Foundations and Trends® in Communications and Information Theory

Volume 2 Issue 5, 2005

Editorial Board

Editorial Scope

Foundations and Trends® in Communications and Information Theory will publish survey and tutorial articles in the following topics:

- Coded modulation
- Coding theory and practice
- Communication complexity
- Communication system design
- Cryptology and data security
- Data compression
- Data networks
- Demodulation and Equalization
- Denoising
- Detection and estimation
- Information theory and statistics
- Information theory and computer science
- Joint source/channel coding
- Modulation and signal design
- Multiuser detection
- Multiuser information theory
- Optical communication channels
- Pattern recognition and learning
- Quantization
- Quantum information processing
- Rate-distortion theory
- Shannon theory
- Signal processing for communications
- Source coding
- Storage and recording codes
- Speech and Image Compression
- Wireless Communications

Information for Librarians

Foundations and Trends® in Communications and Information Theory, 2005, Volume 2, 4 issues. ISSN paper version 1567-2190. ISSN online version 1567-2328. Also available as a combined paper and online subscription.

Network Coding Theory
Part II: Multiple Source

Raymond W. Yeung

The Chinese University of Hong Kong
Hong Kong, China
whyeung@ie.cuhk.edu.hk

Shuo-Yen Robert Li

The Chinese University of Hong Kong
Hong Kong, China
bob@ie.cuhk.edu.hk

Ning Cai

Xidian University
Xi'an, Shaanxi, China
caining@mail.xidian.edu.cn

Zhen Zhang

University of Southern California
Los Angeles, CA, USA
zzhang@milly.usc.edu

now

the essence of knowledge

Boston – Delft

Foundations and Trends® in Communications and Information Theory

Published, sold and distributed by:
now Publishers Inc.
PO Box 1024
Hanover, MA 02339
USA
Tel. +1-781-985-4510
www.nowpublishers.com
sales@nowpublishers.com

Outside North America:
now Publishers Inc.
PO Box 179
2600 AD Delft
The Netherlands
Tel. +31-6-51115274

A Cataloging-in-Publication record is available from the Library of Congress

The preferred citation for this publication is R.W. Yeung, S.-Y.R. Li, N. Cai, and Z. Zhang, Network Coding Theory Part II: Multiple Source, Foundation and Trends® in Communications and Information Theory, vol 2, no 5, pp 330–381, 2005

Printed on acid-free paper

Foundations and Trends® in
Communications and Information Theory
Vol. 2, No 5 (2005) 330–381

DOI: 10.1561/0100000007II

Network Coding Theory Part II: Multiple Source

Raymond W. Yeung[1], Shuo-Yen Robert Li[2], Ning Cai[3] and Zhen Zhang[4]

[1] *Department of Information Engineering, The Chinese University of Hong Kong, Shatin, N.T., Hong Kong, whyeung@ie.cuhk.edu.hk*
[2] *Department of Information Engineering, The Chinese University of Hong Kong, Shatin, N.T., Hong Kong, bob@ie.cuhk.edu.hk*
[3] *The State Key Lab. of ISN, Xidian University, Xi'an, Shaanxi, 710071, China, caining@mail.xidian.edu.cn*
[4] *Department of Electrical Engineering-Systems, University of Southern California, Los Angeles, CA 90089-2565, USA, zzhang@milly.usc.edu*

Abstract

Store-and-forward had been the predominant technique for transmitting information through a network until its optimality was refuted by network coding theory. Network coding offers a new paradigm for network communications and has generated abundant research interest in information and coding theory, networking, switching, wireless communications, cryptography, computer science, operations research, and matrix theory.

In this issue we review network coding theory for the scenario when there are multiple source nodes each intending to transmit to a different set of destination nodes.

A companion issue reviews the foundational work that has led to the development of network coding theory and discusses the theory for the transmission from a single source node to other nodes in the network.

Publisher's Note

References to 'Part I' and 'Part II' in this issue refer to *Foundations and Trends® in Communications and Information Technology* Volume 2 Numbers 4 and 5 respectively.

Contents

1

Superposition Coding and Max-Flow Bound

In Part I of this tutorial, we have discussed the single-source network coding problem in an algebraic setting. Each communication channel in the network is assumed to have unit capacity. The maximum rate at which information can be multicast has a simple characterization in terms of the maximum flows in the network. In Part II, we consider the more general multi-source network coding problem in which more than one *mutually independent* information sources are generated at possibly different nodes, where each information source is transmitted to a certain set of nodes in the network. We continue to assume that the communication channels in the network are free of error.

The *achievable information rate region* for a multi-source network coding problem, which will be formally defined in Section 2, refers to the set of all possible rate tuples at which multiple information sources can be multicast simultaneously on a network. In a single-source network coding problem, a primary goal is to characterize the maximum rate at which information can be multicast from the source node to all the sink nodes. In a multi-source network coding problem, we are interested in characterizing the achievable information rate region.

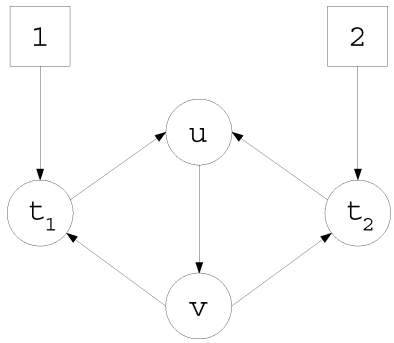

Fig. 1.1 A network for which superposition coding is suboptimal.

Multi-source network coding turns out *not* to be a simple extension of single-source network coding. In the rest of this section, we discuss two characteristics of multi-source networking coding which differentiate it from single-source network coding. In all the examples, the unit of information is the bit.

In Part I, nodes are labelled by capital letters. In Part II, since captical letters are reserved for random variables, nodes will instead be labelled by small letters.

1.1 Superposition coding

Let us first revisit the network in Figure 1.2(b) of Part I which is reproduced here as Figure 1.1 in a slightly different manner. Here, we assume that each channel has unit capacity. For $i = 1, 2$, the source node i generates a bit b_i which is sent to the node t_i. We have shown in Example 1.3 of Part I that in order for the nodes t_1 and t_2 to exchange the two bits b_1 and b_2, network coding must be performed at the node u. This example in fact has a very intriguing implication. Imagine that on the Internet a message in English and a message in Chinese are generated at two different locations. These two messages are to be transmitted from one point to another point within the network, and we can assume that there is no correlation between the two messages. Then this example shows that we may have to perform joint coding of the two messages in the network in order to achieve bandwidth optimality!

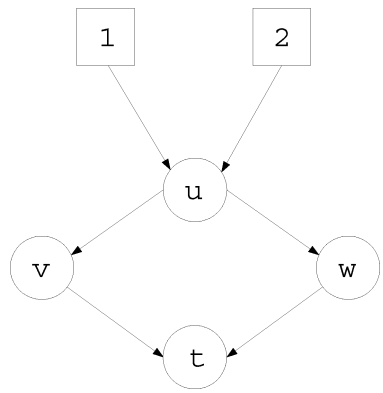

Fig. 1.2 A network for which superposition coding is optimal.

We refer to the method of coding individual information sources separately as *superposition coding*. The above example simply shows that superposition coding can be suboptimal.

We now give an example for which superposition coding does achieve optimality. Consider the network in Figure 1.2. To simply the discussion, we set the capacities of the channels $1u$ and $2u$ to infinity so that the information generated at both source nodes are directly available to the node u. For all the other channels, we set the capacity to 1. We want to multicast the information generated at the source node 1 to the nodes v, w and t, and to transmit the information generated at the source node 2 to the node t.

Let X_1 and X_2 be independent random variables representing the information generated respectively at the source nodes 1 and 2 for one unit time. The rate of the information generated at the source node s is given by $\omega_s = H(X_s)$ for $s = 1, 2$. Let U_{ij} be the random variable sent on the channel ij, where $H(U_{ij}) \leq 1$ due to the bit rate constraint for the channel. Then for any coding scheme achieving the prescribed communication goals, we have

$$\begin{aligned} 2\omega_1 + \omega_2 &= 2H(X_1) + H(X_2) \\ &= 2H(X_1) + H(X_2|X_1) \\ &\overset{a)}{\leq} 2H(X_1) + H(U_{vt}, U_{wt}|X_1) \end{aligned}$$

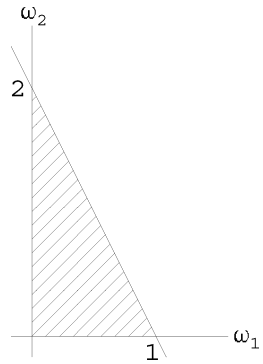

Fig. 1.3 The information rate region for the network in Figure 1.2.

$$
\begin{aligned}
&\overset{b)}{\leq} 2H(X_1) + H(U_{uv}, U_{uw}|X_1) \\
&\leq 2H(X_1) + H(U_{uv}|X_1) + H(U_{uw}|X_1) \\
&= H(U_{uv}, X_1) + H(U_{uw}, X_1) \\
&\overset{c)}{=} H(U_{uv}) + H(U_{uw}) \\
&\leq 2,
\end{aligned}
$$

where a) follows because X_2 is a function of U_{vt} and U_{wt}, b) follows because U_{vt} is a function of U_{uv} and U_{wt} is a function of U_{uw}, and c) follows because X_1 is a function of U_{uv} and a function of U_{uw}.

This region is illustrated in Figure 1.3. To see that the whole region is achievable by superposition coding, let $r_{ij}^{(s)}$ be the bit rate on the channel ij for transmitting the information generated at the source node s. Due to the bit rate constraint for each channel ij, the following must be satisfied:

$$r_{ij}^{(1)} + r_{ij}^{(2)} \leq 1.$$

Then the rate pair $(\omega_1, \omega_2) = (1, 0)$ is achieved by taking

$$r_{uv}^{(1)} = r_{uw}^{(1)} = r_{vt}^{(1)} = 1$$

and

$$r_{wt}^{(1)} = r_{uv}^{(2)} = r_{uw}^{(2)} = r_{vt}^{(2)} = r_{wt}^{(2)} = 0,$$

while the rate pair $(0,2)$ is achieved by taking

$$r_{uv}^{(1)} = r_{uw}^{(1)} = r_{vt}^{(1)} = r_{wt}^{(1)} = 0$$

and

$$r_{uv}^{(2)} = r_{uw}^{(2)} = r_{vt}^{(2)} = r_{wt}^{(2)} = 1.$$

Then the whole information rate region depicted in Figure 1.3 is seen to be achievable via a time-sharing argument.

From the above two examples, we see that superposition coding is sometimes but not always optimal. Optimality of superposition coding for certain classes of multilevel diversity coding problems (special cases of multi-source network coding) has been reported in [17], [14], [21]. For a class of multilevel diversity coding problems (special cases of multi-source network coding) studied in [8], superposition coding is optimal for 86 out of 100 configurations. In any case, superposition coding always induces an inner bound on the information rate region.

1.2 The max-flow bound

In this section, we revisit the two examples in the last section from a different angle. First, for the network in Figure 1.1, we already have seen that superposition coding is suboptimal. Now consideration of the max-flows from t_1 to t_2 and from t_2 to t_1 gives

$$\omega_1, \omega_2 \leq 1.$$

This outer bound on the information rate region, referred to as the *max-flow bound*, is depicted in Figure 1.4. Here the rate pair $(1,1)$ is achieved by using network coding at the node u as we have discussed, which implies the achievability of the whole region. Therefore, the max-flow bound is tight.

We now consider the network in Figure 1.2. Consideration of the max-flow at either node v or w gives

$$\omega_1 \leq 1, \tag{1.1}$$

while consideration of the max-flow at node t gives

$$\omega_1 + \omega_2 \leq 2. \tag{1.2}$$

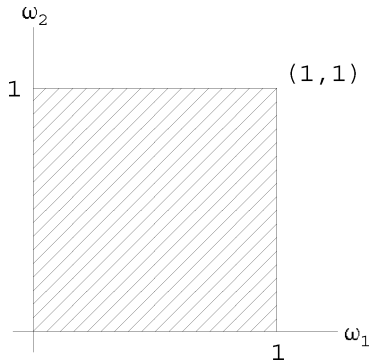

Fig. 1.4 The max-flow bound for the network in Figure 1.1.

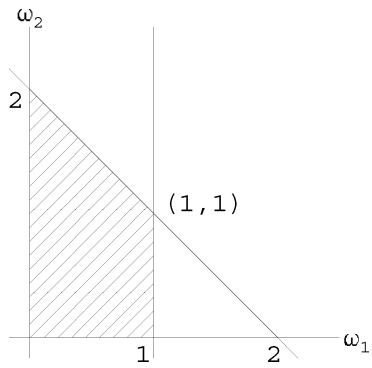

Fig. 1.5 The max-flow bound for the network in Figure 1.2.

Figure 1.5 is an illustration of the region of all (ω_1, ω_2) satisfying these bounds, which constitute the max-flow bound. Comparing with the achievable information rate region shown in Figure 1.3, we see that the max-flow bound is not tight. From these two examples, we see that like superposition coding, the max-flow bound is sometimes but not always tight. Nevertheless, it always gives an outer bound on the information rate region. It has been shown in [6][10] that the max-flow bound is tight for networks with two sink nodes.

2

Network Codes for Acyclic Networks

2.1 Achievable information rate region

In Part I, the capacity of direct transmission from a node to its neighbor is determined by the multiplicity of the channels between them. This is to facilitate the discussion of linear codes. In this section, codes not necessarily linear are considered and we assume that the capacity of a channel can take any positive real number. We, however, continue to allow multiple channels between a pair of nodes to facilitate subsequent comparison with linear codes.

Convention. The following convention applies to every acyclic communication network in this section.

- The set of all nodes and the set of all channels are denoted by V and E, respectively.
- The nodes are ordered in a way such that if there exists a channel from a node i to a node j, then the node i precedes the node j. This is possible by the acyclicity of the network.
- The capacity of a channel e is denoted by R_e.

- An independent information source X_s is generated at a source node s.
- A source node has no input channels.
- The set of all the source nodes in the network is denoted by S, which is a subset of V.
- The set of all sink nodes is denoted by T, where a sink node receives at least one information source[1]. The set of information sources received by a sink node i is denoted by $\beta(i)$.

In the above setup, the decoding requirements are described by the functions $\beta(i), i \in T$. Equivalently, we may think of each information source X_s being multicast to the set of nodes

$$\{i \in T : s \in \beta(i)\}.$$

We now consider a block code with length n. The information source X_s is a random variable which takes values in the set

$$\mathcal{X}_s = \{1, 2, \cdots, \lceil 2^{n\tau_s} \rceil\}$$

according to the uniform distribution. The rate of the information source X_s is τ_s. According to our assumption, the random variables $X_s, s \in S$ are mutually independent.

Definition 2.1. An

$$(n, (\eta_e : e \in E), (\tau_s : s \in S))$$

code on a given communication network is defined by

1) for all source node $s \in S$ and all channel $e \in \mathrm{Out}(s)$, a local encoding mapping

$$\tilde{k}_e : \mathcal{X}_s \to \{1, \cdots, \eta_e\}; \tag{2.1}$$

2) for all node $i \in V \backslash S$ and all channel $e \in \mathrm{Out}(i)$, a local encoding mapping

$$\tilde{k}_e : \prod_{d \in \mathrm{In}(i)} \{1, \cdots, \eta_d\} \to \{1, \cdots, \eta_e\}; \tag{2.2}$$

[1] Since a source node has no input channels, it cannot be a sink node.

3) for all sink node $i \in T$, a decoding mapping

$$g_i : \prod_{d \in \mathrm{In}(i)} \{1 \cdots, \eta_d\} \to \prod_{s \in \beta(i)} \mathcal{X}_s.$$

In a coding session, if a node i precedes a node j, then the encoding mappings $\tilde{k}_e, e \in \mathrm{Out}(i)$ are applied before the encoding mappings $\tilde{k}_e, e \in \mathrm{Out}(j)$. If $e, e' \in \mathrm{Out}(i)$, then $\tilde{k}_e$ and $\tilde{k}_{e'}$ can be applied in any order. Since a node i precedes a node j if there exists a channel from the node i to the node j, a node does not encode until all the necessary information is received on the input channels.

Introduce the notation $X_{S'}$ for $(X_s : s \in S')$, where $S' \subset S$. For all $i \in T$, define

$$\Delta_i = \Pr\left\{\hat{g}_i(X_S) \neq X_{\beta(i)}\right\},$$

where $\hat{g}_i(X_S)$ denotes the value of g_i as a function of X_S. Δ_i is the probability that the set of information sources $X_{\beta(i)}$ is decoded incorrectly at the node i.

In the subsequent discussion, all the logarithms are in the base 2.

Definition 2.2. An information rate tuple

$$\boldsymbol{\omega} = (\omega_s : s \in S),$$

where $\boldsymbol{\omega} \geq 0$ (componentwise), is asymptotically achievable if for any $\epsilon > 0$, there exists for sufficiently large n an

$$(n, (\eta_e : e \in E), (\tau_s : s \in S))$$

code such that

$$n^{-1} \log \eta_e \leq R_e + \epsilon$$

for all $e \in E$, where $n^{-1} \log \eta_e$ is the average bit rate of the code on the channel e,

$$\tau_s \geq \omega_s - \epsilon$$

for all $s \in S$, and

$$\Delta_i \leq \epsilon$$

for all $i \in T$. For brevity, an asymptotically achievable information rate tuple will be referred to as an achievable information rate tuple.

Definition 2.3. The achievable information rate region, denoted by $\mathcal{R}$, is the set of all achievable information rate tuples $\boldsymbol{\omega}$.

Remark 2.4. It follows from the definition of the achievability of an information rate tuple that if $\boldsymbol{\omega}$ is achievable, then $\boldsymbol{\omega}'$ is achievable for all $0 \leq \boldsymbol{\omega}' \leq \boldsymbol{\omega}$. Also, for any sequence of achievable rate tuples $\boldsymbol{\omega}^{(k)}$, $k \geq 1$, it can be proved that

$$\boldsymbol{\omega} = \lim_{k \to \infty} \boldsymbol{\omega}^{(k)},$$

if exists, is also achievable, i.e., $\mathcal{R}$ is closed. It can then be shown by invoking a time-sharing argument that $\mathcal{R}$ is closed and convex.

In this section, we discuss characterizations of the information rate region of a general multi-source network coding problem. Unlike single-source network coding which already has explicit algebraic code constructions, the current understanding of multi-source network coding is quite far from being complete. Specifically, only inner and outer bounds on the achievable information rate region $\mathcal{R}$ are known for *acyclic* networks, and only existence proof of codes by random coding technique is available. The tools we shall use are mainly probabilistic instead of algebraic.

We note that the definition of a network code in this section does not reduce directly to the definitions of a network code in Part I when there is only one information source. It is because in Part I, a network code is defined in a way such that various notions specific to linear codes for a single information source (namely linear broadcast, linear dispersion, and generic network code) can be incorporated. Essentially, the definition of a network code here is the local description of a network code for multicast.

2.2 Inner bound $\mathcal{R}_{in}$

In this section, we discuss an inner bound on the achievable information rate region $\mathcal{R}$ for acyclic networks. We start with some standard definitions and properties of *strong typicality*, a fundamental tool in information theory. For proofs and further details, We refer the reader to [1], [2], [19]. Here, we adopt the convention in [19].

2.2.1 Typical sequences

Consider an information source $\{X_k, k \geq 1\}$ where X_k are i.i.d. with distribution $p(x)$. We use X to denote the generic random variable, $\mathcal{S}_X$ to denote the support of X, and $H(X)$ to denote the common entropy for all X_k, where $H(X) < \infty$. Let $\mathbf{X} = (X_1, X_2, \cdots, X_n)$.

Definition 2.5. The strongly typical set $T^n_{[X]\delta}$ with respect to $p(x)$ is the set of sequences $\mathbf{x} = (x_1, x_2, \cdots, x_n) \in \mathcal{X}^n$ such that $N(x;\mathbf{x}) = 0$ for $x \notin \mathcal{S}_X$, and

$$\sum_x \left| \frac{1}{n} N(x;\mathbf{x}) - p(x) \right| \leq \delta, \tag{2.3}$$

where $N(x;\mathbf{x})$ is the number of occurrences of x in the sequence $\mathbf{x}$, and δ is an arbitrarily small positive real number. The sequences in $T^n_{[X]\delta}$ are called strongly δ-typical sequences.

Theorem 2.6. (Strong asymptotic equipartition property) In the following, η is a small positive quantity such that $\eta \to 0$ as $\delta \to 0$.

1) If $\mathbf{x} \in T^n_{[X]\delta}$, then

$$2^{-n(H(X)+\eta)} \leq p(\mathbf{x}) \leq 2^{-n(H(X)-\eta)}. \tag{2.4}$$

2) For n sufficiently large,

$$\Pr\{\mathbf{X} \in T^n_{[X]\delta}\} > 1 - \delta.$$

3) For n sufficiently large,

$$(1-\delta)2^{n(H(X)-\eta)} \leq |T^n_{[X]\delta}| \leq 2^{n(H(X)+\eta)}. \tag{2.5}$$

Next, we discuss strong joint typicality with respect to a bivariate distribution. Generalization to a multivariate distribution is straightforward.

Consider a bivariate information source $\{(X_k, Y_k), k \geq 1\}$ where (X_k, Y_k) are i.i.d. with distribution $p(x,y)$. We use (X,Y) to denote the pair of generic random variables, and assume that $H(X,Y) < \infty$.

Definition 2.7. The strongly jointly typical set $T^n_{[XY]\delta}$ with respect to $p(x,y)$ is the set of $(\mathbf{x},\mathbf{y}) \in \mathcal{X}^n \times \mathcal{Y}^n$ such that $N(x,y;\mathbf{x},\mathbf{y}) = 0$ for $(x,y) \notin \mathcal{S}_{XY}$, and

$$\sum_x \sum_y \left| \frac{1}{n} N(x,y;\mathbf{x},\mathbf{y}) - p(x,y) \right| \leq \delta, \tag{2.6}$$

where $N(x,y;\mathbf{x},\mathbf{y})$ is the number of occurrences of (x,y) in the pair of sequences $(\mathbf{x},\mathbf{y})$, and δ is an arbitrarily small positive real number. A pair of sequences $(\mathbf{x},\mathbf{y})$ is called strongly jointly δ-typical if it is in $T^n_{[XY]\delta}$.

Strong typicality satisfies the following *consistency* and *preservation* properties.

Theorem 2.8. (Consistency) If $(\mathbf{x},\mathbf{y}) \in T^n_{[XY]\delta}$, then $\mathbf{x} \in T^n_{[X]\delta}$ and $\mathbf{y} \in T^n_{[Y]\delta}$.

Theorem 2.9. (Preservation) Let $Y = f(X)$. If

$$\mathbf{x} = (x_1, x_2, \cdots, x_n) \in T^n_{[X]\delta},$$

then

$$f(\mathbf{x}) = (y_1, y_2, \cdots, y_n) \in T^n_{[Y]\delta}, \tag{2.7}$$

where $y_i = f(x_i)$ for $1 \leq i \leq n$. ([19], Lemma 15.10.)

For a bivariate i.i.d. source $\{(X_k, Y_k)\}$, we have the *strong joint asymptotic equipartition property* (strong JAEP), which can readily be obtained by applying the strong AEP to the source $\{(X_k, Y_k)\}$.

Theorem 2.10. (Strong JAEP) Let

$$(\mathbf{X}, \mathbf{Y}) = ((X_1, Y_1), (X_2, Y_2), \cdots, (X_n, Y_n)),$$

where (X_i, Y_i) are i.i.d. with generic pair of random variables (X, Y). In the following, λ is a small positive quantity such that $\lambda \to 0$ as $\delta \to 0$.

1) If $(\mathbf{x}, \mathbf{y}) \in T^n_{[XY]\delta}$, then

$$2^{-n(H(X,Y)+\lambda)} \leq p(\mathbf{x}, \mathbf{y}) \leq 2^{-n(H(X,Y)-\lambda)}.$$

2) For n sufficiently large,

$$\Pr\{(\mathbf{X}, \mathbf{Y}) \in T^n_{[XY]\delta}\} > 1 - \delta.$$

3) For n sufficiently large,

$$(1 - \delta)2^{n(H(X,Y)-\lambda)} \leq |T^n_{[XY]\delta}| \leq 2^{n(H(X,Y)+\lambda)}.$$

2.2.2 First example

Consider a point-to-point communication system, the simplest possible example of a communication network:

$$V = \{1, a\},\ E = \{1a\},\ S = \{1\},\ T = \{a\},\ \beta(a) = \{1\}.$$

This network is illustrated in Figure 2.1, and we call this network G_1. By the source coding theorem [15], the information rate ω_1 is achievable if and only if $\omega_1 \leq R_{1a}$. The following theorem can be regarded as an alternative form of the direct part of the source coding theorem.

Fig. 2.1 The network G_1 for the first example.

Theorem 2.11. For the network G_1, an information rate ω_1 is achievable if there exists auxiliary random variables Y_1 and U_{1a} such that

$$H(Y_1) > \omega_1 \tag{2.8}$$

$$H(U_{1a}|Y_1) = 0 \tag{2.9}$$

$$H(U_{1a}) < R_{1a} \tag{2.10}$$

$$H(Y_1|U_{1a}) = 0. \tag{2.11}$$

We first note that (2.9) and (2.11) together imply that the random variables Y_1 and U_{1a} determines each other, so we write

$$U_{1a} = u_{1a}(Y_1)$$

and

$$Y_1 = y_1(U_{1a}),$$

which imply

$$Y_1 = y_1(u_{1a}(Y_1)). \tag{2.12}$$

Moreover,

$$H(Y_1) = H(U_{1a}).$$

Then for any ω_1 satisfying (2.8) to (2.11) for some auxiliary random variables Y_1 and U_{1a}, we have

$$R_{1a} > H(U_{1a}) = H(Y_1) > \omega_1,$$

which is essentially the direct part of the source coding theorem except that the inequality is strict here. By invoking the remark following Definition 2.3, we see that the rate

$$R_{1a} = \omega_1$$

is indeed achievable.

We should think of Y_1 and U_{1a} as random variables representing the information source X_1 and the codeword sent on the channel $1a$, respectively. Accordingly, we have (2.8) as the entropy constraint on Y_1, and (2.10) corresponds to the capacity constraint for the channel $1a$.

Proof of Theorem 2.11. Let δ to be a small positive real number to be specified later. For given random variables Y_1 and U_{1a} satisfying (2.8) to (2.11), we construct a random code by the following procedure:

1. Generate $2^{n\omega_1}$ sequences of length n independently according to $p^n(y_1)$.
2. If the message is i, map it to the ith sequence generated in Step 1. Denote this sequence by $\mathbf{y}_1$.
3. If $\mathbf{y}_1 \in T^n_{[Y_1]\delta}$, obtain the sequence

$$\mathbf{u}_{1a} = u_{1a}(\mathbf{y}_1)$$

   (recall the notation $f(\mathbf{x})$ in Theorem 2.9). By Theorem 2.9, $\mathbf{u}_{1a} \in T^n_{[U_{1a}]\delta}$. Otherwise, let $\mathbf{u}_{1a}$ be a constant sequence in $T^n_{[U_{1a}]\delta}$.
4. Output the index of $\mathbf{u}_{1a}$ in $T^n_{[U_{1a}]\delta}$ as the codeword and send on the channel $1a$.
5. At the node b, upon receiving the index of $\mathbf{u}_{1a} \in T^n_{[U_{1a}]\delta}$, recover $\mathbf{u}_{1a}$ and obtain

$$\tilde{\mathbf{y}}_1 = y_1(\mathbf{u}_{1a}).$$

   If $\tilde{\mathbf{y}}_1 = \mathbf{y}_1$ and $\mathbf{y}_1$ is unique among all the sequences generated in Step 1 of the random coding procedure, then the message i can be decoded correctly.

A decoding error is said to occur if the message i is decoded incorrectly. Note that the total number of codewords is upper bounded by

$$|T^n_{[U_{1a}]\delta}| < 2^{n(H(U_{1a})+\eta)}$$

(cf. (2.5)), so that the rate of the code is at most

$$H(U_{1a}) + \eta < R_{1a} + \eta.$$

We now analyze the probability of decoding error of this random code. Consider

$$\begin{aligned}
&\Pr\{\text{decoding error}\} \\
&= \Pr\{\text{decoding error}|\mathbf{y}_1 \notin T^n_{[Y_1]\delta}\}\Pr\{\mathbf{y}_1 \notin T^n_{[Y_1]\delta}\} \\
&\quad + \Pr\{\text{decoding error}|\mathbf{y}_1 \in T^n_{[Y_1]\delta}\}\Pr\{\mathbf{y}_1 \in T^n_{[Y_1]\delta}\} \\
&\le 1 \cdot \Pr\{\mathbf{y}_1 \notin T^n_{[Y_1]\delta}\} + \Pr\{\text{decoding error}|\mathbf{y}_1 \in T^n_{[Y_1]\delta}\} \cdot 1 \\
&= \Pr\{\mathbf{y}_1 \notin T^n_{[Y_1]\delta}\} + \Pr\{\text{decoding error}|\mathbf{y}_1 \in T^n_{[Y_1]\delta}\}.
\end{aligned}$$

By the strong AEP,

$$\Pr\{\mathbf{y}_1 \notin T^n_{[Y_1]\delta}\} \to 0$$

as $n \to \infty$. So it remains to show that

$$\Pr\{\text{decoding error}|\mathbf{y}_1 \in T^n_{[Y_1]\delta}\} \to 0$$

as $n \to \infty$ with an appropriate choice of δ. Toward this end, we observe that if $\mathbf{y}_1 \in T^n_{[Y_1]\delta}$, then

$$\mathbf{u}_{1a} = u_{1a}(\mathbf{y}_1)$$

(instead of being a constant sequence in $T^n_{[U_{1a}]\delta}$), so that

$$\tilde{\mathbf{y}}_1 = y_1(\mathbf{u}_{1a}) = y_1(u_{1a}(\mathbf{y}_1)).$$

Then from (2.12), we see that

$$\tilde{\mathbf{y}}_1 = \mathbf{y}_1.$$

In other words, if $\mathbf{y}_1 \in T^n_{[Y_1]\delta}$, a decoding error occurs if and only if the sequence $\mathbf{y}_1$ is drawn more than once in Step 1. Thus,

$$\begin{aligned}
&\Pr\{\text{decoding error}|\mathbf{y}_1 \in T^n_{[Y_1]\delta}\} \\
&= \Pr\{\mathbf{y}_1 \text{ drawn more than once}|\mathbf{y}_1 \in T^n_{[Y_1]\delta}\} \\
&= \Pr\left\{\cup_{j\neq i}\{\text{obtain } \mathbf{y}_1 \text{ in the } j\text{th drawing}|\mathbf{y}_1 \in T^n_{[Y_1]\delta}\}\right\}
\end{aligned}$$

$$
\begin{aligned}
&\leq \sum_{j \neq i} \Pr\{\text{obtain } \mathbf{y}_1 \text{ in the } j\text{th drawing} | \mathbf{y}_1 \in T^n_{[Y_1]\delta}\} \\
&< 2^{n\omega_1} \cdot \Pr\{\text{obtain } \mathbf{y}_1 \text{ in any drawing} | \mathbf{y}_1 \in T^n_{[Y_1]\delta}\} \\
&< 2^{n\omega_1} \cdot 2^{-n(H(U_{1a})-\eta)} \\
&= 2^{-n(H(U_{1a})-\omega_1-\eta)} \\
&= 2^{-n(H(Y_1)-\omega_1-\eta)},
\end{aligned}
$$

where we have invoked the strong AEP in the last inequality. Since $H(Y_1) > \omega_1$ and $\eta \to 0$ as $\delta \to 0$, by taking δ to be sufficiently small, we have $H(Y_1) - \omega_1 - \eta > 0$, and hence

$$\Pr\{\text{decoding error} | \mathbf{y}_1 \in T^n_{[Y_1]\delta}\} \to 0$$

as $n \to \infty$.

It appears that Theorem 2.11 only complicates the direct part of the source coding theorem, but as we shall see, it actually prepares us to obtain a characterization of the achievable information rate region for more general networks.

2.2.3 Second example

In the next section, we shall state without proof an inner bound on the achievable information rate region $\mathcal{R}$ for a general acyclic network. We already have proved a special case of this inner bound in Theorem 2.11 for a point-to-point communication system. In this section, we prove this inner bound for another network considerably more complicated than the one in the last section. Although this network is still far from being general, the proof of the inner bound for this network contains all the essential ingredients. Besides, the ideas are more transparent without the overwhelming notation in the general proof.

The second network we consider here is the network in Figure 2.2 with the following specification:

$$
\begin{aligned}
&V = \{1, 2, a, b, c, d\},\ E = \{1a, 2b, ab, ac, bc, bd, cd\} \\
&S = \{1, 2\},\ T = \{c, d\},\ \beta(c) = \{1\},\ \beta(d) = \{1, 2\}.
\end{aligned}
$$

Call this network G_2.

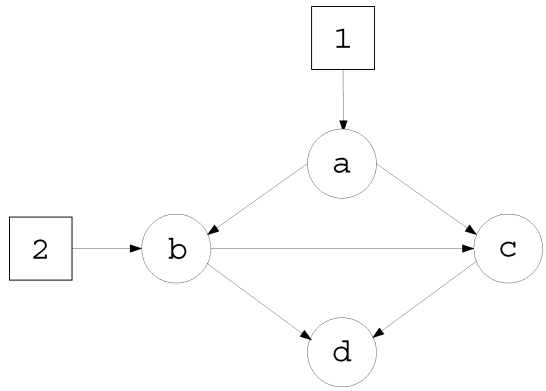

Fig. 2.2 The network G_2 for the second example.

For the network G_2, we first make the observation that the source nodes 1 and 2 each has only one output channel. By the source coding theorem, if either $R_{1a} < \omega_1$ or $R_{2b} < \omega_2$, the sink node d cannot possibly receive both X_1 and X_2. Therefore, in order to make the problem meaningful, we make the assumptions that $R_{1a} \geq \omega_1$ and $R_{2b} \geq \omega_2$, so that we can regard X_1 and X_2 as being directly available to the nodes a and b, respectively.

Theorem 2.12. For the network G_2, an information rate pair (ω_1, ω_2) is achievable if there exist auxiliary random variables $Y_s, s \in S$ and $U_e, e \in E$ such that

$$H(Y_1, Y_2) = H(Y_1) + H(Y_2) \tag{2.13}$$

$$H(Y_s) > \omega_s, \quad s \in S \tag{2.14}$$

$$H(U_{ab}, U_{ac}|Y_1) = 0 \tag{2.15}$$

$$H(U_{bc}, U_{bd}|Y_2, U_{ab}) = 0 \tag{2.16}$$

$$H(U_{cd}|U_{ac}, U_{bc}) = 0 \tag{2.17}$$

$$H(U_e) < R_e, \quad e \in E \tag{2.18}$$

$$H(Y_1|U_{ac}, U_{bc}) = 0 \tag{2.19}$$

$$H(Y_1, Y_2|U_{bd}, U_{cd}) = 0. \tag{2.20}$$

The interpretations of (2.13) to (2.20) are as follows. Similar to our discussion on the network in the last section, Y_s and U_e are random variables representing the information source X_s and the codeword sent on the channel e, respectively. The equality in (2.13) says that the information sources 1 and 2 are independent. The inequality (2.14) is the entropy constraint on the auxiliary random variable Y_s. The equality (2.15) says that the codewords sent on the channels ab and ac depend only on the information source X_1. The equality (2.16) says that the codewords sent on the channels bc and bd depend only on the information source X_2 and the codeword sent on the channel ab. The equality (2.17) says that the codeword sent on the channel cd depends only on the codeword sent on the channels ac and bc. The inequality (2.18) is the capacity constraint for the channel e. The equality (2.19) says that the information source 1 can be recovered (at the sink node c) from the codewords sent on the channels ac and bc, and finally the equality (2.20) says that both the information sources X_1 and X_2 can be recovered (at the sink node d) from the codewords sent on the channels bd and cd.

From (2.15), we see that U_{ab} and U_{ac} are both functions of Y_1. Thus we write

$$U_{ab} = u_{ab}(Y_1) \tag{2.21}$$

and

$$U_{ac} = u_{ac}(Y_1). \tag{2.22}$$

In the same way, from (2.16), (2.17), (2.19), and (2.20), we write

$$U_{bc} = u_{bc}(Y_2, U_{ab}) \tag{2.23}$$

$$U_{bd} = u_{bd}(Y_2, U_{ab}) \tag{2.24}$$

$$U_{cd} = u_{cd}(U_{ac}, U_{bc}) \tag{2.25}$$

$$Y_1 = y_1^{(c)}(U_{ac}, U_{bc}) \tag{2.26}$$

$$Y_1 = y_1^{(d)}(U_{bd}, U_{cd}) \tag{2.27}$$

$$Y_2 = y_2^{(d)}(U_{bd}, U_{cd}). \tag{2.28}$$

In (2.26) to (2.28), the superscript denotes the sink node with which the function is associated.

Proof of Theorem 2.12. Let δ to be a small positive real number to be specified later. For given random variables $Y_s, s \in S$ and $U_e, e \in E$ satisfying (2.13) to (2.20), we construct a random code by the following procedure:

1. For the information source j $(= 1, 2)$,
   a) Generate $2^{n\omega_j}$ sequences of length n independently according to $p^n(y_j)$.
   b) If the message is i_j, map it to the i_j-th sequence generated in Step 1a). Call this sequence $\mathbf{y}_j$.
2. If $\mathbf{y}_1 \in T^n_{[Y_1]}$, obtain the sequences
$$\mathbf{u}_{ab} = u_{ab}(\mathbf{y}_1) \in T^n_{[U_{ab}]\delta}$$
and
$$\mathbf{u}_{ac} = u_{ac}(\mathbf{y}_1) \in T^n_{[U_{ac}]\delta}$$
(cf. (2.21) for the definition of $u_{ab}(\cdot)$, etc, and Theorem 2.9 for the notation $f(\mathbf{x})$). Here, $u_{ab}(\mathbf{y}_1) \in T^n_{[U_{ab}]\delta}$ and $u_{ac}(\mathbf{y}_1) \in T^n_{[U_{ac}]\delta}$ as follow from Theorem 2.8. Otherwise, let $\mathbf{u}_{ab}$ and $\mathbf{u}_{ac}$ be constant sequences in $T^n_{[U_{ab}]\delta}$ and $T^n_{[U_{ac}]\delta}$, respectively.
3. Output the indices of $\mathbf{u}_{ab}$ in $T^n_{[U_{ab}]\delta}$ and $\mathbf{u}_{ac}$ in $T^n_{[U_{ac}]\delta}$ as codewords and send on the channels ab and ac, respectively.
4. If $(\mathbf{y}_2, \mathbf{u}_{ab}) \in T^n_{[Y_2U_{ab}]\delta}$, obtain the sequences
$$\mathbf{u}_{bc} = u_{bc}(\mathbf{y}_2, \mathbf{u}_{ab}) \in T^n_{[U_{bc}]}$$
and
$$\mathbf{u}_{bd} = u_{bd}(\mathbf{y}_2, \mathbf{u}_{ab}) \in T^n_{[U_{bd}]}.$$
Otherwise, let $\mathbf{u}_{bc}$ and $\mathbf{u}_{bd}$ be constant sequences in $T^n_{[U_{bc}]\delta}$ and $T^n_{[U_{bd}]\delta}$, respectively.
5. Output the indices of $\mathbf{u}_{bc}$ in $T^n_{[U_{bc}]\delta}$ and $\mathbf{u}_{bd}$ in $T^n_{[U_{bd}]\delta}$ as codewords and send on the channels bc and bd, respectively.
6. If $(\mathbf{u}_{ac}, \mathbf{u}_{bc}) \in T^n_{[U_{ab}U_{bc}]\delta}$, obtain the sequence
$$\mathbf{u}_{cd} = u_{cd}(\mathbf{u}_{ab}, \mathbf{u}_{bc}) \in T^n_{[U_{cd}]}.$$
Otherwise, let $\mathbf{u}_{cd}$ be a constant sequence in $T^n_{[U_{cd}]\delta}$.

7. Output the index of $\mathbf{u}_{cd}$ in $T^n_{[U_{cd}]\delta}$ as the codeword and send on the channel cd.
8. At the node c, upon receiving the indices of $\mathbf{u}_{ac} \in T^n_{[U_{ac}]\delta}$ and $\mathbf{u}_{bc} \in T^n_{[U_{bc}]\delta}$, $\mathbf{u}_{ac}$ and $\mathbf{u}_{bc}$ can be recovered. Then obtain

$$\tilde{\mathbf{y}}_1^{(c)} = y_1^{(c)}(\mathbf{u}_{ac}, \mathbf{u}_{bc}). \tag{2.29}$$

   If $\tilde{\mathbf{y}}_1^{(c)} = \mathbf{y}_1$ and $\mathbf{y}_1$ is unique among all the sequences generated in Step 1a) for $j = 1$, then the message i_1 can be decoded correctly.
9. At the node d, upon receiving the indices of $\mathbf{u}_{bd} \in T^n_{[U_{bd}]\delta}$ and $\mathbf{u}_{cd} \in T^n_{[U_{cd}]\delta}$, $\mathbf{u}_{bd}$ and $\mathbf{u}_{cd}$ can be recovered. For $j = 1, 2$, obtain

$$\tilde{\mathbf{y}}_j^{(d)} = y_j^{(d)}(\mathbf{u}_{bd}, \mathbf{u}_{cd}).$$

   If $\tilde{\mathbf{y}}_j^{(d)} = \mathbf{y}_j$ and $\mathbf{y}_j$ is unique among all the sequences generated in Step 1a), then the message i_j can be decoded correctly.

If either i_1 is decoded incorrectly at the node c or (i_1, i_2) is decoded incorrectly at the node d, we say that a decoding error occurs. Note that for each channel $e \in E$, the total number of codewords is upper bounded by

$$|T^n_{[U_e]\delta}| < 2^{nH(U_e)+\eta}$$

(cf. (2.5)), so that the rate on the channel e is at most

$$H(U_e) + \eta < R_e + \eta.$$

We now analyze the probability of decoding error of this random code. Analogous to the proof of Theorem 2.11 in the last section, we have

$$\begin{aligned}&\Pr\{\text{decoding error}\}\\ &\le \Pr\{(\mathbf{y}_1, \mathbf{y}_2) \notin T^n_{[Y_1Y_2]\delta}\} + \Pr\{\text{decoding error}|(\mathbf{y}_1, \mathbf{y}_2) \in T^n_{[Y_1]\delta}\}.\end{aligned}$$

Since the pair of sequence $(\mathbf{y}_1, \mathbf{y}_2)$ is generated according to

$$p^n(y_1)p^n(y_2) = p^n(y_1, y_2),$$

by the strong JAEP,

$$\Pr\{(\mathbf{y}_1, \mathbf{y}_2) \notin T^n_{[Y_1Y_2]\delta}\} \to 0$$

as $n \to \infty$, so it suffices to show that

$$\Pr\{\text{decoding error}|(\mathbf{y}_1, \mathbf{y}_2) \in T^n_{[Y_1Y_2]\delta}\} \to 0$$

as $n \to \infty$ with an appropriate choice of δ. Toward this end, we analyze the random coding procedure when $(\mathbf{y}_1, \mathbf{y}_2) \in T^n_{[Y_1Y_2]\delta}$:

- By Theorem 2.8, we have $\mathbf{y}_j \in T^n_{[Y_j]\delta}$, $j = 1, 2$.
- In Step 2, since $\mathbf{y}_1 \in T^n_{[Y_1]\delta}$, we have

$$\mathbf{u}_{ab} = u_{ab}(\mathbf{y}_1) \tag{2.30}$$

  (instead of a constant sequence in $T^n_{[U_{ab}]\delta}$) and

$$\mathbf{u}_{ac} = u_{ac}(\mathbf{y}_1). \tag{2.31}$$

- In Step 4, by (2.30), we have

$$(\mathbf{y}_2, \mathbf{u}_{ab}) = (\mathbf{y}_2, u_{ab}(\mathbf{y}_1)).$$

  Since $(\mathbf{y}_1, \mathbf{y}_2) \in T^n_{[Y_1Y_2]\delta}$,

$$(\mathbf{y}_2, u_{ab}(\mathbf{y}_1)) \in T^n_{[Y_2U_{ab}]\delta}$$

  by Theorem 2.9. Therefore,

$$\mathbf{u}_{bc} = u_{bc}(\mathbf{y}_2, \mathbf{u}_{ab}) \tag{2.32}$$

  and

$$\mathbf{u}_{bd} = u_{bd}(\mathbf{y}_2, \mathbf{u}_{ab}). \tag{2.33}$$

- In Step 6, by applying (2.31), (2.32) and (2.30), we have

$$\begin{aligned}(\mathbf{u}_{ac}, \mathbf{u}_{bc}) &= (u_{ac}(\mathbf{y}_1), u_{bc}(\mathbf{y}_2, \mathbf{u}_{ab})) \\ &= (u_{ac}(\mathbf{y}_1), u_{bc}(\mathbf{y}_2, u_{ab}(\mathbf{y}_1))). \end{aligned} \tag{2.34}$$

  Again, since $(\mathbf{y}_1, \mathbf{y}_2) \in T^n_{[Y_1Y_2]\delta}$,

$$(\mathbf{u}_{ac}, \mathbf{u}_{bc}) \in T^n_{[U_{ac}U_{bc}]\delta}$$

  by Theorem 2.9. Therefore,

$$\mathbf{u}_{cd} = u_{cd}(\mathbf{u}_{ac}, \mathbf{u}_{bc}).$$

- By (2.26), (2.22), (2.23), and (2.21), we can write

$$\begin{aligned}Y_1 &= y_1^{(c)}(U_{ac}, U_{bc}) \\ &= y_1^{(c)}(u_{ac}(Y_1), u_{bc}(Y_2, U_{ab})) \\ &= y_1^{(c)}(u_{ac}(Y_1), u_{bc}(Y_2, u_{ab}(Y_1))). \end{aligned} \tag{2.35}$$

  On the other hand, from (2.29) and (2.34), we have

$$\begin{aligned}\tilde{\mathbf{y}}_1^{(c)} &= y_1^{(c)}(\mathbf{u}_{ac}, \mathbf{u}_{bc}) \\ &= y_1^{(c)}(u_{ac}(\mathbf{y}_1), u_{bc}(\mathbf{y}_2, u_{ab}(\mathbf{y}_1))). \end{aligned} \tag{2.36}$$

  A comparison of (2.35) and (2.36) reveals that

$$\tilde{\mathbf{y}}_1^{(c)} = \mathbf{y}_1. \tag{2.37}$$

  Similarly, it can be shown that

$$\tilde{\mathbf{y}}_1^{(d)} = \mathbf{y}_1. \tag{2.38}$$

  and

$$\tilde{\mathbf{y}}_2^{(d)} = \mathbf{y}_2. \tag{2.39}$$

In conclusion, whenever $(\mathbf{y}_1, \mathbf{y}_2) \in T^n_{[Y_1Y_2]\delta}$, (2.37) to (2.39) hold. By the strong AEP,

$$\Pr\{(\mathbf{y}_1, \mathbf{y}_2) \in T^n_{[Y_1Y_2]\delta}\} \to 1$$

as $n \to \infty$. Therefore, if $(\mathbf{y}_1, \mathbf{y}_2) \in T^n_{[Y_1Y_2]\delta}$, a decoding error occurs if and only if either $\mathbf{y}_1$ or $\mathbf{y}_2$ is drawn more than once in Step 1a).

By means of an argument similar to the one in the proof of Theorem 2.11, it can be shown that

$$\Pr\{\text{decoding error}|(\mathbf{y}_1, \mathbf{y}_2) \in T^n_{[Y_1Y_2]\delta}\} \to 0$$

as $n \to \infty$ with an appropriate choice of δ. The details are omitted here.

2.2.4 General acyclic networks

In this section, we present an inner bound $\mathcal{R}_{in}$ on the information rate region for a general acyclic network. The reader should have no problem understanding the meaning of $\mathcal{R}_{in}$ after studying the special cases in the previous two sections. In the sequel, we will use the abbreviations Y_S, $U_{\mathrm{In}(i)}$ respectively for $\{Y_s : s \in S\}$, $\{U_e : e \in \mathrm{In}(i)\}$, etc.

Definition 2.13. Let $\mathcal{R}'$ be the set of all information rate tuples $\boldsymbol{\omega}$ such that there exist auxiliary random variables $Y_s, s \in S$ and $U_e, e \in E$ which satisfy the following conditions:

$$H(Y_S) = \sum_{s \in S} H(Y_s) \tag{2.40}$$

$$H(Y_s) > \omega_s, \quad s \in S \tag{2.41}$$

$$H(U_{\mathrm{Out}(s)}|Y_s) = 0, \quad s \in S \tag{2.42}$$

$$H(U_{\mathrm{Out}(i)}|U_{\mathrm{In}(i)}) = 0, \quad i \in V \backslash S \tag{2.43}$$

$$H(U_e) < R_e, \quad e \in E \tag{2.44}$$

$$H(Y_{\beta(i)}|U_{\mathrm{In}(i)}) = 0, \quad i \in T. \tag{2.45}$$

Theorem 2.14. $\mathcal{R}' \subset \mathcal{R}$.

The proof of Theorem 2.14 involves a set of techniques originally developed in [20] and [16]. The proof of Theorem 2.12 in the last section, though a special case of Theorem 2.16 here, contains all the essential ingredients necessary for proving Theorem 2.14.

Definition 2.15. Let $\mathcal{R}_{in} = \overline{\mathrm{con}}(\mathcal{R}')$, the convex closure of $\mathcal{R}'$.

Theorem 2.16. $\mathcal{R}_{in} \subset \mathcal{R}$.

Theorem 2.16 can readily be obtained from Theorem 2.14 as a corollary by invoking the remark following Definition 2.3. Specifically, by taking the convex closure on both sides in

$$\mathcal{R}' \subset \mathcal{R},$$

we have

$$\overline{\text{con}}(\mathcal{R}') \subset \overline{\text{con}}(\mathcal{R}) = \mathcal{R}.$$

For a complete proof of Theorem 2.16, we refer the reader to [16] and [19], Ch. 15[2]. The inner bound proved in [16] is for zero-error variable-length network codes.

2.2.5 $\mathcal{R}_{in}$ recasted

In this section, $\mathcal{R}_{in}$ will be recasted in the framework of information inequalities developed in [18]. As we shall see, this alternative characterization of $\mathcal{R}_{in}$, developed in [20] and [16], enables the region to be described on the same footing for different multi-source network coding problems.

Let $\mathcal{N}$ be a collection of discrete random variables whose joint distribution is unspecified, and let

$$\mathcal{Q}_{\mathcal{N}} = 2^{\mathcal{N}} \backslash \{\emptyset\},$$

the set of all nonempty subsets of random variables in $\mathcal{N}$. Then

$$|\mathcal{Q}_{\mathcal{N}}| = 2^{|\mathcal{N}|} - 1.$$

Let $\mathcal{H}_{\mathcal{N}}$ be the $|\mathcal{Q}_{\mathcal{N}}|$-dimensional Euclidean space with the coordinates labeled by $h_A, A \in \mathcal{Q}_{\mathcal{N}}$. We will refer to $\mathcal{H}_{\mathcal{N}}$ as the entropy space for the set of random variables $\mathcal{N}$. A vector

$$\mathbf{h} = (h_A : A \in \mathcal{Q}_{\mathcal{N}}) \in \mathcal{H}_{\mathcal{N}} \tag{2.46}$$

[2] The proof given in Section 2.2.3 is a simplified version of the proofs in [19] and [16].

is said to be an *entropy function* if there exists a joint distribution for $(Z : Z \in \mathcal{N})$ such that

$$h_A = H(Z : Z \in A)$$

for all $A \in \mathcal{Q}_{\mathcal{N}}$. We then define the region

$$\Gamma^*_{\mathcal{N}} = \{\mathbf{h} \in \mathcal{H}_{\mathcal{N}} : \mathbf{h} \text{ is an entropy function}\}.$$

To simplify notation in the sequel, for any nonempty $A, A' \in \mathcal{Q}_{\mathcal{N}}$, we define

$$h_{A|A'} = h_{AA'} - h_{A'}, \tag{2.47}$$

where we use juxtaposition to denote the union of two sets. In using the above notation, we do not distinguish elements and singletons of $\mathcal{N}$, i.e., for a random variable $Z \in \mathcal{N}$, h_Z is the same as $h_{\{Z\}}$. Note that (2.47) corresponds to the information-theoretic identity

$$H(A|A') = H(AA') - H(A').$$

To describe $\mathcal{R}_{in}$ in terms of the above framework, we let

$$\mathcal{N} = \{Y_s : s \in S; U_e : e \in E\}.$$

Observe that the constraints (2.40) to (2.45) in the definition of $\mathcal{R}'$ correspond to the following constraints in $\mathcal{H}_{\mathcal{N}}$, respectively:

$$h_{Y_S} = \sum_{s \in S} h_{Y_s} \tag{2.48}$$

$$h_{Y_s} > \omega_s, \quad s \in S \tag{2.49}$$

$$h_{U_{\mathrm{Out}(s)}|Y_s} = 0, \quad s \in S \tag{2.50}$$

$$h_{U_{\mathrm{Out}(i)}|U_{\mathrm{In}(i)}} = 0, \quad i \in V \backslash S \tag{2.51}$$

$$h_{U_e} < R_e, \quad e \in E \tag{2.52}$$

$$h_{Y_{\beta(i)}|U_{\mathrm{In}(i)}} = 0, \quad i \in T. \tag{2.53}$$

Then we have the following alternative definition of $\mathcal{R}'$.

Definition 2.17. Let $\mathcal{R}'$ be the set of all information rate tuples $\boldsymbol{\omega}$ such that there exists $\mathbf{h} \in \Gamma^*_{\mathcal{N}}$ which satisfies (2.48) to (2.53).

Although the original definition of $\mathcal{R}'$ as given in Definition 2.13 is more intuitive, the region so defined appears to be totally different from one problem to another problem. On the other hand, the alternative definition of $\mathcal{R}'$ above enables the region to be described on the same footing for all cases. Moreover, if $\tilde{\Gamma}_{\mathcal{N}}$ is an explicit inner bound on $\Gamma^*_{\mathcal{N}}$, upon replacing $\Gamma^*_{\mathcal{N}}$ by $\tilde{\Gamma}_{\mathcal{N}}$ in the above definition of $\mathcal{R}'$, we immediately obtain an explicit inner bound on $\mathcal{R}_{in}$ for all cases. We shall see further advantage of this alternative definition when we discuss an explicit outer bound on $\mathcal{R}$ in the next section.

2.3 Outer bound $\mathcal{R}_{out}$

In this section, we prove an outer bound $\mathcal{R}_{out}$ on $\mathcal{R}$. This outer bound is in terms of $\overline{\Gamma}^*_{\mathcal{N}}$, the closure of $\Gamma^*_{\mathcal{N}}$.

Definition 2.18. Let $\mathcal{R}_{out}$ be the set of all information rate tuples $\boldsymbol{\omega}$ such that there exists $\mathbf{h} \in \overline{\Gamma}^*_{\mathcal{N}}$ which satisfies the following constraints:

$$h_{Y_S} = \sum_{s \in S} h_{Y_s} \tag{2.54}$$

$$h_{Y_s} \geq \omega_s, \quad s \in S \tag{2.55}$$

$$h_{U_{\mathrm{Out}(s)}|Y_s} = 0, \quad s \in S \tag{2.56}$$

$$h_{U_{\mathrm{Out}(i)}|U_{\mathrm{In}(i)}} = 0, \quad i \in V \backslash S \tag{2.57}$$

$$h_{U_e} \leq R_e, \quad e \in E \tag{2.58}$$

$$h_{Y_{\beta(i)}|U_{\mathrm{In}(i)}} = 0, \quad i \in T. \tag{2.59}$$

The definition of $\mathcal{R}_{out}$ is the same as the alternative definition of $\mathcal{R}'$ (Definition 2.17) except that

1. $\Gamma^*_{\mathcal{N}}$ is replaced by $\overline{\Gamma}^*_{\mathcal{N}}$.
2. The inequalities in (2.49) and (2.52) are strict, while the inequalities in (2.55) and (2.58) are nonstrict.

From the definitions of $\mathcal{R}'$ and $\mathcal{R}_{out}$, it is clear that

$$\mathcal{R}' \subset \mathcal{R}_{out}. \tag{2.60}$$

It is also easy to verify that the convexity of $\overline{\Gamma}^*_{\mathcal{N}}$ ([19], Theorem 14.5) implies the convexity of $\mathcal{R}_{out}$. Then upon taking convex closure in (2.60), we see that

$$\mathcal{R}_{in} = \overline{\text{con}}(\mathcal{R}') \subset \overline{\text{con}}(\mathcal{R}_{out}) = \mathcal{R}_{out},$$

where the last equality follows because $\mathcal{R}_{out}$ is close and convex. However, it is not apparent that the two regions $\mathcal{R}_{in}$ and $\mathcal{R}_{out}$ coincide in general. This will be further discussed in the next section. We first prove that $\mathcal{R}_{out}$ is indeed an outer bound on $\mathcal{R}$.

Theorem 2.19. $\mathcal{R} \subset \mathcal{R}_{out}$.

Proof. Let $\boldsymbol{\omega}$ be an achievable information rate tuple and n be a sufficiently large integer. Then for any $\epsilon > 0$, there exists an

$$(n, (\eta_e : e \in E), (\tau_s : s \in S))$$

code on the network such that

$$n^{-1} \log \eta_e \leq R_e + \epsilon \tag{2.61}$$

for all $e \in E$,

$$\tau_s \geq \omega_s - \epsilon \tag{2.62}$$

for all $s \in S$, and

$$\Delta_i \leq \epsilon \tag{2.63}$$

for all $i \in T$.

We consider such a code for a fixed ϵ and a sufficiently large n. Since the information sources $X_s, s \in S$ are mutually independent, we have

$$H(X_S) = \sum_{s \in S} H(X_s). \tag{2.64}$$

For all $s \in S$, from (2.62),

$$H(X_s) = \log |\mathcal{X}_s| = \log \lceil 2^{n\tau_s} \rceil \geq n\tau_s \geq n(\omega_s - \epsilon). \tag{2.65}$$

For $e \in E$, let U_e be the codeword sent on the channel e. For all $s \in S$ and $e \in \mathrm{Out}(s)$, since U_e is a function of the information source X_s,

$$H(U_{\mathrm{Out}(s)}|X_s) = 0. \tag{2.66}$$

Similarly, for all $i \in V \backslash S$,

$$H(U_{\mathrm{Out}(i)}|U_{\mathrm{In}(i)}) = 0. \tag{2.67}$$

From (2.1), (2.2), and (2.61), for all $e \in E$,

$$H(U_e) \leq \log |U_e| = \log(\eta_e + 1) \leq n(R_e + 2\epsilon). \tag{2.68}$$

For $i \in T$, by Fano's inequality (cf. [19], Corollary 2.48), we have

$$\begin{aligned} H(X_{\beta(i)}|U_{\mathrm{In}(i)}) &\leq 1 + \Delta_i \log \left(\prod_{s \in \beta(i)} |\mathcal{X}_s| \right) \\ &= 1 + \Delta_i H(X_{\beta(i)}) \qquad (2.69) \\ &\leq 1 + \epsilon H(X_{\beta(i)}), \qquad (2.70) \end{aligned}$$

where (2.69) follows because X_s distributes uniformly on $\mathcal{X}_s$ and X_s, $s \in S$ are mutually independent, and (2.70) follows from (2.63). Then

$$\begin{aligned} H(X_{\beta(i)}) &= I(X_{\beta(i)}; U_{\mathrm{In}(i)}) + H(X_{\beta(i)}|U_{\mathrm{In}(i)}) \\ &\overset{a)}{\leq} I(X_{\beta(i)}; U_{\mathrm{In}(i)}) + 1 + \epsilon H(X_{\beta(i)}) \\ &\leq H(U_{\mathrm{In}(i)}) + 1 + \epsilon H(X_{\beta(i)}) \\ &\overset{b)}{\leq} \left(\sum_{e \in \mathrm{In}(i)} \log \eta_e \right) + 1 + \epsilon H(X_{\beta(i)}) \\ &\overset{c)}{\leq} \left(\sum_{e \in \mathrm{In}(i)} n(R_e + \epsilon) \right) + 1 + \epsilon H(X_{\beta(i)}), \qquad (2.71) \end{aligned}$$

where

a) follows from (2.70);
b) follows from $H(Z) \leq \log |\mathcal{Z}|$, cf. [19], Theorem 2.43;
c) follows from (2.61).

Rearranging the terms in (2.71), we obtain

$$\begin{aligned} H(X_{\beta(i)}) &\leq \frac{n}{1-\epsilon}\left(\sum_{e\in\mathrm{In}(i)}(R_e+\epsilon)+\frac{1}{n}\right) \\ &< 2n\sum_{e\in\mathrm{In}(i)}(R_e+\epsilon) \end{aligned} \tag{2.72}$$

for sufficiently small ϵ and sufficiently large n. Substituting (2.72) into (2.70), we have

$$\begin{aligned} H(X_{\beta(i)}|U_{\mathrm{In}(i)}) &< n\left(\frac{1}{n}+2\epsilon\sum_{e\in\mathrm{In}(i)}(R_e+\epsilon)\right) \\ &= n\phi_i(n,\epsilon), \end{aligned} \tag{2.73}$$

where

$$\phi_i(n,\epsilon)=\left(\frac{1}{n}+2\epsilon\sum_{e\in\mathrm{In}(i)}(R_e+\epsilon)\right)\to 0$$

as $n\to\infty$ and then $\epsilon\to 0$. Thus for this code, from (2.64), (2.65), (2.67), (2.68), and (2.73), we have

$$H(X_S)=\sum_{s\in S}H(X_s) \tag{2.74}$$

$$H(X_s)\geq n(\omega_s-\epsilon),\quad s\in S \tag{2.75}$$

$$H(U_{\mathrm{Out}(s)}|X_s)=0,\quad s\in S \tag{2.76}$$

$$H(U_{\mathrm{Out}(i)}|U_{\mathrm{In}(i)})=0,\quad i\in V\backslash S \tag{2.77}$$

$$H(U_e)\leq n(R_e+2\epsilon),\quad e\in E \tag{2.78}$$

$$H(X_{\beta(i)}|U_{\mathrm{In}(i)})\leq n\phi_i(n,\epsilon),\quad i\in T. \tag{2.79}$$

We note the one-to-one correspondence between (2.74) to (2.79) and (2.54) to (2.59). By letting $Y_s=X_s$ for all $s\in S$, we see that there exists $\mathbf{h}\in\Gamma^*_{\mathcal{N}}$ such that

$$h_{Y_S}=\sum_{s\in S}h_{Y_s} \tag{2.80}$$

$$h_{Y_s}\geq n(\omega_s-\epsilon),\quad s\in S \tag{2.81}$$

$$h_{U_{\mathrm{Out}(s)}|Y_s} = 0, \quad s \in S \tag{2.82}$$

$$h_{U_{\mathrm{Out}(i)}|U_{\mathrm{In}(i)}} = 0, \quad i \in V \backslash S \tag{2.83}$$

$$h_{U_e} \leq n(R_e + 2\epsilon), \quad e \in E \tag{2.84}$$

$$h_{Y_{\beta(i)}|U_{\mathrm{In}(i)}} \leq n\phi_i(n, \epsilon), \quad i \in T. \tag{2.85}$$

By Theorem 14.5 in [19], $\overline{\Gamma}^*_{\mathcal{N}}$ is a convex cone. Therefore, if $\mathbf{h} \in \Gamma^*_{\mathcal{N}}$, then $n^{-1}\mathbf{h} \in \overline{\Gamma}^*_{\mathcal{N}}$. Dividing (2.80) through (2.85) by n and replacing $n^{-1}\mathbf{h}$ by $\mathbf{h}$, we see that there exists $\mathbf{h} \in \overline{\Gamma}^*_{\mathcal{N}}$ such that

$$\begin{aligned}
h_{Y_S} &= \sum_{s \in S} h_{Y_s} \\
h_{Y_s} &\geq \omega_s - \epsilon, \quad s \in S \\
h_{U_{\mathrm{Out}(s)}|Y_s} &= 0, \quad s \in S \\
h_{U_{\mathrm{Out}(i)}|U_{\mathrm{In}(i)}} &= 0, \quad i \in V \backslash S \\
h_{U_e} &\leq R_e + 2\epsilon, \quad e \in E \\
h_{Y_{\beta(i)}|U_{\mathrm{In}(i)}} &\leq \phi_i(n, \epsilon), \quad i \in T.
\end{aligned}$$

We then let $n \to \infty$ and then $\epsilon \to 0$ to conclude that there exists $\mathbf{h} \in \overline{\Gamma}^*_{\mathcal{N}}$ which satisfies (2.54) to (2.59). Hence, $\mathcal{R} \subset \mathcal{R}_{out}$, and the theorem is proved. □

2.4 $\mathcal{R}_{LP}$ – An explicit outer bound

In Section 2.2.5, we stated the inner bound $\mathcal{R}_{in}$ on $\mathcal{R}$ in terms of $\Gamma^*_{\mathcal{N}}$, and in Section 2.3, we proved the outer bound $\mathcal{R}_{out}$ on $\mathcal{R}$ in terms of $\overline{\Gamma}^*_{\mathcal{N}}$. So far, there exists no full characterization of either $\Gamma^*_{\mathcal{N}}$ or $\overline{\Gamma}^*_{\mathcal{N}}$. Therefore, these bounds cannot be evaluated explicitly. In this section, we give a geometrical interpretation of these bounds which leads to an explicit outer bound on $\mathcal{R}$ called the LP bound (LP for *linear programming*).

Let A be a subset of $\mathcal{Q}_{\mathcal{N}}$. For a vector $\mathbf{h} \in \mathcal{H}_{\mathcal{N}}$, let

$$\mathbf{h}_A = (h_Z : Z \in A).$$

For a subset $\mathcal{B}$ of $\mathcal{H}_{\mathcal{N}}$, let

$$\mathrm{proj}_A(\mathcal{B}) = \{\mathbf{h}_A : \mathbf{h} \in \mathcal{B}\}$$

be the projection of the set $\mathcal{B}$ on the coordinates $h_Z, Z \in A$. For a subset $\mathcal{B}$ of $\mathcal{H}_{\mathcal{N}}$, define

$$\Lambda(\mathcal{B}) = \{\mathbf{h} \in \mathcal{H}_{\mathcal{N}} : 0 \leq \mathbf{h} < \mathbf{h}' \text{ for some } \mathbf{h}' \in \mathcal{B}\}$$

and

$$\bar{\Lambda}(\mathcal{B}) = \{\mathbf{h} \in \mathcal{H}_{\mathcal{N}} : 0 \leq \mathbf{h} \leq \mathbf{h}' \text{ for some } \mathbf{h}' \in \mathcal{B}\}.$$

A vector $\mathbf{h} \geq 0$ is in $\Lambda(\mathcal{B})$ if and only if it is *strictly* inferior to some vector $\mathbf{h}'$ in $\mathcal{B}$, and is in $\bar{\Lambda}(\mathcal{B})$ if and only if it is inferior to some vector $\mathbf{h}'$ in $\mathcal{B}$.

Define the following subsets of $\mathcal{H}_{\mathcal{N}}$:

$$\begin{aligned}
\mathcal{C}_1 &= \left\{\mathbf{h} \in \mathcal{H}_{\mathcal{N}} : h_{Y_S} = \sum_{s \in S} h_{Y_s}\right\} \\
\mathcal{C}_2 &= \left\{\mathbf{h} \in \mathcal{H}_{\mathcal{N}} : h_{U_{\mathrm{Out}(s)}|Y_s} = 0 \text{ for all } s \in S\right\} \\
\mathcal{C}_3 &= \left\{\mathbf{h} \in \mathcal{H}_{\mathcal{N}} : h_{U_{\mathrm{Out}(i)}|U_{\mathrm{In}(i)}} = 0 \text{ for all } i \in V \backslash S\right\} \\
\mathcal{C}_4 &= \{\mathbf{h} \in \mathcal{H}_{\mathcal{N}} : h_{U_e} < R_e \text{ for all } e \in E\} \\
\mathcal{C}_5 &= \left\{\mathbf{h} \in \mathcal{H}_{\mathcal{N}} : h_{Y_{\beta(i)}|U_{\mathrm{In}(i)}} = 0 \text{ for all } i \in T\right\}.
\end{aligned}$$

These sets contain points in $\mathcal{H}_{\mathcal{N}}$ that satisfy the constraints in (2.48) and (2.50) to (2.53), respectively. The set $\mathcal{C}_1$ is a hyperplane in $\mathcal{H}_{\mathcal{N}}$. Each of the sets $\mathcal{C}_2$, $\mathcal{C}_3$, and $\mathcal{C}_5$ is the intersection of a collection of hyperplanes in $\mathcal{H}_{\mathcal{N}}$. The set $\mathcal{C}_4$ is the intersection of a collection of open half-spaces in $\mathcal{H}_{\mathcal{N}}$. Then from the alternative definition of $\mathcal{R}'$ (Definition 2.17), we see that

$$\mathcal{R}' = \Lambda(\mathrm{proj}_{Y_S}(\Gamma^*_{\mathcal{N}} \cap \mathcal{C}_1 \cap \mathcal{C}_2 \cap \mathcal{C}_3 \cap \mathcal{C}_4 \cap \mathcal{C}_5)).$$

and

$$\mathcal{R}_{in} = \overline{\mathrm{con}}(\Lambda(\mathrm{proj}_{Y_S}(\Gamma^*_{\mathcal{N}} \cap \mathcal{C}_1 \cap \mathcal{C}_2 \cap \mathcal{C}_3 \cap \mathcal{C}_4 \cap \mathcal{C}_5))).$$

Similarly, we see that

$$\mathcal{R}_{out} = \bar{\Lambda}(\mathrm{proj}_{Y_S}(\overline{\Gamma}^*_{\mathcal{N}} \cap \mathcal{C}_1 \cap \mathcal{C}_2 \cap \mathcal{C}_3 \cap \mathcal{C}_4 \cap \overline{\mathcal{C}_5})). \tag{2.86}$$

It can be shown that if $\Gamma^*_{\mathcal{N}} \cap (\mathcal{C}_1 \cap \mathcal{C}_2 \cap \mathcal{C}_3 \cap \mathcal{C}_5)$ is dense in $\overline{\Gamma}^*_{\mathcal{N}} \cap (\mathcal{C}_1 \cap \mathcal{C}_2 \cap \mathcal{C}_3 \cap \mathcal{C}_5)$, i.e.,

$$\overline{\Gamma^*_{\mathcal{N}} \cap (\mathcal{C}_1 \cap \mathcal{C}_2 \cap \mathcal{C}_3 \cap \mathcal{C}_5)} = \overline{\Gamma}^*_{\mathcal{N}} \cap (\mathcal{C}_1 \cap \mathcal{C}_2 \cap \mathcal{C}_3 \cap \mathcal{C}_5),$$

then

$$\mathcal{R}_{out} = \overline{\mathcal{R}'} \subset \overline{\mathrm{con}}(\mathcal{R}') = \mathcal{R}_{in},$$

which implies

$$\mathcal{R}_{in} = \mathcal{R}_{out}.$$

Note that $(\mathcal{C}_1 \cap \mathcal{C}_2 \cap \mathcal{C}_3 \cap \mathcal{C}_5)$ is a closed subset of $\mathcal{H}_{\mathcal{N}}$. However, while

$$\overline{\Gamma^*_{\mathcal{N}} \cap \mathcal{C}} \subset \overline{\Gamma}^*_{\mathcal{N}} \cap \mathcal{C}$$

for any closed subset $\mathcal{C}$ of $\mathcal{H}_{\mathcal{N}}$, it is not in general true that

$$\overline{\Gamma^*_{\mathcal{N}} \cap \mathcal{C}} = \overline{\Gamma}^*_{\mathcal{N}} \cap \mathcal{C}.$$

As a counterexample, it has been shown in [22] (also see [19], Theorem 14.2) that $\overline{\Gamma^*_3 \cap \tilde{\mathcal{C}}}$ is a proper subset of $\overline{\Gamma}^*_3 \cap \tilde{\mathcal{C}}$, where Γ^*_n denotes $\Gamma^*_{\mathcal{N}}$ for

$$\mathcal{N} = \{X_1, X_2, \cdots, X_n\}$$

and

$$\tilde{\mathcal{C}} = \left\{ \mathbf{h} \in \Gamma^*_3 : h_{X_j} + h_{X_k} = h_{\{X_j, X_k\}}, 1 \le j < k \le 3 \right\}.$$

To facilitate our discussion, we further define

$$i_{A;A'} = h_A - h_{A|A'} \tag{2.87}$$

and

$$i_{A;A'|A''} = h_{A|A''} - h_{A|A'A''} \tag{2.88}$$

for $A, A', A'' \in \mathcal{Q}_{\mathcal{N}}$. Note that (2.87) and (2.88) correspond to the information-theoretic identities

$$I(A;A') = H(A) - H(A|A')$$

and

$$I(A;A'|A'') = H(A|A'') - H(A|A'A''),$$

respectively. Let $\Gamma_{\mathcal{N}}$ be the set of $\mathbf{h} \in \mathcal{H}_{\mathcal{N}}$ such that $\mathbf{h}$ satisfies all the *basic inequalities* involving some or all of the random variables in $\mathcal{N}$, i.e., for all $A, A', A'' \in \mathcal{Q}_{\mathcal{N}}$,

$$\begin{aligned} h_A &\geq 0 \\ h_{A|A'} &\geq 0 \\ i_{A;A'} &\geq 0 \\ i_{A;A'|A''} &\geq 0. \end{aligned}$$

These inequalities are equivalent to the nonnegativity of all Shannon's information measures (entropy, conditional entropy, mutual information, and conditional mutual information). The significance of the region $\Gamma_{\mathcal{N}}$ is that it fully characterizes all the *Shannon-type information inequalities* involving the random variables in $\mathcal{N}$, namely those inequalities implied by the above set of basic inequalities. Since the basic inequalities are satisfied by all joint distributions (i.e., $\mathbf{h} \in \Gamma^*_{\mathcal{N}}$ implies $\mathbf{h} \in \Gamma_{\mathcal{N}}$) and that $\Gamma_{\mathcal{N}}$ is closed, we have $\overline{\Gamma}^*_{\mathcal{N}} \subset \Gamma_{\mathcal{N}}$. Then upon replacing $\overline{\Gamma}^*_{\mathcal{N}}$ by $\Gamma_{\mathcal{N}}$ in the definition of $\mathcal{R}_{out}$, we immediately obtain an outer bound on $\mathcal{R}_{out}$. This is called the *LP bound*, denoted by $\mathcal{R}_{LP}$. In other words, $\mathcal{R}_{LP}$ is obtained by replacing $\overline{\Gamma}^*_{\mathcal{N}}$ by $\Gamma_{\mathcal{N}}$ on the right hand side of (2.86), i.e.,

$$\mathcal{R}_{LP} = \bar{\Lambda}(\mathrm{proj}_{Y_S}(\Gamma_{\mathcal{N}} \cap \mathcal{C}_1 \cap \mathcal{C}_2 \cap \mathcal{C}_3 \cap \mathcal{C}_4 \cap \overline{\mathcal{C}_5})).$$

Since all the constraints defining $\mathcal{R}_{LP}$ are linear, $\mathcal{R}_{LP}$ can in principle be evaluated explicitly, although the computation involved can be nontrivial.

However, it has been shown in [23] by means of the discovery of what is known as a *non-Shannon-type information inequality* that $\overline{\Gamma}^*_n \neq \Gamma_n$ for $n \geq 4$, so there is a potential gap between $\mathcal{R}_{out}$ and $\mathcal{R}_{LP}$. In short, a non-Shannon-type information inequality is an outer bound on $\Gamma^*_{\mathcal{N}}$ which is not implied by the basic inequalities. Specifically, it is proved

in [23] that for any 4 random variables X_1, X_2, X_3, and X_4,

$$\begin{aligned} 2I(X_3;X_4) &\leq I(X_1;X_2) + I(X_1;X_3,X_4) \\ &\quad + 3I(X_3;X_4|X_1) + I(X_3;X_4|X_2). \end{aligned} \tag{2.89}$$

We refer the reader to [19], Ch. 14, for a detailed discussion.

Now return to the question of whether there is indeed a gap between $\mathcal{R}_{out}$ and $\mathcal{R}_{LP}$. This important question has recently been answered in [3], where it is shown by means of the non-Shannon-type inequality (2.89) that $\mathcal{R}_{LP}$ is not tight for a particular multi-source network coding problem constructed from matroid theory. This result implies that $\mathcal{R}_{out}$ is generally tighter than $\mathcal{R}_{LP}$.

Nonetheless, it has been proved in [19], Ch. 15, and [20] that $\mathcal{R}_{LP}$ is tight for all special cases of multi-source network coding for which the achievable information rate region is known. These include single-source network coding discussed in Part I as well as the models described in [17][8][14][21][20]. Since $\mathcal{R}_{LP}$ encompasses all Shannon-type information inequalities and the converse proofs of the achievable information rate region for all these special cases do not involve non-Shannon-type inequalities, the tightness of $\mathcal{R}_{LP}$ for all these cases is not surprising.

3

Fundamental Limits of Linear Codes

In Part I, we have shown that for single-source network coding, linear codes are sufficient for achieving asymptotic optimality. It is not clear whether this continues to hold for multi-source network coding. In this section, we present a framework for discussion and explore a potential gap between the asymptotic performance of linear codes and nonlinear codes.

3.1 Linear network codes for multiple sources

We first generalize the global description of a linear network code in Definition 2.5 of Part I for multiple sources. As in Part I, to facilitate our discussion of linear codes, we assume that each channel has unit capacity. Let F be a finite field,

$$\boldsymbol{\omega} = (\omega_s : s \in S)$$

be a tuple of positive integers, and

$$\Omega = \sum_{s \in S} \omega_s.$$

Consider the space F^{Ω}. The information source generated at a source node s is regarded as an ω_s-dimensional subspace of F^{Ω}, denoted by W_s, and it is assumed that the subspaces for different information sources are linearly independent, i.e.,

$$W_s \cap W_{s'} = 0 \quad \text{for } s \neq s', \tag{3.1}$$

where 0 denotes the zero vector.

As in Part I, the information source generated at a source node s is modelled by ω_s imaginary channels terminating at the node s. We adopt the convention that these channels are labeled by $s(1), s(2), \cdots, s(\omega_s)$.

Definition 3.1. (Global Description of a Linear Network Code) Let F be a finite field, and $\boldsymbol{\omega} = (\omega_s : s \in S)$ be a tuple of positive integers. For $s \in S$, let W_s be an ω_s-dimensional subspace of F^{Ω} such that $W_s \cap W_{s'} = 0$ for $s \neq s'$. An $\boldsymbol{\omega}$-dimensional F-valued linear network code on an acyclic network with respect to $\{W_s\}$ consists of a scalar $k_{d,e}$ for every adjacent pair (d, e) in the network as well as an Ω-dimensional column vector f_e for every channel e such that:

(7.2) $f_e = \sum_{d \in \mathrm{In}(i)} k_{d,e} f_d$, where $e \in \mathrm{Out}(i)$.

(7.3) For $s \in S$, the vectors $f_{s(1)}, f_{s(2)}, \cdots, f_{s(\omega_s)}$ for the ω_s imaginary channels terminating at the node source node s constitute a basis for the subspace W_s.

The scalar $k_{d,e}$ is called the *local encoding kernel* for the adjacent pair (d, e), while the vector f_e is called the *global encoding kernel* for the channel e.

We note that in the above definition, for given $\omega_s, s \in S$, the specific choice of the set of subspaces $\{W_s\}$ is not important. While it is convenient to choose W_s for $s \in S$ and f_e for all imaginary channels e such that the latter form the natural basis for F^{Ω}, in order to keep the definition general and to facilitate subsequent discussion, we do not impose this requirement. In fact, a linear network code as defined in Definition 3.1 that does not satisfy this requirement can readily be converted into one by means of a linear transformation.

Introduce the notations

$$f_s = \left[\, f_{s(1)}\ f_{s(2)} \cdots f_{s(\omega_s)} \,\right] \tag{3.4}$$

for $s \in S$ and

$$f_{E'} = [f_e]_{e \in E'} \tag{3.5}$$

for $E' \subset E$. In (3.5), the matrix elements f_e are put in juxtaposition. This convention will be adopted throughout this section.

Definition 3.2. An information rate tuple

$$\boldsymbol{\omega} = (\omega_s : s \in S)$$

is *linearly achievable* if for some base field F, there exists an $\boldsymbol{\omega}'$-dimensional linear code on the network, where $\boldsymbol{\omega}' \geq \boldsymbol{\omega}$ (componentwise), satisfying: For all $i \in T$, for all $s \in \beta(i)$, there exists an $|\mathrm{In}(i)| \times \omega'_s$ matrix $G_i(s)$ such that

$$f_s = f_{\mathrm{In}(i)} \cdot G_i(s). \tag{3.6}$$

The matrix $G_i(s)$ is called the *decoding kernel* at the node i for the information source generated at the source node s.

3.2 Entropy and the rank function

In this section, we establish a fundamental relation (Theorem 3.4) between entropy and the *rank function* of matrices. This relation is instrumental for the discussion in the next section, where we explore the asymptotic limitation of linear network codes for multiple sources.

Theorem 3.3. Let F be a finite field, Y be an Ω-dimensional random row vector that distributes uniformly on F^Ω, and A be an F-valued $\Omega \times l$ matrix. Let $Z = g(Y)$, where $g(Y) = Y \cdot A$. Then $H(Z) = \mathrm{rank}(A) \log |F|$.

Proof. Let $\mathbf{y} \in F^\Omega$ and $\mathbf{z} \in F^l$ be row vectors. Consider the system of simultaneous equations

$$\mathbf{y} \cdot A = \mathbf{z}$$

with $\mathbf{y}$ being unknown and $\mathbf{z}$ fixed, and let $S_{\mathbf{z}}$ denote the solution set for a particular $\mathbf{z}$. It is readily seen that S_0, where 0 denotes the zero vector, is a linear subspace of F^{Ω}.

For a particular $\mathbf{z}$, $S_{\mathbf{z}}$ may or may not be empty. For distinct $\mathbf{z}_1, \mathbf{z}_2 \in \text{range}(g)$, i.e., both $S_{\mathbf{z}_1}$ and $S_{\mathbf{z}_2}$ are nonempty, it is readily seen that

$$S_{\mathbf{z}_1} \cap S_{\mathbf{z}_2} = \emptyset. \tag{3.7}$$

Now regard the vectors in F^{Ω} together with vector addition as a group, and hence S_0 is a subgroup of F^{Ω}. For a fixed $\mathbf{z}$ such that $S_{\mathbf{z}}$ is nonempty, consider any $\tilde{\mathbf{y}} \in S_{\mathbf{z}}$. Then it is easy to verify that

$$S_{\mathbf{z}} = \{\tilde{\mathbf{y}} + \mathbf{y} : \mathbf{y} \in S_0\}.$$

Thus $S_{\mathbf{z}}$ is a coset of S_0 with respect to $\tilde{\mathbf{y}}$, and by the Lagrange theorem (see for example [7]), $|S_{\mathbf{z}}| = |S_0|$. It follows that $|S_{\mathbf{z}}|$ is equal to a constant for all $\mathbf{z} \in \text{range}(g)$.

Finally, for all $\mathbf{z} \in \text{range}(g)$,

$$\begin{aligned} \Pr\{Z = z\} &= \Pr\{Y \in S_z\} \\ &= \frac{|S_z|}{|F|^{\Omega}} \\ &= \frac{|S_0|}{|F|^{\Omega}}, \end{aligned}$$

which does not depend on z. Thus Z has a uniform distribution on range(g). Since range(g) is a subspace of F^l with dimension rank(A), it follows that

$$H(Z) = \log |F|^{\text{rank}(A)} = \text{rank}(A) \log |F|.$$

The theorem is proved. □

Before we proceed further, we first define a region in the entropy space $\mathcal{H}_{\mathcal{N}}$ which is closely related to the region $\Gamma^*_{\mathcal{N}}$, where we recall from Section 2.2.5 that

$$\mathcal{N} = \{Y_s : s \in S; U_e : e \in E\}.$$

Let Ω be any integer such that $\Omega \geq 1$. For each $e \in E$, associate with the random variable U_e an unspecified Ω-dimensional column vector

denoted by v_{U_e}, and for each $s \in S$, associate with the random variable Y_s an unspecified $\Omega \times \omega_s$ matrix denoted by v_{Y_s} (here v_{Y_s} is regarded as a collection of ω_s Ω-dimensional column vectors). The use of these unspecified vectors/matrices will become clear shortly. For $A \in \mathcal{Q}_{\mathcal{N}}$, let

$$v_A = [v_Z]_{Z \in A}.$$

A vector

$$\mathbf{h} = (h_A : A \in \mathcal{Q}_{\mathcal{N}})$$

as defined in (2.46) is a *rank function* for a finite base field F if there exists a collection of column vectors $\{v_Z : Z \in \mathcal{N}\}$ in F such that

$$h_A = \text{rank}(v_A) \tag{3.8}$$

for all $A \in \mathcal{Q}_{\mathcal{N}}$. We then define the region

$$\Psi^*_{\mathcal{N}} = \{\mathbf{h} \in \mathcal{H}_{\mathcal{N}} : \mathbf{h} \text{ is a rank function for some base field } F \text{ and some } \Omega \geq 1\}.$$

The possible gap between the asymptotic performance between linear and nonlinear codes, as we shall see, hinges on a gap between the region $\Psi^*_{\mathcal{N}}$ and $\Gamma^*_{\mathcal{N}}$ characterized by an inequality on the rank function known as the Ingleton inequality [9]. We first establish the following fundamental theorem.

Theorem 3.4. $\text{con}(\Psi^*_{\mathcal{N}}) \subset \overline{\Gamma}^*_{\mathcal{N}}$, where $\text{con}(\Psi^*_{\mathcal{N}})$ denotes the convex hull of $\Psi^*_{\mathcal{N}}$.

Proof. Consider $\mathbf{h} \in \Psi^*_{\mathcal{N}}$. Then for some finite base field F and some $\Omega \geq 1$, there exists a collection of vectors $\{v_Z : Z \in \mathcal{N}\}$ such that (3.8) is satisfied. Let

$$Y = \left[Y_1 \; Y_2 \cdots Y_\Omega \right]$$

be an Ω-dimensional row vector, where Y_i, $1 \leq i \leq \Omega$ are i.i.d. random variables each distributing uniformly on F, so that Y distributes uniformly on F^Ω. Define the random variable

$$Z = Y \cdot v_Z$$

for every $Z \in \mathcal{N}$, so that for every $A \in \mathcal{Q}_{\mathcal{N}}$,

$$[Z]_{Z \in A} = Y \cdot v_A.$$

Then by Theorem 3.3,

$$H(Z : Z \in A) = \text{rank}(v_A) \log |F|. \tag{3.9}$$

From (3.8) and (3.9), we have

$$h_A = \text{rank}(v_A) = (\log |F|)^{-1} H(Z : Z \in A),$$

or

$$(\log |F|) h_A = H(Z : Z \in A).$$

This implies that $(\log |F|)\mathbf{h}$ is an entropy function, or

$$(\log |F|)\mathbf{h} \in \Gamma^*_{\mathcal{N}}.$$

Since $\overline{\Gamma}^*_{\mathcal{N}}$ is a convex cone,

$$\mathbf{h} \in \overline{\Gamma}^*_{\mathcal{N}}.$$

Therefore, we conclude that

$$\Psi^*_{\mathcal{N}} \subset \overline{\Gamma}^*_{\mathcal{N}}.$$

The proof is then completed by taking the convex hull in the above. □

3.3 Can nonlinear codes be better asymptotically?

Recall the notation

$$f_{E'} = [f_e]_{e \in E'}$$

for $E' \subset E$ and introduce a similar notation

$$f_{S'} = [f_s]_{s \in S'}$$

for $S' \subset S$. For a linear code as defined in Definition 3.1, we observe that the assumption (3.1) is equivalent to

$$\text{rank}(f_S) = \sum_{s \in S} \text{rank}(f_s),$$

while the requirement (7.2) is equivalent to

$$\mathrm{rank}(f_{\mathrm{In}(i)\cup\mathrm{Out}(i)}) = \mathrm{rank}(f_{\mathrm{In}(i)}).$$

Furthermore, in Definition 3.2, the decoding requirement prescribed in (3.6) is equivalent to

$$\mathrm{rank}(f_{\beta(i)\cup\mathrm{In}(i)}) = \mathrm{rank}(f_{\mathrm{In}(i)}).$$

Letting

$$v_{Y_s} = f_s$$

for $s \in S$ and

$$v_{U_e} = f_e$$

for $e \in E$, and following Definitions 3.1 and 3.2 and the foregoing, we see that an information rate tuple $\boldsymbol{\omega}$ is linearly achievable if and only if for some finite base field F, there exists a collection of Ω-dimensional column vectors $\{v_Z : Z \in \mathcal{N}\}$, where $\Omega = \sum_{s\in S}\omega_s$, which satisfies the following conditions:

$$\mathrm{rank}(v_{Y_S}) = \sum_{s\in S}\mathrm{rank}(v_{Y_s}) \tag{3.10}$$

$$\mathrm{rank}(v_{Y_s}) \geq \omega_s, \quad s \in S \tag{3.11}$$

$$\mathrm{rank}(v_{U_{\mathrm{Out}(s)}\cup Y_s}) = \mathrm{rank}(v_{Y_s}), \quad s \in S \tag{3.12}$$

$$\mathrm{rank}(v_{U_{\mathrm{In}(i)\cup\mathrm{Out}(i)}}) = \mathrm{rank}(v_{U_{\mathrm{In}(i)}}), \quad i \in V\backslash S \tag{3.13}$$

$$\mathrm{rank}(v_{U_e}) \leq 1, \quad e \in E \tag{3.14}$$

$$\mathrm{rank}(v_{Y_{\beta(i)}\cup U_{\mathrm{In}(i)}}) = \mathrm{rank}(v_{U_{\mathrm{In}(i)}}), \quad i \in T. \tag{3.15}$$

In other words, there exists $\mathbf{h} \in \Psi^*_{\mathcal{N}}$ which satisfy the following conditions:

$$h_{Y_S} = \sum_{s\in S} h_{Y_s} \tag{3.16}$$

$$h_{Y_s} \geq \omega_s, \quad s \in S \tag{3.17}$$

$$h_{U_{\mathrm{Out}(s)}|Y_s} = 0, \quad s \in S \tag{3.18}$$

$$h_{U_{\mathrm{Out}(i)}|U_{\mathrm{In}(i)}} = 0, \quad i \in V\backslash S \tag{3.19}$$

$$h_{U_e} \leq 1, \quad e \in E \tag{3.20}$$

$$h_{Y_{\beta(i)}|U_{\mathrm{In}(i)}} = 0, \quad i \in T, \tag{3.21}$$

where (3.18), (3.19), and (3.21) follow because these equalities are equivalent to

$$h_{U_{\mathrm{Out}(s)} \cup Y_s} = h_{Y_s}$$
$$h_{U_{\mathrm{Out}(i)} \cup \mathrm{In}(i)} = h_{U_{\mathrm{In}(i)}}$$

and

$$h_{Y_{\beta(i)} \cup U_{\mathrm{In}(i)}} = h_{U_{\mathrm{In}(i)}},$$

which correspond to (3.12), (3.13), and (3.15), respectively. If we allow time-sharing of linear codes, then we simply replace the region $\Psi^*_{\mathcal{N}}$ by the region $\mathrm{con}(\Psi^*_{\mathcal{N}})$. The discussion above is summarized by the following definition and theorem.

Definition 3.5. Let $\mathcal{R}_{linear}$ be the set of all information rate tuple $\boldsymbol{\omega}$ such that there exists $\mathbf{h} \in \mathrm{con}(\Psi^*_{\mathcal{N}})$ satisfying (3.16) to (3.21).

Theorem 3.6. An information rate tuple is achievable by time-sharing of linear codes, possibly defined on base fields with different characteristics, if and only if $\boldsymbol{\omega} \in \mathcal{R}_{linear}$.

By setting $R_e = 1$ in (2.58), (3.16) to (3.21) become exactly the same as (2.54) to (2.59). Invoking Theorem 3.4, we see that

$$\mathcal{R}_{linear} \subset \mathcal{R}_{out},$$

which is expected.

The regions $\mathcal{R}_{in}$ and $\mathcal{R}_{out}$ are in terms of $\Gamma^*_{\mathcal{N}}$ and $\overline{\Gamma}^*_{\mathcal{N}}$, respectively, while the region $\mathcal{R}_{linear}$ is in terms of $\mathrm{con}(\Psi^*_{\mathcal{N}})$. Let A and B be any collections of vectors. It is well known that the rank function satisfies the following properties:

P1. $0 \leq \mathrm{rank}(A) \leq |A|$.
P2. $\mathrm{rank}(A) \leq \mathrm{rank}(B)$ if $A \subset B$.
P3. $\mathrm{rank}(A) + \mathrm{rank}(B) \geq \mathrm{rank}(A \cup B) + \mathrm{rank}(A \cap B)$.

In addition, a rank function also satisfies the *Ingleton inequality* [9]: For any collections of vectors $A_i, i = 1, 2, 3, 4$,

$$\begin{aligned}&\operatorname{rank}(A_{13}) + \operatorname{rank}(A_{14}) + \operatorname{rank}(A_{23}) + \operatorname{rank}(A_{24}) + \operatorname{rank}(A_{34})\\ &\quad \geq \operatorname{rank}(A_3) + \operatorname{rank}(A_4) + \operatorname{rank}(A_{12}) + \operatorname{rank}(A_{134}) + \operatorname{rank}(A_{234}),\end{aligned}$$

where A_{13} denotes $A_1 \cup A_3$, etc.

It has been shown in [23] that there exists entropy functions involving 4 random variables which do not satisfy the corresponding Ingleton inequality for entropy functions. The gap between $\operatorname{con}(\Psi^*_{\mathcal{N}})$ and $\Gamma^*_{\mathcal{N}}$ so implied indicates that for certain multi-source network coding problems, $\mathcal{R}_{Out}$ may be strictly larger than $\mathcal{R}_{Linear}$, opening up the possibility that nonlinear codes can outperform linear codes asymptotically.

In fact, examples have been reported by various authors that nonlinear codes can outperform linear codes [12][13][4][11][5]. In particular, it is shown in [5] that there exist multi-source network coding problems for which nonlinear codes can outperform very general forms of linear codes, including mixtures of linear codes discussed here. This shows that there is indeed a gap between $\mathcal{R}_{Linear}$ and $\mathcal{R}_{Out}$.

Appendix A

Global Linearity versus Nodal Linearity

In this appendix, we define *global linearity* and *local linearity* of a network code based on the first principle. We shall show that global linearity implies local linearity. This justifies the generality of the local and global descriptions of a linear network code on an acyclic network in Definitions 2.4 and 2.5 of Part I.

Definition A.1. (Global Linearity) A network code on an acyclic network is globally linear if the global encoding mappings $\tilde{f}_e, e \in E$ are all linear, i.e.,

$$\tilde{f}_e(a_1x_1 + a_2x_2) = a_1\tilde{f}_e(x_1) + a_2\tilde{f}_e(x_2), \tag{A.1}$$

where x_1 and x_2 are row vectors in F^ω and $a_1, a_2 \in F$.

Definition A.2. (Local Linearity) A network code on an acyclic network is locally linear if the local encoding mappings $\tilde{k}_e, e \in E$ are all linear.

It can easily be seen by induction that local linearity implies global linearity, but the converse is not immediate. We shall prove that this is indeed the case.

We shall need a few preliminary results. We begin with the following lemma whose proof is elementary, but we nevertheless include it so that the reader can compare it with the proof of the next lemma.

Lemma A.3. Let $g : F^m \to F$, where F^m denotes the linear space of F-valued m-dimensional row vectors. Then g is linear if and only if there exists an F-valued m-dimensional column vector a such that

$$g(y) = y \cdot a$$

for all $y \in F^m$.

Proof. It is clear that if $g(y) = y \cdot a$ for all $y \in F^m$, then g is linear. We only need to prove the converse. Let u_k denote the row vector in F^m such that the kth component is equal to 1 while all other components are equal to 0. Write

$$y = \sum_k y_k u_k,$$

where y_k is the kth component of y. Then

$$\begin{aligned} g(y) &= g\left(\sum_k y_k u_k\right) \\ &= \sum_k y_k g(u_k). \end{aligned}$$

Upon letting a be the column vector $[g(u_k)]$, we have

$$g(y) = y \cdot a,$$

proving the lemma. □

This lemma has the following less trivial generalization.

Lemma A.4. Let $g : S \to F$, where S denotes a subspace of row vectors in F^m. Then g is linear if and only if there exists an F-valued

m-dimensional column vector k such that

$$g(y) = y \cdot k$$

for all $y \in S$.

Proof. Again, it is clear that if $g(y) = y \cdot k$ for all $y \in S$, then g is linear. So we only prove the converse.

Denote the dimension of S by κ. Let $\{u_1, \cdots, u_\kappa\}$ be a basis for S and let U be the $\kappa \times m$ matrix with the rows being $u_1, \cdots, u_\kappa$ in this order. Then $y \in S$ if and only if

$$y = w \cdot U$$

for some row vector $w \in F^\kappa$. Since U is full rank by construction, it's right inverse, denoted by U_r^{-1} $(m \times \kappa)$, exists, and we can write

$$w = y \cdot U_r^{-1}.$$

Define a function $\tilde{g} : F^\kappa \to F$ such that

$$\tilde{g}(w) = g(w \cdot U).$$

Since g is linear, it can readily be verified that so is $\tilde{g}$. Then by Lemma A.3,

$$\tilde{g}(w) = w \cdot a$$

for some column vector $a \in F^\kappa$. Hence,

$$\begin{aligned} g(y) &= g(w \cdot U) \\ &= \tilde{g}(w) \\ &= w \cdot a \\ &= (y \cdot U_r^{-1}) \cdot a \\ &= y \cdot (U_r^{-1} \cdot a). \end{aligned}$$

Upon letting $k = U_r^{-1} \cdot a$, we have

$$g(y) = y \cdot k,$$

proving the lemma. □

This lemma has the following immediate matrix generalization.

Corollary A.5. Let $g : S \to F^l$, where S denotes a subspace of row vectors in F^m. Then g is a linear transformation if and only if there exists an F-valued matrix K with dimension $m \times l$ such that

$$g(y) = y \cdot K$$

for all $y \in S$.

Now consider a globally linear network code and any non-source node i. Let $\tilde{K}_i$ be the local encoding mapping at i, i.e.,

$$(\tilde{f}_d(x), d \in \mathrm{In}(i)) \mapsto (\tilde{f}_e(x), e \in \mathrm{Out}(i)).$$

Introduce the notations

$$\tilde{f}_{\mathrm{In}(i)}(x) = [\tilde{f}_d(x)]_{d \in \mathrm{In}(i)}$$

and

$$f_{\mathrm{In}(i)} = [f_d]_{d \in \mathrm{In}(i)},$$

where $\tilde{f}_{\mathrm{In}(i)}(x)$ and $f_{\mathrm{In}(i)}$ are row vectors, and recall that f_d denotes the global encoding kernel of the channel d. In a similar fashion, $\tilde{f}_{\mathrm{Out}(i)}(x)$ and $f_{\mathrm{Out}(i)}$ are defined. It is easy to see that $\{\tilde{f}_{\mathrm{In}(i)}(x) : x \in F^\omega\}$ forms a subspace (of row vectors) in $F^{|\mathrm{In}(i)|}$. In other words, $\tilde{K}_i$ is a mapping from a subspace of $F^{|\mathrm{In}(i)|}$ to $F^{|\mathrm{Out}(i)|}$.

We now show that encoding mapping $\tilde{K}_i$ is linear. Let

$$y_j = \tilde{f}_{\mathrm{In}(i)}(x_j)$$

for $j = 1, 2$. Then for any $c_1, c_2 \in F$,

$$\begin{aligned}
\tilde{K}_i(c_1 y_1 + c_2 y_2) &= \tilde{K}_i(c_1 \tilde{f}_{\mathrm{In}(T)}(x_1) + c_2 \tilde{f}_{\mathrm{In}(T)}(x_2)) \\
&= \tilde{K}_i(\tilde{f}_{\mathrm{In}(T)}(c_1 x_1 + c_2 x_2)) \\
&= \tilde{f}_{\mathrm{Out}(T)}(c_1 x_1 + c_2 x_2) \\
&= c_1 \tilde{f}_{\mathrm{Out}(T)}(x_1) + c_2 \tilde{f}_{\mathrm{Out}(T)}(x_2) \\
&= c_1 \tilde{K}_i(\tilde{f}_{\mathrm{In}(T)}(x_1)) + c_2 \tilde{K}_i(\tilde{f}_{\mathrm{In}(T)}(x_2)) \\
&= c_1 \tilde{K}_i(y_1) + c_2 \tilde{K}_i(y_2).
\end{aligned}$$

Thus $\tilde{K}_i$ is linear. Hence, global linearity implies local linearity.

Now since $\tilde{K}_i$ is linear, by Corollary A.5, there exists an $|\mathrm{In}(i)| \times |\mathrm{Out}(i)|$ matrix K_i (encoding kernel for the node i) such that

$$g_i(y) = y \cdot K_i$$

for all $\{\tilde{f}_{\mathrm{In}(i)}(x) : x \in F^\omega\}$. Then for any row vector $x \in F^\omega$, we have

$$\begin{aligned} x \cdot f_{\mathrm{Out}(i)} &= \tilde{f}_{\mathrm{Out}(i)}(x) \\ &= \tilde{K}_i(\tilde{f}_{\mathrm{In}(i)}(x)) \\ &= \tilde{f}_{\mathrm{In}(i)}(x) \cdot K_i \\ &= (x \cdot f_{\mathrm{In}(i)}) \cdot K_i \\ &= x \cdot (f_{\mathrm{In}(i)} \cdot K_i). \end{aligned}$$

Since the above holds for every $x \in F^\omega$, it implies that

$$f_{\mathrm{Out}(i)} = f_{\mathrm{In}(i)} \cdot K_i,$$

or for every $e \in \mathrm{Out}(T)$,

$$f_e = \sum_{d \in \mathrm{In}(T)} k_{d,e} f_e.$$

This justifies Definition 2.5, and we have shown that this definition as well as Definition 2.4 define the most general linear network code on an acyclic network.

Acknowledgements

The authors would like to thank Chung Ping Kwong and David Tse for the useful discussions, and Siu-Wai Ho for converting part of the manuscript from Word to LaTeX. They also would like to thank Ken Zeger for clarifying their results in [5]. The work of Raymond Yeung and Bob Li were partially supported by grants from the Research Grant Council of the Hong Kong Special Administrative Region, China (RGC Ref. No. CUHK4214/03E and 414005).

References

[1] T. Berger, "Multiterminal source coding," in *The Information Theory Approach to Communications*, (G. Longo, ed.), 1978. CISM Courses and Lectures #229, Springer-Verlag, New York.

[2] T. M. Cover and J. A. Thomas, *Elements of information theory.* 1991.

[3] R. Dougherty, C. Freiling, and K. Zeger, "Matroids, networks, and non-shannon information inequalities," submitted to *IEEE Trans. Inform. Theory.*

[4] R. Dougherty, C. Freiling, and K. Zeger, "Linearity and solvability in multicast networks," in *38th Annual Conference on Information Sciences and Systems*, (Princeton, NJ), March 17–19 2004.

[5] R. Dougherty, C. Freiling, and K. Zeger, "Insufficiency of linear coding in network information flow," *IEEE Trans. Inform. Theory*, vol. IT-51, pp. 2745–2759, 2005.

[6] E. Erez and M. Feder, "Capacity region and network codes for two receivers multicast with private and common data," in *Workshop on Coding, Cryptography and Combinatorics*, 2003.

[7] J. B. Fraleigh, *A first course in abstract algebra.* 7th ed., 2003.

[8] K. P. Hau, *Multilevel diversity coding with independent data streams.* June 1995. M.Phil. thesis, The Chinese University of Hong Kong.

[9] A. W. Ingleton, "Representation of matroids," in *Combinatorial Mathematics and its Applications*, (D. J. A. Welsh, ed.), (London), pp. 149–167, Academic Press, 1971.

[10] C. K. Ngai and R. W. Yeung, "Multisource network coding with two sinks," in *International Conference on Communications, Circuits and Systems (ICCCAS)*, (Chengdu, China), June 27–29 2004.

[11] A. Rasala-Lehman, *Network coding*. Massachusetts Institute of Technology, Department of Electrical Engineering and Computer Science, February 2005.

[12] A. Rasala-Lehman and E. Lehman, "Complexity classification of network information flow problems," in *41st Annual Allerton Conference on Communication Control and Computing*, (Monticello, IL), October 2003.

[13] S. Riis, "Linear versus non-linear boolean functions in network flow," preprint, November 2003.

[14] J. R. Roche, R. W. Yeung, and K. P. Hau, "Symmetrical multilevel diversity coding," *IEEE Trans. Inform. Theory*, vol. IT-43, pp. 1059–1064, 1997.

[15] C. E. Shannon, "A mathematical theory of communication," *Bell Sys. Tech. Journal*, vol. 27, pp. 379–423, 623–656, 1948.

[16] L. Song, R. W. Yeung, and N. Cai, "Zero-error network coding for acyclic networks," *IEEE Trans. Inform. Theory*, vol. IT-49, pp. 3129–3139, 2003.

[17] R. W. Yeung, "Multilevel diversity coding with distortion," *IEEE Trans. Inform. Theory*, vol. IT-41, pp. 412–422, 1995.

[18] R. W. Yeung, "A framework for linear information inequalities," *IEEE Trans. Inform. Theory*, vol. IT-43, pp. 1924–1934, 1997.

[19] R. W. Yeung, *A first course in information theory*. Kluwer Academic/Plenum Publishers, 2002.

[20] R. W. Yeung and Z. Zhang, "Distributed source coding for satellite communications," *IEEE Trans. Inform. Theory*, vol. IT-45, pp. 1111–1120, 1999.

[21] R. W. Yeung and Z. Zhang, "On symmetrical multilevel diversity coding," *IEEE Trans. Inform. Theory*, vol. IT-45, pp. 609–621, 1999.

[22] Z. Zhang and R. W. Yeung, "A non-shannon-type conditional inequality of information quantities," *IEEE Trans. Inform. Theory*, vol. IT-43, pp. 1982–1986, 1997.

[23] Z. Zhang and R. W. Yeung, "On characterization of entropy function via information inequalities," *IEEE Trans. Inform. Theory*, vol. IT-44, pp. 1440–1452, 1998.