Nested Lattice Codes for Arbitrary Continuous Sources and Channels

Aria G. Sahebi and S. Sandeep Pradhan
Department of Electrical Engineering and Computer Science,
University of Michigan, Ann Arbor, MI 48109, USA.
Email: ariaghs@umich.edu, pradhanv@umich.edu

***Abstract*—In this paper, we show that nested lattice codes achieve the capacity of arbitrary continuous channels with or without non-causal state information at the transmitter. We also show that nested lattice codes are optimal for source coding with or without non-causal side information at the receiver for arbitrary continuous sources. We show the optimality of lattice codes for the Gelfand-Pinsker and Wyner-Ziv problems in their most general settings.**

I. Introduction

Lattice codes for continuous sources and channels are the analogue of linear codes for discrete sources and channels and play an important role in information theory and communications. Linear/lattice and nested linear/lattice codes have been used in many communication settings to improve upon the existing random coding bounds [2], [10]–[13], [16], [18], [20].

In [2] and [12] the existence of lattice codes satisfying Shannon's bound has been shown. These results have been generalized and the close relation between linear and lattice codes has been pointed out in [13]. In [24], several results regarding lattice quantization noise in high resolution has been derived and the problem of constructing lattices with an arbitrary quantization noise distribution has been studied in [7].

Nested lattice codes were introduced in [26] where the concept of structured binning is presented. Nested linear/lattice codes are important because in many communication problems, specially multi-terminal settings, such codes can be superior in average performance compared to random codes [11]. It has been shown in [25] that nested lattice codes are optimal for the Wyner-Ziv problem when the source and side information are jointly Gaussian. The dual problem of channel coding with state information has been addressed in [4]–[6], [21] and the optimality of lattice codes for Gaussian channels has been shown. In [3], it was shown that random linear codes provide good binning schemes for general Slepian-Wolf coding.

In a recent work [17], it has been shown that nested linear codes are optimal for arbitrary discrete memoryless channels with state information at the transmitter. In this paper we focus on two problems: 1) The point to point channel coding with state information at the encoder (the Gelfand-Pinsker problem [8]) and 2) Lossy source coding with side information at the decoder (the Winer-Ziv problem [22], [23]). We consider these two problems in their most general settings i.e. when the source and the channel are arbitrary but memoryless. We use nested lattice codes with *joint typicality decoding/encoding* rather than lattice decoding. We consider a lattice code ensemble with simpler random dithers and without modulo-Λ transformations as used in [25]. We show that in both settings, from an information-theoretic point of view, nested lattice codes are optimal.

This work was supported by NSF grants CCF-0915619 and CCF-1116021.

The paper is organized as follows: in Section II we present the required preliminaries and introduce our notation. In Section III we show the optimality of nested lattice codes for channels with state information (the Gelfand-Pinsker problem). We briefly present the optimality of nested lattice codes for source coding with side information (the Wyner-Ziv problem) in Section IV and we conclude in Section V.

II. Preliminaries

1) Channel Model: We associate two sets $\mathcal{X}$ and $\mathcal{Y}$ with the channel as the channel input and output alphabets. The set of channel states is denoted by $\mathcal{S}$ and it is assumed that the channel state is distributed over $\mathcal{S}$ according to P_S. When the state of the channel S is $s \in \mathcal{S}$, the input-output relation of the channel is characterized by a transition kernel $W_{Y|XS}(y|x,s)$ for $x \in \mathcal{X}$ and $y \in \mathcal{Y}$. We assume the state of the channel is known at the transmitter non-causally. The channel is specified by $(\mathcal{X}, \mathcal{Y}, \mathcal{S}, P_S, W_{Y|XS}, w)$ where w is the cost function.

2) Source Model: The source is modeled as a discrete-time random process X with each sample taking values in a fixed set $\mathcal{X}$ called alphabet. Assume X is distributed jointly with the random variable S according to the measure P_{XS} over $\mathcal{X} \times \mathcal{S}$ where $\mathcal{S}$ is an arbitrary set. We assume that the side information S is known to the receiver non-causally. The reconstruction alphabet is denoted by $\mathcal{U}$ and the quality of reconstruction is measured by a single-letter distortion functions $d : \mathcal{X} \times \mathcal{U} \to \mathbb{R}^+$. We denote such sources by $(\mathcal{X}, \mathcal{S}, \mathcal{U}, P_{XS}, d)$.

3) Linear and Coset Codes Over $\mathbb{Z}_p$: For a prime number p, a linear code over $\mathbb{Z}_p$ of length n and rate $R = \frac{k}{n} \log p$ is a collection of p^k codewords of length n which is closed under mod-p addition. Any such code can be characterized by its generator matrix $G \in \mathbb{Z}_p^{k \times n}$. The set of all message tuples for this code is $\mathbb{Z}_p^k$ and the set of all codewords is the range of the matrix G. The linear encoder maps a message

978-1-4673-2579-0/12/$31.00 ©2012 IEEE

tuple $u \in \mathbb{Z}_p^k$ to the codeword x where $x = uG$ and the operations are done mod-p. A coset code over $\mathbb{Z}_p$ is a shift of a linear code by a fixed vector. A coset code of length n and rate $R = \frac{k}{n}\log p$ is characterized by its generator matrix $G \in \mathbb{Z}_p^{k\times n}$ and it's shift vector (dither) $B \in \mathbb{Z}_p^n$. The encoding rule for the corresponding coset code is given by $x = uG+B$, where u is the message tuple and x is the codeword.

4) Lattice Codes and Shifted Lattice Codes: A lattice code of length n is a collection of codewords in $\mathbb{R}^n$ which is closed under real addition. A shifted lattice code is any translation of a lattice code by a real vector. In this paper, we use coset codes to construct (shifted) lattice codes as follows: Given a coset code $\mathbb{C}$ of length n over $\mathbb{Z}_p$ and a *step size* γ, define $\Lambda(\mathbb{C},\gamma,p) = \gamma(\mathbb{C} - \frac{p-1}{2})$. Then the corresponding mod-p lattice code $\bar{\Lambda}(\mathbb{C},\gamma,p)$ is the disjoint union of shifts of Λ by vectors in $\gamma p\mathbb{Z}^n$. i.e. $\bar{\Lambda}(\mathbb{C},\gamma,p) = \bigcup_{v\in p\mathbb{Z}^n} (\gamma v + \Lambda)$. Note that $\Lambda(\mathbb{C},\gamma,p) \subseteq \bar{\Lambda}(\mathbb{C},\gamma,p)$ is a scaled and shifted copy of the linear code $\mathbb{C}$.

5) Nested Linear Codes: A nested linear code consists of two linear codes, with the property than one of the codes (the *inner linear code*) is a subset of the other code (the *outer linear code*). For positive integers k and l, let the outer and inner codes $\mathbb{C}_i$ and $\mathbb{C}_o$ be linear codes over $\mathbb{Z}_p$ characterized by their generator matrices $G \in \mathbb{Z}_p^{l\times n}$ and $G' \in \mathbb{Z}_p^{(k+l)\times n}$ and their shift vectors $B \in \mathbb{Z}_p^n$ and $B' \in \mathbb{Z}_p^n$ respectively. Furthermore, assume $G' = \begin{bmatrix} G \\ \Delta G \end{bmatrix}$ for some $\Delta G \in \mathbb{Z}_p^{k\times n}$ and $B' = B$. In this case,

$$\mathbb{C}_o = \left\{aG + m\Delta G + B | a \in \mathbb{Z}_p^l, m \in \mathbb{Z}_p^k\right\}, \quad (1)$$

$$\mathbb{C}_i = \left\{aG + B | a \in \mathbb{Z}_p^l\right\} \quad (2)$$

It is clear that the inner code is contained in the outer code. Furthermore, the inner code induces a partition of the outer code through its shifts. For $m \in \mathbb{Z}_p^k$ define the mth *bin* of $\mathbb{C}_i$ in $\mathbb{C}_o$ as $\mathbb{B}_m = \left\{aG + m\Delta G + B | a \in \mathbb{Z}_p^l\right\}$. The outer code is the disjoint union of all the bins and each bin index $m \in \mathbb{Z}_p^k$ is considered as a message. We denote a nested linear code by a pair $(\mathbb{C}_i, \mathbb{C}_o)$.

6) Nested Lattice Codes: Given a nested linear code $(\mathbb{C}_i, \mathbb{C}_o)$ over $\mathbb{Z}_p$ and a step size γ, define

$$\Lambda_i(\mathbb{C}_i,\gamma,p) = \gamma(\mathbb{C}_i - \frac{p-1}{2}), \quad (3)$$

$$\Lambda_o(\mathbb{C}_o,\gamma,p) = \gamma(\mathbb{C}_o - \frac{p-1}{2}) \quad (4)$$

Then the corresponding nested lattice code consists of an inner lattice code and an outer lattice code

$$\bar{\Lambda}_i(\mathbb{C}_i,\gamma,p) = \cup_{v\in p\mathbb{Z}^n}(\gamma v + \Lambda_i) \quad (5)$$

$$\bar{\Lambda}_o(\mathbb{C}_o,\gamma,p) = \cup_{v\in p\mathbb{Z}^n}(\gamma v + \Lambda_o) \quad (6)$$

In this case as well, the inner lattice code induces a partition of the outer lattice code. For $m \in \mathbb{Z}_p^k$, define $\mathfrak{B}_m = \gamma(\mathbb{B}_m - \frac{p-1}{2})$ where $\mathbb{B}_m$ is the mth bin of $\mathbb{C}_i$ in $\mathbb{C}_o$. The mth bin of the inner lattice code in the outer lattice code is defined by:

$$\bar{\mathfrak{B}}_m = \cup_{v\in p\mathbb{Z}^n}(\gamma v + \mathfrak{B}_m)$$

The set of messages consists of the set of all bins of $\bar{\Lambda}_i$ in $\bar{\Lambda}_o$. We denote a nested lattice code by a pair $(\bar{\Lambda}_i, \bar{\Lambda}_o)$.

7) Achievability for Channel Coding: A transmission system with parameters (n, M, Γ, τ) for reliable communication over a given channel $(\mathcal{X}, \mathcal{Y}, \mathcal{S}, P_S, W_{Y|XS})$ with cost function $w : \mathcal{X} \to \mathbb{R}^+$ consists of an encoding mapping and a decoding mapping $e : \mathcal{S}^n \times \{1, 2, \ldots, M\} \to \mathcal{X}^n$, $f : \mathcal{Y}^n \to \{1, 2, \ldots, M\}$ such that for all $m = 1, 2, \ldots, M$, if $x = (x_1, \cdots, x_n) = e(m)$ then $\frac{1}{n}\sum_{i=1}^{n} w(x_i) < \Gamma$ and $\sum_{m=1}^{M} \frac{1}{M} Pr\left(f(Y^n) \neq m | X^n = e(m)\right) \leq \tau$.
Given a channel $(\mathcal{X}, \mathcal{Y}, \mathcal{S}, f_S, f_{Y|XS})$, a pair of non negative numbers (R, W) is said to be achievable if for all $\epsilon > 0$ and for all sufficiently large n, there exists a transmission system for reliable communication with parameters (n, M, Γ, τ) such that $\frac{1}{n}\log M \geq R - \epsilon$, $\Gamma \leq W + \epsilon$ and $\tau \leq \epsilon$.

8) Achievability for Source Coding: A transmission system with parameters $(n, \Theta, \Delta, \tau)$ for compressing a given source $(\mathcal{X}, \mathcal{S}, \mathcal{U}, P_{XS}, d(\cdot))$ consists of an encoding mapping $e : \mathcal{X}^n \to \{1, 2, \cdots, \Theta\}$ and a decoding mapping $g : \mathcal{S}^n \times \{1, 2, \cdots, \Theta\} \to \mathcal{U}^n$ such that $P\left(d(X^n, g(e(X^n))) > \Delta\right) \leq \tau$ where X^n is the random vector of length n generated by the source. In this transmission system, n denotes the block length, $\log\Theta$ denotes the number of channel uses, Δ denotes the distortion level and τ denotes the probability of exceeding the distortion level Δ.
Given a source, a pair of non-negative real numbers (R, D) is said to be achievable if there exists for every $\epsilon > 0$, and for all sufficiently large numbers n a transmission system with parameters $(n, \Theta, \Delta, \tau)$ for compressing the source such that $\frac{1}{n}\log\Theta \leq R + \epsilon$, $\Delta \leq D + \epsilon$ and $\tau \leq \epsilon$.

9) Mutual Information and Kullback-Leibler divergence: Let $Q = \{A_1, A_2, \cdots, A_r\}$ be a finite measurable partition of $\mathbb{R}^d$. For random variables U and Y on $\mathbb{R}^d$ with measure P_{UY} define the quantized random variables U_Q and Y_Q on Q with measure $P_{U_QY_Q}(A_i, A_j) = P_{UY}(A_i, A_j)$. The Kullback-Leibler divergence between U and Y is defined as $D(U\|Y) = \sup_Q D(U_Q\|Y_Q)$ where $D(U_Q\|Y_Q)$ is the discrete Kullback-Leibler divergence and the supremum is taken over all finite partitions Q of $\mathbb{R}^d$. Similarly, the mutual information between U and Y is defined as $I(U;Y) = \sup_Q I(U_Q;Y_Q)$ where $I(U_Q;Y_Q)$ is the discrete mutual information between the two random variables and the supremum is taken over all finite partitions Q of $\mathbb{R}^d$.

10) Typicality: We use the notion of weak* typicality with Prokhorov metric introduced in [14]. Let $M(\mathbb{R}^d)$ be the set of probability measures on $\mathbb{R}^d$. For a subset A of $\mathbb{R}^d$ define its ϵ-neighborhood by $A^\epsilon = \{x \in \mathbb{R}^d | \exists y \in A \text{ such that } \|x - y\| < \epsilon\}$ where $\|\cdot\|$ denotes the Euclidean norm in $\mathbb{R}^d$. The Prokhorov distance between two probability measures $P_1, P_2 \in M(\mathbb{R}^d)$ is defined as follows:

$$\pi_d(P_1, P_2) = \inf\{\epsilon > 0 | P_1(A) > P_2(A^\epsilon) + \epsilon \text{ and } P_2(A) > P_1(A^\epsilon) + \epsilon \quad \forall \text{ Borel set } A \text{ in } \mathbb{R}^d\}$$

Consider two random variables X and Y with joint distribution $P_{XY}(\cdot,\cdot)$ over $\mathcal{X}\times\mathcal{Y}\subseteq\mathbb{R}^2$. Let n be an integer and ϵ be a positive real number. For the sequence pair (x,y) belonging to $\mathcal{X}^n\times\mathcal{Y}^n$ where $x=(x_1,\cdots,x_n)$ and $y=(y_1,\cdots,y_n)$ define the empirical joint distribution by

$$P_{xy}(A,B)=\frac{1}{n}\sum_{i=1}^{n}\mathbb{1}_{\{x_i\in A,y_i\in B\}}$$

for Borel sets A and B. Let P_x and P_y be the corresponding marginal probability measures. It is said that the sequence x is weakly* ϵ-typical with respect to P_X if $\pi_1(P_X,P_x)<\epsilon$. We denote the set of all weakly* ϵ-typical sequences of length n by $A_\epsilon^n(X)$. Similarly, x and y are said to be jointly weakly* ϵ-typical with respect to P_{XY} if $\pi_2(P_{xy},P_{XY})<\epsilon$. We denote the set of all weakly* ϵ-typical sequence pairs of length n by $A_\epsilon^n(XY)$. We define $A_\epsilon^n(Y|x)=\{y\in\mathcal{Y}^n\,|(x,y)\in A_\epsilon^n(X,Y)\}$.

11) Notation: In our notation, $O(\epsilon)$ is any function of ϵ such that $\lim_{\epsilon\to0}O(\epsilon)=0$ and for a set G, $|G|$ denotes the cardinality (size) of G.

III. Channel Coding

We show the achievability of the rate $R=I(U;Y)-I(U;S)$ for the Gelfand-Pinsker channel using nested lattice code for U.

Theorem III.1. *For the channel $(\mathcal{X},\mathcal{Y},\mathcal{S},P_S,W_{Y|XS})$ where $\mathcal{X},\mathcal{X},\mathcal{X}\subseteq\mathbb{R}$, let $w:\mathcal{X}\to\mathbb{R}^+$ be a continuous cost function. Let $\mathcal{U}$ be an arbitrary set and let $SUXY$ be distributed over $\mathcal{S}\times\mathcal{U}\times\mathcal{X}\times\mathcal{Y}$ according to $P_SP_{U|S}W_{X|US}W_{Y|SX}$ where $P_{U|S}$ and $W_{X|US}$ are such that $\mathbb{E}\{w(X)\}\le W$. Then the pair (R,W) is achievable using nested lattice codes where $R=I(U;Y)-I(U;S)$.*

A. Discrete U and Bounded Continuous Cost Function

In this section we prove the theorem for the case when $U=\hat{U}$ takes values from the discrete set $\gamma(\mathbb{Z}_p-\frac{p-1}{2})$ where p is a prime and γ is a positive number. We use a random coding argument over the ensemble of mod-p lattice codes to prove the achievability. Let $\mathbb{C}_o$ and $\mathbb{C}_i$ be defined as (1) and (2) where G is a random matrix in $\mathbb{Z}_p^{l\times n}$, ΔG is a random matrix in $\mathbb{Z}_p^{k\times n}$ and B is a random vector in $\mathbb{Z}_p^n$, all uniformly distributed over their respective domains. Define $\bar{\Lambda}_i(\mathbb{C}_i,\gamma,p)$ and $\bar{\Lambda}_o(\mathbb{C}_o,\gamma,p)$ accordingly. The ensemble of nested lattice codes consists of all lattices of the form (3) and (4). The set of messages consists of all bins $\mathfrak{B}_m$ indexed by $m\in\mathbb{Z}_p^k$. The encoder observes the message $m\in\mathbb{Z}_p^k$ and the channel state $s\in\mathcal{S}^n$ and looks for a vector u in the mth bin $\mathfrak{B}_m$ which is jointly weak* typical with s and encodes the message m to x according to $W_{X|SU}$. The encoder declares error if it does not find such a vector.

After receiving $y\in\mathcal{Y}^n$, the decoder decodes it to $m\in\mathbb{Z}_p^k$ if m is the unique tuple such that the mth bin $\mathfrak{B}_m$ contains a sequence jointly typical with y. Otherwise it declares error.

1) Encoding Error: Let $S'=[\frac{-\gamma p}{2},\frac{\gamma p}{2}]^n\cap\gamma\mathbb{Z}^n$. For $a\in\mathbb{Z}_p^k$, $m\in\mathbb{Z}_p^l$, define

$$g(a,m)=\gamma\left((aG+m\Delta G+B)-\frac{(p-1)}{2}\right)$$

$g(a,m)$ has the following properties:

Lemma III.2. *For $a\in\mathbb{Z}_p^l$ and $m\in\mathbb{Z}_p^k$, $g(a,m)$ is uniformly distributed over S'.*

Proof: Follows from the fact that B is independent of G and ΔG and is uniformly distributed. ∎

Lemma III.3. *For $a,\tilde{a}\in\mathbb{Z}_p^l$ and $m\in\mathbb{Z}_p^k$ if $a\neq\tilde{a}$ then $g(a,m)$ and $g(\tilde{a},m)$ are pairwise independent.*

Proof: Follows from a counting argument. ∎

For message $m\in\mathbb{Z}_p^k$ and state $s\in\mathcal{S}^n$, the encoder declares error if there is no sequence in $\mathfrak{B}_m$ jointly typical with s. Define

$$\theta(s)=\sum_{u\in\mathfrak{B}_m}\mathbb{1}_{\{u\in A_\epsilon^n(\hat{U}|s)\}}=\sum_{a\in\mathbb{Z}_p^l}\mathbb{1}_{\{g(a,m)\in A_\epsilon^n(\hat{U}|s)\}}$$

Let Z be a uniform random variable over S'. Then we have

$$\mathbb{E}\{\theta(s)\}=\sum_{a\in\mathbb{Z}_p^l}P\Big(Z^n\in A_\epsilon^n(\hat{U}|s)\Big)$$

we need the following lemmas to proceed:

Lemma III.4. *Let P_{XY} be a joint distribution on $\mathbb{R}^2$ and P_X and P_Y denote its marginals. Let y be a sequence and Z^n a random sequence drawn according to P_Z^n. If $D(P_{XY}\|P_ZP_Y)$ is finite then for each $\delta>0$, there exist $\epsilon(\delta)$ and $\bar{\epsilon}(\delta)$ such that if $\epsilon<\epsilon(\delta)$, $\bar{\epsilon}<\bar{\epsilon}(\delta)$ and $y\in A_{\bar{\epsilon}}^n(P_Y)$ then*

$$\limsup\frac{1}{n}\log P_Z^n((Z^n,y)\in A_\epsilon^n(P_{XY})\le-D(P_{XY}\|P_ZP_Y)+\delta$$

Proof: This lemma is a generalization of theorem 21 of [14]. The complete proof can be found in a more complete version of this work [19]. ∎

Lemma III.5. *Let P_{XY} be a joint distribution on $\mathbb{R}^2$ and P_X and P_Y denote its marginals. Let y be a sequence and Z^n a random sequence drawn according to P_Z^n. Then for each $\epsilon,\delta>0$, there exist $\bar{\epsilon}(\epsilon,\delta)$ such that if $y\in A_{\bar{\epsilon}}^n(P_Y)$ then*

$$\liminf\frac{1}{n}\log P_Z^n((Z^n,y)\in A_\epsilon^n(P_{XY})\ge-D(P_{XY}\|P_ZP_Y)-\delta$$

Proof: This lemma is a generalization of theorem 22 of [14]. The complete proof can be found in [19]. ∎

Using these lemmas we get

$$\mathbb{E}\{\theta(s)\}=p^k2^{-n[D(P_{\hat{U}S}\|P_ZP_S)+O(\epsilon)]}$$

Similarly, let $Z^n=g(a,m)$ and $\tilde{Z}^n=g(\tilde{a},m)$. Note that Z^n and $\tilde{Z}^n$ are equal if $a=\tilde{a}$ and are independent if $a\neq\tilde{a}$. We

have

$$\begin{aligned}\mathbb{E}\{\theta(s)^2\} &= \sum_{a,\tilde{a}\in\mathbb{Z}_p^l} P\Big(Z^n,\tilde{Z}^n \in A_\epsilon^n(\hat{U}|s)\Big)\\ &= \sum_{a\in\mathbb{Z}_p^l} P\Big(Z^n \in A_\epsilon^n(\hat{U}|s)\Big)\\ &+ \sum_{\substack{a,\tilde{a}\in\mathbb{Z}_p^l\\ a\neq\tilde{a}}} P\Big(Z^n \in A_\epsilon^n(\hat{U}|s)\Big)^2\\ &= p^k 2^{-n[D(P_{\hat{U}S}\|P_ZP_S)+O(\epsilon)]}\\ &+ p^k(p^k-1)2^{-2n[D(P_{\hat{U}S}\|P_ZP_S)+O(\epsilon)]}\end{aligned}$$

Therefore, using Chebyshev's inequality we obtain

$$P(\theta(s)=0) \le \frac{\text{var}\{\theta(s)\}}{\mathbb{E}\{\theta(s)\}^2} \le p^k 2^{-n[D(P_{\hat{U}S}\|P_ZP_S]+O(\epsilon)]}$$

Therefore if $\frac{l}{n}\log p > D(P_{\hat{U}S}\|P_ZP_S)$ then the probability of encoding error goes to zero as the block length increases.

2) Decoding Error: The decoder declares error if there is no bin $\mathfrak{B}_m$ containing a sequence jointly typical with the channel output y or if there are multiple bins containing sequences jointly typical with y. Assume that the message m has been encoded to x according to $W_{X|SU}$ where $u = g(a,m)$ and the channel state is s. The channel output y is jointly typical with u with high probability. It can be shown [19] that the probability of decoding error is upper bounded by

$$P_{err} \le p^l p^k 2^{-n[D(P_{\hat{U}Y}\|P_ZP_Y)+O(\epsilon)]}$$

Hence the probability of decoding error goes to zero if $\frac{k+l}{n}\log p < D(P_{\hat{U}Y}\|P_ZP_Y)$. If we choose $\frac{l}{n}\log p$ sufficiently close to $D(P_{\hat{U}S}\|P_ZP_S)$ and $\frac{k+l}{n}\log p$ sufficiently close to $D(P_{\hat{U}S}\|P_ZP_S)$ we can achieve the rate

$$\begin{aligned}R &= \frac{k}{n}\log p \approx D(P_{\hat{U}Y}\|P_ZP_Y) - D(P_{\hat{U}S}\|P_ZP_S)\\ &= I(\hat{U};Y) - I(\hat{U};S)\end{aligned}$$

B. Arbitrary U and Bounded Continuous Cost Function

We have shown in Section III-A that for discrete random variables the region given in Theorem III.1 is achievable. In this part, we make a quantization argument to generalize this result to arbitrary auxiliary random variables. Let S,U,X,Y be distributed according to $P_SP_{U|S}P_{X|US}W_{Y|X}$ where in this case U is an arbitrary random variable. We start with the following theorem:

Theorem III.6. *Let $\mathcal{F}_1 \subseteq \mathcal{F}_2 \subseteq \cdots$ be an increasing sequence of σ-algebras on a measurable set A. Let $\mathcal{F}_\infty$ denote the σ-algebra generated by the union $\cup_{n=1}^\infty \mathcal{F}_n$. Let P and Q be probability measures on A. Then*

$$D(P|_{\mathcal{F}_n}\|Q|_{\mathcal{F}_n}) \to D(P|_{\mathcal{F}_\infty}\|Q|_{\mathcal{F}_\infty}) \text{ as } n\to\infty$$

where $P|_{\mathcal{F}}$ denotes the restriction of P on $\mathcal{F}$.

Proof: Provided in [9] and [1]. ∎

For a prime $p > 2$ and a real positive number γ and for $i = 0\cdots,p-1$ define $a_i = \frac{-\gamma(p-1)}{2} + \gamma i$ and define the quantization $Q_{\gamma,p}$ as $Q_{\gamma,p} = \{A_0, A_2, \cdots, A_{p-1}\}$ where $A_0 = (-\infty, a_0]$, $A_{p-1} = (a_{p-2}, +\infty)$ and $A_i = (a_{i-1}, a_i]$ for $i = 1,\cdots,p-2$. Let the random variable $\hat{U}_{\gamma,p}$ take values from $\{a_0,\cdots,a_{p-1}\}$ according to joint measure

$$P_{S\hat{U}XY}(\hat{U}=a_i, SXY\in B) = P_{SUXY}(U\in A_i, SXY\in B) \quad (7)$$

For all Borel sets $B \subseteq \mathbb{R}^3$. For a fixed γ, let $p \le q$ be two primes. Then the σ-algebra induced by $Q_{\gamma,p}$ is included in the σ-algebra induced by $Q_{\gamma,q}$. Therefore, for a fixed γ, we can use the above theorem to get

$$I(U|_{\mathcal{F}_{\gamma,p}};Y|_{\mathcal{F}_{\gamma,p}}) \to I(U|_{\mathcal{F}_{\gamma,\infty}};Y|_{\mathcal{F}_{\gamma,\infty}}) \text{ as } p\to\infty \quad (8)$$

where $U|_{\mathcal{F}_{\gamma,\infty}}$ is a random variable over $Q_{\gamma,\infty} = \{A_i | i\in\mathbb{Z}\}$ where $A_i = \frac{\gamma}{2} + (\gamma i, \gamma(i+1)]$ with measure $P_{U|_{\mathcal{F}_{\gamma,\infty}}}(A_i) = P_U(A_i)$.

Let $\gamma_0 = 1$ and define $\gamma_n = \frac{1}{2^n}$. Note that if $m > n$ then $\mathcal{F}_{\gamma_n,\infty}$ is included in $\mathcal{F}_{\gamma_m,\infty}$. Also, since dyadic intervals generate the Borel Sigma field ([15] for example), the restriction of U to the sigma algebra generated by $\cup_{n=1}^\infty \mathcal{F}_{\gamma_n,\infty}$ is U itself. We can use Theorem III.6 to get

$$I(U|_{\mathcal{F}_{\gamma_n,\infty}};Y|_{\mathcal{F}_{\gamma_n,\infty}}) \to I(U;Y) \text{ as } n\to\infty \quad (9)$$

Combining (8) and (9) we conclude that for all $\epsilon > 0$, there exist Γ and P such that if $\gamma \le \Gamma$ and $p \ge \Gamma$ then

$$\Big|I(U|_{\mathcal{F}_{\gamma,p}};Y|_{\mathcal{F}_{\gamma,p}}) - I(U;Y)\Big| < \epsilon$$

Since quantization reduces the mutual information ($X_Q \to X \to Y$), we have

$$I(U|_{\mathcal{F}_{\gamma,p}};Y|_{\mathcal{F}_{\gamma,p}}) \le I(U|_{\mathcal{F}_{\gamma,p}};Y) \le I(U;Y)$$

Therefore $\Big|I(U|_{\mathcal{F}_{\gamma,p}};Y) - I(U;Y)\Big| < \epsilon$. Also note that $I(U|_{\mathcal{F}_{\gamma,p}};Y) = I(\hat{U}_{\gamma,p};Y)$ since we define the joint measure to be the same. Therefore

$$\Big|I(\hat{U}_{\gamma,p};Y) - I(U;Y)\Big| \le \epsilon \quad (10)$$

With a similar argument, $\forall\epsilon > 0$ there exist γ and p such that

$$\Big|I(\hat{U}_{\gamma,p};S) - I(U;S)\Big| \le \epsilon \quad (11)$$

if we take the maximum of the two p's and the minimum of the two γ's, we can say for all $\epsilon > 0$ there exist γ and p such that both (10) and (11) happen.

Lemma III.7. *The sequence $P_{S\hat{U}_{\gamma_p,p}X}$ converges to P_{SUX} in the weak* sense as $p\to\infty$.*

Proof: Provided in a more complete version [19]. ∎

The above lemma implies $\mathbb{E}_{P_{S\hat{U}_{\gamma_p,p}X}}\{w(X)\}$ converges to $\mathbb{E}_{P_{SUX}}\{w(X)\} \le W$ since w is assumed to be bounded continuous.

We have shown that for arbitrary $P_{U|S}$ and $W_{X|SU}$, one can find $P_{\hat{U}|S}$ and $W_{X|S\hat{U}}$ induced from (7) such that $\hat{U}$ is a discrete variable and

$$\begin{aligned}&I(\hat{U};Y) - I(\hat{U};S) \approx I(U;Y) - I(U;S)\\ &\mathbb{E}_{P_{S\hat{U}X}}\{w(X)\} \approx \mathbb{E}_{P_{SUX}}\{w(X)\}\end{aligned}$$

Hence, using the result of section III-A, we have shown the achievability of the rate region given in Theorem III.1 for arbitrary auxiliary random variables when the cost function is bounded and continuous.

C. Arbitrary U and Continuous Cost Function

For a positive number l, define the bounded random variable $\hat{X}$ by $\hat{X} = \text{sign}(X)\min(l, |X|)$ and let $\hat{Y}$ be distributed according to $W_{\hat{Y}|\hat{X}}(\cdot, \hat{x}) = W_{Y|X}(\cdot, \hat{x})$.

Lemma III.8. *As $l \to \infty$, $I(U;\hat{Y}) \to I(U;Y)$.*

Proof: Proved in a more complete version [19]. ■

Since $\hat{X}$ is bounded and w is assumed to be continuous, w is also bounded. This completes the proof.

IV. Source Coding

We show the achievability of the rate $R = I(U;X) - I(U;S)$ for the Wyner-Ziv problem using nested lattice codes.

Theorem IV.1. *For the source $(\mathcal{X}, \mathcal{S}, \hat{\mathcal{X}}, P_{XS}, d(\cdot))$ assume $\mathcal{X}, \mathcal{S}, \hat{\mathcal{X}} \subseteq \mathbb{R}$ and $d(\cdot)$ is continuous. Let U be a random variable taking values from the set $\mathcal{U}$ jointly distributed with X and S according to $P_{XS}W_{U|X}$ where $W_{U|X}(\cdot|\cdot)$ is a transition kernel. Further assume that there exists a measurable function $f : \mathcal{S} \times \mathcal{U} \to \hat{\mathcal{X}}$ such that $\mathbb{E}\{d(X, f(S,U))\} \leq D$. Then the rate $R^*(D) = I(X;U) - I(S;U)$ is achievable using nested lattice codes.*

Here we present a sketch of the proof of this theorem. The complete proof is similar to the channel coding problem and is provided in [19]. The ensemble of codes used for source coding is based on the parity check matrix representation of linear and lattice codes. For a prime number p, the linear code over $\mathbb{Z}_p$ corresponding to the parity check matrix $H \in \mathbb{Z}_p^{k\times n}$ is the kernel of the matrix H; i.e. $\mathbb{C} = \left\{u \in \mathbb{Z}_p^n | Hu = 0\right\}$ where the operations are done mod-p. The coset code corresponding to the parity check matrix $H \in \mathbb{Z}_p^{k\times n}$ and the bias vector $c \in \mathbb{Z}_p^k$ is defined as $\mathbb{C} = \left\{u \in \mathbb{Z}_p^n | Hu = c\right\}$.
Nested linear codes based on the parity check representation of linear codes can be defined as follows

$$\mathbb{C}_o = \left\{u \in \mathbb{Z}_p^n | Hu = c\right\},$$
$$\mathbb{C}_i = \left\{u \in \mathbb{Z}_p^n | Hu = c, \Delta Hu = \Delta c\right\}$$

where $H \in \mathbb{Z}_p^{k\times n}$ and $\Delta H \in \mathbb{Z}_p^{k\times n}$. The ensemble of lattice codes are then constructed using (1) and (4). The encoding and decoding rules are as follows: For $m \in \mathbb{Z}_p^k$, Let $\mathfrak{B}_m$ be the mth bin of Λ_i in Λ_o. The encoder observes the source sequence $x \in \mathcal{X}^n$ and looks for a vector u in the outer code Λ_o which is typical with x and encodes the sequence x to the bin of Λ_i in Λ_o containing u. The encoder declares error if it does not find such a vector.
Having observed the index of the bin m and the side information s, the decoder looks for a unique sequence u in the mth bin which is jointly typical with s and outputs $f(u,s)$. Otherwise it declares error. Similar to the channel coding problem, this theorem is first proved for discrete auxiliary random variables and then it is generalized to arbitrary sources.

V. Conclusion

We have shown that nested lattice codes are optimal for the Gelfand-Pinsker problem as well as the Wyner-Ziv problem.

References

[1] A. R. Barron. Limits of Information, Markov Chains, and Projection. *Proceedings of IEEE International Symposium on Information Theory*, 2000. Sorrento, Italy.

[2] R. De Buda. Some optimal codes have structure. *IEEE Journal on Selected Areas in Communications*, 7:893–899, 1989.

[3] I. Csiszar. Linear Codes for Sources and Source Networks: Error Exponents, Universal Coding.

[4] U. Erez and R. Zamir. Achieving $\frac{1}{2}\log(1 + \text{SNR})$ on the AWGN Channel With Lattice Encoding and Decoding. *IEEE Transactions on Information Theory*, 50:2293–2314, 2004.

[5] U. Erez and R. Zamir. Capacity and Lattice Strategies for Canceling Known Interference. *IEEE Transactions on Information Theory*, 51:3820–3833, 2005.

[6] U. Erez and R. Zamir. Lattices Which Are Good for (Almost) Everything. *IEEE Transactions on Information Theory*, 51:3401–3416, 2005.

[7] T. Gariby and U. Erez. On General Lattice Quantization Noise. *Proceedings of IEEE International Symposium on Information Theory*, 2008. Toronto, Canada.

[8] S. I. Gelfand and M. S. Pinsker. Coding for channel with random parameters. *Problems of Control and Information Theory*, 9:19–31, 1980.

[9] P Harremos and K Khler Holst. Convergence of Markov Chains in Information Divergence. *Journal of Theoretical Probability*, 22(1):186–202, 2011.

[10] J. Korner and K. Marton. How to encode the modulo-two sum of binary sources. *IEEE Transactions on Information Theory*, IT-25:219–221, Mar. 1979.

[11] D. Krithivasan and S. S. Pradhan. Distributed source coding using abelian group codes. 2011. IEEE Transactions on Information Theory(57)1495-1519.

[12] T. Linder and C. Schlegel. Corrected Proof of de Buda's Theorem. *IEEE Transactions on Information Theory*, 39:1735–1737, 1993.

[13] H. A. Loeliger. Averaging bounds for lattices and linear codes. *IEEE Transactions on Information Theory*, 43:1767–1773, 1997.

[14] P. Mitran. Typical Sequences for Polish Alphabets. 2010. Online: http://arxiv.org/abs/1005.2321.

[15] P. Mrters, O. Schramm Y. Peres, and W. Werner. *Brownian Motion*. Cambridge University Press, 2010.

[16] B. A. Nazer and M. Gastpar. Computation over multiple-access channels. *IEEE Trans. on Inf. Th.*, 53, Oct. 2007.

[17] A. Padakandla and S. S. Pradhan. Nested linear codes achieve martons inner bound for general broadcast channels. *Proc. IEEE Int. Symp. Information Theory*, 2011. Saint Petersburg, Russia.

[18] T. Philosof, A. Kishty, U. Erez, and R. Zamir. Lattice strategies for the dirty multiple access channel. *Proceedings of IEEE International Symposium on Information Theory*, July 2007. Nice, France.

[19] A. G. Sahebi and S. Sandeep Pradhan. Nested lattice codes for arbitrary continuous sources and channels. 2012. Online: http://arxiv.org/submit/410552.

[20] S. Sridharan, A. Jafarian, S. Vishwanath, S. A. Jafar, and S. Shamai. A layered lattice coding scheme for a class of three user gaussian interference channels. 2008. Online: http://arxiv.org/abs/0809.4316.

[21] R. Urbanke and B. Rimoldi. Lattice Codes Can Achieve Capacity on the AWGN Channel. *IEEE Trans. on Inf. Th.*, 44:273–278, 1998.

[22] A. Wyner. The rate distortion function for source coding with side information at the decoder-ii. *Information and Control*, 38:60–80, 1978.

[23] A. Wyner and J. Ziv. The rate distortion function for source coding with side information at the decoder. *IEEE Transactions on Information Theory*, 22:1–10, 1976.

[24] R. Zamir and M. Feder. On lattice quantization noise. *IEEE Transactions on Information Theory*, 42:1152–1159, 1996.

[25] R. Zamir and S. Shamai. Nested linear/lattice codes for wyner-ziv encoding. *ITW*, 1998. Ireland.

[26] R. Zamir, S. Shamai, and U. Erez. Nested linear/lattice codes for structured multiterminal binning. *IEEE Transactions on Information Theory*, 48(6):1250–1276, 2002.