log \frac{Q(\hat{x}|x)}{t(\hat{x})} - s \sum_x \sum_{\hat{x}} p(x) Q(\hat{x}|x) d(x, \hat{x}),$$

where $\mathbf{u}_1 \leftarrow \mathbf{Q}$ and $\mathbf{u}_2 \leftarrow \mathbf{t}$.

2. Let

$$A_1 = \Big\{ (Q(\hat{x}|x), (x, \hat{x}) \in \mathcal{X} \times \hat{\mathcal{X}}) : Q(\hat{x}|x) > 0, \sum_{\hat{x}} Q(\hat{x}|x) = 1 \text{ for all } x \in \mathcal{X} \Big\} \subset \Re^{|\mathcal{X}||\hat{\mathcal{X}}|}$$

The BA Algorithm for Computing $R(D_s)$ - sD_s

Recall the double infimum in Section 9.1:

$$\inf_{\mathbf{u}_1 \in A_1} \inf_{\mathbf{u}_2 \in A_2} f(\mathbf{u}_1, \mathbf{u}_2).$$

where

1. A_i is a convex subset of $\Re^{n_i}$ for $i = 1, 2$.

2. $f : A_1 \times A_2 \to \Re$ is bounded from below and is such that

- f is continuous and has continuous partial derivatives on $A_1 \times A_2$
- For all $\mathbf{u}_2 \in A_2$, there exists a unique $c_1(\mathbf{u}_2) \in A_1$ such that

$$f(c_1(\mathbf{u}_2), \mathbf{u}_2) = \min_{\mathbf{u}_1' \in A_1} f(\mathbf{u}_1', \mathbf{u}_2),$$

and for all $\mathbf{u}_1 \in A_1$, there exists a unique $c_2(\mathbf{u}_1) \in A_2$ such that

$$f(\mathbf{u}_1, c_2(\mathbf{u}_1)) = \min_{\mathbf{u}_2' \in A_2} f(\mathbf{u}_1, \mathbf{u}_2').$$

Cast the computation of $R(D_s) - sD_s$ into this optimization problem:

1. Let

$$f(\mathbf{Q}, \mathbf{t}) = \sum_x \sum_{\hat{x}} p(x) Q(\hat{x}|x) \log \frac{Q(\hat{x}|x)}{t(\hat{x})} - s \sum_x \sum_{\hat{x}} p(x) Q(\hat{x}|x) d(x, \hat{x}),$$

where $\mathbf{u}_1 \leftarrow \mathbf{Q}$ and $\mathbf{u}_2 \leftarrow \mathbf{t}$.

2. Let

$$\begin{aligned} A_1 &= \Bigg\{ (Q(\hat{x}|x), (x, \hat{x}) \in \mathcal{X} \times \hat{\mathcal{X}}) : Q(\hat{x}|x) > 0, \\ &\quad \sum_{\hat{x}} Q(\hat{x}|x) = 1 \text{ for all } x \in \mathcal{X} \Bigg\} \subset \Re^{|\mathcal{X}||\hat{\mathcal{X}}|} \end{aligned}$$

and

$$A_2 = \{(t(\hat{x}), \hat{x} \in \hat{\mathcal{X}}) : t(\hat{x}) > 0\}.$$

The BA Algorithm for Computing $R(D_s)$ - sD_s

Recall the double infimum in Section 9.1:

$$\inf_{\mathbf{u}_1 \in A_1} \inf_{\mathbf{u}_2 \in A_2} f(\mathbf{u}_1, \mathbf{u}_2).$$

where

1. A_i is a convex subset of $\Re^{n_i}$ for $i = 1, 2$.

2. $f : A_1 \times A_2 \to \Re$ is bounded from below and is such that

- f is continuous and has continuous partial derivatives on $A_1 \times A_2$
- For all $\mathbf{u}_2 \in A_2$, there exists a unique $c_1(\mathbf{u}_2) \in A_1$ such that

$$f(c_1(\mathbf{u}_2), \mathbf{u}_2) = \min_{\mathbf{u}'_1 \in A_1} f(\mathbf{u}'_1, \mathbf{u}_2),$$

and for all $\mathbf{u}_1 \in A_1$, there exists a unique $c_2(\mathbf{u}_1) \in A_2$ such that

$$f(\mathbf{u}_1, c_2(\mathbf{u}_1)) = \min_{\mathbf{u}'_2 \in A_2} f(\mathbf{u}_1, \mathbf{u}'_2).$$

Cast the computation of $R(D_s) - sD_s$ into this optimization problem:

1. Let

$$f(\mathbf{Q}, \mathbf{t}) = \sum_x \sum_{\hat{x}} p(x) Q(\hat{x}|x) \log \frac{Q(\hat{x}|x)}{t(\hat{x})} - s \sum_x \sum_{\hat{x}} p(x) Q(\hat{x}|x) d(x, \hat{x}),$$

where $\mathbf{u}_1 \leftarrow \mathbf{Q}$ and $\mathbf{u}_2 \leftarrow \mathbf{t}$.

2. Let

$$A_1 = \Big\{ (Q(\hat{x}|x), (x, \hat{x}) \in \mathcal{X} \times \hat{\mathcal{X}}) : Q(\hat{x}|x) > 0, \sum_{\hat{x}} Q(\hat{x}|x) = 1 \text{ for all } x \in \mathcal{X} \Big\} \subset \Re^{|\mathcal{X}||\hat{\mathcal{X}}|}$$

and

$$A_2 = \{ (t(\hat{x}), \hat{x} \in \hat{\mathcal{X}}) : t(\hat{x}) > 0 \}.$$

The BA Algorithm for Computing $R(D_s)$ - sD_s

Recall the double infimum in Section 9.1:

$$\inf_{\mathbf{u}_1 \in A_1} \inf_{\mathbf{u}_2 \in A_2} f(\mathbf{u}_1, \mathbf{u}_2).$$

where

1. A_i is a convex subset of $\Re^{n_i}$ for $i = 1, 2$.

2. $f : A_1 \times A_2 \to \Re$ is bounded from below and is such that

- f is continuous and has continuous partial derivatives on $A_1 \times A_2$
- For all $\mathbf{u}_2 \in A_2$, there exists a unique $c_1(\mathbf{u}_2) \in A_1$ such that

$$f(c_1(\mathbf{u}_2), \mathbf{u}_2) = \min_{\mathbf{u}_1' \in A_1} f(\mathbf{u}_1', \mathbf{u}_2),$$

and for all $\mathbf{u}_1 \in A_1$, there exists a unique $c_2(\mathbf{u}_1) \in A_2$ such that

$$f(\mathbf{u}_1, c_2(\mathbf{u}_1)) = \min_{\mathbf{u}_2' \in A_2} f(\mathbf{u}_1, \mathbf{u}_2').$$

Cast the computation of $R(D_s) - sD_s$ into this optimization problem:

1. Let

$$f(\mathbf{Q}, \mathbf{t}) = \sum_x \sum_{\hat{x}} p(x) Q(\hat{x}|x) \log \frac{Q(\hat{x}|x)}{t(\hat{x})} - s \sum_x \sum_{\hat{x}} p(x) Q(\hat{x}|x) d(x, \hat{x}),$$

where $\mathbf{u}_1 \leftarrow \mathbf{Q}$ and $\mathbf{u}_2 \leftarrow \mathbf{t}$.

2. Let

$$A_1 = \Big\{ (Q(\hat{x}|x), (x, \hat{x}) \in \mathcal{X} \times \hat{\mathcal{X}}) : Q(\hat{x}|x) > 0, \sum_{\hat{x}} Q(\hat{x}|x) = 1 \text{ for all } x \in \mathcal{X} \Big\} \subset \Re^{|\mathcal{X}||\hat{\mathcal{X}}|}$$

and

$$A_2 = \{ (t(\hat{x}), \hat{x} \in \hat{\mathcal{X}}) : t(\hat{x}) > 0 \}.$$

Remarks

The BA Algorithm for Computing $R(D_s)$ - sD_s

Recall the double infimum in Section 9.1:

$$\inf_{\mathbf{u}_1 \in A_1} \inf_{\mathbf{u}_2 \in A_2} f(\mathbf{u}_1, \mathbf{u}_2).$$

where

1. A_i is a convex subset of $\Re^{n_i}$ for $i = 1, 2$.

2. $f : A_1 \times A_2 \to \Re$ is bounded from below and is such that

- f is continuous and has continuous partial derivatives on $A_1 \times A_2$

- For all $\mathbf{u}_2 \in A_2$, there exists a unique $c_1(\mathbf{u}_2) \in A_1$ such that

$$f(c_1(\mathbf{u}_2), \mathbf{u}_2) = \min_{\mathbf{u}'_1 \in A_1} f(\mathbf{u}'_1, \mathbf{u}_2),$$

and for all $\mathbf{u}_1 \in A_1$, there exists a unique $c_2(\mathbf{u}_1) \in A_2$ such that

$$f(\mathbf{u}_1, c_2(\mathbf{u}_1)) = \min_{\mathbf{u}'_2 \in A_2} f(\mathbf{u}_1, \mathbf{u}'_2).$$

Cast the computation of $R(D_s) - sD_s$ into this optimization problem:

1. Let

$$f(\mathbf{Q}, \mathbf{t}) = \sum_x \sum_{\hat{x}} p(x) Q(\hat{x}|x) \log \frac{Q(\hat{x}|x)}{t(\hat{x})} - s \sum_x \sum_{\hat{x}} p(x) Q(\hat{x}|x) d(x, \hat{x}),$$

where $\mathbf{u}_1 \leftarrow \mathbf{Q}$ and $\mathbf{u}_2 \leftarrow \mathbf{t}$.

2. Let

$$A_1 = \Big\{ (Q(\hat{x}|x), (x, \hat{x}) \in \mathcal{X} \times \hat{\mathcal{X}}) : Q(\hat{x}|x) > 0, \sum_{\hat{x}} Q(\hat{x}|x) = 1 \text{ for all } x \in \mathcal{X} \Big\} \subset \Re^{|\mathcal{X}||\hat{\mathcal{X}}|}$$

and

$$A_2 = \{ (t(\hat{x}), \hat{x} \in \hat{\mathcal{X}}) : t(\hat{x}) > 0 \}.$$

Remarks

1. Both A_1 and A_2 are convex.

The BA Algorithm for Computing $R(D_s)$ - sD_s

Recall the double infimum in Section 9.1:

$$\inf_{\mathbf{u}_1 \in A_1} \inf_{\mathbf{u}_2 \in A_2} f(\mathbf{u}_1, \mathbf{u}_2).$$

where

✓ 1. A_i is a convex subset of $\Re^{n_i}$ for $i = 1, 2$.

2. $f : A_1 \times A_2 \to \Re$ is bounded from below and is such that

- f is continuous and has continuous partial derivatives on $A_1 \times A_2$
- For all $\mathbf{u}_2 \in A_2$, there exists a unique $c_1(\mathbf{u}_2) \in A_1$ such that

$$f(c_1(\mathbf{u}_2), \mathbf{u}_2) = \min_{\mathbf{u}_1' \in A_1} f(\mathbf{u}_1', \mathbf{u}_2),$$

and for all $\mathbf{u}_1 \in A_1$, there exists a unique $c_2(\mathbf{u}_1) \in A_2$ such that

$$f(\mathbf{u}_1, c_2(\mathbf{u}_1)) = \min_{\mathbf{u}_2' \in A_2} f(\mathbf{u}_1, \mathbf{u}_2').$$

Cast the computation of $R(D_s) - sD_s$ into this optimization problem:

1. Let

$$f(\mathbf{Q}, \mathbf{t}) = \sum_x \sum_{\hat{x}} p(x) Q(\hat{x}|x) \log \frac{Q(\hat{x}|x)}{t(\hat{x})} - s \sum_x \sum_{\hat{x}} p(x) Q(\hat{x}|x) d(x, \hat{x}),$$

where $\mathbf{u}_1 \leftarrow \mathbf{Q}$ and $\mathbf{u}_2 \leftarrow \mathbf{t}$.

2. Let

$$A_1 = \left\{ (Q(\hat{x}|x), (x, \hat{x}) \in \mathcal{X} \times \hat{\mathcal{X}}) : Q(\hat{x}|x) > 0, \sum_{\hat{x}} Q(\hat{x}|x) = 1 \text{ for all } x \in \mathcal{X} \right\} \subset \Re^{|\mathcal{X}||\hat{\mathcal{X}}|}$$

and

$$A_2 = \{ (t(\hat{x}), \hat{x} \in \hat{\mathcal{X}}) : t(\hat{x}) > 0 \}.$$

Remarks

1. Both A_1 and A_2 are convex.

The BA Algorithm for Computing $R(D_s)$ - sD_s

Recall the double infimum in Section 9.1:

$$\inf_{\mathbf{u}_1 \in A_1} \inf_{\mathbf{u}_2 \in A_2} f(\mathbf{u}_1, \mathbf{u}_2).$$

where

1. A_i is a convex subset of $\Re^{n_i}$ for $i = 1, 2$.

2. $f : A_1 \times A_2 \to \Re$ is bounded from below and is such that

- f is continuous and has continuous partial derivatives on $A_1 \times A_2$
- For all $\mathbf{u}_2 \in A_2$, there exists a unique $c_1(\mathbf{u}_2) \in A_1$ such that

$$f(c_1(\mathbf{u}_2), \mathbf{u}_2) = \min_{\mathbf{u}_1' \in A_1} f(\mathbf{u}_1', \mathbf{u}_2),$$

and for all $\mathbf{u}_1 \in A_1$, there exists a unique $c_2(\mathbf{u}_1) \in A_2$ such that

$$f(\mathbf{u}_1, c_2(\mathbf{u}_1)) = \min_{\mathbf{u}_2' \in A_2} f(\mathbf{u}_1, \mathbf{u}_2').$$

Cast the computation of $R(D_s) - sD_s$ into this optimization problem:

1. Let

$$f(\mathbf{Q}, \mathbf{t}) = \sum_x \sum_{\hat{x}} p(x) Q(\hat{x}|x) \log \frac{Q(\hat{x}|x)}{t(\hat{x})} - s \sum_x \sum_{\hat{x}} p(x) Q(\hat{x}|x) d(x, \hat{x}),$$

where $\mathbf{u}_1 \leftarrow \mathbf{Q}$ and $\mathbf{u}_2 \leftarrow \mathbf{t}$.

2. Let

$$\begin{aligned} A_1 \quad &= \quad \Big\{ (Q(\hat{x}|x), (x, \hat{x}) \in \mathcal{X} \times \hat{\mathcal{X}}) : Q(\hat{x}|x) > 0, \\ &\qquad \sum_{\hat{x}} Q(\hat{x}|x) = 1 \text{ for all } x \in \mathcal{X} \Big\} \quad \subset \quad \Re^{|\mathcal{X}||\hat{\mathcal{X}}|} \end{aligned}$$

and

$$A_2 \quad = \quad \{(t(\hat{x}), \hat{x} \in \hat{\mathcal{X}}) : t(\hat{x}) > 0\}.$$

Remarks

1. Both A_1 and A_2 are convex.
2. f is bounded from below because $s \leq 0$.

The BA Algorithm for Computing $R(D_s)$ - sD_s

Recall the double infimum in Section 9.1:

$$\inf_{\mathbf{u}_1 \in A_1} \inf_{\mathbf{u}_2 \in A_2} f(\mathbf{u}_1, \mathbf{u}_2).$$

where

✓ 1. A_i is a convex subset of $\Re^{n_i}$ for $i = 1, 2$.

2. $f : A_1 \times A_2 \to \Re$ is bounded from below and is such that

- f is continuous and has continuous partial derivatives on $A_1 \times A_2$
- For all $\mathbf{u}_2 \in A_2$, there exists a unique $c_1(\mathbf{u}_2) \in A_1$ such that

$$f(c_1(\mathbf{u}_2), \mathbf{u}_2) = \min_{\mathbf{u}_1' \in A_1} f(\mathbf{u}_1', \mathbf{u}_2),$$

and for all $\mathbf{u}_1 \in A_1$, there exists a unique $c_2(\mathbf{u}_1) \in A_2$ such that

$$f(\mathbf{u}_1, c_2(\mathbf{u}_1)) = \min_{\mathbf{u}_2' \in A_2} f(\mathbf{u}_1, \mathbf{u}_2').$$

Cast the computation of $R(D_s) - sD_s$ into this optimization problem:

1. Let

$$f(\mathbf{Q}, \mathbf{t}) = \sum_x \sum_{\hat{x}} p(x) Q(\hat{x}|x) \log \frac{Q(\hat{x}|x)}{t(\hat{x})} - s \underline{\sum_x \sum_{\hat{x}} p(x) Q(\hat{x}|x) d(x, \hat{x})},$$

where $\mathbf{u}_1 \leftarrow \mathbf{Q}$ and $\mathbf{u}_2 \leftarrow \mathbf{t}$.

2. Let

$$\begin{aligned} A_1 &= \Big\{ (Q(\hat{x}|x), (x, \hat{x}) \in \mathcal{X} \times \hat{\mathcal{X}}) : Q(\hat{x}|x) > 0, \\ &\qquad \sum_{\hat{x}} Q(\hat{x}|x) = 1 \text{ for all } x \in \mathcal{X} \Big\} \subset \Re^{|\mathcal{X}||\hat{\mathcal{X}}|} \end{aligned}$$

and

$$A_2 = \{ (t(\hat{x}), \hat{x} \in \hat{\mathcal{X}}) : t(\hat{x}) > 0 \}.$$

Remarks

1. Both A_1 and A_2 are convex.
2. f is bounded from below because $s \leq 0$.

The BA Algorithm for Computing $R(D_s)$ - sD_s

Recall the double infimum in Section 9.1:

$$\inf_{\mathbf{u}_1 \in A_1} \inf_{\mathbf{u}_2 \in A_2} f(\mathbf{u}_1, \mathbf{u}_2).$$

where

✓ 1. A_i is a convex subset of $\Re^{n_i}$ for $i = 1, 2$.

2. $f : A_1 \times A_2 \to \Re$ is bounded from below and is such that

- f is continuous and has continuous partial derivatives on $A_1 \times A_2$
- For all $\mathbf{u}_2 \in A_2$, there exists a unique $c_1(\mathbf{u}_2) \in A_1$ such that

$$f(c_1(\mathbf{u}_2), \mathbf{u}_2) = \min_{\mathbf{u}_1' \in A_1} f(\mathbf{u}_1', \mathbf{u}_2),$$

and for all $\mathbf{u}_1 \in A_1$, there exists a unique $c_2(\mathbf{u}_1) \in A_2$ such that

$$f(\mathbf{u}_1, c_2(\mathbf{u}_1)) = \min_{\mathbf{u}_2' \in A_2} f(\mathbf{u}_1, \mathbf{u}_2').$$

Cast the computation of $R(D_s) - sD_s$ into this optimization problem:

1. Let

$$f(\mathbf{Q}, \mathbf{t}) = \underbrace{\sum_x \sum_{\hat{x}} p(x) Q(\hat{x}|x) \log \frac{Q(\hat{x}|x)}{t(\hat{x})}}_{\geq I(\mathbf{p}, \mathbf{Q}) \geq 0} - s \underline{\sum_x \sum_{\hat{x}} p(x) Q(\hat{x}|x) d(x, \hat{x})},$$

where $\mathbf{u}_1 \leftarrow \mathbf{Q}$ and $\mathbf{u}_2 \leftarrow \mathbf{t}$.

2. Let

$$A_1 = \left\{ (Q(\hat{x}|x), (x, \hat{x}) \in \mathcal{X} \times \hat{\mathcal{X}}) : Q(\hat{x}|x) > 0, \sum_{\hat{x}} Q(\hat{x}|x) = 1 \text{ for all } x \in \mathcal{X} \right\} \subset \Re^{|\mathcal{X}||\hat{\mathcal{X}}|}$$

and

$$A_2 = \{(t(\hat{x}), \hat{x} \in \hat{\mathcal{X}}) : t(\hat{x}) > 0\}.$$

Remarks

1. Both A_1 and A_2 are convex.
2. f is bounded from below because $s \leq 0$.

The BA Algorithm for Computing $R(D_s)$ - sD_s

Recall the double infimum in Section 9.1:

$$\inf_{\mathbf{u}_1 \in A_1} \inf_{\mathbf{u}_2 \in A_2} f(\mathbf{u}_1, \mathbf{u}_2).$$

where

✓ 1. A_i is a convex subset of $\Re^{n_i}$ for $i = 1, 2$.

2. $f : A_1 \times A_2 \to \Re$ is bounded from below and is such that

- f is continuous and has continuous partial derivatives on $A_1 \times A_2$
- For all $\mathbf{u}_2 \in A_2$, there exists a unique $c_1(\mathbf{u}_2) \in A_1$ such that

$$f(c_1(\mathbf{u}_2), \mathbf{u}_2) = \min_{\mathbf{u}_1' \in A_1} f(\mathbf{u}_1', \mathbf{u}_2),$$

and for all $\mathbf{u}_1 \in A_1$, there exists a unique $c_2(\mathbf{u}_1) \in A_2$ such that

$$f(\mathbf{u}_1, c_2(\mathbf{u}_1)) = \min_{\mathbf{u}_2' \in A_2} f(\mathbf{u}_1, \mathbf{u}_2').$$

Cast the computation of $R(D_s) - sD_s$ into this optimization problem:

1. Let

$\geq I(\mathbf{p}, \mathbf{Q}) \geq 0$

$$f(\mathbf{Q}, \mathbf{t}) = \sum_x \sum_{\hat{x}} p(x) Q(\hat{x}|x) \log \frac{Q(\hat{x}|x)}{t(\hat{x})} - s \underline{\sum_x \sum_{\hat{x}} p(x) Q(\hat{x}|x) d(x, \hat{x})},$$

where $\mathbf{u}_1 \leftarrow \mathbf{Q}$ and $\mathbf{u}_2 \leftarrow \mathbf{t}$.

2. Let

$$\begin{aligned} A_1 &= \Big\{ (Q(\hat{x}|x), (x, \hat{x}) \in \mathcal{X} \times \hat{\mathcal{X}}) : Q(\hat{x}|x) > 0, \\ &\quad \sum_{\hat{x}} Q(\hat{x}|x) = 1 \text{ for all } x \in \mathcal{X} \Big\} \subset \Re^{|\mathcal{X}||\hat{\mathcal{X}}|} \end{aligned}$$

and

$$A_2 = \{ (t(\hat{x}), \hat{x} \in \hat{\mathcal{X}}) : t(\hat{x}) > 0 \}.$$

Remarks

1. Both A_1 and A_2 are convex.
2. f is bounded from below because $\underline{s \leq 0}$.

The BA Algorithm for Computing $R(D_s)$ - sD_s

Recall the double infimum in Section 9.1:

$$\inf_{\mathbf{u}_1 \in A_1} \inf_{\mathbf{u}_2 \in A_2} f(\mathbf{u}_1, \mathbf{u}_2).$$

where

1. A_i is a convex subset of $\Re^{n_i}$ for $i = 1, 2$.
2. $f : A_1 \times A_2 \to \Re$ is bounded from below and is such that

- f is continuous and has continuous partial derivatives on $A_1 \times A_2$
- For all $\mathbf{u}_2 \in A_2$, there exists a unique $c_1(\mathbf{u}_2) \in A_1$ such that

$$f(c_1(\mathbf{u}_2), \mathbf{u}_2) = \min_{\mathbf{u}_1' \in A_1} f(\mathbf{u}_1', \mathbf{u}_2),$$

and for all $\mathbf{u}_1 \in A_1$, there exists a unique $c_2(\mathbf{u}_1) \in A_2$ such that

$$f(\mathbf{u}_1, c_2(\mathbf{u}_1)) = \min_{\mathbf{u}_2' \in A_2} f(\mathbf{u}_1, \mathbf{u}_2').$$

Cast the computation of $R(D_s) - sD_s$ into this optimization problem:

1. Let

$\geq I(\mathbf{p}, \mathbf{Q}) \geq 0$

$$f(\mathbf{Q}, \mathbf{t}) = \sum_x \sum_{\hat{x}} p(x) Q(\hat{x}|x) \log \frac{Q(\hat{x}|x)}{t(\hat{x})} - s \sum_x \sum_{\hat{x}} p(x) Q(\hat{x}|x) d(x, \hat{x}),$$

where $\mathbf{u}_1 \leftarrow \mathbf{Q}$ and $\mathbf{u}_2 \leftarrow \mathbf{t}$.

2. Let

$$A_1 = \Big\{ (Q(\hat{x}|x), (x, \hat{x}) \in \mathcal{X} \times \hat{\mathcal{X}}) : Q(\hat{x}|x) > 0, \sum_{\hat{x}} Q(\hat{x}|x) = 1 \text{ for all } x \in \mathcal{X} \Big\} \subset \Re^{|\mathcal{X}||\hat{\mathcal{X}}|}$$

and

$$A_2 = \{ (t(\hat{x}), \hat{x} \in \hat{\mathcal{X}}) : t(\hat{x}) > 0 \}.$$

Remarks

1. Both A_1 and A_2 are convex.
2. f is bounded from below because $s \leq 0$.

The BA Algorithm for Computing $R(D_s)$ - sD_s

Recall the double infimum in Section 9.1:

$$\inf_{\mathbf{u}_1 \in A_1} \inf_{\mathbf{u}_2 \in A_2} f(\mathbf{u}_1, \mathbf{u}_2).$$

where

✓ 1. A_i is a convex subset of $\Re^{n_i}$ for $i = 1, 2$.

✓ 2. $f : A_1 \times A_2 \to \Re$ is bounded from below and is such that

- f is continuous and has continuous partial derivatives on $A_1 \times A_2$
- For all $\mathbf{u}_2 \in A_2$, there exists a unique $c_1(\mathbf{u}_2) \in A_1$ such that

$$f(c_1(\mathbf{u}_2), \mathbf{u}_2) = \min_{\mathbf{u}'_1 \in A_1} f(\mathbf{u}'_1, \mathbf{u}_2),$$

and for all $\mathbf{u}_1 \in A_1$, there exists a unique $c_2(\mathbf{u}_1) \in A_2$ such that

$$f(\mathbf{u}_1, c_2(\mathbf{u}_1)) = \min_{\mathbf{u}'_2 \in A_2} f(\mathbf{u}_1, \mathbf{u}'_2).$$

Cast the computation of $R(D_s) - sD_s$ into this optimization problem:

1. Let

$$f(\mathbf{Q}, \mathbf{t}) = \sum_x \sum_{\hat{x}} p(x) Q(\hat{x}|x) \log \frac{Q(\hat{x}|x)}{t(\hat{x})} - s \sum_x \sum_{\hat{x}} p(x) Q(\hat{x}|x) d(x, \hat{x}),$$

where $\mathbf{u}_1 \leftarrow \mathbf{Q}$ and $\mathbf{u}_2 \leftarrow \mathbf{t}$.

2. Let

$$A_1 = \Big\{ (Q(\hat{x}|x), (x, \hat{x}) \in \mathcal{X} \times \hat{\mathcal{X}}) : Q(\hat{x}|x) > 0, \sum_{\hat{x}} Q(\hat{x}|x) = 1 \text{ for all } x \in \mathcal{X} \Big\} \subset \Re^{|\mathcal{X}||\hat{\mathcal{X}}|}$$

and

$$A_2 = \{ (t(\hat{x}), \hat{x} \in \hat{\mathcal{X}}) : t(\hat{x}) > 0 \}.$$

Remarks

1. Both A_1 and A_2 are convex.
2. f is bounded from below because $s \le 0$.
3. In $f(\mathbf{Q}, \mathbf{t})$, since $\mathbf{Q} \in A_1$ and $\mathbf{t} \in A_2$, $Q(\hat{x}|x) > 0$ and $t(\hat{x}) > 0$ for all x and $\hat{x}$ in the double summations.

The BA Algorithm for Computing $R(D_s)$ - sD_s

Recall the double infimum in Section 9.1:

$$\inf_{\mathbf{u}_1 \in A_1} \inf_{\mathbf{u}_2 \in A_2} f(\mathbf{u}_1, \mathbf{u}_2).$$

where

✓ 1. A_i is a convex subset of $\Re^{n_i}$ for $i = 1, 2$.

✓ 2. $f : A_1 \times A_2 \rightarrow \Re$ is bounded from below and is such that

- f is continuous and has continuous partial derivatives on $A_1 \times A_2$
- For all $\mathbf{u}_2 \in A_2$, there exists a unique $c_1(\mathbf{u}_2) \in A_1$ such that

$$f(c_1(\mathbf{u}_2), \mathbf{u}_2) = \min_{\mathbf{u}_1' \in A_1} f(\mathbf{u}_1', \mathbf{u}_2),$$

and for all $\mathbf{u}_1 \in A_1$, there exists a unique $c_2(\mathbf{u}_1) \in A_2$ such that

$$f(\mathbf{u}_1, c_2(\mathbf{u}_1)) = \min_{\mathbf{u}_2' \in A_2} f(\mathbf{u}_1, \mathbf{u}_2').$$

Cast the computation of $R(D_s) - sD_s$ into this optimization problem:

1. Let

$$f(\mathbf{Q}, \mathbf{t}) = \sum_x \sum_{\hat{x}} p(x) Q(\hat{x}|x) \log \frac{Q(\hat{x}|x)}{t(\hat{x})} - s \sum_x \sum_{\hat{x}} p(x) Q(\hat{x}|x) d(x, \hat{x}),$$

where $\mathbf{u}_1 \leftarrow \mathbf{Q}$ and $\mathbf{u}_2 \leftarrow \mathbf{t}$.

2. Let

$$A_1 = \Big\{ (Q(\hat{x}|x), (x, \hat{x}) \in \mathcal{X} \times \hat{\mathcal{X}}) : Q(\hat{x}|x) > 0, \sum_{\hat{x}} Q(\hat{x}|x) = 1 \text{ for all } x \in \mathcal{X} \Big\} \subset \Re^{|\mathcal{X}||\hat{\mathcal{X}}|}$$

and

$$A_2 = \{ (t(\hat{x}), \hat{x} \in \hat{\mathcal{X}}) : t(\hat{x}) > 0 \}.$$

Remarks

1. Both A_1 and A_2 are convex.

2. f is bounded from below because $s \le 0$.

3. In $f(\mathbf{Q}, \mathbf{t})$, since $\mathbf{Q} \in A_1$ and $\mathbf{t} \in A_2$, $Q(\hat{x}|x) > 0$ and $t(\hat{x}) > 0$ for all x and $\hat{x}$ in the double summations.

4. Therefore, f is continuous and has continuous partial derivatives on $A = A_1 \times A_2$.

The BA Algorithm for Computing $R(D_s)$ - sD_s

Recall the double infimum in Section 9.1:

$$\inf_{\mathbf{u}_1 \in A_1} \inf_{\mathbf{u}_2 \in A_2} f(\mathbf{u}_1, \mathbf{u}_2).$$

where

✓ 1. A_i is a convex subset of $\Re^{n_i}$ for $i = 1, 2$.

✓ 2. $f : A_1 \times A_2 \to \Re$ is bounded from below and is such that

✓ • f is continuous and has continuous partial derivatives on $A_1 \times A_2$

• For all $\mathbf{u}_2 \in A_2$, there exists a unique $c_1(\mathbf{u}_2) \in A_1$ such that

$$f(c_1(\mathbf{u}_2), \mathbf{u}_2) = \min_{\mathbf{u}_1' \in A_1} f(\mathbf{u}_1', \mathbf{u}_2),$$

and for all $\mathbf{u}_1 \in A_1$, there exists a unique $c_2(\mathbf{u}_1) \in A_2$ such that

$$f(\mathbf{u}_1, c_2(\mathbf{u}_1)) = \min_{\mathbf{u}_2' \in A_2} f(\mathbf{u}_1, \mathbf{u}_2').$$

Cast the computation of $R(D_s) - sD_s$ into this optimization problem:

1. Let

$$f(\mathbf{Q}, \mathbf{t}) = \sum_x \sum_{\hat{x}} p(x) Q(\hat{x}|x) \log \frac{Q(\hat{x}|x)}{t(\hat{x})} - s \sum_x \sum_{\hat{x}} p(x) Q(\hat{x}|x) d(x, \hat{x}),$$

where $\mathbf{u}_1 \leftarrow \mathbf{Q}$ and $\mathbf{u}_2 \leftarrow \mathbf{t}$.

2. Let

$$A_1 = \left\{ (Q(\hat{x}|x), (x, \hat{x}) \in \mathcal{X} \times \hat{\mathcal{X}}) : Q(\hat{x}|x) > 0, \sum_{\hat{x}} Q(\hat{x}|x) = 1 \text{ for all } x \in \mathcal{X} \right\} \subset \Re^{|\mathcal{X}||\hat{\mathcal{X}}|}$$

and

$$A_2 = \{(t(\hat{x}), \hat{x} \in \hat{\mathcal{X}}) : t(\hat{x}) > 0\}.$$

Remarks

1. Both A_1 and A_2 are convex.

2. f is bounded from below because $s \leq 0$.

3. In $f(\mathbf{Q}, \mathbf{t})$, since $\mathbf{Q} \in A_1$ and $\mathbf{t} \in A_2$, $Q(\hat{x}|x) > 0$ and $t(\hat{x}) > 0$ for all x and $\hat{x}$ in the double summations.

4. Therefore, f is continuous and has continuous partial derivatives on $A = A_1 \times A_2$.

The BA Algorithm for Computing $R(D_s)$ - sD_s

Recall the double infimum in Section 9.1:

$$\inf_{\mathbf{u}_1 \in A_1} \inf_{\mathbf{u}_2 \in A_2} f(\mathbf{u}_1, \mathbf{u}_2).$$

where

1. A_i is a convex subset of $\Re^{n_i}$ for $i = 1, 2$.
2. $f : A_1 \times A_2 \to \Re$ is bounded from below and is such that

- f is continuous and has continuous partial derivatives on $A_1 \times A_2$
- For all $\mathbf{u}_2 \in A_2$, there exists a unique $c_1(\mathbf{u}_2) \in A_1$ such that

$$f(c_1(\mathbf{u}_2), \mathbf{u}_2) = \min_{\mathbf{u}_1' \in A_1} f(\mathbf{u}_1', \mathbf{u}_2),$$

and for all $\mathbf{u}_1 \in A_1$, there exists a unique $c_2(\mathbf{u}_1) \in A_2$ such that

$$f(\mathbf{u}_1, c_2(\mathbf{u}_1)) = \min_{\mathbf{u}_2' \in A_2} f(\mathbf{u}_1, \mathbf{u}_2').$$

Cast the computation of $R(D_s) - sD_s$ into this optimization problem:

1. Let

$$f(\mathbf{Q}, \mathbf{t}) = \sum_x \sum_{\hat{x}} p(x) Q(\hat{x}|x) \log \frac{Q(\hat{x}|x)}{t(\hat{x})} - s \sum_x \sum_{\hat{x}} p(x) Q(\hat{x}|x) d(x, \hat{x}),$$

where $\mathbf{u}_1 \leftarrow \mathbf{Q}$ and $\mathbf{u}_2 \leftarrow \mathbf{t}$.

2. Let

$$A_1 = \Big\{ (Q(\hat{x}|x), (x, \hat{x}) \in \mathcal{X} \times \hat{\mathcal{X}}) : Q(\hat{x}|x) > 0, \sum_{\hat{x}} Q(\hat{x}|x) = 1 \text{ for all } x \in \mathcal{X} \Big\} \subset \Re^{|\mathcal{X}||\hat{\mathcal{X}}|}$$

and

$$A_2 = \{ (t(\hat{x}), \hat{x} \in \hat{\mathcal{X}}) : t(\hat{x}) > 0 \}.$$

The BA Algorithm for Computing $R(D_s)$ - sD_s

Recall the double infimum in Section 9.1:

$$\inf_{\mathbf{u}_1 \in A_1} \inf_{\mathbf{u}_2 \in A_2} f(\mathbf{u}_1, \mathbf{u}_2).$$

where

✓ 1. A_i is a convex subset of $\Re^{n_i}$ for $i = 1, 2$.

✓ 2. $f : A_1 \times A_2 \to \Re$ is bounded from below and is such that

✓ • f is continuous and has continuous partial derivatives on $A_1 \times A_2$

• For all $\mathbf{u}_2 \in A_2$, there exists a unique $c_1(\mathbf{u}_2) \in A_1$ such that

$$f(c_1(\mathbf{u}_2), \mathbf{u}_2) = \min_{\mathbf{u}'_1 \in A_1} f(\mathbf{u}'_1, \mathbf{u}_2),$$

and for all $\mathbf{u}_1 \in A_1$, there exists a unique $c_2(\mathbf{u}_1) \in A_2$ such that

$$f(\mathbf{u}_1, c_2(\mathbf{u}_1)) = \min_{\mathbf{u}'_2 \in A_2} f(\mathbf{u}_1, \mathbf{u}'_2).$$

Cast the computation of $R(D_s) - sD_s$ into this optimization problem:

1. Let

$$f(\mathbf{Q}, \mathbf{t}) = \sum_x \sum_{\hat{x}} p(x) Q(\hat{x}|x) \log \frac{Q(\hat{x}|x)}{t(\hat{x})} - s \sum_x \sum_{\hat{x}} p(x) Q(\hat{x}|x) d(x, \hat{x}),$$

where $\mathbf{u}_1 \leftarrow \mathbf{Q}$ and $\mathbf{u}_2 \leftarrow \mathbf{t}$.

2. Let

$$A_1 = \left\{ (Q(\hat{x}|x), (x, \hat{x}) \in \mathcal{X} \times \hat{\mathcal{X}}) : Q(\hat{x}|x) > 0, \sum_{\hat{x}} Q(\hat{x}|x) = 1 \text{ for all } x \in \mathcal{X} \right\} \subset \Re^{|\mathcal{X}||\hat{\mathcal{X}}|}$$

and

$$A_2 = \{ (t(\hat{x}), \hat{x} \in \hat{\mathcal{X}}) : t(\hat{x}) > 0 \}.$$

3. The double infimum now becomes

$$\inf_{\mathbf{Q} \in A_1} \inf_{\mathbf{t} \in A_2} f(\mathbf{Q}, \mathbf{t}) = \inf_{\mathbf{Q} \in A_1} \inf_{\mathbf{t} \in A_2} \left[\sum_x \sum_{\hat{x}} p(x) Q(\hat{x}|x) \log \frac{Q(\hat{x}|x)}{t(\hat{x})} - s \sum_x \sum_{\hat{x}} p(x) Q(\hat{x}|x) d(x, \hat{x}) \right],$$

The BA Algorithm for Computing $R(D_s)$ - sD_s

Recall the double infimum in Section 9.1:

$$\inf_{\mathbf{u}_1 \in A_1} \inf_{\mathbf{u}_2 \in A_2} f(\mathbf{u}_1, \mathbf{u}_2).$$

where

1. A_i is a convex subset of $\Re^{n_i}$ for $i = 1, 2$.
2. $f : A_1 \times A_2 \to \Re$ is bounded from below and is such that

- f is continuous and has continuous partial derivatives on $A_1 \times A_2$
- For all $\mathbf{u}_2 \in A_2$, there exists a unique $c_1(\mathbf{u}_2) \in A_1$ such that

$$f(c_1(\mathbf{u}_2), \mathbf{u}_2) = \min_{\mathbf{u}_1' \in A_1} f(\mathbf{u}_1', \mathbf{u}_2),$$

and for all $\mathbf{u}_1 \in A_1$, there exists a unique $c_2(\mathbf{u}_1) \in A_2$ such that

$$f(\mathbf{u}_1, c_2(\mathbf{u}_1)) = \min_{\mathbf{u}_2' \in A_2} f(\mathbf{u}_1, \mathbf{u}_2').$$

Cast the computation of $R(D_s) - sD_s$ into this optimization problem:

1. Let

$$f(\mathbf{Q}, \mathbf{t}) = \sum_x \sum_{\hat{x}} p(x) Q(\hat{x}|x) \log \frac{Q(\hat{x}|x)}{t(\hat{x})} - s \sum_x \sum_{\hat{x}} p(x) Q(\hat{x}|x) d(x, \hat{x}),$$

where $\mathbf{u}_1 \leftarrow \mathbf{Q}$ and $\mathbf{u}_2 \leftarrow \mathbf{t}$.

2. Let

$$A_1 = \Big\{ (Q(\hat{x}|x), (x, \hat{x}) \in \mathcal{X} \times \hat{\mathcal{X}}) : Q(\hat{x}|x) > 0, \sum_{\hat{x}} Q(\hat{x}|x) = 1 \text{ for all } x \in \mathcal{X} \Big\} \subset \Re^{|\mathcal{X}||\hat{\mathcal{X}}|}$$

and

$$A_2 = \{ (t(\hat{x}), \hat{x} \in \hat{\mathcal{X}}) : t(\hat{x}) > 0 \}.$$

3. The double infimum now becomes

$$\inf_{\mathbf{Q} \in A_1} \inf_{\mathbf{t} \in A_2} f(\mathbf{Q}, \mathbf{t}) = \inf_{\mathbf{Q} \in A_1} \inf_{\mathbf{t} \in A_2} \left[\sum_x \sum_{\hat{x}} p(x) Q(\hat{x}|x) \log \frac{Q(\hat{x}|x)}{t(\hat{x})} - s \sum_x \sum_{\hat{x}} p(x) Q(\hat{x}|x) d(x, \hat{x}) \right],$$

where the infimum over all $\mathbf{t} \in A_2$ is in fact a minimum, and

The BA Algorithm for Computing $R(D_s)$ - sD_s

Recall the double infimum in Section 9.1:

$$\inf_{\mathbf{u}_1 \in A_1} \inf_{\mathbf{u}_2 \in A_2} f(\mathbf{u}_1, \mathbf{u}_2).$$

where

1. A_i is a convex subset of $\Re^{n_i}$ for $i = 1, 2$.
2. $f : A_1 \times A_2 \rightarrow \Re$ is bounded from below and is such that

- f is continuous and has continuous partial derivatives on $A_1 \times A_2$
- For all $\mathbf{u}_2 \in A_2$, there exists a unique $c_1(\mathbf{u}_2) \in A_1$ such that

$$f(c_1(\mathbf{u}_2), \mathbf{u}_2) = \min_{\mathbf{u}_1' \in A_1} f(\mathbf{u}_1', \mathbf{u}_2),$$

and for all $\mathbf{u}_1 \in A_1$, there exists a unique $c_2(\mathbf{u}_1) \in A_2$ such that

$$f(\mathbf{u}_1, c_2(\mathbf{u}_1)) = \min_{\mathbf{u}_2' \in A_2} f(\mathbf{u}_1, \mathbf{u}_2').$$

Cast the computation of $R(D_s) - sD_s$ into this optimization problem:

1. Let

$$f(\mathbf{Q}, \mathbf{t}) = \sum_x \sum_{\hat{x}} p(x) Q(\hat{x}|x) \log \frac{Q(\hat{x}|x)}{t(\hat{x})} - s \sum_x \sum_{\hat{x}} p(x) Q(\hat{x}|x) d(x, \hat{x}),$$

where $\mathbf{u}_1 \leftarrow \mathbf{Q}$ and $\mathbf{u}_2 \leftarrow \mathbf{t}$.

2. Let

$$\begin{aligned} A_1 &= \Big\{ (Q(\hat{x}|x), (x, \hat{x}) \in \mathcal{X} \times \hat{\mathcal{X}}) : Q(\hat{x}|x) > 0, \\ & \qquad \sum_{\hat{x}} Q(\hat{x}|x) = 1 \text{ for all } x \in \mathcal{X} \Big\} \subset \Re^{|\mathcal{X}||\hat{\mathcal{X}}|} \end{aligned}$$

and

$$A_2 = \{ (t(\hat{x}), \hat{x} \in \hat{\mathcal{X}}) : t(\hat{x}) > 0 \}.$$

3. The double infimum now becomes

$$\begin{aligned} \inf_{\mathbf{Q} \in A_1} \inf_{\mathbf{t} \in A_2} f(\mathbf{Q}, \mathbf{t}) &= \inf_{\mathbf{Q} \in A_1} \overbrace{\inf_{\mathbf{t} \in A_2} \Bigg[\sum_x \sum_{\hat{x}} p(x) Q(\hat{x}|x) \log \frac{Q(\hat{x}|x)}{t(\hat{x})}}^{I(\mathbf{p}, \mathbf{Q})} \\ & \qquad - s \sum_x \sum_{\hat{x}} p(x) Q(\hat{x}|x) d(x, \hat{x}) \Bigg], \end{aligned}$$

where the infimum over all $\mathbf{t} \in A_2$ is in fact a minimum, and

The BA Algorithm for Computing $R(D_s)$ - sD_s

Recall the double infimum in Section 9.1:

$$\inf_{\mathbf{u}_1 \in A_1} \inf_{\mathbf{u}_2 \in A_2} f(\mathbf{u}_1, \mathbf{u}_2).$$

where

✓ 1. A_i is a convex subset of $\Re^{n_i}$ for $i = 1, 2$.

✓ 2. $f : A_1 \times A_2 \to \Re$ is bounded from below and is such that

✓ • f is continuous and has continuous partial derivatives on $A_1 \times A_2$

• For all $\mathbf{u}_2 \in A_2$, there exists a unique $c_1(\mathbf{u}_2) \in A_1$ such that

$$f(c_1(\mathbf{u}_2), \mathbf{u}_2) = \min_{\mathbf{u}_1' \in A_1} f(\mathbf{u}_1', \mathbf{u}_2),$$

and for all $\mathbf{u}_1 \in A_1$, there exists a unique $c_2(\mathbf{u}_1) \in A_2$ such that

$$f(\mathbf{u}_1, c_2(\mathbf{u}_1)) = \min_{\mathbf{u}_2' \in A_2} f(\mathbf{u}_1, \mathbf{u}_2').$$

Cast the computation of $R(D_s) - sD_s$ into this optimization problem:

1. Let

$$f(\mathbf{Q}, \mathbf{t}) = \sum_x \sum_{\hat{x}} p(x) Q(\hat{x}|x) \log \frac{Q(\hat{x}|x)}{t(\hat{x})} - s \sum_x \sum_{\hat{x}} p(x) Q(\hat{x}|x) d(x, \hat{x}),$$

where $\mathbf{u}_1 \leftarrow \mathbf{Q}$ and $\mathbf{u}_2 \leftarrow \mathbf{t}$.

2. Let

$$A_1 = \Big\{ (Q(\hat{x}|x), (x, \hat{x}) \in \mathcal{X} \times \hat{\mathcal{X}}) : Q(\hat{x}|x) > 0, \sum_{\hat{x}} Q(\hat{x}|x) = 1 \text{ for all } x \in \mathcal{X} \Big\} \subset \Re^{|\mathcal{X}||\hat{\mathcal{X}}|}$$

and

$$A_2 = \{ (t(\hat{x}), \hat{x} \in \hat{\mathcal{X}}) : t(\hat{x}) > 0 \}.$$

3. The double infimum now becomes

$$\inf_{\mathbf{Q} \in A_1} \inf_{\mathbf{t} \in A_2} f(\mathbf{Q}, \mathbf{t}) = \inf_{\mathbf{Q} \in A_1} \inf_{\mathbf{t} \in A_2} \Bigg[\sum_x \sum_{\hat{x}} p(x) Q(\hat{x}|x) \log \frac{Q(\hat{x}|x)}{t(\hat{x})} - s \sum_x \sum_{\hat{x}} p(x) Q(\hat{x}|x) d(x, \hat{x}) \Bigg],$$

where the infimum over all $\mathbf{t} \in A_2$ is in fact a minimum, and

$$f^* = \inf_{\mathbf{Q} \in A_1} \inf_{\mathbf{t} \in A_2} f(\mathbf{Q}, \mathbf{t}) = R(D_s) - sD_s.$$

Algorithm Details

$$f^* = \inf_{\mathbf{Q} \in A_1} \inf_{\mathbf{t} \in A_2} f(\mathbf{Q}, \mathbf{t})$$

$$f(\mathbf{Q}, \mathbf{t}) = \sum_x \sum_{\hat{x}} p(x) Q(\hat{x}|x) \log \frac{Q(\hat{x}|x)}{t(\hat{x})} - s \sum_x \sum_{\hat{x}} p(x) Q(\hat{x}|x) d(x, \hat{x}),$$

Algorithm Details

$$f^* = \inf_{\mathbf{Q} \in A_1} \inf_{\mathbf{t} \in A_2} f(\mathbf{Q}, \mathbf{t})$$

$$f(\mathbf{Q}, \mathbf{t}) = \sum_x \sum_{\hat{x}} p(x) Q(\hat{x}|x) \log \frac{Q(\hat{x}|x)}{t(\hat{x})} - s \sum_x \sum_{\hat{x}} p(x) Q(\hat{x}|x) d(x, \hat{x}),$$

1. By Lemma 9.3, for any given $\mathbf{Q} \in A_1$, the unique $\mathbf{t} \in A_2$ that minimizes f is given by

Algorithm Details

$$\boxed{f^* = \inf_{\mathbf{Q} \in A_1} \inf_{\mathbf{t} \in A_2} f(\mathbf{Q}, \mathbf{t})}$$

$$\boxed{f(\mathbf{Q}, \mathbf{t}) = \sum_x \sum_{\hat{x}} p(x) Q(\hat{x}|x) \log \frac{Q(\hat{x}|x)}{t(\hat{x})} - s \sum_x \sum_{\hat{x}} p(x) Q(\hat{x}|x) d(x, \hat{x}),}$$

1. By Lemma 9.3, for any given $\mathbf{Q} \in A_1$, the unique $\mathbf{t} \in A_2$ that minimizes f is given by

$$t^*(\hat{x}) = \sum_x p(x) Q(\hat{x}|x). \tag{1}$$

Algorithm Details

$$f^* = \inf_{\mathbf{Q} \in A_1} \inf_{\mathbf{t} \in A_2} f(\mathbf{Q}, \mathbf{t})$$

$$f(\mathbf{Q}, \mathbf{t}) = \sum_x \sum_{\hat{x}} p(x) Q(\hat{x}|x) \log \frac{Q(\hat{x}|x)}{t(\hat{x})} - s \sum_x \sum_{\hat{x}} p(x) Q(\hat{x}|x) d(x, \hat{x}),$$

1. By Lemma 9.3, for any given $\mathbf{Q} \in A_1$, the unique $\mathbf{t} \in A_2$ that minimizes f is given by

$$t^*(\hat{x}) = \sum_x p(x) Q(\hat{x}|x). \tag{1}$$

2. By Lagrange multipliers, it can be shown that for a given $\mathbf{t} \in A_2$, the transition matrix $\mathbf{Q}$ that minimizes f is given by

Algorithm Details

$$f^* = \inf_{\mathbf{Q} \in A_1} \inf_{\mathbf{t} \in A_2} f(\mathbf{Q}, \mathbf{t})$$

$$f(\mathbf{Q}, \mathbf{t}) = \sum_x \sum_{\hat{x}} p(x) Q(\hat{x}|x) \log \frac{Q(\hat{x}|x)}{t(\hat{x})} - s \sum_x \sum_{\hat{x}} p(x) Q(\hat{x}|x) d(x, \hat{x}),$$

1. By Lemma 9.3, for any given $\mathbf{Q} \in A_1$, the unique $\mathbf{t} \in A_2$ that minimizes f is given by

$$t^*(\hat{x}) = \sum_x p(x) Q(\hat{x}|x). \tag{1}$$

2. By Lagrange multipliers, it can be shown that for a given $\mathbf{t} \in A_2$, the transition matrix $\mathbf{Q}$ that minimizes f is given by

$$Q(\hat{x}|x) = \frac{t(\hat{x}) e^{s d(x, \hat{x})}}{\sum_{\hat{x}'} t(\hat{x}') e^{s d(x, \hat{x}')}} > 0, \tag{2}$$

Algorithm Details

$$f^* = \inf_{\mathbf{Q} \in A_1} \inf_{\mathbf{t} \in A_2} f(\mathbf{Q}, \mathbf{t})$$

$$f(\mathbf{Q}, \mathbf{t}) = \sum_x \sum_{\hat{x}} p(x) Q(\hat{x}|x) \log \frac{Q(\hat{x}|x)}{t(\hat{x})} - s \sum_x \sum_{\hat{x}} p(x) Q(\hat{x}|x) d(x, \hat{x}),$$

1. By Lemma 9.3, for any given $\mathbf{Q} \in A_1$, the unique $\mathbf{t} \in A_2$ that minimizes f is given by

$$t^*(\hat{x}) = \sum_x p(x) Q(\hat{x}|x). \tag{1}$$

2. By Lagrange multipliers, it can be shown that for a given $\mathbf{t} \in A_2$, the transition matrix $\mathbf{Q}$ that minimizes f is given by

$$Q(\hat{x}|x) = \frac{t(\hat{x}) e^{s d(x, \hat{x})}}{\sum_{\hat{x}'} t(\hat{x}') e^{s d(x, \hat{x}')}} > 0, \tag{2}$$

and so $\mathbf{Q} \in A_1$.

Algorithm Details

$$f^* = \inf_{\mathbf{Q} \in A_1} \inf_{\mathbf{t} \in A_2} f(\mathbf{Q}, \mathbf{t})$$

$$f(\mathbf{Q}, \mathbf{t}) = \sum_x \sum_{\hat{x}} p(x) Q(\hat{x}|x) \log \frac{Q(\hat{x}|x)}{t(\hat{x})} - s \sum_x \sum_{\hat{x}} p(x) Q(\hat{x}|x) d(x, \hat{x}),$$

1. By Lemma 9.3, for any given $\mathbf{Q} \in A_1$, the unique $\mathbf{t} \in A_2$ that minimizes f is given by

$$t^*(\hat{x}) = \sum_x p(x) Q(\hat{x}|x). \tag{1}$$

2. By Lagrange multipliers, it can be shown that for a given $\mathbf{t} \in A_2$, the transition matrix $\mathbf{Q}$ that minimizes f is given by

$$Q(\hat{x}|x) = \frac{t(\hat{x}) e^{s d(x, \hat{x})}}{\sum_{\hat{x}'} t(\hat{x}') e^{s d(x, \hat{x}')}} > 0, \tag{2}$$

and so $\mathbf{Q} \in A_1$.

Algorithm Details

$$f^* = \inf_{\mathbf{Q} \in A_1} \inf_{\mathbf{t} \in A_2} f(\mathbf{Q}, \mathbf{t})$$

$$f(\mathbf{Q}, \mathbf{t}) = \sum_x \sum_{\hat{x}} p(x) Q(\hat{x}|x) \log \frac{Q(\hat{x}|x)}{t(\hat{x})} - s \sum_x \sum_{\hat{x}} p(x) Q(\hat{x}|x) d(x, \hat{x}),$$

1. By Lemma 9.3, for any given $\mathbf{Q} \in A_1$, the unique $\mathbf{t} \in A_2$ that minimizes f is given by

$$t^*(\hat{x}) = \sum_x p(x) Q(\hat{x}|x). \tag{1}$$

2. By Lagrange multipliers, it can be shown that for a given $\mathbf{t} \in A_2$, the transition matrix $\mathbf{Q}$ that minimizes f is given by

$$Q(\hat{x}|x) = \frac{t(\hat{x}) e^{s d(x,\hat{x})}}{\sum_{\hat{x}'} t(\hat{x}') e^{s d(x,\hat{x}')}} > 0, \tag{2}$$

and so $\mathbf{Q} \in A_1$.

3. Let $\mathbf{Q}^{(0)}$ be an arbitrarily chosen strictly positive transition matrix in A_1. Then $\mathbf{t}^{(0)} \in A_2$ can be determined accordingly. This forms $(\mathbf{Q}^{(0)}, \mathbf{t}^{(0)})$.

Algorithm Details

$$f^* = \inf_{\mathbf{Q}\in A_1} \inf_{\mathbf{t}\in A_2} f(\mathbf{Q}, \mathbf{t})$$

$$f(\mathbf{Q}, \mathbf{t}) = \sum_x \sum_{\hat{x}} p(x)Q(\hat{x}|x) \log \frac{Q(\hat{x}|x)}{t(\hat{x})} - s \sum_x \sum_{\hat{x}} p(x)Q(\hat{x}|x)d(x, \hat{x}),$$

1. By Lemma 9.3, for any given $\mathbf{Q} \in A_1$, the unique $\mathbf{t} \in A_2$ that minimizes f is given by

$$t^*(\hat{x}) = \sum_x p(x)Q(\hat{x}|x). \qquad (1)$$

2. By Lagrange multipliers, it can be shown that for a given $\mathbf{t} \in A_2$, the transition matrix $\mathbf{Q}$ that minimizes f is given by

$$Q(\hat{x}|x) = \frac{t(\hat{x})e^{sd(x,\hat{x})}}{\sum_{\hat{x}'} t(\hat{x}')e^{sd(x,\hat{x}')}} > 0, \qquad (2)$$

and so $\mathbf{Q} \in A_1$.

3. Let $\mathbf{Q}^{(0)}$ be an arbitrarily chosen strictly positive transition matrix in A_1. Then $\mathbf{t}^{(0)} \in A_2$ can be determined accordingly. This forms $(\mathbf{Q}^{(0)}, \mathbf{t}^{(0)})$.

4. Compute $\mathbf{Q}^{(1)}, \mathbf{t}^{(1)}, \mathbf{Q}^{(2)}, \mathbf{t}^{(2)}, \cdots$ iteratively by applying (2) and (1) alternately.

Algorithm Details

$$f^* = \inf_{\mathbf{Q}\in A_1} \inf_{\mathbf{t}\in A_2} f(\mathbf{Q},\mathbf{t})$$

$$f(\mathbf{Q},\mathbf{t}) = \sum_x \sum_{\hat{x}} p(x)Q(\hat{x}|x)\log\frac{Q(\hat{x}|x)}{t(\hat{x})} - s\sum_x\sum_{\hat{x}} p(x)Q(\hat{x}|x)d(x,\hat{x}),$$

1. By Lemma 9.3, for any given $\mathbf{Q} \in A_1$, the unique $\mathbf{t} \in A_2$ that minimizes f is given by

$$t^*(\hat{x}) = \sum_x p(x)Q(\hat{x}|x). \tag{1}$$

2. By Lagrange multipliers, it can be shown that for a given $\mathbf{t} \in A_2$, the transition matrix $\mathbf{Q}$ that minimizes f is given by

$$Q(\hat{x}|x) = \frac{t(\hat{x})e^{sd(x,\hat{x})}}{\sum_{\hat{x}'} t(\hat{x}')e^{sd(x,\hat{x}')}} > 0, \tag{2}$$

and so $\mathbf{Q} \in A_1$.

3. Let $\mathbf{Q}^{(0)}$ be an arbitrarily chosen strictly positive transition matrix in A_1. Then $\mathbf{t}^{(0)} \in A_2$ can be determined accordingly. This forms $(\mathbf{Q}^{(0)}, \mathbf{t}^{(0)})$.

4. Compute $\mathbf{Q}^{(1)}$, $\mathbf{t}^{(1)}$, $\mathbf{Q}^{(2)}$, $\mathbf{t}^{(2)}$, $\cdots$ iteratively by applying (2) and (1) alternately.

5. It will be shown in Section 9.3 that $f^{(k)} = f(\mathbf{Q}^{(k)}, \mathbf{t}^{(k)}) \to f^* = R(D_s) - sD_s$.