Universal Quantization for Separate Encodings and Joint Decoding of Correlated Sources

Avraham Reani and Neri Merhav
Department of Electrical Engineering
Technion - Israel Institute of Technology
Technion City, Haifa 32000, Israel
Emails: [avire@tx,merhav@ee].technion.ac.il

***Abstract*—TO BE CONSIDERED FOR THE IEEE JACK KEIL WOLF ISIT STUDENT PAPER AWARD. We consider the multi-user lossy source-coding problem for continuous alphabet sources. In previous work, Ziv proposed a universal coding scheme which uses uniform quantization with dither, followed by a lossless source encoder (entropy coder). In this paper, we generalize Ziv's scheme to the multi-user setting. For this generalized scheme, upper bounds are derived on the redundancies, defined as the differences between the actual rates and the closest corresponding rates on the boundary of the rate region. It is shown that this scheme can achieve redundancies of no more than 0.754 bits per sample, for each user. These results are obtained without knowledge of the multi-user rate region, which is an open problem in general.**

I. Introduction

Consider the case where two correlated sources are observed separately by two non-cooperative encoders which communicate with one decoder. The decoder needs to reconstruct both sources and the distortions between the reconstructions and the corresponding sources should not exceed some given values. The general version of this problem has remained open for several decades, even under the assumption of memoryless sources. However, many special cases have been solved. When no distortion is allowed, this is the problem considered by Slepian and Wolf [1]. Their well-known result states that two discrete sources X_1 and X_2 can be losslessly reproduced if and only if

$$\begin{aligned} R_1 &\geq H(X_1|X_2) \\ R_2 &\geq H(X_2|X_1) \\ R_1+R_2 &\geq H(X_1,X_2) \end{aligned} \tag{1}$$

where R_1 is the rate of the encoder observing X_1 and R_2 is the rate of the encoder observing X_2. The setting in which one of the variables is known to the decoder, is the original Wyner-Ziv problem [2]. Other examples include the source coding problem with side information of Wyner [3] and Ahlswede-Korner [4], where an arbitrary distortion is allowed for one of the sources and the other source should be reconstructed losslessly. Berger and Yeung [5] considered a setting where one of the sources is to be perfectly reconstructed and the other source should be reconstructed with a distortion constraint (their setting subsumes all previous examples). Zamir and Berger [6] characterized the rate-distortion region in the high-SNR limit. Wagner and Anantharan [7] presented a new outer bound which is better than the previous outer bounds in the literature.

Recent results for specific sources and distortion measures include the work of Wagner, Tavildar, and Viswanath [8], who determined the rate region for the quadratic Gaussian multi-terminal source coding problem, by showing that the Berger-Tung [9] inner bound is tight. In addition, characterization of the rate region under logarithmic loss was given by Courtade and Weissman [10]. Finally, a version of this problem, where both users and the decoder must operate with zero-delay, was considered by Kaspi and Merhav [11], who characterized the rate region in this case.

In [12], Ziv presented a universal coding scheme for the single-user case. This scheme is composed of a uniform, one-dimensional quantizer with dither, followed by a noiseless variable-rate encoder (entropy encoder). He then showed that this scheme can yield a rate that is, for any n, no more than 0.754 bits per sample higher than the best possible rate associated with the optimal n-dimensional quantizer. This result was later revisited and developed by Zamir and Feder [13], [14], who also gave a redundancy upper bound which depends on the source distribution. However, their derivation of the global upper bound relies on the known formula of the single-user rate-distortion function.

In this paper, we investigate a generalized scheme for the multi-user setting. In this scheme, each user uses dithered quantizer followed by Slepian-Wolf encoder. We show that the rates achieved by this scheme are no more than 0.754 bits per sample away from the boundary of the achievable rate region, for each user. This is done regardless of the characterization of the achievable region, which is, as mentioned before, unknown in general. As a direct consequence of these results, inner and outer bounds on the achievable region are obtained.

A. Problem formulation

Throughout the paper, random variables will be denoted by capital letters and their alphabets will be denoted by calligraphic letters. Random vectors (all of length n) will be denoted by capital letters in the bold face font.

We first define the multi-user rate region and the dithered coding scheme we use. A rate pair (R_1^*, R_2^*) is (D_1, D_2)-achievable under the mean-square error distortion measure d for a memoryless source $(\mathcal{X}_1, \mathcal{X}_2, P_{X_1X_2})$ if for any $\delta > 0$

and sufficiently large n, there exists a code of block length n consisting of two encoders f_1, f_2 and a decoder g, defined as

$$\begin{aligned} f_1 &: \mathcal{X}_1^n \to I_{M_1}, f_2 : \mathcal{X}_2^n \to I_{M_2} \\ g &: I_{M_1} \times I_{M_2} \to \hat{\mathcal{X}}_1^n \times \hat{\mathcal{X}}_2^n \end{aligned} \tag{2}$$

such that

$$\begin{aligned} &\frac{1}{n}\mathbb{E}d(\mathbf{X}_1, \hat{\mathbf{X}}_1) \leq D_1 + \delta, \quad \frac{1}{n}\log M_1 \leq R_1^* + \delta \\ &\frac{1}{n}\mathbb{E}d(\mathbf{X}_2, \hat{\mathbf{X}}_2) \leq D_2 + \delta, \quad \frac{1}{n}\log M_2 \leq R_2^* + \delta \end{aligned} \tag{3}$$

where $I_{M_i} \triangleq \{1, 2, \ldots, M_i\}$, $i \in \{1, 2\}$. The convex hull of the set of (D_1, D_2)-achievable rate pairs, is denoted by $\mathcal{R}^*(D_1, D_2)$. From now on, we assume that $\mathcal{X}_i = \hat{\mathcal{X}}_i = \mathbb{R}$, $i \in \{1, 2\}$.

Our scheme works as follows. We have two encoders $\tilde{f}_1$, $\tilde{f}_2$ and a decoder g

$$\begin{aligned} \tilde{f}_1 : &\quad \mathcal{X}_1^n \times [-\sqrt{3D_1}, \sqrt{3D_1}] \to I_{M_1} \\ \tilde{f}_2 : &\quad \mathcal{X}_2^n \times [-\sqrt{3D_2}, \sqrt{3D_2}] \to I_{M_2} \\ \tilde{g} : &\quad I_{M_1} \times I_{M_2} \times [-\sqrt{3D_1}, \sqrt{3D_1}] \\ &\quad \times [-\sqrt{3D_2}, \sqrt{3D_2}] \to \hat{\mathcal{X}}_1^n \times \hat{\mathcal{X}}_2^n \end{aligned} \tag{4}$$

Each encoder $\tilde{f}_i$, $i \in \{1, 2\}$, uses a one-dimensional uniform quantizer Q_i, $Q_i : \mathbb{R} \to \{0, \pm 2\sqrt{3D_i}, \pm 2 \cdot 2\sqrt{3D_i}, \ldots\}$ and a dither random variable (RV) Z_i, uniformly distributed over $[-\sqrt{3D_i}, \sqrt{3D_i}]$, to produce $\mathbf{Q}_i(\mathbf{X}_i + \mathbf{Z}_i) \triangleq [Q_i(X_{i,1} + Z_i), Q_i(X_{i,2} + Z_i), \ldots, Q_i(X_{i,n} + Z_i)]$, where $\mathbf{Z}_i$ denotes a vector of size n composed of n repetitions of the same realization of Z_i. For convenience, $Q_i(X_i + Z_i)$ and $\mathbf{Q}_i(\mathbf{X}_i + \mathbf{Z}_i)$ will be denoted by Q_i and $\mathbf{Q}_i$, respectively. The dither RV's $\{Z_i\}$ are available to both encoders and to the decoder and are independent. As shown in [12, Lemma 1],

$$\frac{1}{n}\mathbb{E}||\mathbf{Q}_i - \mathbf{Z}_i - \mathbf{X}_i||^2 = D_i, \quad i \in \{1, 2\} \tag{5}$$

where the expectation is taken over Z_i. The entropy encoders use Slepian-Wolf encoding[1], with a rate pair (R_1, R_2), for losslessly compressing $\{\mathbf{Q}_i\}_{i=1}^2$. Complying with Eq. (1), the rate pair satisfies the following:

$$\begin{aligned} R_1 &\geq \frac{1}{n}H(\mathbf{Q}_1|\mathbf{Q}_2, Z_1, Z_2) \\ R_2 &\geq \frac{1}{n}H(\mathbf{Q}_2|\mathbf{Q}_1, Z_1, Z_2) \\ R_1 + R_2 &\geq \frac{1}{n}H(\mathbf{Q}_1, \mathbf{Q}_2|Z_1, Z_2) \end{aligned} \tag{6}$$

This rate region is achievable for n sufficiently large and it is denoted by $\mathcal{R}(D_1, D_2)$. Notice that the interesting range of R_1 is $\overline{\mathcal{R}}_1(D_1, D_2) \triangleq \left[n^{-1}H(\mathbf{Q}_1|\mathbf{Q}_2, Z_1, Z_2), n^{-1}H(\mathbf{Q}_1|Z_1)\right]$ and similarly for R_2. The decoder first decodes $\{\mathbf{Q}_i\}_{i=1}^2$ (correctly with high probability), and then subtracts the dithers to obtain the reconstruction vectors $\hat{\mathbf{X}}_1$, $\hat{\mathbf{X}}_2$:

$$\hat{\mathbf{X}}_i = \mathbf{Q}_i - \mathbf{Z}_i. \tag{7}$$

[1] Since Slepian-Wolf encoding works for a finite-source alphabet, we assume for simplicity that the sources have bounded supports.

We begin with a simple result.

Theorem 1: For any rate pair $(R_1^*, R_2^*) \in \mathcal{R}^*(D_1, D_2)$ and any rate pair (R_1, R_2) on the boundary of $\mathcal{R}(D_1, D_2)$, with $R_1 \in \overline{\mathcal{R}}_1(D_1, D_2)$, we have

$$R_1 + R_2 \leq R_1^* + R_1^* + 2c \tag{8}$$

Moreover, if $R_1^* \in \overline{\mathcal{R}}_1(D_1, D_2)$, then there exists a rate pair $(R_1, R_2) \in \mathcal{R}(D_1, D_2)$ such that

$$\begin{aligned} R_1 &= R_1^* \\ R_2 &\leq R_2^* + 2c \end{aligned} \tag{9}$$

where $c = 0.754$.

Proof of Theorem 1: We have

$$\begin{aligned} &\frac{1}{n}H(\mathbf{Q}_1, \mathbf{Q}_2|Z_1, Z_2) \\ \leq\ &\frac{1}{n}H(\mathbf{Q}_1, \mathbf{Q}_2, m_1, m_2|Z_1, Z_2) \\ \leq\ &\frac{1}{n}H(m_1, m_2) + \frac{1}{n}H(\mathbf{Q}_1, \mathbf{Q}_2|m_1, m_2, Z_1, Z_2) \\ \leq\ &R_1^* + R_2^* + \frac{1}{n}H(\mathbf{Q}_1, \mathbf{Q}_2|m_1, m_2, Z_1, Z_2) \\ \leq\ &R_1^* + R_2^* + \frac{1}{n}H(\mathbf{Q}_1, \mathbf{Q}_2|g(m_1, m_2), Z_1, Z_2) \\ =\ &R_1^* + R_2^* + \frac{1}{n}H\left(\mathbf{Q}_1, \mathbf{Q}_2|\hat{\mathbf{X}}_1^{opt}, \hat{\mathbf{X}}_2^{opt}, Z_1, Z_2\right) \\ \leq\ &R_1^* + R_2^* + \frac{1}{n}H\left(\mathbf{Q}_1|\hat{\mathbf{X}}_1^{opt}, Z_1\right) \\ &+ \frac{1}{n}H\left(\mathbf{Q}_2|\hat{\mathbf{X}}_2^{opt}, Z_2\right) \\ \leq\ &R_1^* + R_2^* + 2c \end{aligned} \tag{10}$$

where m_1, m_2 are the outputs of encoders f_1, f_2, respectively, and $(R_1^*, R_2^*) \in \mathcal{R}^*(D_1, D_2)$. The last inequality can be obtained in the same way as in [12]. The left-hand side is achievable for sufficiently large n. Therefore, for any rate pair $(R_1, R_2) \in \mathcal{R}(D_1, D_2)$, which lies on the straight line $R_1 + R_2 = n^{-1}H(\mathbf{Q}_1, \mathbf{Q}_2|Z_1, Z_2)$, we have

$$R_1 + R_2 \leq R_1^* + R_2^* + 2c \tag{11}$$

Moreover, if $R_1^* \in \overline{\mathcal{R}}_1(D_1, D_2)$, we can always take $R_1 = R_1^*$ and obtain:

$$R_2 \leq R_2^* + 2c \tag{12}$$

The same can be done, of course, if the roles of the two users are interchanged. This completes the proof. ■

The following theorem suggests another result, regarding the relation between the boundary of $\mathcal{R}(D_1, D_2)$ and that of $\mathcal{R}^*(D_1, D_2)$.

Theorem 2: For any rate pair (R_1, R_2) on the boundary of $\mathcal{R}(D_1, D_2)$, with $R_1 \in \overline{\mathcal{R}}_1(D_1, D_2)$, there exists a rate pair $(R_1^*, R_2^*) \in \mathcal{R}^*(D_1, D_2)$ such that:

$$\begin{aligned} R_1 &\leq R_1^* + c \\ R_2 &\leq R_2^* + c \end{aligned} \tag{13}$$

Notice that Theorems 1 and 2 provide a characterization of $\mathcal{R}^*(D_1, D_2)$. Theorem 1 asserts that the straight line $R_1 + R_2 = n^{-1}H(\mathbf{Q}_1, \mathbf{Q}_2|Z_1, Z_2) - 2c$ defines an outer bound for $\mathcal{R}^*(D_1, D_2)$. In addition, Theorem 2 bounds the distance between the boundary of $\mathcal{R}(D_1, D_2)$ and that of $\mathcal{R}^*(D_1, D_2)$, in each coordinate. This result will be useful in Theorem 4 below, where different upper bounds are obtained for each coordinate. Also notice that using methods similar to those of [14], we can derive a distribution-dependent upper bound for the redundancy of the sum of rates. However, this upper bound contains the divergence between the source and the Gaussian distribution and thus it is not universal as the bound of Theorem 1.

We first prove a very simple auxiliary result regarding the source-coding problem where side information is available only to the encoders but not to the decoder. The setting is as follows: A rate pair (R_1, R_2) is achievable for a memoryless source $(\mathcal{X}_1, \mathcal{X}_2, P_{X_1,X_2})$ and some side information $S \in \mathcal{S}$ which depends statistically on (X_1^n, X_2^n) through the joint probability distributions $P_{X_1^n,X_2^n,S}$, if for any $\delta > 0$ and sufficiently large n, there exists a block code of length n consisting of two encoders f_1, f_2 and a decoder g

$$\begin{aligned} f_1: &\quad \mathcal{X}_1^n \times \mathcal{S} \to I_{M_1} \\ f_2: &\quad \mathcal{X}_2^n \times \mathcal{S} \to I_{M_2} \\ g: &\quad I_{M_1} \times I_{M_2} \to \mathcal{X}_1^n \times \mathcal{X}_2^n \end{aligned} \tag{14}$$

such that

$$\Pr\{g\left(f_1(X_1^n, S), f_2(X_2^n, S)\right) \neq (X_1^n, X_2^n)\} \leq \delta \tag{15}$$

and

$$\frac{1}{n}\log M_1 \leq R_1 + \delta, \quad \frac{1}{n}\log M_2 \leq R_2 + \delta \tag{16}$$

The convex hull of the set of achievable rate pairs is denoted by $\tilde{\mathcal{R}}$. The regular Slepian-Wolf region (without side information) is denoted by $\mathcal{R}_{SW}$. Obviously, $\mathcal{R}_{SW} \subseteq \tilde{\mathcal{R}}$. We have the following:

Lemma 1: Any rate pair $(\tilde{R}_1, \tilde{R}_2) \in \tilde{\mathcal{R}}$ must satisfy the following constraint:

$$\tilde{R}_1 + \tilde{R}_2 \geq H(X_1, X_2). \tag{17}$$

Therefore, side information available only to the encoders cannot improve the performance if $\tilde{R}_1 \in [H(X_1|X_2), H(X_1)]$ or $\tilde{R}_2 \in [H(X_2|X_1), H(X_2)]$.

Proof of Lemma 1: The proof follows directly from the fact that even one encoder, which has access to $(\mathbf{X}_1, \mathbf{X}_2, S)$, cannot do better than $H(\mathbf{X}_1, \mathbf{X}_2)$, when the side information S is not available to the decoder. ∎

The generalization of Lemma 1 to our case where, in addition, a dither is available to the encoders and decoder, is straightforward. We can now prove Theorem 2.

Proof of Theorem 2: Assume that the optimal code (f_1, f_2, g), which achieves the rate pair (R_1^*, R_2^*), is known, and that the encoders of the dithered scheme, which transmit $\mathbf{Q}_1$, $\mathbf{Q}_2$ at rates (R_1, R_2) to the decoder, have access to $f_1(\mathbf{X}_1)$, $f_2(\mathbf{X}_2)$. According to Lemma 1, this side information does not change the fact that any rate pair $(R_1, R_2) \in \mathcal{R}(D_1, D_2)$ must satisfy $R_1 + R_2 \geq n^{-1}H(\mathbf{Q}_1, \mathbf{Q}_2|Z_1, Z_2)$. Consider the following auxiliary coding scheme: User i compresses $m_i = f_i(\mathbf{X}_i)$ using nR_i^* bits, $i \in \{1, 2\}$. Then, the first user uses Slepian-Wolf coding to compress $\mathbf{Q}_1$ given $\{m_1, m_2, Z_1\}$ into $H(\mathbf{Q}_1|m_1, m_2, Z_1)$ bits. The second user uses Slepian-Wolf coding to compress $\mathbf{Q}_2$ given $\{\mathbf{Q}_1, m_1, m_2, Z_1, Z_2\}$ into $H(\mathbf{Q}_2|\mathbf{Q}_1, m_1, m_2, Z_1, Z_2)$ bits. The decoder, which has access to $\{m_1, m_2, Z_1, Z_2\}$, first decodes $\mathbf{Q}_1$, using $\{m_1, m_2, Z_1\}$. Then, it decodes $\mathbf{Q}_2$ using $\{\mathbf{Q}_1, m_1, m_2, Z_1, Z_2\}$. The rate pair of this scheme, (R_1, R_2), satisfies

$$\begin{aligned} R_1 &= R_1^* + \frac{1}{n}H\left(\mathbf{Q}_1|m_1, m_2, Z_1\right) \\ &\leq R_1^* + \frac{1}{n}H\left(\mathbf{Q}_1|g(m_1, m_2), Z_1\right) \\ &= R_1^* + \frac{1}{n}H\left(\mathbf{Q}_1|\hat{\mathbf{X}}_1^{opt}, \hat{\mathbf{X}}_2^{opt}, Z_1\right) \\ &\leq R_1^* + \frac{1}{n}H\left(\mathbf{Q}_1|\hat{\mathbf{X}}_1^{opt}, Z_1\right) \end{aligned} \tag{18}$$

and

$$\begin{aligned} R_2 &= R_2^* + \frac{1}{n}H\left(\mathbf{Q}_2|\mathbf{Q}_1, m_1, m_2, Z_1, Z_2\right) \\ &\leq R_2^* + \frac{1}{n}H\left(\mathbf{Q}_2|\mathbf{Q}_1, g(m_1, m_2), Z_1, Z_2\right) \\ &= R_2^* + \frac{1}{n}H\left(\mathbf{Q}_2|\mathbf{Q}_1, \hat{\mathbf{X}}_1^{opt}, \hat{\mathbf{X}}_2^{opt}, Z_1, Z_2\right) \\ &\leq R_2^* + \frac{1}{n}H\left(\mathbf{Q}_2|\hat{\mathbf{X}}_2^{opt}, Z_2\right) \end{aligned} \tag{19}$$

The upper bounds for $H(\mathbf{Q}_i|\hat{\mathbf{X}}_i^{opt}, Z_i)$ can be obtained in the same way as in [12]. Notice that the Slepian-Wolf coding part requires long blocks of $(m_1, m_2, \mathbf{Q}_1, \mathbf{Q}_2)$ so the results are asymptotic in nature. Now, since $\mathcal{R}(D_1, D_2) \subseteq \mathcal{R}^*(D_1, D_2)$, we can always find a rate pair $(R_1^*, R_2^*) \in \mathcal{R}^*(D_1, D_2)$ such that $R_1^* + c \in \overline{\mathcal{R}}_1(D_1, D_2)$. From the above, the rate pair $(R_1, R_2) = (R_1^* + c, R_2^* + c)$ can be achieved by the auxiliary scheme. Therefore, it can also be achieved by the dithered scheme, since $R_1 \in \overline{\mathcal{R}}_1(D_1, D_2)$, and in this range the regions of the auxiliary scheme and the dithered scheme coincide. Notice that any rate pair in $\mathcal{R}(D_1, D_2)$ can be achieved in practice by time-sharing the two edge points of $\mathcal{R}(D_1, D_2)$. ∎

B. Revisiting the upper bound for $H\left(\mathbf{Q}|\hat{\mathbf{X}}^{opt}, \mathbf{Z}\right)$

The goal of this subsection is to revisit the proof of [12] for the upper bound on $H\left(\mathbf{Q}|\hat{\mathbf{X}}^{opt}, \mathbf{Z}\right)$. The new proof is easier for generalization. The width of the quantization cell is denoted by $\Delta \triangleq 2\sqrt{3D} \Rightarrow D = \Delta^2/12$. Notice that in this subsection we deal with only one source X. First, it can be easily shown [16] that for each coordinate X_k, $k \in \{1, \ldots, n\}$,

$$\mathbb{E}\left[X_k - \hat{X}_k^{opt} + Z\right] = 0. \tag{20}$$

We now rederive the upper bound for $H\left(\mathbf{Q}|\mathbf{X}^{opt},\mathbf{Z}\right)$. Using a method similar to [13], the following can be shown [16]:

$$\begin{aligned} H\left(Q_k|\hat{X}_k^{opt},Z\right) &= I\left(X_k;X_k+Z|\hat{X}_k^{opt}\right) \\ &= h\left(X_k+Z|\hat{X}_k^{opt}\right)-h\left(Z\right) \end{aligned} \quad (21)$$

Now, we can upper bound $h\left(X_k+Z|\hat{X}_k^{opt}\right)$ in the following way:

$$\begin{aligned} & h\left(X_k+Z|\hat{X}_k^{opt}\right) \\ = \; & h\left(X_k-\hat{X}_k^{opt}+Z|\hat{X}_k^{opt}\right) \\ = \; & \sum_{q\in\hat{\mathcal{X}}} P_{\hat{X}_k^{opt}}(q)h\left(X_k-\hat{X}_k^{opt}+Z|\hat{X}_k^{opt}=q\right) \\ \leq \; & \frac{1}{2}\log\left(2\pi e\mathbb{E}\left[\left(X_k-\hat{X}_k^{opt}+Z\right)^2\right]\right) \end{aligned} \quad (22)$$

using Jensen inequality and the maximum-entropy property of the Gaussian random variable. As a result, we can upper bound $H\left(\mathbf{Q}|\hat{\mathbf{X}}^{opt},\mathbf{Z}\right)$ in the following way:

$$\begin{aligned} & H\left(\mathbf{Q}|\hat{\mathbf{X}}^{opt},\mathbf{Z}\right) \\ \leq \; & \sum_{k=1}^{n} H\left(Q_k|\hat{X}_k^{opt},Z\right) \\ \leq \; & \sum_{k=1}^{n}\frac{1}{2}\log\left(2\pi e\mathbb{E}\left[\left(X_k-\hat{X}_k^{opt}+Z\right)^2\right]\right) \\ & -nh(Z) \\ \leq \; & \frac{n}{2}\log\left(2\pi e\frac{1}{n}\sum_{k=1}^{n}\mathbb{E}\left[\left(X_k-\hat{X}_k^{opt}+Z\right)^2\right]\right) \\ & -n\log\Delta \\ = \; & \frac{n}{2}\log\left(2\pi e\frac{1}{n}\mathbb{E}\left\|\mathbf{X}-\hat{\mathbf{X}}^{opt}+\mathbf{Z}\right\|^2\right) \\ & -n\log\Delta \\ \leq \; & \frac{n}{2}\log\left(2\pi e2D\right)-n\log\Delta \\ = \; & \frac{n}{2}\log\left(2\pi e2D\right)-\frac{n}{2}\log(\Delta^2) \\ = \; & \frac{n}{2}\log\left(4\pi eD\right)-\frac{n}{2}\log\left(12D\right) \\ = \; & \frac{n}{2}\log\left(\frac{\pi e}{3}\right) \end{aligned} \quad (23)$$

where the third inequality is due to Jensen, and in the fourth we used the following:

$$\begin{aligned} \frac{1}{n}\mathbb{E}\left\|\mathbf{X}-\hat{\mathbf{X}}^{opt}+\mathbf{Z}\right\|^2 &= \frac{1}{n}\mathbb{E}\left\|\mathbf{X}-\hat{\mathbf{X}}^{opt}\right\|^2+\frac{1}{n}\mathbb{E}\left\|\mathbf{Z}\right\|^2 \\ &\leq 2D \end{aligned} \quad (24)$$

which stems from the independence of X and Z. This completes the proof of the basic result.

Notice that this proof can be adapted to any other difference distortion measure, by replacing the upper bound of Eq. (22) with an appropriate maximum-entropy bound.

C. Improving the bounds by adding an estimation stage

The goal of this subsection is to enhance the results of Subsection A by improving the coding scheme described in the beginning of Subsection 2.1. The idea is to decrease the distortion by adding an estimation stage at the decoder side. The new scheme works as follows. After producing $\mathbf{Q}_1,\mathbf{Q}_2$ and instead of just using them as outputs, the decoder uses them to estimate each one of the source vectors $(\mathbf{X}_1,\mathbf{X}_2)$. For simplicity, we assume that the estimation is done on a symbol-by-symbol basis.

We begin with the following lemma:

Lemma 2: For the multi-terminal setting described in Subsection A, we have ($i\in\{1,2\}$):

$$\mathbb{E}[Q_i-Z_i] = \mathbb{E}[X_i] \quad (25)$$
$$\mathbb{E}[(Q_i-Z_i)^2] = \mathbb{E}[{X_i}^2]+D \quad (26)$$
$$\mathbb{E}[X_i(Q_i-Z_i)] = \mathbb{E}[X_i^2] \quad (27)$$
$$\mathbb{E}\left[(Q_1-Z_1)(Q_2-Z_2)\right] = \mathbb{E}\left[X_1X_2\right] \quad (28)$$
$$\mathbb{E}\left[X_1(Q_2-Z_2)\right] = \mathbb{E}\left[X_1X_2\right] \quad (29)$$
$$\mathbb{E}\left[X_2(Q_1-Z_1)\right] = \mathbb{E}\left[X_1X_2\right] \quad (30)$$

Notice that the results above are true for each coordinate $k\in\{1,\ldots,n\}$. The proof appears in [16].

The improved decoder described below requires the knowledge of the second-order statistics of the source. However, as Lemma 2 shows, these statistics can be estimated from $\{\mathbf{Q}_i\}_{i=1}^2$, so universality can still be maintained.

The decoder of the multi-terminal setting uses the optimal linear estimator, under the MMSE criterion, of X_1 given $(Q(X_1+Z_1)-Z_1,Q(X_2+Z_2)-Z_2,Z_1,Z_2)$. The estimation error is calculated by using the results of Lemma 2. From now on, without loss of generality, we assume that $\mathbb{E}[X_1]=\mathbb{E}[X_2]=0$. The covariance matrix of $\underline{Y}\triangleq[Q(X_1+Z_1)-Z_1,Q(X_2+Z_2)-Z_2]$ is:

$$\Lambda=\begin{pmatrix} \mathbb{E}[X_1^2]+D_1 & \mathbb{E}[X_1X_2] \\ \mathbb{E}[X_1X_2] & \mathbb{E}[X_2^2]+D_2 \end{pmatrix} \quad (31)$$

and the inverse matrix is:

$$\Lambda^{-1}=\frac{1}{|\Lambda|}\begin{pmatrix} \mathbb{E}[X_2^2]+D_2 & -\mathbb{E}[X_1X_2] \\ -\mathbb{E}[X_1X_2] & \mathbb{E}[X_1^2]+D_1 \end{pmatrix} \quad (32)$$

The vector $\mathbb{E}\left[X_1\cdot\underline{Y}^\dagger\right]$ is given by:

$$\mathbb{E}\left[X_1\cdot\underline{Y}^\dagger\right]=\begin{pmatrix} \mathbb{E}[X_1^2] \\ \mathbb{E}[X_1X_2] \end{pmatrix} \quad (33)$$

It can be shown by direct calculation that

$$\Lambda^{-1}\mathbb{E}\left[X_1\cdot\underline{Y}^\dagger\right]=\frac{1}{|\Lambda|}\begin{pmatrix} |\Lambda|-D_1(\mathbb{E}[X_2^2]+D_2) \\ \mathbb{E}[X_1X_2]D_1 \end{pmatrix}$$

Therefore, the optimal linear estimator of X_1 given the vector $\underline{Y}$ is:

$$\hat{X}_1=\underline{Y}\cdot\frac{1}{|\Lambda|}\begin{pmatrix} |\Lambda|-D_1(\mathbb{E}[X_2^2]+D_2) \\ \mathbb{E}[X_1X_2]D_1 \end{pmatrix} \quad (34)$$

The error of the optimal linear estimator is given by:

$$D_1^* = \mathbb{E}\left[X_1^2\right] - \mathbb{E}\left[\hat{X}_1^2\right] \tag{35}$$

It can be shown by direct calculation that the estimation error takes the following form:

$$D_1^* = D_1\left(\frac{\mathbb{E}[X_1^2](\mathbb{E}[X_2^2]+D_2) - \mathbb{E}[X_1X_2]^2}{(\mathbb{E}[X_1^2]+D_1)(\mathbb{E}[X_2^2]+D_2) - \mathbb{E}[X_1X_2]^2}\right)$$

Remember that D_1^* is the distortion of X_1 in the multi-terminal setting, when we add the above estimation stage after decoding $[Q(X_1+Z_1), Q(X_2+Z_2)]$. The same can be done, of course, for X_2. Since the distortion of X_i in the improved scheme is D_i^*, we should compare the rate pair (R_1, R_2) of this scheme, to the optimal rate pair (R_1^*, R_2^*) which achieves (D_1^*, D_2^*). This fact immediately improves the results of Theorems 1 and 2. Revisiting the derivation of the upper bound for $H\left(\mathbf{Q}|\hat{\mathbf{X}}^{opt}, \mathbf{Z}\right)$ in Eq. (23), it can be shown that $(i \in \{1,2\})$:

$$H\left(\mathbf{Q}_i|\hat{\mathbf{X}}_i^{opt}, \mathbf{Z}_i\right) \leq \frac{n}{2}\log\left[\frac{\pi e}{6}\left(\frac{D_i^*}{D_i}+1\right)\right] \tag{36}$$

by using the following:

$$\begin{aligned} \frac{1}{n}\mathbb{E}\left\|\mathbf{X}_i - \hat{\mathbf{X}}_i^{opt} + \mathbf{Z}_i\right\|^2 &= \frac{1}{n}\mathbb{E}\left\|\mathbf{X}_i - \hat{\mathbf{X}}_i^{opt}\right\|^2 \\ &+ \frac{1}{n}\mathbb{E}\left\|\mathbf{Z}_i\right\|^2 \\ &\leq D_i^* + D_i \end{aligned} \tag{37}$$

Notice that when X_1 and X_2 are independent, $\mathbb{E}[X_1X_2] = 0$ and we have

$$H\left(\mathbf{Q}_i|\hat{\mathbf{X}}_i^{opt}, \mathbf{Z_i}\right) \leq \frac{n}{2}\log\left[\frac{\pi e}{6}\left(2 - \frac{D_i}{\mathbb{E}[X_i^2]+D_i}\right)\right] \tag{38}$$

The maximum interesting value of D_i^* is, of course, $\mathbb{E}[X_i^2]$. This value is obtained for $D_i \to \infty$. It is not hard to see that the range of the upper bound in (38) is $[0.255, 0.755]$ and that it is a decreasing function of D_1. For the high-SNR limit, i.e., $D_i \to \infty$, it is well known that the redundancy is 0.255 bits/sample (cf. [15]). It can be shown [16] that this result can be derived using the mechanism above. We define $(i \in \{1,2\})$:

$$c_i(D_1, D_2) = \frac{n}{2}\log\left[\frac{\pi e}{6}\left(\frac{D_i^*}{D_i}+1\right)\right] \tag{39}$$

Using the upper bound in (36), we can state Theorems 3 and 4:

Theorem 3: For any rate pair $(R_1^*, R_2^*) \in \mathcal{R}^*(D_1^*, D_2^*)$ and any rate pair (R_1, R_2) on the boundary of $\mathcal{R}(D_1^*, D_2^*)$, with $R_1 \in \overline{\mathcal{R}}_1(D_1, D_2)$, we have

$$R_1 + R_2 \leq R_1^* + R_1^* + c_1(D_1, D_2) + c_2(D_1, D_2) \tag{40}$$

Moreover, if $R_1^* \in \overline{\mathcal{R}}_1(D_1, D_2)$, then there exists a rate pair $(R_1, R_2) \in \mathcal{R}(D_1^*, D_2^*)$ such that:

$$\begin{aligned} R_1 &= R_1^* \\ R_2 &\leq R_2^* + c_1(D_1, D_2) + c_2(D_1, D_2) \end{aligned} \tag{41}$$

Theorem 4: For any rate pair (R_1, R_2) on the boundary of $\mathcal{R}(D_1^*, D_2^*)$, with $R_1 \in \overline{\mathcal{R}}_1(D_1, D_2)$, there exists a rate pair $(R_1, R_2) \in \mathcal{R}(D_1^*, D_2^*)$ such that:

$$\begin{aligned} R_1 &\leq R_1^* + c_1(D_1, D_2) \\ R_2 &\leq R_2^* + c_2(D_1, D_2) \end{aligned} \tag{42}$$

II. Conclusion

We introduced upper bounds on the redundancies of the dithered scheme in the multi-user setting. The mechanism used does not depend on the knowledge of the optimal rate region and can be extended easily to the case of stationary sources. The suggested scheme requires only one realization of the dither RV in each round. As mentioned in Subsection B, the results can also be extended to other difference distortion measures.

Acknowledgment

This research was supported by The Israeli Science Foundation (ISF), grant no. 208/08.

References

[1] D. Slepian and J. K. Wolf, "Noiseless coding of correlated information sources," *IEEE Trans. Inform. Theory*, vol. 19, pp. 471–480, Jul. 1973.

[2] A. D. Wyner and J. Ziv, "The rate-distortion function for source coding with side information at the decoder," *IEEE Trans. Inform. Theory*, vol. 22, pp. 1–10, Jan. 1976.

[3] A. D. Wyner, "The rate-distortion function for source coding with side information at the decoder-II: General sources," *Inform. Contr.*, vol. 38, pp. 60–80, 1978.

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

[5] T. Berger and R. W. Yeung, "Multiterminal source encoding with one distortion criterion," *IEEE Trans. Inform. Theory*, vol. 35, no. 2, pp. 228–236, Mar. 1989.

[6] R. Zamir and T. Berger, "Multiterminal source coding with high resolution," *IEEE Trans. Inf. Theory*, vol. 45, pp. 106–117, Jan. 1999.

[7] A. Wagner and V. Anantharam, "An improved outer bound for multiterminal source coding," *IEEE Trans. Inf. Theory*, vol. 54, no. 5, pp. 1919–1937, May 2008.

[8] A. Wagner, S. Tavildar, and P. Viswanath, "Rate region of the quadratic gaussian two-encoder source-coding problem," *IEEE Trans. Inf. Theory*, vol. 54, no. 5, pp. 1938–1961, May 2008.

[9] T. Berger and S. Y. Tung, "Encoding of correlated analog sources," in *Proc. IEEE–USSR Joint Workshop on Information Theory*, pp. 7–10, 1975.

[10] T. A. Courtade and T. Weissman, "Multiterminal source coding under logarithmic loss," *CoRR*, vol. abs/1110.3069, 2011.

[11] Y. Kaspi and N. Merhav, "Zero-Delay and Causal Single-User and Multi-User Lossy Source Coding with Decoder Side Information," *Submitted to IEEE Trans. Inf. Theory 2013*, *CoRR*, vol. abs/1301.0079, 2013.

[12] J. Ziv, "On universal quantization," *IEEE Trans. Inform. Theory*, vol. 31, pp. 344–347, May 1985.

[13] R. Zamir and M. Feder, "On universal quantization by randomized uniform/lattice quantizer," *IEEE Trans. Inform. Theory*, vol. 38, pp. 428–436, Mar. 1992.

[14] R. Zamir and M. Feder, "Information rates of pre/post filtered dithered quantizers," *IEEE Trans. Inform. Theory*, vol. 42. pp. 1340–1353, Sept. 1996.

[15] H. Gish and N. J. Pierce, "Asymptotically efficient quantization," *IEEE Trans. Inform. Theory*, vol. IT-14, pp. 676–683, Sept. 1968.

[16] A. Reani, N. Merhav, "Universal quantization for separate encodings and joint decoding of correlated sources," in preparation.