

On the Continuity of Achievable Rate Regions for Source Coding over Networks

WeiHsin Gu and Michelle Effros
 Department of Electrical Engineering
 California Institute of Technology
 Pasadena, California, CA 91125, USA
 Email: {wgu, effros}@caltech.edu

Abstract—The continuity property of achievable rate regions for source coding over networks is considered. We show rate-distortion regions are continuous with respect to distortion vectors. Then we focus on the continuity of lossless rate regions with respect to source distribution: First, the proof of continuity for general networks with independent sources is given; then, for the case of dependent sources, continuity is proven both in examples where one-letter characterizations are known and in examples where one-letter characterizations are not known; the proofs in the latter case rely on the concavity of the rate regions for those networks.

I. INTRODUCTION

In the point-to-point network, the entropy and rate-distortion functions are solutions to minimizing the required source coding rates. In the field of network source coding, researchers are trying to extend Shannon's result to more general networks. While complete, one-letter characterizations of the rate-distortion regions or lossless rate regions for some networks are known (for example, see [1], [2], [3], and [4]), the lossless rate regions and rate-distortion regions for source coding of i.i.d. random variables in more general networks remain entirely unsolved.

Throughout this paper, we assume that all the source sequences are drawn i.i.d. according to a known probability mass function of source vector $\mathbf{X} = (X_1, \dots, X_s)$, where each X_i has finite alphabet \mathcal{X}_i . We consider the continuity property of rate-distortion regions and lossless rate regions for more general networks whose lossless and rate-distortion regions are not yet known. There are two continuity properties: continuity of rate-distortion regions with respect to distortion vector and continuity of lossless regions with respect to the source distribution. A motivation behind this is when a network source coding problem is given and the source distribution is unknown, a natural way to build a good performance code is to estimate the source distribution first and then design a good performance code according to this estimate distribution. We don't want the error causes a big gap between the true regions and the estimated ones.

In Section II, we formally define network source coding problem and two objects of interest - "lossless rate region"

⁰This material is based upon work partially supported by NSF Grant No. CCR-0325324 and Caltech's Lee Center for Advanced Networking.

and "rate-distortion regions". The proof of continuity of rate-distortion regions with respect to the distortion vector is given in Section III. Some types of network source coding problems whose lossless rate regions are continuous with respect to the source distribution are discussed in Section IV

II. FORMULATION

Here we define a general network source coding problem, and its lossless rate region and rate-distortion region. A multipath directed network is an ordered pair $(\mathcal{V}, \mathcal{E})$ with vertex set \mathcal{V} and indexed edge set \mathcal{E} . Since multiple edges between two vertices are allowed, we assign a positive integer $M_{v,v'}$ for any distinct vertices v and v' connected by an edge, to indicate the number of edges from v to v' . Hence an edge set \mathcal{E} is a subset of $\mathcal{V} \times \mathcal{V} \times \mathbb{Z}$ consisting of all triples (v, v', i) satisfying $1 \leq i \leq M_{v,v'}$. For each edge $e = (v, v', i) \in \mathcal{E}$, we call v the tail of e and v' the head of e , denoted by $\text{tail}(e)$ and $\text{head}(e)$ respectively. The set of edges that end at vertex v is denoted by $\Gamma_I(v)$ and the set of edges that begin at v is denoted by $\Gamma_O(v)$, i.e.,

$$\begin{aligned}\Gamma_I(v) &:= \{e \in \mathcal{E} : \text{head}(e) = v\} \\ \Gamma_O(v) &:= \{e \in \mathcal{E} : \text{tail}(e) = v\}.\end{aligned}$$

Let $G = (\mathcal{V}, \mathcal{E})$ be a multipath directed network. A network source coding problem is defined to be a 4-tuple $\mathcal{N} = (G, s, \mathcal{S}, \mathcal{T})$. Here integer s is the number of source random variables, and sets $\mathcal{S} \subseteq \mathcal{V} \times \{1, \dots, s\}$ and $\mathcal{T} \subseteq \mathcal{V} \times \{1, \dots, s\}$ describe the source availabilities and demands, respectively. That is, source $i \in \{1, \dots, s\}$ is available at node $v \in \mathcal{V}$ if and only if $(v, i) \in \mathcal{S}$ and node $v' \in \mathcal{V}$ has to decode source $j \in \{1, \dots, s\}$ if and only if $(v', j) \in \mathcal{T}$. For each $v \in \mathcal{V}$ sets $I(v), J(v) \subseteq \{1, 2, \dots, s\}$ summarize v 's source availabilities and demands, respectively, giving $I(v) = \{i : (v, i) \in \mathcal{S}\}$ and $J(v) = \{j : (v, j) \in \mathcal{T}\}$. Let k denote the total number of reproduction demands, i.e., $k = |J(v)|$. We suppose that for each demand pair $(v', j) \in \mathcal{T}$, there exist a pair $(v, i) \in \mathcal{S}$ such that there is a path starting from v to v' .

Let $\mathbf{X} = (X_1, \dots, X_s)$ be an s -dimensional random vector representing the source vector of the network source coding problem $\mathcal{N} = (G, s, \mathcal{S}, \mathcal{T})$ and let $P_{\mathbf{X}}$ denote its probability mass function. Assume that the source coding problem described by \mathcal{N} is supportable, meaning that for any demand

$(v', j) \in \mathcal{T}$ there exists a node $v \in \mathcal{V}$ for which $(v, j) \in \mathcal{S}$ and there exists a path from v to v' in G . Let \mathcal{X}_i denote the alphabet of X_i for each $i \in \{1, \dots, s\}$. We define block codes for \mathcal{N} assuming that G is acyclic. (If the network is cyclic, we can define block codes chronologically, using the approach from [5].)

A length- n block code \mathcal{C} for $(\mathcal{N}, P_{\mathbf{X}})$ with $|\mathcal{E}|$ -dimensional rate vector $\mathbf{R} = (R(e))_{e \in \mathcal{E}}$ contains a set of encoding functions

$$\mathcal{F} := \{f_e \mid e \in \mathcal{E}\}$$

and a set of decoding functions

$$\mathcal{G} := \{g_{v,j} \mid v \in \mathcal{V}, j \in J(v)\}.$$

The encoding and decoding functions are defined as follows

- (i) For each $v \in \mathcal{V}$ with $\Gamma_O(v) \neq \emptyset$, every $e \in \Gamma_O(v)$ has an encoding function

$$f_e : \prod_{e' \in \Gamma_I(\text{tail}(e))} \{1, 2, \dots, 2^{nR(e')}\} \times \prod_{i \in I(\text{tail}(e))} \mathcal{X}_i^n \rightarrow \{1, 2, \dots, 2^{nR(e)}\}.$$

- (ii) For each $v \in \mathcal{V}$ with $J(v) \neq \emptyset$, every $j \in J(v)$ has a decoding function

$$g_{v,j} : \prod_{e \in \Gamma_I(v)} \{1, 2, \dots, 2^{nR(e)}\} \times \prod_{i \in I(v)} \mathcal{X}_i^n \rightarrow \hat{\mathcal{X}}_j^n,$$

where $\hat{\mathcal{X}}_j$ is the reproduction alphabet for X_j .

A rate vector \mathbf{R} is losslessly-achievable if and only if there exists a sequence of rate- \mathbf{R} length- n block codes $\{\mathcal{C}_n\}_{n=1}^\infty$ such that the symbol error probabilities are asymptotically small. More precisely, for any $v \in \mathcal{V}$ with $J(v) \neq \emptyset$ and any $j \in J(v)$, if $\hat{X}_j^n(v)$ denotes the reproduction of source X_j^n at node v , then

$$\lim_{n \rightarrow \infty} \Pr\{X_j^n \neq \hat{X}_j^n(v)\} = 0.$$

The closure of the set consisting of all losslessly-achievable rate vectors is called the lossless rate region for $(\mathcal{N}, P_{\mathbf{X}})$, denoted by $\mathcal{R}_L(\mathcal{N}, P_{\mathbf{X}})$. Throughout this paper, we fix \mathcal{N} so we simply denote by $\mathcal{R}_L(P_{\mathbf{X}})$ the lossless rate region $\mathcal{R}_L(\mathcal{N}, P_{\mathbf{X}})$.

Fix a set of distortion measures $\{d_i\}_{i=1}^s$, where for each i , d_i is a distortion measure $d_i : \mathcal{X}_i \times \hat{\mathcal{X}}_i \rightarrow [0, \infty]$, and, with a slight abuse of notation, let $d_i(x^n, \hat{x}^n) = \sum_{j=1}^n d(x_j, \hat{x}_j)$. Let $\mathbf{D} = (D_{v,j})_{v \in \mathcal{V}, j \in J(v)}$ be a k -dimensional vector whose components are non-negative real numbers. An $|\mathcal{E}|$ -dimensional non-negative real vector \mathbf{R} is said to be \mathbf{D} -achievable if and only if there exists a sequence of rate- \mathbf{R} length- n block codes $\{\mathcal{C}_n\}_{n=1}^\infty$ such that the distortion requirement is asymptotically satisfied, i.e., for any $v \in \mathcal{V}$ with $J(v) \neq \emptyset$ and any $j \in J(v)$, if $\hat{X}_j^n(v)$ denotes the reproduction of source X_j^n at node v , then

$$\limsup_{n \rightarrow \infty} \frac{1}{n} E d_j(X_j^n, \hat{X}_j^n(v)) \leq D_{v,j}.$$

The closure of the set of \mathbf{D} -achievable rate vectors \mathbf{R} is called the rate-distortion region for $(\mathcal{N}, P_{\mathbf{X}})$, denoted by

$\mathcal{R}_d(\mathcal{N}, P_{\mathbf{X}}, \mathbf{D})$. When the network and source are clear from the context, we adopt the notations $\mathcal{R}(P_{\mathbf{X}}, \mathbf{D})$ or $\mathcal{R}(\mathbf{D})$ for simplicity.

Let \mathcal{M}_s denote the space of all joint probability distributions for s -dimensional source vector \mathbf{X} . For every set A , we use 2^A to denote the power set of A , i.e., the set of all subsets of A . Let \mathbb{R} denote the set of nonnegative real numbers. Since for all $P_{\mathbf{X}} \in \mathcal{M}_s$, $\mathcal{R}(P_{\mathbf{X}}, \mathbf{D})$ is a subset of $\mathbb{R}_+^{|\mathcal{E}|}$, $\mathcal{R}(\cdot, \cdot)$ can be considered as a function from $\mathcal{M}_s \times \mathbb{R}_+^{|\mathcal{E}|}$ to $2^{\mathbb{R}_+^{|\mathcal{E}|}}$. In this paper, we discuss the continuity of rate-distortion regions with respect to \mathbf{D} , and of lossless rate region with respect to $P_{\mathbf{X}}$. Before explaining the term of ‘‘continuity’’, we introduce ϵ -close for two subsets of $\mathbb{R}_+^{|\mathcal{E}|}$ as follows.

Definition 1: Let A and B be two subsets of the n -dimensional Euclidean space \mathbb{R}^n . We use $\|\mathbf{x}\|$ to denote the L_2 -norm of $\mathbf{x} \in \mathbb{R}^n$. A and B are said to be ϵ -close ($\epsilon > 0$) if and only if

- (a) For every $a \in A$, there exists $b_{a,\epsilon} \in B$ such that $\|a - b_{a,\epsilon}\| \leq \sqrt{n}\epsilon$.
(b) For every $b \in B$, there exists $a_{b,\epsilon} \in A$ such that $\|b - a_{b,\epsilon}\| \leq \sqrt{n}\epsilon$.

We consider the finite-alphabet source case. Let m be the alphabet size of each source random variable. We use $\mathcal{M}_s(m)$ to denote the set of all probability distributions on $\{1, \dots, m\}^s$. The reason why we fix the alphabet size is that, the continuity property apparently doesn't hold without the constraint on the alphabet size. A counterexample is that $H(X)$, the entropy function of random variable X , is not continuous when X has countably infinite alphabets. Please see Example 1.

III. CONTINUITY OF $\mathcal{R}(P_{\mathbf{X}}, \mathbf{D})$ WITH RESPECT TO \mathbf{D}

In this section, we discuss the continuity of $\mathcal{R}(P_{\mathbf{X}}, \mathbf{D})$ with respect to \mathbf{D} . It means that for a fixed source distribution $P_{\mathbf{X}} \in \mathcal{M}_s(m)$, $\mathcal{R}(P_{\mathbf{X}}, \mathbf{D}_1)$ and $\mathcal{R}(P_{\mathbf{X}}, \mathbf{D}_2)$ are ϵ -close whenever \mathbf{D}_1 and \mathbf{D}_2 are close enough. The next theorem shows this property. The lemmas applied in the proof of Theorem 1 are stated and proved in Appendix.

Theorem 1: For any $\epsilon > 0$, there exists $\delta > 0$ such that for any $P_{\mathbf{X}} \in \mathcal{M}_s(m)$, and any $\mathbf{D}, \mathbf{D}' \in \mathbb{R}_+^k$, if $\|\mathbf{D} - \mathbf{D}'\| < \delta$, then $\mathcal{R}(P_{\mathbf{X}}, \mathbf{D})$ and $\mathcal{R}(P_{\mathbf{X}}, \mathbf{D}')$ are ϵ -close.

Before proving the theorem, we introduce some notation. Let $\mathbf{a} = (a_1, \dots, a_n)$ and $\mathbf{b} = (b_1, \dots, b_n)$ be n -dimensional real vectors, we say that \mathbf{a} is greater than or equal to \mathbf{b} , denoted by $\mathbf{a} \geq \mathbf{b}$, if and only if $a_i \geq b_i$ for all $i \in \{1, \dots, n\}$. For any vector \mathbf{a} and any $t > 0$, $\mathbf{a}_t = (a_{t,1}, \dots, a_{t,n})$ denotes the vector such that for each $i \in \{1, \dots, n\}$,

$$a_{t,i} := \begin{cases} t & , \text{ if } a_i \leq t \\ a_i & , \text{ otherwise.} \end{cases}$$

Proof. (of Theorem 1) We want to claim that for any $\epsilon > 0$, there exists $\delta > 0$ such that for any $P_{\mathbf{X}} \in \mathcal{M}_s(m)$, $\mathcal{R}(P_{\mathbf{X}}, \mathbf{D})$ and $\mathcal{R}(P_{\mathbf{X}}, \mathbf{D} + \delta \cdot \mathbf{1})$ are $\epsilon/2$ -close for any $\mathbf{D} \in \mathbb{R}_+^k$. Assuming this claim, for any $\mathbf{D} = (D_i)_{i=1}^k$ and $\mathbf{D}' = (D'_i)_{i=1}^k$ in \mathbb{R}_+^k such that $\|\mathbf{D} - \mathbf{D}'\| \leq \delta$, let

$$\mathbf{D}_0 := \min(\mathbf{D}, \mathbf{D}') = (\min(D_i, D'_i))_{i=1}^k.$$

It is then easy to see that $\|\mathbf{D} - \mathbf{D}_0\| \leq \delta$ and $\|\mathbf{D}' - \mathbf{D}_0\| \leq \delta$. Then we have

$$\begin{cases} \mathbf{D}_0 \leq \mathbf{D} \leq \mathbf{D}_0 + \delta \cdot \mathbf{1} \\ \mathbf{D}_0 \leq \mathbf{D}' \leq \mathbf{D}_0 + \delta \cdot \mathbf{1} \end{cases}.$$

By claim, $\mathcal{R}(P_{\mathbf{X}}, \mathbf{D}_0)$ and $\mathcal{R}(P_{\mathbf{X}}, \mathbf{D}_0 + \delta \cdot \mathbf{1})$ are $\epsilon/2$ -close. Since

$$\begin{cases} \mathcal{R}(P_{\mathbf{X}}, \mathbf{D}_0) \subseteq \mathcal{R}(P_{\mathbf{X}}, \mathbf{D}) \subseteq \mathcal{R}(P_{\mathbf{X}}, \mathbf{D} + \delta \cdot \mathbf{1}) \\ \mathcal{R}(P_{\mathbf{X}}, \mathbf{D}_0) \subseteq \mathcal{R}(P_{\mathbf{X}}, \mathbf{D}') \subseteq \mathcal{R}(P_{\mathbf{X}}, \mathbf{D}' + \delta \cdot \mathbf{1}) \end{cases},$$

this implies that both $\mathcal{R}(P_{\mathbf{X}}, \mathbf{D})$ and $\mathcal{R}(P_{\mathbf{X}}, \mathbf{D}')$ are $\epsilon/2$ -close to $\mathcal{R}(P_{\mathbf{X}}, \mathbf{D}_0)$. So $\mathcal{R}(P_{\mathbf{X}}, \mathbf{D})$ and $\mathcal{R}(P_{\mathbf{X}}, \mathbf{D}')$ are ϵ -close. Therefore, it suffices to prove the claim. Fix $\epsilon > 0$. We prove the claim in the following three steps :

- (i) (Lemma 2) There exists $\delta_1 > 0$ such that $\mathcal{R}(P_{\mathbf{X}}, \mathbf{D})$ and $\mathcal{R}(P_{\mathbf{X}}, \mathbf{D}_{\delta_1})$ are $\epsilon/4$ -close for all $\mathbf{D} \in \mathbb{R}_+^k$.
- (ii) We claim that there exists δ_2 with $0 < \delta_2 < \delta_1$ such that for any $\mathbf{D} \in \mathbb{R}_+^k$ with $D_i \geq \delta_1$ for all i , $\mathcal{R}(P_{\mathbf{X}}, \mathbf{D})$ and $\mathcal{R}(P_{\mathbf{X}}, \mathbf{D} + \delta_2 \cdot \mathbf{1})$ are $\epsilon/4$ -close:

Fix $\mathbf{R}_0 \in \bigcap_{P_{\mathbf{X}} \in \mathcal{M}_s(m)} \mathcal{R}(P_{\mathbf{X}}, \mathbf{0})$. For any $t > 0$ with $t < \delta_1$, let $\lambda(t) = \frac{t}{\delta_1}$. Then any $\mathbf{D} \in \mathbb{R}_+^k$ with $D_i \geq \delta_1$ for all i satisfies

$$(1 - \lambda(t))(\mathbf{D} + t \cdot \mathbf{1}) + \lambda(t)(t \cdot \mathbf{1}) \leq \mathbf{D}$$

since for any $i \in \{1, \dots, k\}$,

$$\begin{aligned} (1 - \lambda(t))(D_i + t) + \lambda(t)t - D_i \\ = t - \lambda(t)D_i = t(1 - \frac{D_i}{\delta_1}) \leq 0. \end{aligned}$$

By convexity of $\mathcal{R}(P_{\mathbf{X}}, \mathbf{D})$ on \mathbf{D} , we have

$$\begin{aligned} (1 - \lambda(t))\mathcal{R}(P_{\mathbf{X}}, \mathbf{D} + t \cdot \mathbf{1}) + \lambda(t)\mathcal{R}(P_{\mathbf{X}}, t \cdot \mathbf{1}) \\ \subseteq \mathcal{R}(P_{\mathbf{X}}, \mathbf{D}). \end{aligned}$$

In particular,

$$\mathcal{R}(P_{\mathbf{X}}, \mathbf{D} + t \cdot \mathbf{1}) + \lambda(t)\mathbf{R}_0 \subseteq \mathcal{R}(P_{\mathbf{X}}, \mathbf{D}).$$

So $\delta_2 > 0$ can be chosen such that $0 < \delta_2 < \delta_1$ and $\lambda(\delta_2)\mathbf{R}_0 \leq \epsilon/4 \cdot \mathbf{1}$.

- (iii) Pick $\delta = \min(\delta_1, \delta_2)$. For any \mathbf{D} , $\mathcal{R}(P_{\mathbf{X}}, \mathbf{D})$ and $\mathcal{R}(P_{\mathbf{X}}, \mathbf{D}_\delta)$ are $\epsilon/4$ -close by (i). On the other hand, $\mathcal{R}(P_{\mathbf{X}}, \mathbf{D}_\delta)$ and $\mathcal{R}(P_{\mathbf{X}}, \mathbf{D}_\delta + \delta \cdot \mathbf{1})$ are $\epsilon/4$ -close by (ii). Therefore, $\mathcal{R}(P_{\mathbf{X}}, \mathbf{D})$ and $\mathcal{R}(P_{\mathbf{X}}, \mathbf{D}_\delta + \delta \cdot \mathbf{1})$ are $\epsilon/2$ -close. Since

$$\mathcal{R}(P_{\mathbf{X}}, \mathbf{D}) \subseteq \mathcal{R}(P_{\mathbf{X}}, \mathbf{D} + \delta \cdot \mathbf{1}) \subseteq \mathcal{R}(P_{\mathbf{X}}, \mathbf{D}_\delta + \delta \cdot \mathbf{1}),$$

we conclude that $\mathcal{R}(P_{\mathbf{X}}, \mathbf{D})$ and $\mathcal{R}(P_{\mathbf{X}}, \mathbf{D} + \delta \cdot \mathbf{1})$ are $\epsilon/2$ -close. \square

IV. CONTINUITY OF $\mathcal{R}_L(P_{\mathbf{X}})$ WITH RESPECT TO $P_{\mathbf{X}}$

We investigate the continuity of \mathcal{R}_L with respect to $P_{\mathbf{X}} \in \mathcal{M}_s(m)$ in this section. Formally, it means that given $P_{\mathbf{X}} \in \mathcal{M}_s(m)$, for any $\epsilon > 0$, there exists $\delta > 0$ such that $\mathcal{R}_L(P_{\mathbf{X}})$ and $\mathcal{R}_L(Q_{\mathbf{X}})$ are ϵ -close whenever $\|P_{\mathbf{X}} - Q_{\mathbf{X}}\| < \delta$, where $\|\cdot\|$ denotes the L_2 -norm in \mathbb{R}^m . The example that follows shows that when $m = \infty$, the property doesn't hold.

Example 1: Let $T > 0$ be fixed. In this example, we show that for any arbitrarily small $\epsilon > 0$, there always exists a finite-support distribution P_X on \mathbb{Z}_+ , the set of nonnegative integers, such that $|1 - P_X(0)| + \sum_{i=1}^{\infty} |P_X(i)| < \epsilon$ and $H(X) \geq T$. This shows that the entropy for the distribution P_X can be arbitrarily large even if the distance of P_X and the distribution of the constant 0 random variable is small.

Given any probability mass function $\mathbf{q} = (q_1, q_2, \dots)$, consider the random variable X with probability distribution $\mathbf{p} = (p_0, p_1, \dots)$ defined by

$$p_0 = 1 - \epsilon, \quad p_i = \epsilon q_i \quad \text{for } i \geq 1.$$

Then

$$\begin{aligned} H(X) &= (1 - \epsilon) \log \left(\frac{1}{1 - \epsilon} \right) + \epsilon \left(\sum_{i=1}^{\infty} q_i \log \left(\frac{1}{\epsilon q_i} \right) \right) \\ &= H(\epsilon) + \epsilon H(\mathbf{q}). \end{aligned}$$

Let $T' = \max\{T, H(\epsilon) + 1\}$. If we can find some \mathbf{q} such that

$$H(\mathbf{q}) \geq \frac{T' - H(\epsilon)}{\epsilon},$$

then X has entropy greater than or equal to T and $W_H(X) = \epsilon$. To construct such a distribution \mathbf{q} , let $L = \lceil 2^{(T' - H(\epsilon))/\epsilon} \rceil$. Let $q_i = \frac{1}{L}$ for all $1 \leq i \leq L$ and $q_i = 0$ otherwise. \square

A. Independent Sources

When sources are independent, the lossless rate region depends only on their entropies, as shown in the following theorem and corollary.

Theorem 2: Consider a network source coding problem where all the sources are independent. Let $\mathbf{X} = (X_1, \dots, X_m)$ be the source vector and let $\mathcal{R}_L(P_{\mathbf{X}})$ denote the lossless rate region for source distribution $P_{\mathbf{X}} = \prod_{i=1}^m P_{X_i}$. If $Q_{\mathbf{X}}$ is another distribution such that $H(X_i)_{P_{X_i}} \geq H(X_i)_{Q_{X_i}}$ for all $i \in \{1, \dots, m\}$, then $\mathcal{R}_L(P_{\mathbf{X}}) \subseteq \mathcal{R}_L(Q_{\mathbf{X}})$.

Proof. Let $H_P = H(X_1)_{P_{X_1}}$ and $H_Q = H(X_1)_{Q_{X_1}}$. Let $A_\epsilon^{(n)}(P_{\mathbf{X}})$ and $A_\epsilon^{(n)}(Q_{\mathbf{X}})$ denote the typical sets with respect to $P_{\mathbf{X}}$ and $Q_{\mathbf{X}}$ respectively. It suffices to show

$$\mathcal{R}_L(Q_{X_1} \prod_{i=2}^m P_{X_i}) \subseteq \mathcal{R}_L(P_{\mathbf{X}})$$

for two distributions P_{X_1} and Q_{X_1} such that $H_P \geq H_Q > 0$. Given any achievable rate vector $\mathbf{R} \in \mathcal{R}_L(Q_{X_1} \prod_{i=2}^m P_{X_i})$. Let $\epsilon > 0$ be arbitrarily small such that $\epsilon < 1/10$ and $\epsilon < H_Q$. Choose a rate \mathbf{R} -code \mathcal{C} of length n with $P_{\mathbf{X}^n}(\text{error}) < \epsilon$ such that

$$P_{X_1^n}(A_\epsilon^{(n)}(P_{X_1})) \geq 1 - \epsilon \quad (1)$$

and

$$Q_{X_1^n}(A_\epsilon^{(n)}(Q_{X_1})) \geq 1 - \epsilon. \quad (2)$$

For any $x_1^n \in \mathcal{X}_1^n$, let

$$E(x_1^n) \subseteq \prod_{i=2}^m \mathcal{X}_i^n$$

be the error event $\{\text{error} | X_1^n = x_1^n\}$ given x_1^n . Then by definition,

$$\sum_{x_1^n \in A_\epsilon^{(n)}(P_{X_1})} P_{X_1^n}(x_1^n) P_{X_2^n \dots X_m^n}(E(x_1^n)) < \epsilon.$$

Enumerate all the typical sequences $x_1^n(1), \dots, x_1^n(L)$ in $A_\epsilon^{(n)}(P_{X_1})$, where $L = |A_\epsilon^{(n)}(P_{X_1})|$, such that

$$e_1 \leq e_2 \leq \dots \leq e_L,$$

where for each $j \in \{1, \dots, L\}$

$$e_j = P_{X_2^n \dots X_m^n}(E(x_1^n(j))).$$

For each $j \in \{1, \dots, L\}$, set $p_j = P_{X_1^n}(x_1^n(j))$. Pick n large enough such that there exists $1 \leq l \leq L$ satisfying

$$\frac{1}{4} \leq \sum_{j=l+1}^L p_j \leq \frac{1}{2}.$$

Then one has

$$\epsilon > \sum_{j=l+1}^L p_j e_j \geq \left(\sum_{j=l+1}^L p_j \right) e_l,$$

which implies that

$$e_l \leq \frac{\epsilon}{\sum_{j=l+1}^L p_j} \leq 4\epsilon. \quad (3)$$

Now since $2^{-n(H_P+\epsilon)} \leq p_j \leq 2^{-n(H_P-\epsilon)}$ for all $1 \leq j \leq l$,

$$l2^{-n(H_P-\epsilon)} \geq \sum_{j=1}^l p_j \geq \frac{1}{2} \geq \frac{1}{2} \sum_{j=1}^L p_j \geq \frac{L}{2} 2^{-n(H_P+\epsilon)}.$$

Hence,

$$l \geq L2^{-2n\epsilon-1}.$$

Partition the typical set $A_\epsilon^{(n)}(Q_{X_1})$ as

$$A_\epsilon^{(n)}(Q_{X_1}) = A_1 \cup A_2 \cup \dots \cup A_K \quad (4)$$

such that $|A_i| = l$ for $1 \leq i \leq K-1$ and $|A_K| \leq l$, where the unions in (4) are disjoint unions. Then

$$K = \left\lceil \frac{|A_\epsilon^{(n)}(Q_{P_{X_1}})|}{l} \right\rceil \leq \frac{2^{n(H_Q+\epsilon)}}{2^{n(H_P-\epsilon)}} 2^{2n\epsilon+1} \leq 2^{4n\epsilon+1}.$$

Set $B := \{x_1^n(1), \dots, x_1^n(l)\}$. For each $i \in \{1, \dots, K\}$, define a one-to-one function η_i from A_i to B . Now define a function

$$\phi : A_\epsilon^{(n)}(Q_{X_1}) \rightarrow \{1, \dots, K\} \times B$$

as

$$\phi(x_1^n) = (\tau(x_1^n), \eta(x_1^n)),$$

where $\tau(x_1^n) = i$ if and only if $x_1^n \in A_i$, and

$$\eta(x_1^n) = \eta_{\tau(x_1^n)}(x_1^n).$$

So it is obvious that ϕ is a one-to-one function.

Now construct a new code \mathcal{C}' for the source distribution $Q_{X_1} \prod_{i=2}^m P_{X_i}$ as follows. First, for source sequence (x_1^n, \dots, x_m^n) such that $x_1^n \in A_\epsilon^{(n)}(Q_{X_1})$, we apply the same encoding functions of \mathcal{C} on $(\eta(x_1^n), x_2^n, \dots, x_m^n)$, and send the index $\tau(x_1^n)$ through the network to all the nodes. At each node that is requested to reproduce some sources, the decoding functions of \mathcal{C} are then applied. If that node has to reproduce x_1^n , then we first apply the decoding function of \mathcal{C} to get \hat{x}_1^n , the reproduction of $\eta(x_1^n)$ by \mathcal{C} at this node. Then we use \hat{x}_1^n and $\tau(x_1^n)$ to recover x_1^n by using the inverse map from $\phi(A_\epsilon^{(n)}(Q_{X_1}))$ to $A_\epsilon^{(n)}(Q_{X_1})$ when

$$(\tau(x_1^n), \hat{x}_1^n) \in \phi(A_\epsilon^{(n)}(Q_{X_1})).$$

Otherwise, we declare an error.

By the construction of \mathcal{C}' , we can easily bound the error probability Q_{error} for the source distribution $Q_{X_1} \prod_{i=2}^m P_{X_i}$ as

$$Q_{\text{error}} \leq Q_{X_1^n}(A_\epsilon^{(n)}(Q_{X_1})^c) + \sum_{j=1}^l Q_{X_1^n}(x_1^n(j)) e_j \leq 5\epsilon$$

by (2) and (3). On the other hand, \mathcal{C}' has rate

$$\mathbf{R} + (4\epsilon + 1/n) \cdot \mathbf{1}.$$

Since $\epsilon > 0$ is arbitrary, $\mathbf{R} \in \mathcal{R}_L(P_{\mathbf{X}})$. \square

Corollary 1: In a fixed network source coding problem, if two distributions $P_{\mathbf{X}} = \prod_{i=1}^m P_{X_i}$ and $Q_{\mathbf{X}} = \prod_{i=1}^m Q_{X_i}$ satisfy that $H(X_i)_{P_{X_i}} = H(X_i)_{Q_{X_i}} \forall i \in \{1, \dots, m\}$, then $\mathcal{R}_L(P_{\mathbf{X}}) = \mathcal{R}_L(Q_{\mathbf{X}})$.

The proof technic for Theorem 2 can also be applied to show that $\mathcal{R}_L(\prod_{i=1}^m P_{X_i})$ is continuous with respect to the entropy vector of the independent sources X_1, \dots, X_m . Since each entropy vector $H(X_i)_{P_{X_i}}$ is continuous with respect to P_{X_i} , this implies that $\mathcal{R}_L(\prod_{i=1}^m P_{X_i})$ is continuous with respect to the distribution vector $(P_{X_1}, \dots, P_{X_m})$ in the independent source case.

Theorem 3: $\mathcal{R}_L(P_{\mathbf{X}})$ is continuous with respect to the entropy vector

$$(H(X_1)_{P_{X_1}}, \dots, H(X_m)_{P_{X_m}})$$

when $P_{\mathbf{X}} = \prod_{i=1}^m P_{X_i}$. In other words, for any $\epsilon > 0$, there exists $\delta > 0$ such that $\mathcal{R}_L(P_{\mathbf{X}})$ and $\mathcal{R}_L(Q_{\mathbf{X}})$ are ϵ -close whenever $|H(X_i)_{P_{X_i}} - H(X_i)_{Q_{X_i}}| < \delta$ for all $i \in \{1, \dots, m\}$, and $P_{\mathbf{X}} = \prod_{i=1}^m P_{X_i}$ and $Q_{\mathbf{X}} = \prod_{i=1}^m Q_{X_i}$.

Proof. It suffices to consider the case

$$H(X_i)_{P_{X_i}} = H(X_i)_{Q_{X_i}} \forall i \in \{2, \dots, m\}.$$

Let $H_P = H(X_1)_{P_{X_1}}$ and $H_Q = H(X_1)_{Q_{X_1}}$, and suppose that $H_Q \leq H_P \leq H_Q + \delta$. Then $\mathcal{R}_L(P_{\mathbf{X}}) \subseteq \mathcal{R}_L(Q_{\mathbf{X}})$ by Theorem 2.

For any achievable rate vector $\mathbf{R} \in \mathcal{R}_L(Q_{\mathbf{X}})$, let $\{\mathcal{C}_n\}$ be a sequence of length- n codes of rate \mathbf{R} such that the error probability with respect to $P_{\mathbf{X}}$ tends to zero as n grows without bound. For each n and $\tau > 0$, since $H_P \leq H_Q + \delta$, we partition the set $A_\tau^{(n)}(P_{X_1})$ as

$$A_\tau^{(n)}(P_{X_1}) = \cup_{i=1}^{2^{n(\delta+2\tau)}} A_i^{(n)},$$

such that each $A_i^{(n)}$ has size smaller than $A_\tau^{(n)}(Q_{X_1})$. By building injections from $A_i^{(n)}$ to $A_\tau^{(n)}(Q_{X_1})$ as in the proof of Theorem 2 and sending additional rate $\Delta \mathbf{R} \leq (2\tau + \delta) \cdot \mathbf{1}$ to distinguish those the sets $A_1^{(n)}, \dots, A_{2^{n(\delta+2\tau)}}^{(n)}$, we get a sequence of new codes $\{\mathcal{C}'_n\}$ of rate $\mathbf{R} + \Delta \mathbf{R}$ whose error probability with respect to $Q_{\mathbf{X}}$ tend to zero as n grows without bound. That shows $\mathcal{R}_L(P_{\mathbf{X}})$ and $\mathcal{R}_L(Q_{\mathbf{X}})$ are $\sqrt{m}(2\tau + \delta)$ -close. This completes the proof. \square

B. Concavity

The so-called concavity property of $\mathcal{R}_L(P_{\mathbf{X}})$ generalized from the concavity of the entropy function $H(X)$ was discussed in [6]. Concavity property means that the following holds for every $P_{\mathbf{X}}, Q_{\mathbf{X}} \in \mathcal{M}_s(m)$ and $\lambda \in [0, 1]$

$$\mathcal{R}_L(\lambda P_{\mathbf{X}} + (1 - \lambda)Q_{\mathbf{X}}) \subseteq \lambda \mathcal{R}_L(P_{\mathbf{X}}) + (1 - \lambda)\mathcal{R}_L(Q_{\mathbf{X}}).$$

Some example networks that satisfy and a counterexample network that doesn't satisfy the concavity property can be found in [6]. As we can prove continuity of $H(X)$ by using its concavity property (although we don't have to), we can prove the concavity property of \mathcal{R}_L by using the concavity property. We mainly show concavity implies continuity in this subsection.

By definition, $\mathcal{R}_L(P_{\mathbf{X}})$ can be totally determined by its boundary points, namely,

$$\mathcal{R}_L(P_{\mathbf{X}}) = \{\mathbf{R} \mid \mathbf{R} \geq \mathbf{R}^* \text{ for some } \mathbf{R}^* \in \partial \mathcal{R}_L(P_{\mathbf{X}})\},$$

where $\partial \mathcal{R}_L(P_{\mathbf{X}})$ is the set of boundary points in $\mathcal{R}_L(P_{\mathbf{X}})$. Furthermore, $\mathcal{R}_L(P_{\mathbf{X}})$ can be totally determined by the subset $\partial^* \mathcal{R}_L(P_{\mathbf{X}})$ of $\partial \mathcal{R}_L(P_{\mathbf{X}})$ consisting of all the boundary points each of whose components is no larger than $s \log(m)$. That is,

$$\mathcal{R}_L(P_{\mathbf{X}}) = \{\mathbf{R} \mid \mathbf{R} \geq \mathbf{R}^* \text{ for some } \mathbf{R}^* \in \partial^* \mathcal{R}_L(P_{\mathbf{X}})\}. \quad (5)$$

We sketch the proof of (5) as follows. Let $\mathbf{R} = (R(e))_{e \in \mathcal{E}} \in \partial \mathcal{R}_L(P_{\mathbf{X}})$ be a boundary point that has one component r_{e_0} larger than $s \log(m)$. Since the encoding message on the edge e_0 in an asymptotically lossless length- n code at rate $\mathbf{R} + \epsilon \cdot \mathbf{1}$ ($\epsilon > 0$ is an arbitrarily small positive number) is a function of the source vector \mathbf{X}^n , the rate on e_0 can be made no larger than the average maximal possible entropy $H(\mathbf{X}^n)/n$ of \mathbf{X}^n which is $s \log(m)$, by letting n large enough. This implies that the rate vector $\mathbf{R}' = (R(e'))_{e \in \mathcal{E}}$ is also in $\mathcal{R}_L(P_{\mathbf{X}})$, where

$$R(e)' := \begin{cases} R(e), & \text{if } R(e) < s \log(m), \\ s \log(m), & \text{otherwise.} \end{cases}$$

We use this fact to prove the following theorem.

Theorem 4: If a network source coding problem satisfies the concavity property, then its lossless rate region is continuous with respect to the source distribution $P_{\mathbf{X}}$.

Proof. Given a sequence of distributions $\{P_{\mathbf{X}}^{(n)}\} \in \mathcal{M}_s(m)$ such that $P_{\mathbf{X}}^{(n)} \rightarrow P_{\mathbf{X}}$. We want to show for any $\epsilon > 0$, there exists n_0 such that $\mathcal{R}_L(P_{\mathbf{X}}^{(n)})$ and $\mathcal{R}_L(P_{\mathbf{X}})$ are ϵ -close whenever $n \geq n_0$.

For any $\epsilon > 0$, choose $0 < \lambda < 1$ such that

$$(1 - \lambda)s \log(m) < \epsilon.$$

Let

$$Q_n(\mathbf{x}) = \frac{P_{\mathbf{X}}^{(n)}(\mathbf{x}) - \lambda P(\mathbf{x})}{1 - \lambda}$$

for $\mathbf{x} \in \{1, \dots, m\}^s$. Since $\lim_{n \rightarrow \infty} P_{\mathbf{X}}^{(n)} = P_{\mathbf{X}}$, $\{Q_n(\mathbf{x})\}_{\mathbf{x}}$ is a distribution on $\{1, \dots, m\}^s$ when $n \geq n_1$ for some n_1 . Now $P_{\mathbf{X}}^{(n)} = \lambda P_{\mathbf{X}} + (1 - \lambda)Q_n$, hence by concavity of \mathcal{R}_L

$$\mathcal{R}_L(P_{\mathbf{X}}^{(n)}) \subseteq \lambda \mathcal{R}_L(P_{\mathbf{X}}) + (1 - \lambda)\mathcal{R}_L(Q_n).$$

Thus for any $\mathbf{R} \in \mathcal{R}_L(P_{\mathbf{X}}^{(n)})$, there exist $\mathbf{R}_1 \in \mathcal{R}_L(P_{\mathbf{X}})$ and $\mathbf{R}_2 \in \mathcal{R}_L(Q_n)$ such that

$$\mathbf{R} \geq \lambda \mathbf{R}_1 + (1 - \lambda)\mathbf{R}_2 \geq \lambda \mathbf{R}_1 \geq \lambda \mathbf{R}_1^*$$

for some boundary point \mathbf{R}_1^* in $\mathcal{R}_L(P_{\mathbf{X}})$. Since \mathbf{R}_1^* is a boundary point, $\mathbf{R}_1^* \leq s \log(m) \cdot \mathbf{1}$, and so

$$\mathbf{R} + \epsilon \cdot \mathbf{1} > \mathbf{R} + (1 - \lambda)s \log(m) \cdot \mathbf{1} \geq \mathbf{R}_1^*,$$

which implies that $\mathbf{R} + \epsilon \cdot \mathbf{1} \in \mathcal{R}_L(P_{\mathbf{X}})$. Thus

$$\mathcal{R}_L(P_{\mathbf{X}}^{(n)}) + \epsilon \cdot \mathbf{1} \subseteq \mathcal{R}_L(P_{\mathbf{X}})$$

for $n \geq n_1$.

On the other hand, for every $\mathbf{R} = (R(e))_{e \in \mathcal{E}} \in \partial^* \mathcal{R}_L(P_{\mathbf{X}})$, by [6, Lemma 1], let

$$R(e) = \frac{1}{t_{\mathbf{R}}} (H(F_e)_{P_{\mathbf{X}}} + \sum_{w_e} H(X_j^{t_{\mathbf{R}}} | G_{v,j})_{P_{\mathbf{X}}})$$

for some length- $t_{\mathbf{R}}$ block code with encoding messages $\{F_e\}$ and decoding messages $\{G_{v,j}\}$. Since $P_{\mathbf{X}}^{(n)} \rightarrow P_{\mathbf{X}}$ as n grows without bound, by the continuity of the entropy function $H(F_e)_{P_{\mathbf{X}}} + \sum_{w_e} H(X_j^{t_{\mathbf{R}}} | G_{v,j})_{P_{\mathbf{X}}}$, there exists $n_{\mathbf{R}} > 0$ such that

$$\mathbf{R} + \frac{\epsilon}{2} \cdot \mathbf{1} \in \mathcal{R}_L(P_{\mathbf{X}}^{(n)}) \quad (6)$$

whenever $n \geq n_{\mathbf{R}}$. Since $\partial^* \mathcal{R}_L(P_{\mathbf{X}})$ is compact, there exists a finite collection of open sets $\{U_{\mathbf{R}_i}\}_{i=1}^l$ of $\partial^* \mathcal{R}_L(P_{\mathbf{X}})$, each $U_{\mathbf{R}_i}$ is centered at $\mathbf{R}_i \in \partial^* \mathcal{R}_L(P_{\mathbf{X}})$, such that

$$\partial^* \mathcal{R}_L(P_{\mathbf{X}}) = \cup_{i=1}^l U_{\mathbf{R}_i}$$

and that for all $i \in \{1, \dots, l\}$

$$\|\mathbf{R}_i - \mathbf{R}'\| < \frac{\epsilon}{2} \quad \forall \mathbf{R}' \in U_{\mathbf{R}_i}.$$

By letting $n_2 = \max_{1 \leq i \leq l} \{n_{\mathbf{R}_i}\}$, we have that for any $n \geq n_2$, and any $\mathbf{R} \in \partial^* \mathcal{R}_L(P_{\mathbf{X}})$, there exists $\mathbf{R}_n \in \mathcal{R}$ such

that $\|\mathbf{R} - \mathbf{R}_n\| < \sqrt{m}\epsilon$. This implies that for any $\mathbf{R} \in \mathcal{R}_L(P_{\mathbf{X}})$, there exists $\mathbf{R}_n \in \mathcal{R}_L(P_{\mathbf{X}}^{(n)})$ such that

$$\|\mathbf{R} - \mathbf{R}_n\| < \sqrt{m}\epsilon.$$

In conclusion, $\mathcal{R}_L(P_{\mathbf{X}})$ and $\mathcal{R}_L(P_{\mathbf{X}}^{(n)})$ are ϵ -close whenever $n \geq n_0 = \max\{n_1, n_2\}$. \square

APPENDIX

Lemma 1: Let $V^n = (V_1, V_2, \dots, V_n)$ be a random vector in $\{0, 1, \dots, m-1\}^n$ and let w be the per symbol expected Hamming weight of V^n defined as

$$w := \frac{1}{n} E d_H(V^n, \mathbf{0}) = \frac{1}{n} E |\{i \mid V_i \neq 0\}|.$$

Then

$$H(V^n) \leq n(H(w) + w \log(m-1)).$$

Proof. First for each $1 \leq i \leq n$ and each $0 \leq j \leq m-1$, let $p_{i,j}$ denote the probability $\Pr\{V_i = j\}$. Then

$$w = \frac{1}{n} E d_H(V^n, \mathbf{0}) = \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^{m-1} p_{i,j}.$$

Let $w_i := \sum_{j \neq 0} p_{i,j}$. The maximal entropy $H(V_i)$ subject to the constraint $\sum_{j \neq 0} p_{i,j} = w_i$ for every i occurs when $p_{i,0} = 1 - w_i$ and $p_{i,j} = \frac{w_i}{m-1}$ for each $j \neq 0$. Hence

$$\begin{aligned} H(V_i) &\leq H\left(1 - w_i, \frac{w_i}{m-1}, \dots, \frac{w_i}{m-1}\right) \\ &= H(w_i) + w_i \log(m-1). \end{aligned}$$

Therefore, by convexity of entropy function, we have

$$\begin{aligned} H(V^n) &\leq \sum_{i=1}^n H(V_i) \leq \sum_{i=1}^n [H(w_i) + w_i \log(m-1)] \\ &\leq n(H(w) + w \log(m-1)). \end{aligned}$$

\square

Corollary 2: Let V^n and W^n be two random vectors in $\{0, 1, \dots, m-1\}^n$. Then

$$\frac{1}{n} H(V^n | W^n) \leq H(e) + e \log(m-1) = O(e \log(e)),$$

where

$$e := E \left\{ \frac{1}{n} d_H(V^n, W^n) \right\}.$$

Proof. Apply Lemma 1 to the random vector $V^n - W^n$. \square

Lemma 2: Let \mathcal{N} be a network coding problem. For any $\epsilon > 0$, there exists δ such that for any $\mathbf{D} \in \mathbb{R}_+^k$ and any $P_{\mathbf{X}} \in \mathcal{M}_s(m)$, $\mathcal{R}(P_{\mathbf{X}}, \mathbf{D})$ and $\mathcal{R}(P_{\mathbf{X}}, \mathbf{D}_\delta)$ are ϵ -close.

Proof. Let $\epsilon > 0$ be given. We show that there exists a $\delta > 0$ such that for any $\mathbf{D} = (D_1, \dots, D_k) \in \mathbb{R}_+^k$ and any $P_{\mathbf{X}} \in \mathcal{M}_s(m)$, $\mathcal{R}(P_{\mathbf{X}}, \mathbf{D})$ and $\mathcal{R}(P_{\mathbf{X}}, \mathbf{D}'(\delta))$ are $\epsilon/2$ -close, where $\mathbf{D}'(\delta) = (D'_1, \dots, D'_k)$ is the vector such that for every $i \in \{1, \dots, k\}$,

$$D'_i(\delta) = \begin{cases} 0 & , \text{ if } D_i \leq \delta \\ D_i & , \text{ otherwise.} \end{cases}$$

From this we can observe that both $\mathcal{R}(P_{\mathbf{X}}, \mathbf{D})$ and $\mathcal{R}(P_{\mathbf{X}}, \mathbf{D}_\delta)$ are $\epsilon/2$ -close to $\mathcal{R}(P_{\mathbf{X}}, \mathbf{D}'(\delta))$, and hence $\mathcal{R}(P_{\mathbf{X}}, \mathbf{D})$ and $\mathcal{R}(P_{\mathbf{X}}, \mathbf{D}_\delta)$ are ϵ -close. Now $\mathbf{D}'(\delta) \leq \mathbf{D}$ for any $\delta > 0$, so $\mathcal{R}(P_{\mathbf{X}}, \mathbf{D}'(\delta)) \subseteq \mathcal{R}(P_{\mathbf{X}}, \mathbf{D})$. We only need to choose appropriate $\delta > 0$ (independent of \mathbf{X} and \mathbf{D}) such that $\mathcal{R}(P_{\mathbf{X}}, \mathbf{D}) + (\epsilon/2) \cdot \mathbf{1} \subseteq \mathcal{R}(P_{\mathbf{X}}, \mathbf{D}'(\delta))$.

Now k is the total number of reproduction demands, so we number each pair (v, j) for which $j \in J(v)$ as $\sigma(v, j) \in \{1, \dots, k\}$ such that there is a one-to-one correspondence between the set of reproduction requests $\{(v, j) \mid j \in J(v)\}$ and the set $\{1, \dots, k\}$ by σ . For any achievable rate vector $\mathbf{R} \in \mathcal{R}(\mathbf{D})$, pick a length- n block code \mathcal{C} with rate \mathbf{R} such that

$$E d(X_j^n, \hat{X}_j^n(v)) \leq D_{\sigma(v,j)} + e(\epsilon),$$

where for any $j \in \{1, \dots, k\}$ and any $v \in \mathcal{V}$ with $j \in J(v)$, $\hat{X}_j^n(v)$ is the reproduction vector and $e(\epsilon) > 0$ was chosen such that

$$k(H(e(\epsilon)) + e(\epsilon) \log m) < \frac{\epsilon}{2}.$$

By the definition of $\mathbf{D}'(\delta)$, we need to construct another code based on \mathcal{C} so that it is possible to make the error probability of the reproduction for each pair (v, j) with $D_{\sigma(v,j)} \leq \delta$ arbitrarily small. By Corollary 2, for each (v, j) such that $D_{\sigma(v,j)} \leq e(\epsilon)$, we have that

$$H(X_j^n | \hat{X}_j^n(v)) \leq n\epsilon/(2k).$$

Since at node v , we already have the side information $\hat{X}_j^n(v)$, the additional rate that is required to describe X_j^n losslessly is less than $\epsilon/(2k)$. So if we base on \mathcal{C} to add additional description for each X_j^n to the node v satisfying $D_{\sigma(v,j)} \leq e(\epsilon)$ along a path connecting source X_j^n to node v , the total additional rate (for all such pair (v, j)) on each link is less than $\epsilon/2$. Therefore, the rate vector $\mathbf{R} + (\epsilon/2) \cdot \mathbf{1}$ is $\mathbf{D}'(\delta)$ -achievable. Now let $\delta = e(\epsilon)/2$, we will get the desired result. \square

Since each component of any source vector in $\mathcal{M}_s(m)$ has alphabet size no greater than m , it is clear that the vector $k \log(m) \cdot \mathbf{1} \in \mathbb{R}_+^k$ is \mathbf{D}' -achievable for any source random vector $P_{\mathbf{X}} \in \mathcal{M}_s(m)$. Hence we have the following lemma.

Lemma 3: $\bigcap_{P_{\mathbf{X}} \in \mathcal{M}_s(m)} \mathcal{R}(P_{\mathbf{X}}, \mathbf{0}) \neq \emptyset$.

REFERENCES

- [1] R. Ahlswede and J. Körner. Source coding with side information and a converse for degraded broadcast channels. *IEEE Transactions on Information Theory*, IT-21(6):629–637, November 1975.
- [2] R. M. Gray and A. D. Wyner. Source coding for a simple network. *Bell System Technical Journal*, 53(9):1681–1721, November 1974.
- [3] C. Heegard and T. Berger. Rate distortion when side information may be absent. *IEEE Transactions on Information Theory*, IT-31:727–734, November 1985.
- [4] A. D. Wyner. On source coding with side information at the decoder. *IEEE Transactions on Information Theory*, IT-21(3):294–300, November 1975.
- [5] S. Robert Li, R. W. Yeung, and N. Cai. Linear network coding. *IEEE Transactions on Information Theory*, IT-49(2):371–381, February 2003.
- [6] W. Gu and M. Effros. On the concavity of lossless rate regions. In *Proceedings of the IEEE International Symposium on Information Theory*, Seattle, WA, June 2006.