Corollary 12.6. $\tilde{R}$ is continuous on $(0, \infty)$.

**Proof.** It is well known [2, p. 5] that convexity on an interval implies continuity at all interior points of the interval. \[ \square \]

A convex function on an interval need not be continuous at its endpoints. To see this, consider the function $f(t) := 0$ for $t > 0$ and $f(0) := 1$. Fortunately, the function $\tilde{R}$ is continuous at the origin.

Lemma 12.7. $\tilde{R}$ is continuous at $D = 0$.

**Proof.** (Based on [1, p. 125].) It suffices to show that for every nonnegative sequence $D_n \to 0$, we have $\tilde{R}(D_n) \to \tilde{R}(0)$. To show that $\tilde{R}(D_n) \to \tilde{R}(0)$, we show that given any subsequence $\tilde{R}(D_{nk})$, there is a further subsequence such that $\tilde{R}(D_{nk}) \to \tilde{R}(0)$. So suppose $\tilde{R}(D_{nk})$ is any subsequence. Note that this is a bounded sequence of real numbers since $0 \leq I(p \times W) \leq H(p) \leq \log |X|$. Hence, there is a converging sub-subsequence, $\tilde{R}(D_{nk})$. Since the infimum defining $\tilde{R}(D)$ is always achieved, we have

$$\tilde{R}(D_{nk}) = I(p \times W)$$

for some $W$ satisfying

$$\sum_x p(x) \sum_y W(y|x)d(x, y) \leq D_{nk}.$$  \[ (12.1) \]

Furthermore, since the $W$ lie in a sequentially compact set, there is a further subsequence $W_{k_i} \to W_\infty$ for some transition probability $W_\infty$ as $i \to \infty$. By continuity of $I(p \times W)$ in $W$,

$$\lim_{\ell \to \infty} \tilde{R}(D_{nk\ell}) = \lim_{i \to \infty} \tilde{R}(D_{nk\ell}) = \lim_{i \to \infty} I(p \times W_{k\ell}) = I(p \times W_\infty).$$

Also, from (12.1) with $\ell$ replaced by $k\ell$, letting $i \to \infty$ yields

$$\sum_x p(x) \sum_y W_\infty(y|x)d(x, y) \leq 0.$$
Hence,
\[ \tilde{R}(0) := \inf_{W: \mathbb{E}[d(X,Y)] \leq 0} I(p \times W) \leq I(p \times W_{\ell_1}) = \lim_{i \to \infty} I(p \times W_{\ell_i}) = \lim_{i \to \infty} \tilde{R}(D_{nk_{\ell_i}}). \]

Since $\tilde{R}$ is nonincreasing, $\tilde{R}(0) \geq \tilde{R}(D_{nk_{\ell_i}})$. Letting $i \to \infty$ yields the desired result. \(\square\)

### 12.5. The Converse Theorem

**Theorem 12.8.** If $(R,D)$ is achievable for a discrete memoryless source, then $R \geq \tilde{R}(D)$.

**Proof.** It suffices to restrict attention to $D_{\text{min}} \leq D < D_{\text{max}}$, where $\tilde{R}$ is finite, continuous, and convex there. Let $0 < \Delta D < D_{\text{max}} - D$ and $\Delta R > 0$ be given. Suppose $Y = q(X)$, where $q: X^n \to Y^n$ takes $N$ distinct values,

\[ n(R + \Delta R) \geq \log N \geq H(Y) \quad \text{and} \quad \mathbb{E}[d_n(X,q(X))] < D + \Delta D. \]

Then

\[ n(R + \Delta R) > H(Y) \]

\[ \geq H(Y) - H(Y|X), \quad \text{since} \quad H(Y|X) \geq 0, \]

\[ = I(X \land Y) \]

\[ = H(X) - H(X|Y) \]

\[ = \sum_{k=1}^{n} H(X_k) - H(X|Y) \]

\[ = \sum_{k=1}^{n} H(X_k) - H(X_k|Y,X_1,\ldots,X_{k-1}), \quad \text{chain rule for entropy}, \]

\[ \geq \sum_{k=1}^{n} H(X_k) - H(X_k|Y_k), \quad \text{reducing conditioning increases entropy}, \]

\[ = \sum_{k=1}^{n} I(X_k \land Y_k) \quad \ldots \quad \text{to be continued} \]

**References**
