Journal of Zhejiang University SCIENCE A
ISSN 1673-565X (Print); ISSN 1862-1775 (Online)
www.zju.edu.cn/jzus; www.springerlink.com
E-mail: jzus@zju.edu.cn

On Wyner-Ziv Problem for general sources with average distortion criterion*

YANG Sheng-tian†, ZHAO Min-jian, QIU Pei-liang

(*Department of Information Science and Electronic Engineering, Zhejiang University, Hangzhou 310027, China*)

†E-mail: yangshengtian@zju.edu.cn

Received Feb. 27, 2007; revision accepted Apr. 9, 2007

Abstract: The Wyner-Ziv Problem for general sources with average distortion under fixed-length coding is investigated in this paper. To solve the problem, an enhanced covering lemma for a Markov chain is first established. Then based on the lemma, a general formula for the rate-distortion function of the problem is derived, where the distortion is only assumed uniformly bounded and may be nonadditive. Finally, it is further pointed out that such methods can be used to establish more general results on multiterminal source coding problems.

Key words: Rate-distortion with side information, Information spectrum, Fixed-length coding, Average distortion criterion
doi:10.1631/jzus.2007.A1263 **Document code:** A **CLC number:** O236; TN911.2

INTRODUCTION

Wyner and Ziv (1976) first investigated the rate-distortion problem (now usually called the Wyner-Ziv Problem) when the decoder can fully observe the side information about the source (Fig.1). They established the rate-distortion function of the problem for independent and identically distributed sources with an additive distortion. Readers may refer to (Cover and Thomas, 1991; Iwata and Muramatsu, 2002; Gastpar, 2004) and the references therein for the history of classic results on the Wyner-Ziv Problem.

Recently, Iwata and Muramatsu (2002) derived a general formula for the rate-distortion function of the Wyner-Ziv Problem for general sources with a maximum distortion criterion under fixed-length coding by using the information-spectrum approach (Han, 2003). It is an interesting advance, but according to the framework of the general rate-distortion theory in (Han, 2003), it is only one of the four types of Wyner-Ziv Problems. The other three problems are the Wyner-Ziv Problem for general sources with: (1) average distortion criterion under fixed-length coding; (2) maximum distortion criterion under variable-length coding; (3) average distortion criterion under variable-length coding. Evidently, none of these three problems can be solved by simply applying the approach of (Iwata and Muramatsu, 2002). In this paper, we will investigate the first problem.

DEFINITION, NOTATION AND MAIN RESULT

A general source in the information-spectrum methods (Han, 2003) is defined as an infinite sequence

$$\boldsymbol{X}=\{X^n=(X_1^{(n)},X_2^{(n)},\cdots,X_n^{(n)})\}_{n=1}^{\infty}$$

of n-dimensional random variables X^n, where each component random variable $X_i^{(n)}$ ($1\le i\le n$) takes values in the alphabet $\mathcal{X}$ (finite or countably infinite). Analogously, we can define the correlated general sources $\boldsymbol{XY}$ as an infinite sequence

* Project supported by the National Natural Science Foundation of China (Nos. 60472079 and 60602023), and China Postdoctoral Science Foundation (No. 20060400314)

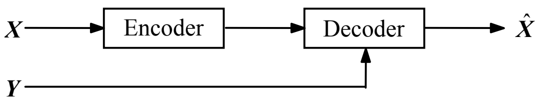

Fig.1 Wyner-Ziv type compression system

$$\boldsymbol{XY}=\{X^nY^n=(X_1^{(n)}Y_1^{(n)},X_2^{(n)}Y_2^{(n)},\cdots,X_n^{(n)}Y_n^{(n)})\}_{n=1}^{\infty}$$

of n-dimensional random variables X^nY^n, where each component random variable $X_i^{(n)}Y_i^{(n)}=(X_i^{(n)},Y_i^{(n)})$ ($1\le i\le n$) takes values in the product alphabet $\mathcal{X}\times\mathcal{Y}$. We denote the sample space and sample sequence of the n-dimensional random variables X^nY^n, X^n and Y^n by $\mathcal{X}^n\times\mathcal{Y}^n$, $\mathcal{X}^n$, $\mathcal{Y}^n$ and x^ny^n, x^n, y^n, respectively.

Throughout the sequel, we assume that the alphabets of all general sources are finite; and for convenience, we also use the notations $P_X(x)$ and $P_{X|Y}(x|y)$ to substitute for $\Pr\{X=x\}$ and $\Pr\{X=x|Y=y\}$, respectively.

Now let us define the Wyner-Ziv coding system for general sources with average distortion criterion under fixed-length coding depicted by Fig.1. At first, we define the source for encoder and the side information for decoder as a pair of correlated general sources $\boldsymbol{XY}$ with the product alphabet $\mathcal{X}\times\mathcal{Y}$. Next, we define the output of decoder as a reproduction source $\hat{\boldsymbol{X}}$ with the reproduction alphabet $\hat{\mathcal{X}}$, and the distortion between $\boldsymbol{X}$ and $\hat{\boldsymbol{X}}$ can then be measured by any given sequence $\{d_n\}_{n=1}^{\infty}$ of mappings $d_n:\mathcal{X}^n\times\hat{\mathcal{X}}^n\rightarrow[0,\infty)$. Furthermore, we define a sequence $\{(\varphi_n,\psi_n)\}_{n=1}^{\infty}$ of fixed-length codes by two mappings:

$$\text{Encoder } \varphi_n:\mathcal{X}^n\rightarrow\mathcal{I}_{L_n},$$

$$\text{Decoder } \psi_n:\mathcal{Y}^n\times\mathcal{I}_{L_n}\rightarrow\hat{\mathcal{X}}^n,$$

where $\mathcal{I}_n\triangleq\{1,2,\cdots,n\}$ and L_n is a positive integer, and the rate of the code (φ_n, ψ_n) is defined by $R(\varphi_n)=(\ln|\mathrm{Im}\varphi_n|)/n$, where $\mathrm{Im}f$ denotes the image of f and $|A|$ denotes the cardinality of the set A.

Then, the Wyner-Ziv system works in the following way. For each n, the n-length source output X^n is encoded into a fixed-length codeword $\varphi_n(X^n)$, and then the decoder observes the codeword and the side information Y^n to reproduce the estimate $\hat{X}^n=\psi_n(Y^n,\varphi_n(X^n))$. The distortion between X^n and $\hat{X}^n$ is given by $d_n(X^n,\hat{X}^n)$. According to (Han, 2003), a rate-distortion pair (R, D) is called fa-achievable with side information if there exists a sequence $\{(\varphi_n,\psi_n)\}_{n=1}^{\infty}$ of fixed-length codes such that

$$\limsup_{n\to\infty} R(\varphi_n)\le R,$$

$$\limsup_{n\to\infty} E[d_n(X^n,\psi_n(Y^n,\varphi_n(X^n)))]\le D.$$

And the fa-rate-distortion function $r_{fa}(D|\boldsymbol{X},\boldsymbol{Y})$ with side information for given general source $\boldsymbol{X}$ and distortion level D is hence defined by

$$r_{fa}(D\,|\,\boldsymbol{X},\boldsymbol{Y})\triangleq\inf\{R\,|\,(R,D)\text{ is } fa\text{-achievable with side information for given } D\}.$$

The aim of this paper is to characterize this fa-rate-distortion function, with the main result being as follows:

Theorem 1 For the Wyner-Ziv system depicted by Fig.1 with the uniformly bounded distortion measure $\{d_n\}_{n=1}^{\infty}$, the rate-distortion pair (R, D) is fa-achievable with side information if and only if there exists an auxiliary general source $\boldsymbol{Z}$ and a sequence $\{f_n\}_{n=1}^{\infty}$ of mappings $f_n:\mathcal{Y}^n\times\mathcal{Z}^n\rightarrow\hat{\mathcal{X}}^n$ such that

$$R\ge\overline{I}(\boldsymbol{X};\boldsymbol{Z})-\underline{I}(\boldsymbol{Y};\boldsymbol{Z}), \tag{1}$$

$$P_{X^nY^nZ^n}=P_{X^nY^n}P_{Z^n|X^n},\ \forall n\ge 1, \tag{2}$$

$$\limsup_{n\to\infty} E[d_n(X^n,f_n(Y^n,Z^n))]\le D. \tag{3}$$

Then the fa-rate-distortion function $r_{fa}(D|\boldsymbol{X},\boldsymbol{Y})$ is given by

$$\inf\{\overline{I}(\boldsymbol{X};\boldsymbol{Z})-\underline{I}(\boldsymbol{Y};\boldsymbol{Z})\}, \tag{4}$$

where inf is over all $\boldsymbol{Z}$ and $\{f_n\}_{n=1}^{\infty}$ satisfying the conditions Eqs.(2) and (3).

The quantities $\overline{I}(\boldsymbol{X};\boldsymbol{Y})$ and $\underline{I}(\boldsymbol{X};\boldsymbol{Y})$ are called the spectral sup-mutual information rate and the spectral inf-mutual information rate respectively in the information-spectrum methods (Han, 2003), and they are defined by

$$\overline{I}(\boldsymbol{X};\boldsymbol{Y}) \triangleq \text{p-}\limsup_{n\to\infty}\frac{1}{n}\ln\frac{P_{X^nY^n}(X^n,Y^n)}{P_{X^n}(X^n)P_{Y^n}(Y^n)} \quad (5)$$

and

$$\underline{I}(\boldsymbol{X};\boldsymbol{Y}) \triangleq \text{p-}\liminf_{n\to\infty}\frac{1}{n}\ln\frac{P_{X^nY^n}(X^n,Y^n)}{P_{X^n}(X^n)P_{Y^n}(Y^n)}, \quad (6)$$

respectively, where

$$\text{p-}\limsup_{n\to\infty} Z_n \triangleq \inf\{\alpha \mid \lim_{n\to\infty}\Pr\{Z_n > \alpha\} = 0\}$$

and

$$\text{p-}\liminf_{n\to\infty} Z_n \triangleq \sup\{\beta \mid \lim_{n\to\infty}\Pr\{Z_n < \beta\} = 0\}$$

denote the limit superior in probability and the limit inferior in probability of the sequence $\{Z_n\}_{n=1}^{\infty}$ of real valued random variables, respectively.

ENHANCED COVERING LEMMA FOR MARKOV CHAIN

The main difficulty in the proof of Theorem 1 is how to deal with the average distortion criterion. We notice that Lemma 1 in (Iwata and Muramatsu, 2002), originally due to (Miyake and Kanaya, 1995), plays an important role in the proof of the general formula for the rate-distortion function in Theorem 1 of (Iwata and Muramatsu, 2002), so in order to prove Theorem 1, we first need to establish a corresponding lemma. Since Lemma 1 in (Iwata and Muramatsu, 2002) may be regarded as a covering lemma for a Markov chain, we call our lemma an enhanced covering lemma for a Markov chain, which is stated as follows:

Lemma 1 Let U^n, V^n and W^n be random variables which take values in finite sets $\mathcal{U}^n$, $\mathcal{V}^n$ and $\mathcal{W}^n$, respectively, and satisfy a Markov condition for each n

$$P_{U^nV^nW^n} = P_{U^nV^n}P_{W^n|V^n}.$$

Now let $\{A_n\}_{n=1}^{\infty}$ be a sequence of arbitrary sets in $\mathcal{U}^n\times\mathcal{W}^n$ satisfying

$$\lim_{n\to\infty}\Pr\{(U^n,W^n)\in A_n\} = 1, \quad (7)$$

and let $\{d_n\}_{n=1}^{\infty}$ be a sequence of mappings d_n: $\mathcal{U}^n\times\mathcal{W}^n\to[0, \infty)$ which are uniformly bounded, that is,

$$D_0 \triangleq \sup_{n\geq 1}\max\{\operatorname{Im} d_n\} < \infty, \quad (8)$$

then for any $\gamma>0$, there exists a sequence $\{F_n\}_{n=1}^{\infty}$ of random functions F_n: $\mathcal{V}^n\to\mathcal{W}^n$ such that

$$\left|\operatorname{Im} F_n\right| \leq \left\lceil \mathrm{e}^{n(\overline{I}(\boldsymbol{V};\boldsymbol{W})+\gamma)}\right\rceil, \quad (9)$$

$$\lim_{n\to\infty}\Pr\{(U^n,F_n(V^n))\in A_n\} = 1, \quad (10)$$

$$\limsup_{n\to\infty}\left\{E[d_n(U^n,F_n(V^n))] - E[d_n(U^n,W^n)]\right\} \leq 0, \quad (11)$$

where $\boldsymbol{V}$ and $\boldsymbol{W}$ denote the sequences $\{V_n\}_{n=1}^{\infty}$ and $\{W_n\}_{n=1}^{\infty}$, respectively.

Proof Define

$$\eta_n^{(1)}(v^n,w^n) \triangleq \sum_{u^n\in\mathcal{U}^n} P_{U^n|V^nW^n}(u^n\mid v^n,w^n)1\{(u^n,w^n)\notin A_n\}$$
$$= \sum_{u^n\in\mathcal{U}^n} P_{U^n|V^n}(u^n\mid v^n)1\{(u^n,w^n)\notin A_n\} \quad (12)$$

and

$$\eta_n^{(2)}(v^n,w^n) \triangleq \sum_{u^n\in\mathcal{U}^n} P_{U^n|V^nW^n}(u^n\mid v^n,w^n)d_n(u^n,w^n)$$
$$= \sum_{u^n\in\mathcal{U}^n} P_{U^n|V^n}(u^n\mid v^n)d_n(u^n,w^n) \quad (13)$$

for $v^n\in\mathcal{V}^n$, $w^n\in\mathcal{W}^n$. Then, it follows from Eqs.(7) and (8) that

$$\lim_{n\to\infty}E[\eta_n^{(1)}(V^n,W^n)] = 0,$$
$$\max_{v^n\in\mathcal{V}^n,w^n\in\mathcal{W}^n}\eta_n^{(2)}(v^n,w^n) \leq D_0,$$
$$E[\eta_n^{(2)}(V^n,W^n)] = E[d_n(U^n,W^n)]. \quad (14)$$

We denote $E[\eta_n^{(1)}(V^n,W^n)]$ by δ_n, and define the set

$$T_n^{(1)} \triangleq \{(v^n,w^n)\in\mathcal{V}^n\times\mathcal{W}^n \mid \eta_n^{(1)}(v^n,w^n)\leq\sqrt{\delta_n}\}. \quad (15)$$

Clearly, by Markov's inequality, we have

$$\Pr\{(V^n,W^n)\notin T_n^{(1)}\} \leq \frac{E[\eta_n^{(1)}(V^n,W^n)]}{\sqrt{\delta_n}} = \sqrt{\delta_n}. \quad (16)$$

Moreover, we define the set

$$T_n^{(2)}(\rho) \triangleq \left\{ (v^n, w^n) \in \mathcal{V}^n \times \mathcal{W}^n \,\middle|\, \frac{1}{n} \ln \frac{P_{W^n|V^n}(w^n \mid v^n)}{P_{W^n}(w^n)} \le \rho \right\}, \quad (17)$$

where ρ is an arbitrary nonnegative real number.

Next, set

$$M_n = \left\lceil e^{n(\bar{I}(V;W)+\gamma)} \right\rceil, \quad (18)$$

and we randomly generate a sequence $\mathcal{C} \triangleq \{W_i^n\}_{i=1}^{M_n}$ of sequences in $\mathcal{W}^n$, where each sequence W_i^n is generated independently according to the probability distribution P_{W^n}. Therefore, for any sample sequence $C = \{w_i^n\}_{i=1}^{M_n}$, we have

$$P_{\mathcal{C}}(C) = \prod_{i=1}^{M_n} P_{W^n}(w_i^n). \quad (19)$$

When no ambiguity is involved, the notations $\mathcal{C}$ and C are also regarded as a set though there are possibly duplicated sequences in them.

Now, let us define the random function $F_n^{\mathcal{C}}: \mathcal{V}^n \to \mathcal{W}^n$ with respect to the random sequence $\mathcal{C}$. For any $v^n \in \mathcal{V}^n$, define

$$F_n^{\mathcal{C}}(v^n) \triangleq \begin{cases} G_{\mathcal{C}}(v^n), & S_{\mathcal{C}}(v^n) \ne \varnothing, \\ G_{\mathcal{C}}'(v^n), & S_{\mathcal{C}}(v^n) = \varnothing, \end{cases} \quad (20)$$

where

$$G_{\mathcal{C}}(v^n) \triangleq \arg \min_{w^n \in S_{\mathcal{C}}(v^n)} \eta_n^{(2)}(v^n, w^n),$$

$$G_{\mathcal{C}}'(v^n) \triangleq \arg \min_{w^n \in \mathcal{C}} \eta_n^{(2)}(v^n, w^n),$$

$$S_{\mathcal{C}}(v^n) \triangleq \mathcal{C} \cap \{w^n \in \mathcal{W}^n \mid (v^n, w^n) \in T_n^{(1)}\}.$$

Clearly, $F_n^{\mathcal{C}}$ satisfies the requirement Eq.(9). Next, let us estimate the upper bound of $\Pr\{(U^n, F_n^{\mathcal{C}}(V^n)) \notin A_n\}$ and $E[d_n(U^n, F_n^{\mathcal{C}}(V^n))]$. First, we have

$$\begin{aligned}
&\Pr\{(U^n, F_n^{\mathcal{C}}(V^n)) \notin A_n\} \\
&= \sum_{v^n \in \mathcal{V}^n} P_{V^n}(v^n) \sum_{C \in (\mathcal{W}^n)^{M_n}} P_{\mathcal{C}}(C) \cdot \sum_{u^n \in \mathcal{U}^n} P_{U^n|V^n}(u^n \mid v^n) 1\{(u^n, F_n^C(v^n)) \notin A_n\} \\
&\overset{(a)}{=} \sum_{v^n \in \mathcal{V}^n} P_{V^n}(v^n) \sum_{C \in (\mathcal{W}^n)^{M_n}} P_{\mathcal{C}}(C) \eta_n^{(1)}(v^n, F_n^C(v^n)) \\
&= \sum_{v^n \in \mathcal{V}^n} P_{V^n}(v^n) \sum_{C \in (\mathcal{W}^n)^{M_n}} P_{\mathcal{C}}(C) \eta_n^{(1)}(v^n, F_n^C(v^n)) \cdot \left\{ 1\{(v^n, F_n^C(v^n)) \in T_n^{(1)}\} + 1\{(v^n, F_n^C(v^n)) \notin T_n^{(1)}\} \right\} \\
&\overset{(b)}{\le} \sqrt{\delta_n} + \sum_{v^n \in \mathcal{V}^n} P_{V^n}(v^n) \sum_{C \in (\mathcal{W}^n)^{M_n}} P_{\mathcal{C}}(C) 1\{(v^n, F_n^C(v^n)) \notin T_n^{(1)}\} \\
&= \sqrt{\delta_n} + \sum_{v^n \in \mathcal{V}^n} P_{V^n}(v^n) \Pr\{(v^n, F_n^{\mathcal{C}}(v^n)) \notin T_n^{(1)}\},
\end{aligned}$$

where (a) follows from Eq.(12), and (b) follows from Eq.(15). Furthermore, we have

$$\begin{aligned}
&\Pr\{(v^n, F_n^{\mathcal{C}}(v^n)) \notin T_n^{(1)}\} \\
&\overset{(a)}{=} \sum_{C \in (\mathcal{W}^n)^{M_n}} P_{\mathcal{C}}(C) 1\{S_C(v^n) = \varnothing\} \\
&\overset{(b)}{=} \sum_{C \in (\mathcal{W}^n)^{M_n}} \prod_{i=1}^{M_n} P_{W^n}(w_i^n) 1\{(v^n, w_i^n) \notin T_n^{(1)}\} \\
&= \left(\sum_{w^n \in \mathcal{W}^n} P_{W^n}(w^n) 1\{(v^n, w^n) \notin T_n^{(1)}\} \right)^{M_n} \\
&= \left(1 - \sum_{w^n \in \mathcal{W}^n} P_{W^n}(w^n) 1\{(v^n, w^n) \in T_n^{(1)}\} \right)^{M_n} \\
&\le \left(1 - \sum_{w^n \in \mathcal{W}^n} P_{W^n}(w^n) 1\{(v^n, w^n) \in T_n^{(1)} \cap T_n^{(2)}(\rho)\} \right)^{M_n} \\
&\overset{(c)}{\le} \left(1 - e^{-n\rho} \sum_{w^n \in \mathcal{W}^n} P_{W^n|V^n}(w^n \mid v^n) 1\{(v^n, w^n) \in T_n^{(1)} \cap T_n^{(2)}(\rho)\} \right)^{M_n} \\
&\overset{(d)}{\le} 1 - \sum_{w^n \in \mathcal{W}^n} P_{W^n|V^n}(w^n \mid v^n) 1\{(v^n, w^n) \in T_n^{(1)} \cap T_n^{(2)}(\rho)\} + e^{-M_n e^{-n\rho}},
\end{aligned}$$

where (a) and (b) follow from Eqs.(19) and (20), (c) follows from Eq.(17), and (d) follows from the inequality $(1-xy)^n \le 1-x+e^{-yn}$ for $0 \le x, y \le 1$, $n \ge 1$. Then we have

$$\begin{aligned}
&\Pr\{(U^n, F_n^{\mathcal{C}}(V^n)) \notin A_n\} \\
&\le \sqrt{\delta_n} + \sum_{v^n \in \mathcal{V}^n} P_{V^n}(v^n) \left\{ 1 + e^{-M_n e^{-n\rho}} - \sum_{w^n \in \mathcal{W}^n} P_{W^n|V^n}(w^n \mid v^n) 1\{(v^n, w^n) \in T_n^{(1)} \cap T_n^{(2)}(\rho)\} \right\} \\
&= \sqrt{\delta_n} + \Pr\{(V^n, W^n) \notin T_n^{(1)} \cap T_n^{(2)}(\rho)\} + e^{-M_n e^{-n\rho}}
\end{aligned}$$

$$\le \sqrt{\delta_n} + \Pr\{(V^n, W^n) \notin T_n^{(1)}\} + \Pr\{(V^n, W^n) \notin T_n^{(2)}(\rho)\} + \mathrm{e}^{-M_n \mathrm{e}^{-n\rho}}$$
$$\overset{(a)}{\le} 2\sqrt{\delta_n} + \Pr\{(V^n, W^n) \notin T_n^{(2)}(\rho)\} + \mathrm{e}^{-M_n \mathrm{e}^{-n\rho}},$$

where (a) follows from Eq.(16). Letting $\rho = \overline{I}(\boldsymbol{V};\boldsymbol{W}) + \gamma/2,$ we have

$$\Pr\{(U^n, F_n^{\mathcal{C}}(V^n)) \notin A_n\}$$
$$\overset{(a)}{\le} 2\sqrt{\delta_n} + \Pr\{(V^n, W^n) \notin T_n^{(2)}(\overline{I}(\boldsymbol{V};\boldsymbol{W}) + \gamma/2)\} + \mathrm{e}^{-\mathrm{e}^{n\gamma/2}}$$
$$\overset{(b)}{\to} 0$$

as $n\to\infty$, where (a) follows from Eq.(18), and (b) follows from Eqs.(5) and (17). This concludes Eq.(10).

Secondly, the expectation $E[d_n(U^n, F_n^{\mathcal{C}}(V^n))]$ can be written as

$$E[d_n(U^n, F_n^{\mathcal{C}}(V^n))]$$
$$= \sum_{v^n \in \mathcal{V}^n} P_{V^n}(v^n) \sum_{C \in (\mathcal{W}^n)^{M_n}} P_{\mathcal{C}}(C) \cdot \sum_{u^n \in \mathcal{U}^n} P_{U^n|V^n}(u^n \mid v^n) d_n(u^n, F_n^C(v^n))$$
$$\overset{(a)}{=} \sum_{v^n \in \mathcal{V}^n} P_{V^n}(v^n) \sum_{C \in (\mathcal{W}^n)^{M_n}} P_{\mathcal{C}}(C) \int_0^{D_0} 1\{\eta_n^{(2)}(v^n, F_n^C(v^n)) \ge \beta\} \mathrm{d}\beta$$
$$= \sum_{v^n \in \mathcal{V}^n} P_{V^n}(v^n) \int_0^{D_0} \Pr\{\eta_n^{(2)}(v^n, F_n^{\mathcal{C}}(v^n)) \ge \beta\} \mathrm{d}\beta,$$

where (a) follows from Eq.(13). Furthermore, we have

$$\Pr\{\eta_n^{(2)}(v^n, F_n^{\mathcal{C}}(v^n)) \ge \beta\}$$
$$\overset{(a)}{\le} \sum_{C \in (\mathcal{W}^n)^{M_n}} \prod_{i=1}^{M_n} P_{W^n}(w_i^n) 1\{\eta_n^{(2)}(v^n, w_i^n) \ge \beta \text{ or } (v^n, w_i^n) \notin T_n^{(1)}\}$$
$$= \left(\sum_{w^n \in \mathcal{W}^n} P_{W^n}(w^n) 1\{\eta_n^{(2)}(v^n, w^n) \ge \beta \text{ or } (v^n, w^n) \notin T_n^{(1)}\} \right)^{M_n}$$
$$= \left(1 - \sum_{w^n \in \mathcal{W}^n} P_{W^n}(w^n) 1\{\eta_n^{(2)}(v^n, w^n) < \beta,\ (v^n, w^n) \in T_n^{(1)}\} \right)^{M_n}$$
$$\le \left(1 - \sum_{w^n \in \mathcal{W}^n} P_{W^n}(w^n) 1\{\eta_n^{(2)}(v^n, w^n) < \beta,\ (v^n, w^n) \in T_n^{(1)} \cap T_n^{(2)}(\rho)\} \right)^{M_n}$$
$$\overset{(b)}{\le} \left(1 - \mathrm{e}^{-n\rho} \sum_{w^n \in \mathcal{W}^n} P_{W^n|V^n}(w^n \mid v^n) 1\{\eta_n^{(2)}(v^n, w^n) < \beta,\ (v^n, w^n) \in T_n^{(1)} \cap T_n^{(2)}(\rho)\} \right)^{M_n}$$
$$\overset{(c)}{\le} 1 - \sum_{w^n \in \mathcal{W}^n} P_{W^n|V^n}(w^n \mid v^n) 1\{\eta_n^{(2)}(v^n, w^n) < \beta,\ (v^n, w^n) \in T_n^{(1)} \cap T_n^{(2)}(\rho)\} + \mathrm{e}^{-M_n \mathrm{e}^{-n\rho}}$$
$$= \sum_{w^n \in \mathcal{W}^n} P_{W^n|V^n}(w^n \mid v^n) 1\{\eta_n^{(2)}(v^n, w^n) \ge \beta \text{ or } (v^n, w^n) \notin T_n^{(1)} \text{ or } (v^n, w^n) \notin T_n^{(2)}(\rho)\} + \mathrm{e}^{-M_n \mathrm{e}^{-n\rho}},$$

where (a) follows from Eqs.(19) and (20), (b) follows from Eq.(17), and (c) also follows from the inequality $(1-xy)^n \le 1-x+\mathrm{e}^{-yn}$ for $0\le x, y\le 1$, $n\ge 1$. Then we have

$$E[d_n(U^n, F_n^{\mathcal{C}}(V^n))]$$
$$\le \sum_{v^n \in \mathcal{V}^n} P_{V^n}(v^n) \int_0^{D_0} \left\{ \sum_{w^n \in \mathcal{W}^n} P_{W^n|V^n}(w^n \mid v^n) 1\{\eta_n^{(2)}(v^n, w^n) \ge \beta \text{ or } (v^n, w^n) \notin T_n^{(1)} \text{ or } (v^n, w^n) \notin T_n^{(2)}(\rho)\} + \mathrm{e}^{-M_n \mathrm{e}^{-n\rho}} \right\} \mathrm{d}\beta$$
$$\le \sum_{v^n \in \mathcal{V}^n, w^n \in \mathcal{W}^n} P_{V^n W^n}(v^n, w^n) \int_0^{D_0} \left\{ 1\{\eta_n^{(2)}(v^n, w^n) \ge \beta\} + 1\{(v^n, w^n) \notin T_n^{(1)}\} + 1\{(v^n, w^n) \notin T_n^{(2)}(\rho)\} \right\} \mathrm{d}\beta + D_0 \mathrm{e}^{-M_n \mathrm{e}^{-n\rho}}$$
$$\le \sum_{v^n \in \mathcal{V}^n, w^n \in \mathcal{W}^n} P_{V^n W^n}(v^n, w^n) \left\{ \eta_n^{(2)}(v^n, w^n) + D_0 1\{(v^n, w^n) \notin T_n^{(1)}\} + D_0 1\{(v^n, w^n) \notin T_n^{(2)}(\rho)\} \right\} + D_0 \mathrm{e}^{-M_n \mathrm{e}^{-n\rho}}$$
$$\overset{(a)}{\le} E[d_n(U^n, W^n)] + D_0 \Pr\{(V^n, W^n) \notin T_n^{(1)}\} + D_0 \Pr\{(V^n, W^n) \notin T_n^{(2)}(\rho)\} + D_0 \mathrm{e}^{-M_n \mathrm{e}^{-n\rho}},$$

where (a) follows from Eq.(14). Letting $\rho = \overline{I}(\boldsymbol{V};\boldsymbol{W}) + \gamma/2$, we have

$$E[d_n(U^n, F_n^{\mathcal{C}}(V^n))] - E[d_n(U^n, W^n)]$$
$$\overset{(a)}{\le} D_0 \Pr\left\{(V^n, W^n) \notin T_n^{(2)}(\overline{I}(\boldsymbol{V};\boldsymbol{W}) + \gamma/2)\right\} + D_0\sqrt{\delta_n} + D_0 \mathrm{e}^{-\mathrm{e}^{n\gamma/2}} \overset{(b)}{\to} 0$$

as $n\to\infty$, where (a) follows from Eqs.(16) and (18), and (b) follows from Eqs.(5) and (17). This concludes Eq.(11) and hence completes the proof.

Remark 1 The main idea of this proof is a combination of the methods in the proofs of Lemma 1 in (Iwata and Muramatsu, 2002) and Theorem 5.5.1 in

(Han, 2003). However, such a method has its own limitation. Because the minimum operation in Eq.(20) should be applied to an ordered set, we can establish a covering lemma with only one average distortion criterion.

By Lemma 1, we can easily obtain the following corollary that is also a generalized version of Lemma 1 in (Iwata and Muramatsu, 2002).

Corollary 1 (Yang and Qiu, 2006) Let U^n, V^n and W^n be random variables which take values in finite sets $\mathcal{U}^n$, $\mathcal{V}^n$ and $\mathcal{W}^n$, respectively, and satisfy a Markov condition

$$P_{U^nV^nW^n}=P_{U^nV^n}P_{W^n|V^n}$$

for each n. Now let $\{B_n\}_{n=1}^{\infty}$ be a sequence of arbitrary sets in $\mathcal{U}^n\times\mathcal{W}^n$ satisfying

$$\liminf_{n\to\infty}\Pr\{(U^n,W^n)\in B_n\}=\varepsilon, \tag{21}$$

then for any $\gamma>0$, there exists a sequence $\{F_n\}_{n=1}^{\infty}$ of random functions F_n: $\mathcal{U}^n\to\mathcal{W}^n$ such that

$$\left|\mathrm{Im}\,F_n\right|\le\left\lceil \mathrm{e}^{n(\bar{I}(\boldsymbol{V};\boldsymbol{W})+\gamma)}\right\rceil, \tag{22}$$

$$\liminf_{n\to\infty}\Pr\{(U^n,F_n(V^n))\in B_n\}\ge\varepsilon. \tag{23}$$

Proof Letting $A_n=\mathcal{U}^n\times\mathcal{W}^n$ and $d_n(u^n,w^n)=1\{(u^n,w^n)\notin B_n\}$, and then applying Lemma 1, we have

$$\limsup_{n\to\infty}\left\{E[d_n(U^n,F_n(V^n))]-E[d_n(U^n,W^n)]\right\}\le 0,$$

where F_n is the random function constructed in the proof of Lemma 1. Then we have

$$\begin{aligned}&\liminf_{n\to\infty}\Pr\{(U^n,F_n(V^n))\in B_n\}\\&=1-\limsup_{n\to\infty}E[d_n(U^n,F_n(V^n))]\\&\ge 1-\Big\{\limsup_{n\to\infty}E[d_n(U^n,W^n)]\\&\quad+\limsup_{n\to\infty}\left\{E[d_n(U^n,F_n(V^n))]-E[d_n(U^n,W^n)]\right\}\Big\}\\&\ge 1-\limsup_{n\to\infty}E[d_n(U^n,W^n)]\\&=\liminf_{n\to\infty}\Pr\{(U^n,W^n)\in B_n\}.\end{aligned}$$

This proves the corollary.

PROOF OF THEOREM 1

Now we start to prove Theorem 1. It suffices to show that Eqs.(1)~(3) are the sufficient and necessary conditions for the *fa*-achievability of any given rate-distortion pair. Therefore, the proof below consists of two parts, i.e., the direct part and the converse part of the theorem.

1. Direct Part

Let γ be an arbitrary positive real number. We define

$$T_n\triangleq\left\{(y^n,z^n)\in\mathcal{Y}^n\times\mathcal{Z}^n\,\middle|\,\frac{1}{n}\ln\frac{P_{Y^n|Z^n}(y^n\mid z^n)}{P_{Y^n}(y^n)}\ge\underline{I}(\boldsymbol{Y};\boldsymbol{Z})-\gamma\right\}. \tag{24}$$

Clearly, by the definition of $\underline{I}(\boldsymbol{Y};\boldsymbol{Z})$, we have

$$\lim_{n\to\infty}\Pr\{(X^n,Y^n,Z^n)\in\mathcal{X}^n\times T_n\}=1.$$

Now let $\mathcal{U}=\mathcal{X}\times\mathcal{Y}$, $\mathcal{V}=\mathcal{X}$, $\mathcal{W}=\mathcal{Z}$, and set $U^n=(X^n, Y^n)$, $V^n=X^n$, and $W^n=Z^n$. It follows from Eq.(2) that U^n, V^n and W^n also form a Markov chain. Then according to Lemma 1, there exists a sequence $\{G_n\}_{n=1}^{\infty}$ of functions G_n: $\mathcal{X}^n\to\mathcal{Z}^n$ such that

$$\left|\mathrm{Im}\,G_n\right|\le\left\lceil \mathrm{e}^{n(\bar{I}(\boldsymbol{X};\boldsymbol{Z})+\gamma)}\right\rceil, \tag{25}$$

$$\lim_{n\to\infty}\Pr\{(Y^n,G_n(X^n))\in T_n\}=1, \tag{26}$$

and

$$\begin{aligned}&\limsup_{n\to\infty}E[d_n(X^n,f_n(Y^n,G_n(X^n)))]\\&\le\limsup_{n\to\infty}E[d_n(X^n,f_n(Y^n,Z^n))]\overset{(a)}{\le}D,\end{aligned} \tag{27}$$

where (a) follows from Eq.(3).

Next, we specify the encoding and decoding procedures. We first define a random binning function Ω_L: $\mathcal{Z}^n\to\mathcal{I}_L$, which assigns each $z^n\in\mathcal{Z}^n$ to one of the elements in $\mathcal{I}_L$ according to a uniform distribution on $\mathcal{I}_L$ independently, where L is a positive integer.

Encoding $\Phi_n:\mathcal{X}^n\to\mathcal{I}_{L_n}$. The random encoder Φ_n is simply defined by

$$\Phi_n(x^n)\triangleq\Omega_{L_n}(G_n(x^n)),$$

where

$$L_n = \left\lceil \mathrm{e}^{n(\bar{I}(\boldsymbol{X};\boldsymbol{Z})-\underline{I}(\boldsymbol{Y};\boldsymbol{Z})+3\gamma)} \right\rceil. \quad (28)$$

Decoding $\Psi_n : \mathcal{Y}^n \times \mathcal{I}_{L_n} \to \hat{\mathcal{X}}^n$. The decoder receives the pair (y^n,$\Phi_n(x^n)$). For (y^n,$\Phi_n(x^n)$), if there exists a unique $z^n \in \mathrm{Im}G_n$ such that $\Omega_{L_n}(z^n) = \Phi_n(x^n)$ and $(y^n,z^n)\in T_n$, then we declare $\Psi_n(y^n,\Phi_n(x^n))= f_n(y^n,z^n)$. Otherwise, $\Psi_n(y^n, \Phi_n(x^n))$ is declared to be an arbitrarily fixed element in $\hat{\mathcal{X}}^n$.

If a pair (x^n, y^n) satisfies the following conditions:

(1) $(y^n, G_n(x^n))\in T_n$,
(2) There is no $z^n\in \mathrm{Im}G_n$ such that $z^n \neq G_n(x^n)$, $(y^n, z^n)\in T_n$ and $\Omega_{L_n}(z^n) = \Phi_n(x^n)$,

then the decoding procedure succeeds, and the output

$$\hat{x}^n = \Psi_n(y^n, \Phi_n(x^n)) = f_n(y^n, G_n(x^n)). \quad (29)$$

Define the sets

$$S_i = \{(x^n, y^n) \in \mathcal{X}^n \times \mathcal{Y}^n \mid (x^n, y^n) \text{ satisfies the } i\text{th condition}\}, \quad (30)$$

where i=1 or 2, and define $S=S_1\cap S_2$, then the average distortion between X^n and the estimate $\hat{X}^n = \Psi_n(Y^n, \Phi_n(X^n))$ can be written as

$$\begin{aligned}
&E[d_n(X^n, \Psi_n(Y^n, \Phi_n(X^n)))] \\
&= E\Big[[d_n(X^n, \Psi_n(Y^n, \Phi_n(X^n)))\big\{1\{(X^n, Y^n) \in S\} + 1\{(X^n, Y^n) \notin S\}\big\}\Big] \\
&\overset{(a)}{\le} E[d_n(X^n, \Psi_n(Y^n, \Phi_n(X^n)))1\{(X^n, Y^n) \in S\}] + D_0 \Pr\{(X^n, Y^n) \notin S\} \\
&\overset{(b)}{\le} E[d_n(X^n, f_n(Y^n, G_n(X^n)))] + D_0 \Pr\{(X^n, Y^n) \notin S\},
\end{aligned} \quad (31)$$

where (a) follows from the uniform boundedness of $\{d_n\}_{n=1}^{\infty}$ and $D_0 = \sup_{n\ge 1} \max\{\mathrm{Im}\, d_n\}$, and (b) follows from Eq.(29) and the fact that $1\{\cdot\}\le 1$.

To further estimate the upper bound of the average distortion, we need to estimate $\Pr\{(X^n,Y^n)\notin S\}$. We first estimate the probability $\Pr\{(X^n,Y^n)\notin S_2\}$, which can be written as

$$\begin{aligned}
&\Pr\{(X^n, Y^n) \notin S_2\} \\
&= \sum_{\omega_{L_n}} P_{\Omega_{L_n}}(\omega_{L_n}) \sum_{(x^n,y^n)\in\mathcal{X}^n\times\mathcal{Y}^n} P_{X^nY^n}(x^n, y^n) 1\{\exists z^n \in \mathrm{Im}\, G_n, \\
&\quad \text{s.t. } z^n \neq G_n(x^n), \omega_{L_n}(z^n) = \omega_{L_n}(G_n(x^n)),\ (y^n, z^n) \in T_n\} \\
&\le \sum_{\omega_{L_n}} P_{\Omega_{L_n}}(\omega_{L_n}) \sum_{\substack{(x^n,y^n,z^n)\in\mathcal{X}^n\times T_n, \\ z^n\in \mathrm{Im}\, G_n\setminus\{G_n(x^n)\}}} P_{X^nY^n}(x^n, y^n) \cdot 1\{\omega_{L_n}(z^n) = \omega_{L_n}(G_n(x^n))\} \\
&= \sum_{\substack{(x^n,y^n,z^n)\in\mathcal{X}^n\times T_n, \\ z^n\in \mathrm{Im}\, G_n\setminus\{G_n(x^n)\}}} P_{X^nY^n}(x^n, y^n) \Pr\{\Omega_{L_n}(z^n) = \Omega_{L_n}(G_n(x^n))\} \\
&\overset{(a)}{=} \frac{1}{L_n} \sum_{\substack{(x^n,y^n,z^n)\in\mathcal{X}^n\times T_n, \\ z^n\in \mathrm{Im}\, G_n\setminus\{G_n(x^n)\}}} P_{X^nY^n}(x^n, y^n) \\
&\le \frac{1}{L_n} \sum_{(y^n,z^n)\in T_n, z^n\in \mathrm{Im}\, G_n} \sum_{x^n\in\mathcal{X}^n} P_{X^nY^n}(x^n, y^n) \\
&= \frac{1}{L_n} \sum_{(y^n,z^n)\in T_n, z^n\in \mathrm{Im}\, G_n} P_{Y^n}(y^n) \\
&\overset{(b)}{\le} \frac{\mathrm{e}^{-n(\underline{I}(\boldsymbol{Y};\boldsymbol{Z})-\gamma)}}{L_n} \sum_{(y^n,z^n)\in T_n, z^n\in \mathrm{Im}\, G_n} P_{Y^n|Z^n}(y^n \mid z^n) \\
&\le \frac{\mathrm{e}^{-n(\underline{I}(\boldsymbol{Y};\boldsymbol{Z})-\gamma)} |\mathrm{Im}\, G_n|}{L_n} \overset{(c)}{\le} \frac{\mathrm{e}^{n(\bar{I}(\boldsymbol{X};\boldsymbol{Z})+\gamma)}+1}{\mathrm{e}^{n(\bar{I}(\boldsymbol{X};\boldsymbol{Z})+2\gamma)}},
\end{aligned} \quad (32)$$

where (a) follows from the property of random binning function, (b) follows from Eq.(24), and (c) from Eqs.(25) and (28). Then we have

$$\begin{aligned}
&\Pr\{(X^n, Y^n) \notin S\} \\
&\le \Pr\{(X^n, Y^n) \notin S_1\} + \Pr\{(X^n, Y^n) \notin S_2\} \\
&\overset{(a)}{\le} \Pr\{(Y^n, G_n(X^n)) \notin T_n\} + \frac{\mathrm{e}^{n(\bar{I}(\boldsymbol{X};\boldsymbol{Z})+\gamma)}+1}{\mathrm{e}^{n(\bar{I}(\boldsymbol{X};\boldsymbol{Z})+2\gamma)}} \\
&\overset{(b)}{\to} 0
\end{aligned}$$

as $n\to\infty$, where (a) follows from Eq.(32), and (b) follows from Eq.(26). Then we have

$$\begin{aligned}
&\limsup_{n\to\infty} E[d_n(X^n, \Psi_n(Y^n, \Phi_n(X^n)))] \\
&\overset{(a)}{\le} \limsup_{n\to\infty} E[d_n(X^n, f_n(Y^n, G_n(X^n)))] \\
&\quad + D_0 \limsup_{n\to\infty} \Pr\{(X^n, Y^n) \notin S\} \\
&\overset{(b)}{\le} D,
\end{aligned}$$

where (a) follows from Eq.(31), and (b) follows from Eq.(27). This means that there exists at least one sequence $\{(\varphi_n,\psi_n)\}_{n=1}^{\infty}$ of codes satisfying

$$\limsup_{n\to\infty} E[d_n(X^n,\psi_n(Y^n,\varphi_n(X^n)))] \le D.$$

On the other hand, by the definition of L_n, we have

$$\limsup_{n\to\infty} R(\varphi_n) \le \limsup_{n\to\infty} \frac{\ln L_n}{n} = \overline{I}(\boldsymbol{X};\boldsymbol{Z}) - \underline{I}(\boldsymbol{Y};\boldsymbol{Z}) + 3\gamma.$$

Finally, by repeating the argument above with replacing γ by a sequence $\{\gamma_i\}_{i=1}^{\infty}$ which satisfies $\gamma_1\ge\gamma_2\ge\ldots>0$ and $\gamma_i\to 0$ as $i\to\infty$, we can conclude by the diagonal method that there exists a sequence $\{(\varphi_n,\psi_n)\}_{n=1}^{\infty}$ of codes satisfying

$$\limsup_{n\to\infty} E[d_n(X^n,\psi_n(Y^n,\varphi_n(X^n)))] \le D,$$
$$\limsup_{n\to\infty} R(\varphi_n) \le \overline{I}(\boldsymbol{X};\boldsymbol{Z}) - \underline{I}(\boldsymbol{Y};\boldsymbol{Z}).$$

This completes the proof of the direct part.

2. Converse Part

Suppose that the rate-distortion pair (R,D) is *fa*-achievable with side information, then according to the definition, there exists a sequence $\{(\varphi_n,\psi_n)\}_{n=1}^{\infty}$ of codes satisfying

$$\limsup_{n\to\infty} R(\varphi_n) \le R,$$
$$\limsup_{n\to\infty} E[d_n(X^n,\psi_n(Y^n,\varphi_n(X^n)))] \le D.$$

Define $\boldsymbol{Z}=\{Z^n\}_{n=1}^{\infty}$ by $Z^n=\varphi_n(X^n)$, then the conditions Eqs.(2) and (3) hold clearly. Furthermore, for sufficiently large n, we have

$$\begin{aligned} R &\ge R(\varphi_n)-\gamma \\ &\overset{(a)}{\ge} \text{p-}\limsup_{n\to\infty}\frac{1}{n}\ln\frac{1}{P_{Z^n}(Z^n)}-2\gamma \\ &\overset{(b)}{\ge} \text{p-}\limsup_{n\to\infty}\frac{1}{n}\ln\frac{P_{Z^n|X^n}(Z^n\mid X^n)}{P_{Z^n}(Z^n)}-2\gamma \\ &= \overline{I}(\boldsymbol{X};\boldsymbol{Z})-2\gamma \\ &\overset{(c)}{\ge} \overline{I}(\boldsymbol{X};\boldsymbol{Z})-\underline{I}(\boldsymbol{Y};\boldsymbol{Z})-2\gamma, \end{aligned}$$

where (a) follows from Lemma 2.6.2 in (Han, 2003), (b) from the fact that $P_{Z^n|X^n}(Z^n\mid X^n)\le 1$, and (c) from the nonnegativity of $\underline{I}(\boldsymbol{Y};\boldsymbol{Z})$. Since γ is arbitrary, we then have

$$R \ge \overline{I}(\boldsymbol{X};\boldsymbol{Z}) - \underline{I}(\boldsymbol{Y};\boldsymbol{Z}).$$

This concludes Eq.(1) and hence completes the proof of the converse part.

CONCLUSION

In this paper, we prove an enhanced covering lemma for a Markov chain, and then show a general formula for the rate-distortion function of the Wyner-Ziv Problem for general sources with average distortion criterion under fixed-length coding. In fact, our method can also be combined with the method in (Yang and Qiu, 2006) to establish more general results on multiterminal source coding problems with one average distortion criterion and multiple maximum distortion criteria.

References

Cover, T.M., Thomas, J.A., 1991. Elements of Information Theory. John Wiley & Sons, New York.

Gastpar, M., 2004. The Wyner-Ziv problem with multiple sources. *IEEE Trans. on Inf. Theory*, **50**(11):2762-2768. [doi:10.1109/TIT.2004.836707]

Han, T.S., 2003. Information-Spectrum Methods in Information Theory. Springer, Berlin.

Iwata, K., Muramatsu, J., 2002. An information-spectrum approach to rate-distortion function with side information. *IEICE Trans. on Fund.*, **E85-A**(6):1387-1395.

Miyake, S., Kanaya, F., 1995. Coding theorems on correlated general sources. *IEICE Trans. on Fund.*, **E78-A**(9):1063-1070.

Wyner, A.D., Ziv, J., 1976. The rate-distortion function for source coding with side information at the decoder. *IEEE Trans. on Inf. Theory*, **22**(1):1-10. [doi:10.1109/TIT.1976.1055508]

Yang, S., Qiu, P., 2006. An information-spectrum approach to multiterminal rate-distortion theory. Submitted to *IEEE Trans. on Inf. Theory*, draft available at http://arxiv.org/abs/cs/0605006