ITW 1998, Killarney, Ireland, June 22 – 26

Nested Linear / Lattice Codes for Wyner-Ziv Encoding

Ram Zamir†, and Shlomo Shamai (Shitz)‡

† Dept. of Elect. Eng. - Systems, Tel-Aviv University, Tel-Aviv 69978 (*zamir@eng.tau.ac.il*).

‡ Dept. of Electrical Engineering, Technion, Haifa 32000 (*sshlomo@ee.technion.ac.il*).

Abstract — **We construct structured codes for lossy compression with uncoded side information at the decoder. These codes aim at achieving the optimal performance, given by the Wyner-Ziv (WZ) rate-distortion function, for two typical settings. For binary symmetric source / side-information, we show that a pair of nested linear binary codes, each individually "good" in the classical sense, approaches the WZ-function relative to the Hamming distortion measure. For the quadratic Gaussian case we show a similar result at high SNR, using a pair of nested lattices and their associated lattice-decoding functions.**

I. Introduction

Wyner and Ziv [6] consider coding a source X, relative to some distortion measure $\rho(x,\hat{x})$, where the decoder, which outputs $\widehat{X}$, (but not the encoder) has access to a correlated source Y, called "side information". They derive a "single letter" expression for the minimum coding rate, $R^{WZ}_{X|Y}(d)$, called the "Wyner-Ziv function", under the constraint that the average distortion $\overline{E\rho(X_n,\hat{X}_n)}$ does not exceed d. This problem generalizes the lossless ($d=0$) setting introduced by Slepian and Wolf [4]. But unlike in the Slepian-Wolf setting, the Wyner-Ziv function suffers a loss in rate relative to the case where both the encoder and the decoder have access to Y [7]. The Wyner-Ziv function plays a fundamental role in the characterization of the achievable rates in systematic lossy source/channel coding [3].

The achievability parts of the Slepian-Wolf and the Wyner-Ziv theorems follow a non constructive random coding argument. Wyner [5] suggests a *linear* binary code interpretation for the Slepian-Wolf code, which achieves the optimum for binary symmetric source / side information. Following Wyner, we propose structured schemes for Wyner-Ziv encoding for the binary symmetric (Hamming distortion) case using *nested* linear binary codes, and for the jointly Gaussian (mean squared distortion) case using *nested* lattices. The binary Wyner-Ziv coding scheme was introduced in a joint work with S. Verdú in [3]. For completeness we describe it in Section II, and then propose the Wyner-Ziv lattice coding scheme in Section III.

II. Binary Symmetric Source/Side-Information

Let X and Y be binary symmetric sources connected via a BSC with transition probability p, i.e., $Y=X\oplus N$, where N is an independent Bernoulli(p) process, and $\oplus$ denotes modulo-2 addition. Let the distortion measure be the Hamming distance. The Wyner-Ziv function $R^{WZ}_{X|Y}(d)$ is given by the lower convex envelope (l.c.e.) of the function $h(p*d)-h(d)$ and the point $(R,d)=(0,p)$, where $h(\cdot)$ denotes the binary entropy function, and $*$ denotes binary convolution [6].

Clearly it is enough to consider achieving the function $h(p*d)-h(d)$. (Any point on the l.c.e. can be achieved via time-sharing with no coding.) Choose two linear codes defined by parity-check matrices $(m_1\times n)$ H_1 and $(m_2\times n)$ H_2, such that $\frac{m_1}{n}\simeq h(d)$, $\frac{m_2}{n}\simeq h(p*d)$, and Code 2 is a sub-code of Code 1. We say that "Code 2 is nested in Code 1". Thus, the parity check matrices H_1 and H_2 satisfy

$$H_2=\left[\begin{array}{c} H_1 \\ \hline H_a \end{array}\right].$$

Every codeword of Code 1 $A=(a_1\cdots a_n)$ satisfies $H_1A^T=0$, where the superscript T stands for the transpose operation. If, in addition, $H_aA^T=0$, then A is also a codeword of Code 2. The decoders for these codes are defined by functions f_1 and f_2 of the corresponding syndromes. According to well-known properties of optimal linear codes, the codebooks and decoding functions of Codes 1 and 2 can be chosen so that $f_1(H_1\,\mathcal{U}_1^T)=\mathcal{U}_1$ and $f_2(H_2\,\mathcal{U}_2^T)=\mathcal{U}_2$ for most of the realizations $\mathcal{U}_1$ of a Bernoulli(d) process, and for most of the realizations $\mathcal{U}_2$ of a Bernoulli$(p*d)$ process. We discuss the existence of such nested good binary codes in Section IV.

The *Wyner-Ziv encoding* of the binary source word consists of two steps:

(a) Among the codewords in Code 1 select X_q which is closest to X in Hamming distance. Let $E_q=X\oplus X_q$.

(b) Output the m_2-m_1 vector $H_aX_q^T$.

Note that the output rate of the Wyner-Ziv encoder is

$$R=\frac{m_2-m_1}{n}\simeq h(p*d)-h(d) \tag{1}$$

as desired.

The *decoder* receives $H_aX_q^T$ and $Y=X\oplus N$. Since X_q is a codeword of Code 1, $H_1X_q^T=0$ and

$$H_2X_q^T=\left[\begin{array}{c} 0 \\ \hline H_aX_q^T \end{array}\right].$$

The decoder outputs the n-codeword

$$\widehat{X}=Y\oplus f_2(H_2X_q^T\oplus H_2Y^T)=Y\oplus f_2(H_2(E_q^T\oplus N^T)). \tag{2}$$

If $E_q\oplus N$ were a Bernoulli$(p*d)$ process, the output of the decoder would yield with high probability

$$\widehat{X}=Y\oplus E_q\oplus N=X\oplus E_q=X_q, \tag{3}$$

in which case the decoder obtains a distorted version of the input X within distortion (bit error rate) d, as desired. Indeed, N is Bernoulli(p) and E_q is independent of N. Furthermore, for increasingly long blocklength and for the appropriate choice of Code 1, the distribution of E_q will resemble those of a Bernoulli(d) process.

III. Jointly Gaussian Source / Side-Information

Let the side-information be $Y=X+N$, where the source X and the noise N are independent Gaussians with variances VAR(X) and VAR(N), respectively, and let the distortion measure be $\rho(\hat{x},x)=(\hat{x}-x)^2$. The WZ-function at distortion level d is equal to, [6],

$$R^{WZ}_{X|Y}(d)=\frac{1}{2}\log\left(\frac{\sigma^2_{X|Y}}{d}\right)\ ,\quad 0<d\le\sigma^2_{X|Y} \qquad (4)$$

where $\sigma^2_{X|Y}=[1/\mathrm{VAR}(X)+1/\mathrm{VAR}(N)]^{-1}$.

The proposed lattice WZ-coding scheme consists of a pair of *unbounded* lattices $\mathcal{L}_1$ and $\mathcal{L}_2$ in $\mathcal{R}^n$, with lattice-decoding ("nearest neighbor quantizing") functions $Q_i:\ \mathcal{R}^n\to\mathcal{L}_i$, where

$$Q(x)\triangleq\arg\min_{\ell\in\mathcal{L}}\|x-\ell\|\ ;$$

basic Voronoi cells $\mathcal{V}_i=\{x:\ Q_i(x)=0\}$; cell volumes $V_i=\int_{\mathcal{V}_i}dx$; second moments $\sigma_i^2=\frac{1}{nV_i}\int_{\mathcal{V}_i}\|x\|^2dx$; and normalized second moments, [2], $G_i\triangleq\sigma_i^2/V_i^{2/n},\ \ i=1,2$; such that

a) $\mathcal{L}_2\subseteq\mathcal{L}_1$ ($\mathcal{L}_2$ is "nested" in $\mathcal{L}_1$)

b) $\sigma_1^2=(1/d-1/\sigma^2_{X|Y})^{-1}$

c) $\sigma_2^2\le\sigma_1^2+\mathrm{VAR}(N)+\epsilon$

where $\epsilon>0$ is arbitrary. Furthermore, we assume that for large enough n we can find nested lattices as above which are *both good for source / channel coding* in the sense of [2, sec. 1-3]. In particular, we assume

d) $\Pr\{Q_2(c+N+Z^*)\ne c\}<\epsilon\quad\forall c\in\mathcal{L}_2$,
where Z^* is an independent Gaussian with variance σ_1^2 (i.e., the variance of $N+Z^*$ is slightly less than σ_2^2); and

e) $\log(2\pi eG_i)<\epsilon\ \ i=1,2.$

Note that for any lattice $G>\frac{1}{2\pi e}$, and for a sequence of "good" lattices G. Poltyrev [8] showed that $G_n\to\frac{1}{2\pi e}$. See the discussion on the existence of nested "good" lattices in Section IV.

Before describing the encoding-decoding operation we introduce some terminology regarding nested lattices and lattice quantization. Given a point $\ell\in\mathcal{L}_1$, let $\ell-Q_2(\ell)$ be the "coset shift of ℓ with respect to $\mathcal{L}_2$", let $i(\ell)\in\{1,\ldots,V_2/V_1\}$ be a function that indexes the coset shifts, and let $g(i(\ell))=\ell-Q_2(\ell)$ be the inverse function of $i(\cdot)$. Let Z (the "dither") be a random vector independent of (X,N) and uniformly distributed over $\mathcal{V}_1$, known to both encoder and decoder. We have $\frac{1}{n}E\|Z\|^2=\sigma_1^2$, and for large n (by Prop. (e)) Z resembles an i.i.d. Gaussian process [8]. Choose coefficients a and b such that $a(X+N)+b(X+Z)$ is an optimal linear estimator of X from $X+N$ and $X+Z$. By Property (b) we have

$$\frac{1}{n}E\|a(X+N)+b(X+Z)-X\|^2=d. \qquad (5)$$

Wyner-Ziv encoding: The encoder calculates $X_q\triangleq Q_1(X+Z)$ and transmits the index $i(X_q)$. By Properties (b),(c) and (e), the coding rate is thus (within at most $1/n$ bit from)

$$R=\frac{1}{n}\log\left(\frac{V_2}{V_1}\right)=\frac{1}{2}\log\left(\frac{\sigma_2^2}{\sigma_1^2}\frac{G_1}{G_2}\right)\le\frac{1}{2}\log\left(\frac{\mathrm{VAR}(N)}{d}\right)+O(\epsilon) \qquad (6)$$

which is close to the WZ-function (4) for high SNR ($\mathrm{VAR}(X)\gg\mathrm{VAR}(N)$) in the side-information channel. The slight rate loss at low and medium SNRs follows since we limit our scheme to lattice-decoding (using the functions $Q_1(\cdot)$ and $Q_2(\cdot)$), instead of a more complicated probabilistic decoding. A similar loss of capacity occurs in using bounded lattices with lattice-decoding for channel coding at low SNR; see De-Buda [1].

Wyner-Ziv decoding: Given $i(X_q)$, Z and $Y=X+N$, the decoder calculates the coset shift $S=g(i(X_q))$, and outputs

$$\widehat{X}=aY+b\big\{Q_2[Y+Z-S]-Z+S\big\}. \qquad (7)$$

Since $S=X_q-Q_2(X_q)$, the argument of $Q_2[\cdot]$ in (7) is equal to $X+N+Z-[X_q-Q_2(X_q)]\triangleq c+E_q+N$, where $c\triangleq Q_2(X_q)$ and $E_q\triangleq X+Z-Q_1(X+Z)$. By the properties of subtractive dithered quantization [8], the quantization error E_q is independent of X and is distributed as Z. Thus $c+E_q+N$ is the sum of a point in $\mathcal{L}_2$ and a "Gaussian-like" noise whose power is $\sigma_1^2+\mathrm{VAR}(N)$, implying by Property (d) that $Q_2[c+E_q+N]=c$ with probability $>1-\epsilon$. Substituting in (7), we see that with high probability

$$\widehat{X}=a(X+N)+b(X-E_q)\ ,$$

implying by (5), and since $\frac{1}{n}E\|E_q\|^2=\frac{1}{n}E\|Z\|^2=\sigma_1^2$, that the resulting distortion satisfies

$$\frac{1}{n}E\|\widehat{X}-X\|^2=d+O(\epsilon)$$

as desired.

IV. Discussion

The two schemes above assume that for dimension n large enough, there exist a pair of nested linear codes (for the binary-Hamming case), and a pair of nested lattices (for the quadratic Gaussian case), such that each of the codes is "good" in the ordinary sense for both source and channel coding. This is a stronger requirement than that of Wyner [5], whose Slepian-Wolf encoding scheme can use any linear binary code which is good for channel coding.

The existence of pairs of good nested codes seems to follow from random coding arguments. Specifically, in the binary case S. Litsyn provided a constructive argument showing that if the coarse code (Code 2) is good, then at least one of its cosets generates a good fine code (Code 1). Beyond the question of existence, it is of interest to find specific constructions for good nested codes.

References

[1] R. De Buda. Some optimal codes have structure. *IEEE Jr. on Selected Areas in Comm.*, 7:893–899, Aug. 1989.

[2] J. H. Conway and N. J. A. Sloane. *Sphere Packings, Lattices and Groups.* Springer-Verlag, New York, N.Y., 1988.

[3] S. Shamai, S. Verdú, and R. Zamir. Systematic lossy source/channel coding. *IEEE Trans. Information Theory*, 44, pp. 564–579, March 1998.

[4] D. Slepian and J.K. Wolf. Noiseless coding of correlated information sources. *IEEE Trans. Information Theory*, IT-19:471–480, July 1973.

[5] A.D. Wyner. Recent results in Shannon theory. *IEEE Trans. Information Theory*, IT-20:2–10, Jan. 1974.

[6] A.D. Wyner and J. Ziv. The rate-distortion function for source coding with side information at the decoder. *IEEE Trans. Information Theory*, IT-22:1–10, Jan., 1976.

[7] R. Zamir. The rate loss in the Wyner-Ziv problem. *IEEE Trans. Information Theory*, pages 2073–2084, Nov. 1996.

[8] R. Zamir and M. Feder. On lattice quantization noise. *IEEE Trans. Information Theory*, pages 1152–1159, July 1996.