2017 IEEE International Symposium on Information Theory (ISIT)

Strong Functional Representation Lemma and Applications to Coding Theorems

Cheuk Ting Li and Abbas El Gamal
Department of Electrical Engineering, Stanford University
Email: ctli@stanford.edu, abbas@ee.stanford.edu

***Abstract*—This paper shows that for any random variables X and Y, it is possible to represent Y as a function of (X, Z) such that Z is independent of X and $I(X;Z|Y) \leq \log(I(X;Y)+1)+4$. We use this strong functional representation lemma (SFRL) to establish a tighter bound on the rate needed for one-shot exact channel simulation than was previously established by Harsha et. al., and to establish achievability results for one-shot variable-length lossy source coding and multiple description coding. We also show that the SFRL can be used to reduce the channel with state noncausally known at the encoder to a point-to-point channel, which provides a simple achievability proof of the Gelfand-Pinsker theorem. Finally we present an example in which the SFRL inequality is tight to within 5 bits.**

***Index Terms*—Functional representation lemma, channel simulation, one-shot achievability, lossy source coding, channel with state.**

I. Introduction

The functional representation lemma [1, p. 626] states that for any random variables X and Y, there exists a random variable Z independent of X such that Y can be represented as a function of X and Z. This result has been used to establish several results in network information theory beginning with the early work of Hajek and Pursley on the broadcast channel [2] and Willems and van der Meulen on the multiple access channel with cribbing encoders [3]. In this paper, we strengthen this result by showing that for any X and Y, there exists a Z independent of X such that Y is a function of X and Z, *and*

$$I(X;Z|Y) \leq \log(I(X;Y)+1)+4.$$

We can use this strong functional representation lemma (SFRL) together with an optimal prefix code such as Huffman code to establish one-shot, *variable-length* coding results for channel simulation [4], Shannon's lossy source coding [5], multiple description coding [6], [7] and lossy Gray–Wyner system [8] (which is discussed in [9] for space limitation). We then show how the SFRL can be used to reduce the channel with state known at the encoder to a point-to-point channel, providing a simple proof to the Gelfand-Pinsker theorem [10]. The asymptotic block coding counterparts of these one-shot results can be readily obtained by converting the variable-length code into a block code and incurring an error probability that vanishes as block length approaches infinity.

A weaker form of the SFRL for discrete random variables can be obtained using the result by Harsha et. al. [4] on the one-shot exact channel simulation with unlimited common randomness. Assuming the input X has a given pmf, then [4] implies that $I(X;Z|Y) \leq (1+\epsilon)\log(I(X;Y)+1)+c_\epsilon$ is achievable, where $\epsilon > 0$ and c_ϵ is a function of ϵ. SFRL strengthens this result in two way; first it provides a tighter bound, and second it generalizes the bound to random variables with arbitrary distributions. This is obtained via a new proof that uses a construction we refer to as the *Poisson functional representation* instead of using rejection sampling as in [4]. Perhaps more importantly, we are the first to show that the result in [4] can be considered as a strengthened functional representation lemma, which led us to explore applications in source and channel coding.

One-shot achievability results using fixed length (random) coding have been recently established for lossy source coding and several setting in network information theory. In [11], Liu, Cuff and Verdú established a one-shot achievability result for lossy source coding using channel resolvability. One-shot quantum lossy source coding settings were investigated by Datta et. al. [12]. In [13], Verdú introduced non-asymptotic packing and covering lemmas and used them to establish one-shot achievability results for several settings including Gelfand-Pinsker. In [14], Liu, Cuff and Verdú proved a one-shot mutual covering lemma and used it to establish a one-shot achievability result for the broadcast channel. In [15], Watanabe, Kuzuoka and Tan established several one-shot achievability results for coding with side-information (including Gelfand-Pinsker). In [16], Yassaee, Aref and Gohari established several one-shot achievability results, including Gelfand-Pinsker and multiple description coding. Most of these results are stated in terms of information density and various other quantities. In contrast, our one-shot achievability results using variable-length codes are all stated in terms only of mutual information. Moreover, given the SFRL, our proofs are generally simpler.

Several variable-length lossy source coding settings have been previously studied, e.g., see [17], [18], [19], [20]. Some of these works concern the universal setting in which the distribution of the source is unknown, hence the use of variable-length codes is justified. In contrast, the reason we consider variable-length code in this paper is that it allows us to give one-shot results that subsume their asymptotic fixed-length counterparts.

The omitted proofs and derivations can be found in [9].

Notation

We assume $\log$ is base 2 and the entropy H is in bits. We write $X_a^b = (X_a, \ldots, X_b)$, $X^n = X_1^n$ and $[a:b] = [a,b] \cap \mathbb{Z}$.

For discrete X, we write the probability mass function as p_X. For continuous X, we write the probability density function as f_X. For general random variable X, we write the probability measure (push-forward measure by X) as P_X.

II. Strong Functional Representation Lemma

The main result in this paper is given in the following.

Theorem 1 (Strong functional representation lemma)**.** *For any pair of random variables $(X,Y) \sim \mathsf{P}_{XY}$ with $I(X;Y) < \infty$, there exists a random variable Z independent of X such that Y can be expressed as a function $g(X,Z)$ of X and Z, and*

$$I(X;Z|Y) \le \log(I(X;Y)+1) + 4.$$

Moreover, if X and Y are discrete with cardinalities $|\mathcal{X}|$ and $|\mathcal{Y}|$, respectively, then $|\mathcal{Z}| \le |\mathcal{X}|(|\mathcal{Y}|-1)+2$.

Note that SFRL can be applied conditionally; given $\mathsf{P}_{XY|U}$, we can represent Y as a function $g(X,Z,U)$ such that Z is independent of (X,U) and

$$I(X;Z|Y,U) \le \log\left(I(X;Y|U)+1\right) + 4.$$

The reason we can have a Z independent of U is that by the functional representation lemma, we can represent Z as a function of U and $\tilde{Z}$ such that $\tilde{Z}$ is independent of U and use $\tilde{Z}$ instead of Z.

Note that SFRL applies to general distributions P_{XY}. Although $H(Y)$ may be infinite, the cardinality of Y conditioned on Z can still be countable and $H(Y|Z)$ can be finite. Note that $Z \perp\!\!\!\perp X$ and $H(Y|X,Z) = 0$ imply that $I(X;Z|Y) = H(Y|Z) - I(X;Y)$. Hence the SFRL implies the existence of $Z \perp\!\!\!\perp X$ such that $H(Y|Z)$ is close to $I(X;Y)$. For simplicity of presentation, we first prove the SFRL for discrete Y.

First consider the following construction of Z and the function g which we use in the proof of the SFRL for discrete Y.

Definition 1 (Exponential functional representation)**.** Let X and Y be random variables, where $Y \in \{1,\ldots,|\mathcal{Y}|\}$ and $|\mathcal{Y}|$ is finite or countably infinite. The exponential functional representation of Y given X is defined as $Y = g_{X\to Y}(X, Z^{|\mathcal{Y}|})$, where $Z^{|\mathcal{Y}|}$ is a sequence of i.i.d. $\mathrm{Exp}(1)$ random variables independent of X, and

$$g_{X\to Y}(x, z^{|\mathcal{Y}|}) = \arg\min_{y\in\mathcal{Y}} \frac{z_y}{p_{Y|X}(y|x)}.$$

Since the $\arg\min$ of independent exponential random variables with different rates has a pmf proportional to the rates, we have $g_{X\to Y}(x, Z^{|\mathcal{Y}|}) \sim p_{Y|X}(\cdot|x)$.

We now proceed to prove Theorem 1 for discrete Y by showing that the exponential functional representation satisfies the constraints.

Proof: Let $\Phi_y = Z_y/p_Y(y)$,

$$\Theta = \inf_y \frac{Z_y}{p_{Y|X}(y|X)},$$

and K be the index of Φ_Y in $\{\Phi_y\}_{y\in\mathcal{Y}}$ sorted in ascending order (hence $|\{y : \Phi_y < \Phi_Y\}| = K-1$ with probability 1). Since Y is a function of $Z^{|\mathcal{Y}|}$ and K, we have $H(Y|Z^{|\mathcal{Y}|}) \le H(K)$. We now proceed to bound $H(K)$. Let $r(x,y) = \frac{p_{Y|X}(y|x)}{p_Y(y)}$. Since $\Theta|\{X=x, Y=y\} \sim \mathrm{Exp}(1)$,

$$\begin{aligned}
&\mathsf{E}\left[\log K | X=x\right] \\
&= \sum_y p_{Y|X}(y|x)\mathsf{E}\left[\log K \mid X=x, Y=y\right] \\
&= \sum_y p(y|x) \int_0^\infty e^{-\theta}\mathsf{E}\left[\log K \mid X=x, Y=y, \Theta=\theta\right] d\theta.
\end{aligned}$$

Consider

$$\begin{aligned}
&\mathsf{E}[\log K \mid X=x, Y=y, \Theta=\theta] \\
&= \mathsf{E}\left[\log\left(\left|\left\{y' \ne y : \Phi_{y'} < \frac{\theta p(y|x)}{p(y)}\right\}\right| + 1\right) \,\middle|\, \frac{Z_{y'}}{p(y'|x)} \ge \theta \;\forall\, y'\right] \\
&\le \log\left(\mathsf{E}\left[\left|\left\{y' \ne y : \Phi_{y'} < \frac{\theta p(y|x)}{p(y)}\right\}\right| \,\middle|\, \frac{Z_{y'}}{p(y'|x)} \ge \theta \;\forall\, y'\right] + 1\right) \\
&= \log\left(\sum_{y'\ne y} \mathsf{P}\left\{\Phi_{y'} < \frac{\theta p(y|x)}{p(y)} \;\middle|\; \frac{Z_{y'}}{p(y'|x)} \ge \theta\right\} + 1\right) \\
&= \log\left(\sum_{y' :\, r(x,y')<r(x,y)} \left(1 - e^{-\theta p(y')\left(r(x,y)-r(x,y')\right)}\right) + 1\right) \\
&\le \log\left(\sum_{y' :\, r(x,y')<r(x,y)} \theta p(y')\left(r(x,y) - r(x,y')\right) + 1\right) \\
&\le \log\left(\theta r(x,y) + 1\right).
\end{aligned}$$

Hence

$$\begin{aligned}
&\mathsf{E}\left[\log K | X=x\right] \\
&\le \sum_y p(y|x) \int_0^\infty e^{-\theta} \log\left(\theta r(x,y)+1\right) d\theta \\
&\le \sum_y p(y|x) \log\left(r(x,y)+1\right) \\
&= \Bigg(\sum_{y:\, r(x,y)\ge 1} p(y|x)\log\left(r(x,y)+1\right) \\
&\quad + \sum_{y:\, r(x,y)<1} p(y|x)\log\left(r(x,y)+1\right)\Bigg) \\
&\le \left(\sum_{y:\, r(x,y)\ge 1} p(y|x)\left(\log r(x,y)+1\right) + \sum_{y:\, r(x,y)<1} p(y|x)\right) \\
&= \sum_{y:\, r(x,y)\ge 1} p(y|x)\log r(x,y) + 1 \\
&= D(p_{Y|X}(\cdot|x)\|\, p_Y) - \sum_{y:\, r(x,y)<1} p(y|x)\log r(x,y) + 1 \\
&\le D(p_{Y|X}(\cdot|x)\|\, p_Y) + e^{-1}\log e + 1,
\end{aligned}$$

where the last inequality follows from Appendix A in [4]. Therefore $\mathsf{E}\left[\log K\right] \le I(X;Y) + e^{-1}\log e + 1$. By the maximum entropy distribution subject to a given $\mathsf{E}\left[\log K\right]$,

$$H(K) \le \mathsf{E}\left[\log K\right] + \log\left(\mathsf{E}\left[\log K\right] + 1\right) + 1.$$

Hence

$$\begin{aligned} H(K) &\le I(X;Y) + e^{-1}\log e + 2 \\ &\quad + \log\big(I(X;Y) + e^{-1}\log e + 2\big) \\ &< I(X;Y) + \log\left(I(X;Y)+1\right) + 4. \end{aligned}$$

To prove the cardinality bound, first note that if $|\mathcal{X}|$, $|\mathcal{Y}|$ are finite, then $|\mathcal{Z}| \le |\mathcal{Y}|^{|\mathcal{X}|}$ can be assumed to be finite since it is the number of different functions $x \mapsto g_{X\to Y}(x,z)$ for different z. To further reduce the cardinality, we apply Carathéodory's theorem on the $(|\mathcal{X}|(|\mathcal{Y}|-1)+1)$-dimensional vectors with entries $H(Y|Z=z)$ and $p(x,y|z)$ for $x \in \{1,\dots,|\mathcal{X}|\}$, $y \in \{1,\dots,|\mathcal{Y}|-1\}$; see [21], [22]. ∎

Remark 1. To extend the proof of Theorem 1 to arbitrary random variables, we use the following Poisson functional representation. Let $0 \le T_1 \le T_2 \le \cdots$ be a Poisson point process with rate 1 (i.e., the increments $T_i - T_{i-1}$ are i.i.d. $\mathrm{Exp}(1)$ for $i = 1,2,\dots$ with $T_0 = 0$), and $\tilde{Y}_1, \tilde{Y}_2, \dots$ be i.i.d. with $\tilde{Y}_1 \sim \mathsf{P}_Y$. Take $Z = \{(T_i, \tilde{Y}_i)\}_{i=1,2,\dots}$, i.e., a marked Poisson point process. Define

$$g_{X\to Y}(x,\, \{(t_i,\tilde{y}_i)\}) = \tilde{y}\left(\arg\min_i\, t_i \cdot \frac{d\,\mathsf{P}_Y}{d\,\mathsf{P}_{Y|X}(\cdot|x)}(\tilde{y}_i)\right),$$

where we write $\tilde{y}(k) = \tilde{y}_k$ for readability. Theorem 1 continues to hold for Poisson functional representation; see [9] for the proof.

A natural question to ask is whether the SFRL is tight for some X, Y, and whether the log term is necessary. The following proposition shows that there exists a sequence of (X,Y) for which SFRL is tight within 5 bits. The proof of this proposition can be found in [9].

Proposition 1. *For every $\alpha \ge 0$, there exists discrete X, Y such that $I(X;Y) \ge \alpha$ and*

$$\inf_{Z:\, Z \perp\!\!\!\perp X,\, H(Y|X,Z)=0} I(X;Z|Y) \ge \log(I(X;Y)+1) - 1.$$

III. One-shot Channel Simulation

Channel simulation aims to find the minimum amount of communication over a noiseless channel needed to simulate a memoryless channel $\mathsf{P}_{Y|X}$. Several settings of this problem have been studied, e.g., see [23], [24], [25]. Consider the one-shot channel simulation with unlimited common randomness setup [4] in which Alice and Bob share unlimited common randomness W. Alice observes $X \sim \mathsf{P}_X$ and sends a prefix-free description M to Bob such that Bob can generate Y (from M and W) according to a prescribed conditional distribution $\mathsf{P}_{Y|X}$. The problem is to find the minimum expected description length of M, $\mathsf{E}\,[L(M)]$, needed. It is straightforward to show that $\mathsf{E}\,[L(M)] \ge I(X;Y)$. In [4], it is shown that for X and Y discrete, the following is achievable:

$$\mathsf{E}\,[L(M)] \le I(X;Y) + (1+\epsilon)\log(I(X;Y)+1) + c_\epsilon,$$

where $\epsilon > 0$ and c_ϵ is a function of ϵ. It is possible to remove ϵ by using a power-law code instead of a universal code in [4], though the unspecified constant in [4] appears to be large.

We now show that the SFRL provides a tighter upper bound on $\mathsf{E}\,[L(M)]$ that applies to arbitrary (not only discrete) memoryless channels. By the SFRL, there exists a Z independent of X such that $Y = g_{X\to Y}(X,Z)$ and

$$H(Y\,|\,Z) \le I(X;Y) + \log(I(X;Y)+1) + 4.$$

We use $W = Z$ as the common randomness. Upon observing $X = x$, Alice generates $Y \sim \mathsf{P}_{Y|X}(\cdot|x)$ and encodes Y using a Huffman code for the pmf $p_{Y|Z}(\cdot|Z)$ into the description M (Y can be arbitrary but by the SFRL $Y|\{Z=z\}$ is discrete). Bob recovers Y from M and Z. The expected length is

$$\mathsf{E}\,[L(M)] \le I(X;Y) + \log(I(X;Y)+1) + 5.$$

Remark 2. In [4], the setting in which $X = x$ is an arbitrary input (instead of $X \sim p_X$) is studied. It is shown that

$$\mathsf{E}\,[L(M)] \le C + (1+\epsilon)\log(C+1) + c_\epsilon$$

for all $x \in \mathcal{X}$ is achievable, where C is the capacity of the channel $p_{Y|X}$ and c_ϵ is a function of ϵ.

The exponential functional representation can still be applied to this setting. If we encode K (defined in the proof of the SFRL) into M using the optimal prefix-free code for the power-law distribution $q(k) \propto k^{-\lambda}$, where $\lambda = 1 + 1/(C + e^{-1}\log e + 1)$, then by the same argument in the proof of the SFRL, and Claim 3.1 in [4], we can achieve

$$\mathsf{E}\,[L(M)] \le C + \log(C+1) + 5.$$

IV. Lossy Source Coding and Multiple Description Coding

We use the SFRL to establish one-shot achievability results for two source coding settings.

A. Lossy source coding

Consider the following one-shot variable-length lossy source coding problem. We are given a random variable (source) $X \in \mathcal{X}$ with $X \sim \mathsf{P}_X$, a reproduction alphabet $\mathcal{Y}$, and a distortion function $d : \mathcal{X} \times \mathcal{Y} \to [0,\infty]$ (note that X, Y can be arbitrary, and $d(x,y)$ can be infinite). Given X, the encoder selects $\tilde{Y} \in \mathcal{Y}$ and encodes it using a prefix-free code into $M \in \{0,1\}^*$. The decoder recovers $\tilde{Y}$ from M. Let $\bar{R} = \mathsf{E}[L(M)]$ be the expected length of the description M and $\mathsf{E}[d(X,\tilde{Y})]$ be the average distortion of representing X by $\tilde{Y}$. An expected length-distortion pair $(\bar{R}, D)$ is achievable if there exists a variable-length code with expected description length $\bar{R}$ such that $\mathsf{E}[d(X,\tilde{Y})] \le D$. In the following we use the SFRL to establish a set of achievable $(\bar{R}, D)$ pairs.

Theorem 2. *A pair $(\bar{R}, D)$ is achievable for the one-shot variable-length lossy source coding problem with source $X \sim \mathsf{P}_X$, reproduction alphabet $\mathcal{Y}$, and distortion measure $d(x,y)$ if*

$$\bar{R} > R(D) + \log(R(D)+1) + 6,$$

where

$$R(D) = \inf_{\mathsf{P}_{Y|X}:\, \mathsf{E}[d(X,Y)] \le D} I(X;Y)$$

is the (asymptotic) rate-distortion function [5].

Proof: Let Y be the random variable that attains $\mathsf{E}[d(X,Y)] \le D$ and $I(X;Y) \le R(D)+\epsilon$. By SFRL, there exists Z independent of X such that $Y = g_{X\to Y}(X,Z)$ and

$$H(g_{X\to Y}(X,Z)|Z) \le I(X;Y)+\eta,$$

where $\eta = \log(I(X;Y)+1)+4$. Consider the set

$$A = \{(H(g_{X\to Y}(X,z)),\, \mathsf{E}_X\left[d(X,\, g_{X\to Y}(X,z))\right]) \,:\, z\in\mathcal{Z}\}.$$

Since $(H(g_{X\to Y}(X,Z)|Z),\, \mathsf{E}\left[d(X,Y)\right])$ is a weighted average of the points in A, it is in the convex hull of A. By Carathéodory's theorem, there exists z_0, z_1 and $\lambda\in[0,1]$ with

$$\begin{aligned}
&(1-\lambda)H(g_{X\to Y}(X,z_0)) + \lambda H(g_{X\to Y}(X,z_1))\\
&\le H(g_{X\to Y}(X,Z)|Z) \le I(X;Y)+\eta,\\
&(1-\lambda)\mathsf{E}_X[d(X,\, g_{X\to Y}(X,z_0))]+\lambda\mathsf{E}_X[d(X,\, g_{X\to Y}(X,z_1))]\\
&\le \mathsf{E}\left[d(X,Y)\right].
\end{aligned}$$

Note that to satisfy the above inequalities, we need one point less than stated in Carathéodory's theorem. Take $Q\sim \mathrm{Bern}(\lambda)$, $\tilde{Y} = g_{X\to Y}(X,z_Q)$. Then

$$H(\tilde{Y}) \le H(\tilde{Y}|Q)+1 \le I(X;Y)+\eta+1.$$

We use a Huffman code to encode $\tilde{Y}$ and obtain an expected length $\bar{R}\le H(\tilde{Y})+1$. The result follows by letting $\epsilon\to 0$. ∎

Although the above achievability proof does not use random coding, it can be interpreted as using the following *soft random coding* scheme (for discrete Y).

Soft codebook generation. The random variable Z produced by the SFRL in the proof of Theorem 2 represents the choice of the codebook. We select a "soft codebook" by fixing $Z^{|\mathcal{Y}|} = z^{|\mathcal{Y}|}$. Unlike conventional codebook $\mathcal{C}\subseteq Y$ in which each y can either be in $\mathcal{C}$ or not, a soft codebook $z^{|\mathcal{Y}|}$ assigns to each y a weight $w_y = p_Y(y)/z_y$, which indicates the likelihood that y is used.

Encoding. The encoder observes x and finds

$$\tilde{y} = \arg\max_y\, w_y\cdot\frac{p_{Y|X}(y|x)}{p_Y(y)}.$$

It then finds k, the index of $w_{\tilde{y}}$ in $\{w_y\}_{y\in\mathcal{Y}}$ sorted in descending order, and encodes k using an optimal prefix-free code for the power-law distribution $q(k)\propto k^{-\lambda}$, where $\lambda = 1+1/(I(X;Y)+e^{-1}\log e+1)$. This is analogous to the case in conventional codebook generation in which we find the closest $\tilde{y}\in\mathcal{C}$ to x and encodes it into its index in $\mathcal{C}$. Here we use a prefix-free code over the positive integers to encode the index into the description m because the number of possible codewords $\tilde{y}$ (which is typically the entire $\mathcal{Y}$) is large, but those with large w_y are more likely to be used so they are assigned shorter descriptions.

Decoding. The decoder receives m, recovers k, then finds $\tilde{y}$ at the index k in $\{w_y\}_{y\in\mathcal{Y}}$ sorted in descending order.

A related one-shot variable-length lossy source coding setting with a constraint on the probability that the distortion exceed certain level (instead of average distortion) was studied in [20]. In [26], a result similar to Theorem 2 is given in the context of epsilon entropy.

The finite blocklength variable-length lossy source coding problem [17] concerns the case in which the source is memoryless and average per symbol distortion $d(x^n,y^n) = (1/n)\sum_i d(x_i,y_i)$. In [27] it is shown that the expected per symbol description length $\bar{R}/n = R(D)+(1+o(1))(1/n)\log n$ is achievable via d-semifaithful codes [28] with $d(X^n,\tilde{Y}^n)\le D$ surely. Applying Theorem 2 to X^n,

$$\begin{aligned}
\bar{R}/n &= R(D)+(1/n)(\log(nR(D)+1)+6)\\
&= R(D)+(1+o(1))(1/n)\log n.
\end{aligned}$$

Hence we achieve the same redundancy as [27] albeit under the expected distortion constraint instead of the stronger sure distortion constraint using the d-semifaithful codes.

We can use Theorem 2 to establish the achievability of Shannon's (asymptotic) lossy source coding theorem [5], assuming there exists a symbol $y_0\in\mathcal{Y}$ with finite $d(x,y_0)$ for all x. First note that the redundancy $(1+o(1))(1/n)\log n$ in the finite block length extension can be made arbitrarily small, hence $\bar{R}/n$ can be made arbitrarily close to $R(D)$. Now we use the finite block length scheme over l blocks of n source symbols each of length n (for a total block length of nl). By the law of large numbers, the probability that the total description length is greater than $nl(R(D)+\epsilon)$ tends to 0 as $l\to\infty$. Hence, we can construct a fixed length code out of the variable-length code by simply discarding descriptions longer than $nl(R(D)+\epsilon)$ and assigning the reconstruction sequence $(y_0,\dots,y_0)$ to the discarded descriptions.

B. Multiple Description Coding

In this section, we use the SFRL to establish a one-shot inner bound for the variable-length multiple description coding problem, which yields an alternative proof of the El Gamal-Cover inner bound [6] and the Zhang-Berger inner bound [7], [29] in the asymptotic regime. The encoder observes $X\sim \mathsf{P}_X$ and produces two prefix-free descriptions $M_1, M_2\in\{0,1\}^*$. Decoder i observes M_i and generates $\tilde{Y}_i$ with distortion $d_i(X,\tilde{Y}_i)$ $(i=1,2)$. Decoder 0 observes M_1 and M_2 and produces $\tilde{Y}_0$ with distortion $d_0(X,\tilde{Y}_0)$. A tuple $(\bar{R}_1,\bar{R}_2,D_0,D_1,D_2)$ is said to be achievable if there exists a scheme with $\mathsf{E}[L(M_i)]\le\bar{R}_i$ and $\mathsf{E}[d_i(X,\tilde{Y}_i)]\le D_i$.

Theorem 3. *The tuple $(\bar{R}_1,\bar{R}_2,D_0,D_1,D_2)$ is achievable if*

$$\begin{aligned}
&\bar{R}_i \ge I(X;Y_i,U)+2\eta \quad \textit{for } i=1,2,\\
&\bar{R}_1+\bar{R}_2 \ge I(X;Y_0,Y_1,Y_2|U)+2I(X;U)+I(Y_1;Y_2|U)+5\eta,\\
&D_i \ge \mathsf{E}[d_i(X,Y_i)] \quad \textit{for } i=0,1,2
\end{aligned}$$

for some $\mathsf{P}_{U,Y_0,Y_1,Y_2|X}$, where

$$\eta = \log\left(I(X;Y_0,Y_1,Y_2,U)+I(Y_1;Y_2|U)+1\right)+7.$$

The proof can be found in [9]. Note that the only difference between the above region and Zhang-Berger inner bound is the addition of η, which grows like $\log n$ if we consider X^n.

V. Achievability of Gelfand–Pinsker

In this section, we use the SFRL to prove the achievability part of the Gelfand-Pinsker theorem [10] for discrete memoryless channels with discrete memoryless state $p_S p_{Y|X,S}$, where the state is noncausally available at the encoder. The asymptotic capacity of this setting is

$$C_{\mathrm{GP}} = \max_{p_{U|S},\, x(u,s)} \left(I(U;Y) - I(U;S) \right).$$

We show the achievability of any rate below C_{GP} directly by using the SFRL to reduce the channel to a point-to-point memoryless channel. Fix $p_{U|S}$ and $x(u,s)$ that attain the capacity. Applying the SFRL to S, U, there exists a random variable $V \perp\!\!\!\perp S$ such that

$$H(U|V) \le I(U;S) + \log(I(U;S)+1) + 4.$$

Note that

$$\begin{aligned} I(V;Y) &= I(U;Y) - I(U;Y|V) + I(V;Y|U) \\ &\ge I(U;Y) - I(U;S) - \log(I(U;S)+1) - 4. \end{aligned}$$

Hence we have constructed a memoryless point-to-point channel $p_{Y|V}$ with achievable rate close to $I(U;Y) - I(U;S)$.

For n channel uses, let $U^n | \{S^n = s^n\} \sim \prod_i p_{U|S}(u_i|s_i)$. The SFRL applied to S^n, U^n gives

$$I(V;Y^n) \ge nI(U;Y) - nI(U;S) - \log(nI(U;S)+1) - 4.$$

Now we use the channel $p_{Y^n|V}$ l times (for a total block length of nl). By the channel coding theorem, we can communicate $l(nI(U;Y) - nI(U;S) - \log(nI(U;S)+1) - 4) - o(l)$ bits with error probability that tends to 0 as $l \to \infty$. Letting $n \to \infty$ completes the proof.

In the above proof, we see that the SFRL can be used to convert a channel with state into a point-to-point channel by "orthogonalizing" the auxiliary input U and the state S. The point-to-point channel can be constructed explicitly via exponential functional representation. This construction can be useful for designing codes for channels with state based on codes for point-to-point channels. It is interesting to note that this reduction makes the achievability proof for the Gelfand–Pinsker quite similar to that for the causal case in which the channel is reduced to a point-to-point channel using the "Shannon strategy" (see [1, p. 176]).

Marton's inner bound for broadcast channels with private messages [30] can be proved using the SFRL similarly.

References

[1] A. El Gamal and Y.-H. Kim, *Network information theory*. Cambridge university press, 2011.

[2] B. Hajek and M. Pursley, "Evaluation of an achievable rate region for the broadcast channel," *IEEE Transactions on Information Theory*, vol. 25, no. 1, pp. 36–46, Jan 1979.

[3] F. Willems and E. van der Meulen, "The discrete memoryless multiple-access channel with cribbing encoders," *IEEE Transactions on Information Theory*, vol. 31, no. 3, pp. 313–327, May 1985.

[4] P. Harsha, R. Jain, D. McAllester, and J. Radhakrishnan, "The communication complexity of correlation," *IEEE Trans. Info. Theory*, vol. 56, no. 1, pp. 438–449, Jan 2010.

[5] C. E. Shannon, "Coding theorems for a discrete source with a fidelity criterion," in *IRE Int. Conv. Rec.*, 1959, vol. 7, part 4, pp. 142–163, reprint with changes (1960). In R. E. Machol (ed.) *Information and Decision Processes*, pp. 93–126. McGraw-Hill, New York.

[6] A. El Gamal and T. M. Cover, "Achievable rates for multiple descriptions," *IEEE Trans. Inf. Theory*, vol. 28, no. 6, pp. 851–857, 1982.

[7] Z. Zhang and T. Berger, "New results in binary multiple descriptions," *IEEE Trans. Inf. Theory*, vol. 33, no. 4, pp. 502–521, 1987.

[8] R. M. Gray and A. D. Wyner, "Source coding for a simple network," *Bell Syst. Tech. J.*, vol. 53, no. 9, pp. 1681–1721, 1974.

[9] C. T. Li and A. El Gamal, "Strong functional representation lemma and applications to coding theorems," *arXiv preprint*, 2017. [Online]. Available: http://arxiv.org/abs/1701.02827

[10] S. I. Gelfand and M. S. Pinsker, "Coding for channel with random parameters," *Probl. Control Inf. Theory*, vol. 9, no. 1, pp. 19–31, 1980.

[11] J. Liu, P. Cuff, and S. Verdu, "Resolvability in E_γ with applications to lossy compression and wiretap channels," in *2015 IEEE International Symposium on Information Theory (ISIT)*, June 2015, pp. 755–759.

[12] N. Datta, J. M. Renes, R. Renner, and M. M. Wilde, "One-shot lossy quantum data compression," *IEEE Transactions on Information Theory*, vol. 59, no. 12, pp. 8057–8076, Dec 2013.

[13] S. Verdú, "Non-asymptotic achievability bounds in multiuser information theory," in *Communication, Control, and Computing (Allerton), 2012 50th Annual Allerton Conference on*, Oct 2012, pp. 1–8.

[14] J. Liu, P. Cuff, and S. Verdú, "One-shot mutual covering lemma and marton's inner bound with a common message," in *2015 IEEE International Symposium on Information Theory (ISIT)*, June 2015, pp. 1457–1461.

[15] S. Watanabe, S. Kuzuoka, and V. Y. F. Tan, "Non-asymptotic and second-order achievability bounds for source coding with side-information," in *2013 IEEE International Symposium on Information Theory*, July 2013, pp. 3055–3059.

[16] M. H. Yassaee, M. R. Aref, and A. Gohari, "A technique for deriving one-shot achievability results in network information theory," in *Information Theory Proceedings (ISIT), 2013 IEEE International Symposium on*, July 2013, pp. 1287–1291.

[17] J. T. Pinkston, *Encoding independent sample information sources*. Research Laboratory of Electronics, Massachusetts Inst. of Technology, 1967.

[18] M. Pursley and L. Davisson, "Variable rate coding for nonergodic sources and classes of ergodic sources subject to a fidelity constraint," *IEEE Transactions on Information Theory*, vol. 22, no. 3, pp. 324–337, May 1976.

[19] K. Mackenthun and M. Pursley, "Variable-rate universal block source coding subject to a fidelity constraint," *IEEE Transactions on Information Theory*, vol. 24, no. 3, pp. 349–360, May 1978.

[20] V. Kostina, Y. Polyanskiy, and S. Verdú, "Variable-length compression allowing errors," *IEEE Transactions on Information Theory*, vol. 61, no. 8, pp. 4316–4330, 2015.

[21] R. Ahlswede and J. Körner, "Source coding with side information and a converse for degraded broadcast channels," *IEEE Trans. Inf. Theory*, vol. 21, no. 6, pp. 629–637, 1975.

[22] A. D. Wyner and J. Ziv, "The rate–distortion function for source coding with side information at the decoder," *IEEE Trans. Inf. Theory*, vol. 22, no. 1, pp. 1–10, 1976.

[23] C. H. Bennett, P. W. Shor, J. Smolin, and A. V. Thapliyal, "Entanglement-assisted capacity of a quantum channel and the reverse Shannon theorem," *IEEE Trans. Info. Theory*, vol. 48, no. 10, pp. 2637–2655, 2002.

[24] P. Cuff, "Distributed channel synthesis," *IEEE Trans. Info. Theory*, vol. 59, no. 11, pp. 7071–7096, 2013.

[25] C. H. Bennett, I. Devetak, A. W. Harrow, P. W. Shor, and A. Winter, "The quantum reverse shannon theorem and resource tradeoffs for simulating quantum channels," *IEEE Trans. Info. Theory*, vol. 60, no. 5, pp. 2926–2959, May 2014.

[26] E. C. Posner and E. R. Rodemich, "Epsilon entropy and data compression," *The Annals of Mathematical Statistics*, pp. 2079–2125, 1971.

[27] Z. Zhang, E. h. Yang, and V. K. Wei, "The redundancy of source coding with a fidelity criterion. 1. known statistics," *IEEE Transactions on Information Theory*, vol. 43, no. 1, pp. 71–91, Jan 1997.

[28] D. S. Ornstein and P. C. Shields, "Universal almost sure data compression," *The Annals of Probability*, pp. 441–452, 1990.

[29] R. Venkataramani, G. Kramer, and V. K. Goyal, "Multiple description coding with many channels," *IEEE Trans. Inf. Theory*, vol. 49, no. 9, pp. 2106–2114, Sep. 2003.

[30] K. Marton, "A coding theorem for the discrete memoryless broadcast channel," *IEEE Trans. Inf. Theory*, vol. 25, no. 3, pp. 306–311, 1979.