Source Coding with Unreliable Side Information in the Finite Blocklength Regime

Siming Sun and Michelle Effros
Department of Electrical Engineering, California Institute of Technology, Pasadena, CA 91125, USA
Email: {ssun5, effros}@caltech.edu

2022 IEEE International Symposium on Information Theory (ISIT) | 978-1-6654-2159-1/22/$31.00 ©2022 IEEE | DOI: 10.1109/ISIT50566.2022.9834490

***Abstract*—This paper studies a special case of the problem of source coding with side information. A single transmitter describes a source to a receiver that has access to a side information observation that is unavailable at the transmitter. While the source and true side information sequences are dependent, stationary, memoryless random processes, the side information observation at the decoder is unreliable, which here means that it may or may not equal the intended side information and therefore may or may not be useful for decoding the source description. The probability of side information observation failure, caused, for example, by a faulty sensor or source decoding error, is non-vanishing but is bounded by a fixed constant independent of the blocklength. This paper proposes a coding system that uses unreliable side information to get efficient source representation subject to a fixed error probability bound. Results include achievability and converse bounds under two different models of the joint distribution of the source, the intended side information, and the side information observation.**

I. Introduction

In source coding, the availability of dependent side information at the decoder improves the efficiency of lossless source sequence description. In this work, we consider the impact of side information uncertainty caused by failures such as faulty sensors or noisy data collection. Given unreliable decoder side information, the generalized Slepian-Wolf code should take advantage of the side information observation when it is useful, ignore it when it is not, and in both cases ensure a bound on the error probability – all with no access to the side information at the encoder and no knowledge of whether the side information observation is good beyond what the decoder can learn from the side information itself.

In this paper, we introduce two models of unreliable side information, the *worst-case model* and the *unknown model*. In the worst-case model, a Bernoulli random variable independent of the source and side information determines whether the decoder sees the side information sequence or an independent random sequence with the same marginal. We call this the worst-case model because any source coding advantage must arise from some sort of statistical dependence, and here there is no dependence available. In the unknown model, the observed side information again equals the intended side information with the same probability; in this case, however, the dependence between the decoder's observation and the source-side information sequence pair is unknown. To derive achievability bounds for both cases, we propose a source code based on the classical random binning code design. Under the worst-case model, the resulting achievability bound matches the converse up to the second order. Under the unknown model, the code achieves first order Slepian-Wolf optimal rate when the side information observation equals the true side information and first order optimal point-to-point rate when the side information is faulty, despite the fact that neither the encoder nor the decoder receives side information to distinguish between these scenarios. We discuss the performance of the code under some example joint distributions.

The body of this paper is organized as follows. Section II provides notation and definitions, describes the two unreliable side information models, and sets up the problem. Sections III and IV show our main result for the worst-case and unknown models, respectively. Section V contains concluding remarks, including a discussion of how our proposed technique for bounding output error probability in the wake of an unreliable input is useful beyond this source coding problem, for example for stopping errors from propagating in successive interference cancellation [1]–[4].

This material is based upon work supported by the National Science Foundation under Grant No. 1817241 and by the Oringer Graduate Fellowship at the California Institute of Technology.

II. Problem Statement

A. Notation and Definitions

For any positive integer k, we denote $\{1, 2, ..., k\}$ by $[k]$.

We denote any random process $(U_1, U_2, \ldots)$ by $\{U\}$.

Let discrete random variables X and Y be distributed according to distribution P_{XY} on alphabet $\mathcal{X} \times \mathcal{Y}$; let P_X and $P_{X|Y}$ denote the marginal distribution of X and the conditional distribution of X given Y, respectively. For any $(x, y) \in \mathcal{X} \times \mathcal{Y}$, define the information density $\imath(x)$ and conditional information density $\imath(x|y)$ as

$$\imath(x) \triangleq \log \frac{1}{P_X(x)} \qquad \imath(x|y) \triangleq \log \frac{1}{P_{X|Y}(x|y)}.$$

The non-conditional and conditional entropy, varentropy, and third centered moment of information density are

$$\begin{aligned} H(X) &\triangleq \mathbb{E}[\imath(X)] \\ H(X|Y) &\triangleq \mathbb{E}[\imath(X|Y)] \\ V(X) &\triangleq \mathrm{Var}[\imath(X)] \\ V(X|Y) &\triangleq \mathrm{Var}[\imath(X|Y)] \\ T(X) &\triangleq \mathbb{E}[|\imath(X) - H(X)|^3] \\ T(X|Y) &\triangleq \mathbb{E}[|\imath(X|Y) - H(X|Y)|^3]. \end{aligned}$$

B. The Worst-Case Model

Consider a pair of stationary, memoryless, dependent random processes $\{X\}$ and $\{Y\}$ with single-letter joint distribution P_{XY} on discrete alphabet $\mathcal{X} \times \mathcal{Y}$. Random processes $\{X\}$ and $\{Y\}$ represent the source and side information, respectively. Since they are dependent by assumption,

$$H(X|Y) < H(X).$$

For any $\epsilon_0 \in [0, 1]$, an ϵ_0-unreliable observation $\{\hat{Y}\}$ of side information process $\{Y\}$ is defined by

$$\hat{Y}_i = W \cdot Y_i + (1 - W) \cdot Z_i, \tag{1}$$

where W is a single Bernoulli random variable with $\mathbb{P}[W = 0] = \epsilon_0$, $\{Z\}$ is a stationary, memoryless random process with single-letter distribution P_Y on $\mathcal{Y}$ (making $\{Z\}$, on its own, indistinguishable from $\{Y\}$), and random process $\{(X, Y, Z, W)\}$ has single-letter distribution

$$P_{XYZW}(x, y, z, w) = P_{XY}(x, y)P_Y(z)P_W(w)$$

for all $(x, y, z, w) \in \mathcal{X} \times \mathcal{Y}^2 \times \{0, 1\}$.

C. The Unknown Model

Under the unknown model, joint random process $\{(X, Y)\}$ is unchanged and (1) continues to hold with $\mathbb{P}[W = 0] = \epsilon_0$, but we make no further assumptions about P_{XYZW}. Specifically, neither the encoder, nor the decoder, nor the code designer knows $P_{ZW|XY}$, and $\{Z\}$ need not be memoryless.

While the worst-case model statistics are the worst possible statistics that can occur under the unknown model, performance may be inferior under the unknown model since not knowing the true statistics complicates code design.

D. Problem Setup

For simplicity, we assume that the codeword symbol alphabet is binary. Define 2-dimensional rate vector

$$\mathbf{R} \triangleq [R_D \quad R_I]'$$

with $R_D < R_I$, and 2-dimensional error vector

$$\boldsymbol{\epsilon} \triangleq [\epsilon_D \quad \epsilon_I]' \in (0, 1)^2.$$

Definition 1: An $(n, \mathbf{R}, \boldsymbol{\epsilon})$ unreliable side-information source code (USSC) for source random process $\{X\}$, true side-information random process $\{Y\}$ and observed side-information random process $\{\hat{Y}\}$ with discrete alphabets $\mathcal{X}, \mathcal{Y}$, and $\mathcal{Y}$ comprises an encoding function $\mathsf{f} : \mathcal{X}^n \to \{0, 1\}^{nR_I}$ and a pair of decoding functions, $\mathsf{g}_D : \{0, 1\}^{nR_D} \times \mathcal{Y}^n \to \mathcal{X}^n \cup \{\mathsf{e}\}$ for some symbol $\mathsf{e} \notin \mathcal{X}^n$ and $\mathsf{g}_I : \{0, 1\}^{nR_I} \to \mathcal{X}^n$ satisfying error probability constraints

$$\mathbb{P}[\mathsf{g}_D(\lfloor \mathsf{f}(X^n) \rfloor_{nR_D}, \hat{Y}^n) \neq X^n | W = 1] \leq \epsilon_D$$

and

$$\mathbb{P}\Big[\{\mathsf{g}_D(\lfloor \mathsf{f}(X^n) \rfloor_{nR_D}, \hat{Y}^n) \in \mathcal{X}^n \backslash \{X^n\}\} \cup \{\{\mathsf{g}_D(\lfloor \mathsf{f}(X^n) \rfloor_{nR_D}, \hat{Y}^n) = \mathsf{e}\} \cap \{\mathsf{g}_I(\mathsf{f}(X^n)) \neq X^n\}\} \Big| W = 0\Big] \leq \epsilon_I.$$

Using the given code, the encoded source description is sent in two phases. The first nR_D bits of $\mathsf{f}(X^n)$, denoted by $\lfloor \mathsf{f}(X^n) \rfloor_{nR_D}$, constitute the first phase. The decoder g_D maps these bits either to a reconstruction from $\mathcal{X}^n$ or to a symbol $\mathsf{e} \notin \mathcal{X}^n$ indicating that it is not yet able to decode. Following the tradition of [5] [6], a single bit of feedback to the encoder specifies whether ($\mathsf{g}_D(\lfloor \mathsf{f}(X^n) \rfloor_{nR_D}, \hat{Y}^n) = \mathsf{e}$) or not ($\mathsf{g}_D(\lfloor \mathsf{f}(X^n) \rfloor_{nR_D}, \hat{Y}^n) \neq \mathsf{e}$) the remaining $R_I - R_D$ bits of $\mathsf{f}(X^n)$ should be sent. An error occurs if either decoder g_D fails to decode correctly and promptly when the side information observation is good ($W = 1$) or, when the side information observation fails ($W = 0$), the source reproduction differs from X^n.

Definition 2: Rate pair $\mathbf{R} = [R_D \quad R_I]'$ is $(n, \boldsymbol{\epsilon})$-achievable if there exists an $(n, \mathbf{R}, \boldsymbol{\epsilon})$ USSC. Rate region $\mathscr{R}^*_{US}(n, \boldsymbol{\epsilon})$ is the closure of the set of $(n, \boldsymbol{\epsilon})$-achievable rate pairs.

Remark 1: While the case $\epsilon_D = \epsilon_I = \epsilon$ may be most useful for many applications, differing ϵ_D and ϵ_I may be useful, for example, when an impending deadline changes the user's priorities regarding acceptable error probability.

III. Main Results for the Worst-Case Model

Consider the third-order optimal point-to-point rate R^*_I and Slepian-Wolf rate R^*_D from [7], [8][1], where

$$R^*_D \triangleq H(X|Y) + \sqrt{\frac{V(X|Y)}{n}} Q^{-1}(\epsilon_D) - \frac{\log n}{2n}$$
$$R^*_I \triangleq H(X) + \sqrt{\frac{V(X)}{n}} Q^{-1}(\epsilon_I) - \frac{\log n}{2n}.$$

Define inner and outer bounding rate regions

$$\mathscr{R}_{\text{in}}(n, \boldsymbol{\epsilon}) \triangleq \left\{\mathbf{R} : R_D \geq R^*_D + \frac{\log n}{2n} + O\left(\frac{1}{n}\right), R_I \geq R^*_I + O\left(\frac{1}{n}\right)\right\}$$

$$\mathscr{R}_{\text{out}}(n, \boldsymbol{\epsilon}) \triangleq \left\{\mathbf{R} : R_D \geq R^*_D - O\left(\frac{1}{n}\right), R_I \geq R^*_I - O\left(\frac{1}{n}\right)\right\}.$$

Theorem 1: Under the worst-case model, if

$$V(X) > 0,\ V(X|Y) > 0,\ T(X) < \infty,\ T(X|Y) < \infty, \tag{2}$$

then

$$\mathscr{R}_{\text{in}}(n, \boldsymbol{\epsilon}) \subseteq \mathscr{R}^*_{US}(n, \boldsymbol{\epsilon}) \subseteq \mathscr{R}_{\text{out}}(n, \boldsymbol{\epsilon}).$$

[1]The second- and third-order optimal rate regions for the Slepian-Wolf scenario where side information is available directly to the decoder are due to [9, Eqn. 29a] and [8, Theorem 7, 8], respectively. The third-order gap between the R_D boundary in our result and the rate region for a code with true side information by Gavalakis and Kontonyiannis [8] is due to the inability of the decoder to distinguish Y^n from Z^n. The gap can be closed if the marginals are different.

A. Proof of Achievability

While it is possible, under the given framework, to first send a subsequence of $\{X\}$ losslessly and then use that subsequence to learn W at the decoder, we here employ instead a code design where the decoder determines W and the encoded message X^n through a single decoding operation. This approach is preferable when W can change across many uses of the code.

1) Codebook Generation: For each $x^n \in \mathcal{X}^n$, randomly and independently draw encoder output $\mathsf{f}(x^n)$ from the uniform distribution on $[2^{nR_I}]$. Each index in $[2^{nR_I}]$ represents a unique binary string of length nR_I.

2) Decoder: The decoder g_D performs threshold decoding using a constant threshold $\log\gamma$ defined below, giving

$$\mathsf{g}_D(c_D, \hat{y}^n) = \begin{cases} \hat{x}^n & \text{if } \lfloor \mathsf{f}(\hat{x}^n)\rfloor_{nR_D} = c_D,\ \imath(\hat{x}^n|\hat{y}^n) \le \log\gamma \\ \mathsf{e} & \text{otherwise.} \end{cases} \quad (3)$$

When more than one source vector $\hat{x}^n$ satisfies the threshold test in (3), we choose one that minimizes $\imath(\hat{x}^n|\hat{y}^n)$. The decoder g_I performs maximum likelihood decoding, giving

$$\mathsf{g}_I(c_I) = \arg \min_{x^n\in\mathcal{X}^n:\mathsf{f}(x^n)=c_I} \imath(x^n).$$

Ties are broken uniformly at random.

3) Error Analysis: When $W = 1$, the dependence of the side information makes it useful for source coding. Since delayed reconstruction implies high rate, we define error event

$$\mathcal{E}_D \triangleq \{\mathsf{g}_D(\lfloor \mathsf{f}(X^n)\rfloor_{nR_D}, \hat{Y}^n) \ne X^n\}.$$

Since Y^n, Z^n, and W are independent under the worst-case model studied in this section and $W = 1$ implies $\hat{Y}^n = Y^n$,

$$\begin{aligned}
&\mathbb{P}[\mathcal{E}_D|W=1] \\
&\le \mathbb{P}[\{\exists \bar{x}^n \in \mathcal{X}^n\backslash\{X^n\} : \lfloor \mathsf{f}(\bar{x}^n)\rfloor_{nR_D} = \lfloor \mathsf{f}(X^n)\rfloor_{nR_D}, \\
&\quad \imath(\bar{x}^n|Y^n) \le \log\gamma\} \cup \{\imath(X^n|Y^n) > \log\gamma\}] \\
&\le \mathbb{P}[\imath(X^n|Y^n) > \log\gamma] \\
&\quad + \frac{1}{2^{nR_D}} \sum_{\bar{x}^n\in\mathcal{X}^n,\ y^n\in\mathcal{Y}^n} P_Y^n(y^n)\mathbf{1}\{\imath(\bar{x}^n|y^n) \le \log\gamma\} \\
&= \mathbb{P}[\imath(X^n|Y^n) > \log\gamma] \\
&\quad + \frac{1}{2^{nR_D}}\mathbb{E}\left[\frac{1}{P_{X|Y}^n(X^n|Y^n)}\mathbf{1}\{\imath(X^n|Y^n) \le \log\gamma\}\right].
\end{aligned}$$

Let

$$\begin{aligned}
\gamma &= 2^{nR_D} \\
nR_D &= nH(X|Y) + \tau\sqrt{nV(X|Y)} \\
\tau &= Q^{-1}\left(\epsilon_D - \frac{B+C}{\sqrt{n}}\right).
\end{aligned}$$

Given (2), we can apply [10, Lemma 47] to show

$$\frac{1}{2^{nR_D}}\mathbb{E}\left[\frac{1}{P_{X|Y}^n(X^n|Y^n)}\mathbf{1}\{\imath(X^n|Y^n) \le \log\gamma\}\right] \le \frac{C}{\sqrt{n}}$$

for some constant C and the Berry Esseen bound to show

$$\mathbb{P}[\imath(X^n|Y^n) > \log\gamma] \le Q(\tau) + \frac{B}{\sqrt{n}}$$

for some constant B, giving $\mathbb{P}[\mathcal{E}_D|W=1] \le \epsilon_D$ when

$$R_D = H(X|Y) + \sqrt{\frac{V(X|Y)}{n}}Q^{-1}(\epsilon_D) + O\left(\frac{1}{n}\right), \quad (4)$$

which relies on the differentiability and Taylor expansion of $Q^{-1}(\cdot)$.

We bound $\mathbb{P}[\mathcal{E}_I|W=0]$ as

$$\mathbb{P}[\mathcal{E}_I|W=0] \le \mathbb{P}[\mathcal{E}_{I1}|W=0] + \mathbb{P}[\mathcal{E}_{I2}|W=0],$$

where

$$\begin{aligned}
\mathcal{E}_{I1} &\triangleq \{\mathsf{g}_D(\lfloor \mathsf{f}(X^n)\rfloor_{nR_D}, \hat{Y}^n) \in \mathcal{X}^n\backslash\{X^n\}\} \\
\mathcal{E}_{I2} &\triangleq \{\mathsf{g}_I(\mathsf{f}(X^n)) \ne X^n\}.
\end{aligned}$$

The probability of decoding to an incorrect source vector $\bar{x}^n$ after nR_D bits is identical when $W = 0$ and when $W = 1$ since in both cases the incorrect codeword is independent of the observed side information. Therefore,

$$\begin{aligned}
&\mathbb{P}[\mathcal{E}_{I1}|W=0] \\
&\le \mathbb{P}[\exists \bar{x}^n \in \mathcal{X}^n\backslash\{X^n\} : \\
&\quad \lfloor \mathsf{f}(\bar{x}^n)\rfloor_{nR_D} = \lfloor \mathsf{f}(X^n)\rfloor_{nR_D}, \imath(\bar{x}^n|Z^n) \le \log\gamma] \\
&\le \frac{1}{2^{nR_D}} \sum_{\bar{x}^n\in\mathcal{X}^n, z^n\in\mathcal{Y}^n} P_Y^n(z^n)\mathbf{1}\{\imath(\bar{x}^n|z^n) \le \log\gamma\} \\
&= \frac{1}{2^{nR_D}}\mathbb{E}\left[\frac{1}{P_{X|Y}^n(X^n|Y^n)}\mathbf{1}\{\imath(X^n|Y^n) \le \log\gamma\}\right] \\
&\le \frac{C}{\sqrt{n}}
\end{aligned}$$

under the same parameter choices. Meanwhile, $\mathbb{P}[\mathcal{E}_{I2}|W=0]$ is analogous to a maximum likelihood decoder for a source code without side information. From [6, Proof of Theorem 5],

$$\mathbb{P}[\mathcal{E}_{I2}|W=0] \le \mathbb{P}[\imath(X^n) > nR_I + \frac{1}{2}\log n - \log C'] + \frac{C'}{\sqrt{n}}.$$

We choose

$$\begin{aligned}
R_I =& H(X) + \sqrt{\frac{V(X)}{n}}Q^{-1}\left(\epsilon_I - \frac{B'+C'+C}{\sqrt{n}}\right) \\
&- \frac{\log n}{2n} + \frac{\log C'}{n}
\end{aligned}$$

(for large enough n) to keep the total error probability bounded as $\mathbb{P}[\mathcal{E}_I|W=0] \le \epsilon_I$. By the differentiability of $Q^{-1}(\cdot)$,

$$R_I = H(X) + \sqrt{\frac{V(X)}{n}}Q^{-1}(\epsilon_I) - \frac{\log n}{2n} + O\left(\frac{1}{n}\right). \quad (5)$$

Combining (4) and (5) shows that

$$\mathscr{R}_{\text{in}}(n, \boldsymbol{\epsilon}) \subseteq \mathscr{R}_{US}^*(n, \boldsymbol{\epsilon}).$$

Remark 2: This code design can be extended to more than two possible joint distributions $P_{XY}^{(1)}, P_{XY}^{(2)}, \ldots$ by increasing the number of stages, allowing $R^{(1)}, R^{(2)}, \ldots$ in place of R_D and R_I. The result is a universal Slepian-Wolf code for a family of possible side information distributions.

B. Proof of Converse

When $W = 1$, the decoder output counts as an error if it employs a rate exceeding R_D. Therefore, when $W = 1$, the decoder must decode at a fixed rate R_D, before the encoder receives any feedback. Applying earlier bounds on fixed-rate lossless source coding with side information and without feedback from [9], [8] (as in footnote 1) gives

$$R_D \geq H(X|Y) + \sqrt{\frac{V(X|Y)}{n}} Q^{-1}(\epsilon_D) - \frac{\log n}{2n} - O\left(\frac{1}{n}\right). \quad (6)$$

In contrast, when $W = 0$, feedback plays an active role.

Since the channel in this source coding problem is noiseless and the decoder is deterministic by assumption,[2] the stop feedback is a deterministic function of the encoder's output and the side information sequence observed at the decoder. Therefore, when $W = 0$, the system can perform no better than a 2-rate system without feedback for which an independent side information sequence is available at both the encoder and the decoder. Since the presence of independent side information does not increase the source coding rate region, and since we here constrain the source code to operate at the same rates R_D and R_I as the source code under investigation, applying Han's converse [11, Lemma 1.3.2] with $\gamma = \frac{\log n}{2}$ gives

$$\begin{aligned}&\log(2^{nR_D} + 2^{nR_I}) \\ &\quad \geq nH(X) + \sqrt{nV(X)} Q^{-1}(\epsilon_I) - \frac{1}{2}\log n - O(1). \quad (7)\end{aligned}$$

Since $R_I \geq R_D$, the left-hand side of (7) is no larger than $nR_I + 1$. Therefore,

$$R_I \geq H(X) + \sqrt{\frac{V(X)}{n}} Q^{-1}(\epsilon_I) - \frac{\log n}{2n} - O\left(\frac{1}{n}\right). \quad (8)$$

Combining (6) with (8) yields

$$\mathscr{R}^*_{US}(n, \boldsymbol{\epsilon}) \subseteq \mathscr{R}_{\text{out}}(n, \boldsymbol{\epsilon}).$$

IV. Main Results for the Unknown Model

Under the unknown model, both the encoder and the decoder know $\mathbb{P}[W = 0] = \epsilon_0$, but neither knows the conditional distribution of $\hat{Y}^n$ given X^n and Y^n. In the proof of the theorem that follows, we re-use the 2-stage code from Theorem 1 with a modified value of γ to bound the achievable rate in this scenario.

Theorem 2: Under an unknown model that satisfies (2), for any $0 < \epsilon_D < 1$ and $0 < \epsilon_I < 1$, the $(n, \boldsymbol{\epsilon})$-rate region $\mathscr{R}^*_{USW}(n, \boldsymbol{\epsilon})$ satisfies

$$\mathscr{R}'_{\text{in}}(n, \boldsymbol{\epsilon}) \subseteq \mathscr{R}^*_{USW}(n, \boldsymbol{\epsilon}),$$

where

$$\begin{aligned}\mathscr{R}'_{\text{in}}(n, \boldsymbol{\epsilon}) \triangleq \Bigg\{\mathbf{R} : &R_D \geq R^*_{DW} + \frac{\log n}{n} + O\left(\frac{1}{n}\right), \\ &R_I \geq R^*_{IW} + O\left(\frac{1}{n}\right)\Bigg\}\end{aligned}$$

$$R^*_{DW} \triangleq H(X|Y) + \sqrt{\frac{V(X|Y)}{n}} Q^{-1}(\epsilon_D(1-\epsilon_0)) - \frac{\log n}{2n}$$
$$R^*_{IW} \triangleq H(X) + \sqrt{\frac{V(X)}{n}} Q^{-1}(\epsilon_I \epsilon_0) - \frac{\log n}{2n}.$$

The theorem implies that, up to the first order, there exists a code that can perform as well as a code with true side information when the observation is correct, and as well as a point-to-point code when the observation is incorrect.

A. Proof of Theorem 2

We use the encoder and decoder from Theorem 1 but modify threshold parameter γ. The error definitions for $\mathcal{E}_D$, $\mathcal{E}_I$, $\mathcal{E}_{I1}$ and $\mathcal{E}_{I2}$ remain unchanged. The conditional error probability $\mathbb{P}[\mathcal{E}_D|W = 1]$ is a function of the joint probability $\mathbb{P}[\mathcal{E}_D, W = 1]$ as

$$\mathbb{P}[\mathcal{E}_D|W = 1] = \frac{1}{1-\epsilon_0}\mathbb{P}[\mathcal{E}_D, W = 1].$$

Since $W = 1$ implies that $\hat{Y}^n = Y^n$,

$$\begin{aligned}&\mathbb{P}[\mathcal{E}_D, W = 1] \\ &\quad = \mathbb{P}[\mathsf{g}(\lfloor \mathsf{f}(X^n)\rfloor_{nR_D}, Y^n) \neq X^n, W = 1] \\ &\quad \leq \mathbb{P}[\mathsf{g}(\lfloor \mathsf{f}(X^n)\rfloor_{nR_D}, Y^n) \neq X^n] \\ &\quad \leq \mathbb{P}[\imath(X^n|Y^n) > \log\gamma] \\ &\qquad + \frac{1}{2^{nR_D}} \sum_{\bar{x}^n \in \mathcal{X}^n,\ y^n \in \mathcal{Y}^n} P^n_Y(y^n)\mathbf{1}\{\imath(\bar{x}^n|y^n) \leq \log\gamma\} \\ &\quad \leq \mathbb{P}[\imath(X^n|Y^n) > \log\gamma] + \frac{1}{2^{nR_D}}\mathbb{U}[\imath(\bar{X}^n|Y^n) \leq \log\gamma],\end{aligned}$$

where $\mathbb{U}$ is a finite measure on $\mathcal{X}^n \times \mathcal{Y}^n$ for which $U_{X^nY^n}(x^n, y^n) = P^n_Y(y^n)$ for all $(x^n, y^n) \in \mathcal{X}^n \times \mathcal{Y}^n$. Extending the Chernoff Bound from probability measures to a finite measure $\mathbb{Q}$ by applying Markov's Inequality for finite measures gives

$$\mathbb{Q}[X \geq a] \leq \frac{\sum_{x\in\mathcal{X}} \mathbb{Q}[X = x]\exp\{x\}}{\exp\{a\}}.$$

Therefore,

$$\begin{aligned}&\mathbb{U}[\imath(\bar{X}^n|Y^n) \leq \log\gamma] \\ &\quad = \mathbb{U}[-\imath(\bar{X}^n|Y^n) \geq -\log\gamma] \\ &\quad \leq \frac{\sum_{\bar{x}^n, y^n} -P^n_Y(y^n)\exp\{-\imath(\bar{x}^n|y^n)\}}{\exp\{-\log\gamma\}} \\ &\quad = \gamma \sum_{\bar{x}^n, y^n} P^n_Y(y^n) P^n_{X|Y}(\bar{x}^n|y^n) \\ &\quad = \gamma.\end{aligned}$$

[2]Allowing a random decoder complicates the argument but does not change the outcome. Every random decoder can be expressed as a deterministic decoder with an additional random input that is independent of the source. Making this random input available to the encoder does not enhance its performance.

Setting $\gamma = \frac{2^{nR_D}}{\sqrt{n}}$ gives

$$\frac{1}{2^{nR_D}}\mathbb{U}[\imath(\bar{X}^n|Y^n) \le \log\gamma] \le \frac{1}{\sqrt{n}},$$

and setting

$$R_D = H(X|Y) + \tau\sqrt{\frac{V(X|Y)}{n}} + \frac{\log n}{2n}$$

gives

$$\log\gamma = nH(X|Y) + \tau\sqrt{nV(X|Y)}.$$

Again recalling that P_{XY} satisfies (2), we apply the Berry Esseen bound to give

$$\mathbb{P}[\imath(X^n|Y^n) > \log\gamma] \le Q(\tau) + \frac{B}{\sqrt{n}}$$

for some constant B. Setting

$$\tau = Q^{-1}\left(\epsilon_D(1-\epsilon_0) - \frac{B+1}{\sqrt{n}}\right), \tag{9}$$

we find that $\mathbb{P}[\mathcal{E}_D, W = 1] \le \epsilon_D(1-\epsilon_0)$, which implies $\mathbb{P}[\mathcal{E}_D|W = 1] \le \epsilon_D$. Again employing the Taylor's series expansion of $Q^{-1}(\cdot)$, this argument proves the achievability of rate

$$R_D = H(X|Y)+\sqrt{\frac{V(X|Y)}{n}}Q^{-1}(\epsilon_D(1-\epsilon_0))+\frac{\log n}{2n}+O\left(\frac{1}{n}\right)$$

with conditional error probability ϵ_D given $W = 1$. Similarly, expanding the conditional error probability $\mathbb{P}[\mathcal{E}_I|W = 0]$ as

$$\begin{aligned}\mathbb{P}[\mathcal{E}_I|W = 0] =& \frac{1}{\epsilon_0}\mathbb{P}[\mathcal{E}_I, W = 0]\\ \le& \frac{1}{\epsilon_0}(\mathbb{P}[\mathcal{E}_{I1}, W = 0] + \mathbb{P}[\mathcal{E}_{I2}, W = 0])\end{aligned}$$

and defining finite measure $\mathbb{V}$ on $\mathcal{X}^n \times \mathcal{Y}^n$ by $V_{X^nY^n}(x^n, y^n) = P_{\hat{Y}^n}(y^n)$, we find

$$\begin{aligned}&\mathbb{P}[\mathcal{E}_{I1}, W = 0]\\ &\le \mathbb{P}[\exists\bar{x}^n \in \mathcal{X}\backslash\{X^n\}, \lfloor\mathsf{f}(\bar{x}^n)\rfloor_{nR_D} = \lfloor\mathsf{f}(X^n)\rfloor_{nR_D},\\ &\quad \imath(\bar{x}^n|\hat{Y}^n) \le \log\gamma]\\ &\le \frac{1}{2^{nR_D}}\mathbb{V}[\imath(\bar{X}^n|\hat{Y}^n) \le \log\gamma]\end{aligned}$$

and

$$\begin{aligned}&\mathbb{V}[\imath(\bar{X}^n|\hat{Y}^n) \le \log\gamma]\\ &\quad = \gamma\sum_{\bar{x}^n, \hat{y}^n} P_{\hat{Y}^n}(\hat{y}^n)P^n_{X|Y}(\bar{x}^n|\hat{y}^n)\\ &\quad = \gamma,\end{aligned}$$

where the last equality follows since $P^n_{X|Y}P_{\hat{Y}^n}$ is *some* pmf on $\mathcal{X}^n \times \mathcal{Y}^n$. Note that measure $\mathbb{V}$ is unknown to the encoder and decoder and is used only for the purpose of analysis. Hence $\gamma = \frac{2^{nR_D}}{\sqrt{n}}$ implies $\mathbb{P}[\mathcal{E}_{I1}, W = 0] \le \frac{1}{\sqrt{n}}$. Using the same analysis for the maximum likelihood decoder and choosing

$$\begin{aligned}R_I =& H(X) + \sqrt{\frac{V(X)}{n}}Q^{-1}\left(\epsilon_I\epsilon_0 - \frac{B' + C' + 1}{\sqrt{n}}\right)\\ &- \frac{\log n}{2n} + \frac{\log C'}{n}\end{aligned}$$

gives $\mathbb{P}[\mathcal{E}_I, W = 0] \le \epsilon_I\epsilon_0$, which implies $\mathbb{P}[\mathcal{E}_I|W = 0] \le \epsilon_I$. Therefore, $\mathscr{R}'_{\text{in}}(n, \boldsymbol{\epsilon}) \subseteq \mathscr{R}^*_{USW}(n, \boldsymbol{\epsilon})$.

Remark 3: If the decoder for the unknown model is used when the distribution of the source and observation follows the worst-case model, the code does as well as the achievability part of Theorem 1 up to the first order. If the decoder for the unknown model is used when the true side information is available at the decoder, with probability $(1 - \epsilon_D)$, the source sequence is reconstructed correctly at a rate that asymptotically equals $H(X|Y)$.

Remark 4: Due to the unknown nature of $P_{ZW|XY}$, to date, $\mathscr{R}^*_{USW}(n, \boldsymbol{\epsilon}) \subseteq \mathscr{R}'_{\text{out}}(n, \boldsymbol{\epsilon})$ can only be guaranteed for an uninteresting outer region

$$\begin{aligned}&\mathscr{R}'_{\text{out}}(n, \boldsymbol{\epsilon})\\ &\triangleq \left\{\mathbf{R} : R_D \ge R^*_D - O\left(\frac{1}{n}\right), R_I \ge R^*_D - O\left(\frac{1}{n}\right)\right\}.\end{aligned}$$

B. Discussion

The achievability result of Theorem 2 only matches that of Theorem 1 up to the first order. The second order gap and the dependence of the dispersion term on ϵ_0 come from the possible but unknown dependence of W and Z^n on the source and the side information. The converse of Theorem 2 does not match the converse in Theorem 1 due to the unknown true joint distribution of the source and the side information. The main contribution of our code under the unknown model is that it allows us to take advantage of dependent side information roughly fraction $1-\epsilon_0$ of the time without allowing flaws in the side information (e.g., a sensor that is faulty fraction ϵ_0 of the time) to cause propagating errors in the source reconstruction.

V. Conclusion

This work introduces the problem of source coding with asymptotically unreliable side information at the decoder and analyzes the performance of a two-stage code under the cases of known and unknown joint distributions of the unreliable side information.

The strategy employed to avoid error propagation in this work may also be useful in systems that rely on successive interference cancellation. There, as here, error propagation is an important problem since errors can accumulate even when the error probability conditioned on correct cancellation is small. Success in avoiding error propagation due to unreliable side information suggests that it may also be possible to design codes that prevent errors from propagating while reaping the benefit from cancellations.

An important problem for future work is to optimize the strategy used to stop such error propagation when more is known about the unreliable side information. For example, we would like to bound the performance if the unreliable nature of the side information results from its representation using a code with bounded error probability and a given decoder.

REFERENCES

[1] R. Gallager, "A perspective on multiaccess channels," *IEEE Transactions on Information Theory*, vol. 31, no. 2, pp. 124–142, 1985.

[2] R. Shu-Kwan Cheng, "Stripping cdma-an asymptotically optimal coding scheme for l-out-of-k white gaussian channels," in *Proceedings of GLOBECOM'96. 1996 IEEE Global Telecommunications Conference*, 1996, pp. 142–146.

[3] D. Tse and P. Viswanath, *Fundamentals of wireless communication*. Cambridge, 2013.

[4] O. Gazi and A. A. Andi, "The effect of error propagation on the performance of polar codes utilizing successive cancellation decoding algorithm," *Advanced Electromagnetics*, vol. 8, no. 2, p. 114–120, Jun. 2019.

[5] S. Draper, "Universal incremental slepian-wolf coding," in *Proc. 42th Annual Allerton Conference on Communications, Control, and Computing*, Sept. 2004.

[6] S. Chen, M. Effros, and V. Kostina, "Lossless source coding in the point-to-point, multiple access, and random access scenarios," *IEEE Transactions on Information Theory*, vol. 66, no. 11, pp. 6688–6722, 2020.

[7] I. Kontoyiannis and S. Verdú, "Optimal lossless data compression: Non-asymptotics and asymptotics," *IEEE Transactions on Information Theory*, vol. 60, no. 2, pp. 777–795, 2014.

[8] L. Gavalakis and I. Kontoyiannis, "Fundamental limits of lossless data compression with side information," *IEEE Transactions on Information Theory*, vol. 67, no. 5, pp. 2680–2692, 2021.

[9] V. Y. F. Tan and O. Kosut, "On the dispersions of three network information theory problems," *IEEE Transactions on Information Theory*, vol. 60, no. 2, pp. 881–903, 2014.

[10] Y. Polyanskiy, H. V. Poor, and S. Verdu, "Channel coding rate in the finite blocklength regime," *IEEE Transactions on Information Theory*, vol. 56, no. 5, pp. 2307–2359, 2010.

[11] T. S. Han, *Information-Spectrum Methods in Information Theory*. Springer-Verlag Berlin Heidelberg, 2003.