Multiterminal Source Coding with High Resolution

Ram Zamir, *Member, IEEE*, and Toby Berger, *Fellow, IEEE*

***Abstract*—We consider separate encoding and joint decoding of correlated continuous information sources, subject to a difference distortion measure. We first derive a multiterminal extension of the Shannon lower bound for the rate region. Then we show that this Shannon outer bound is asymptotically tight for small distortions. These results imply that the loss in the sum of the coding rates due to the separation of the encoders vanishes in the limit of high resolution. Furthermore, lattice quantizers followed by Slepian–Wolf lossless encoding are asymptotically optimal. We also investigate the high-resolution rate region in the *remote* coding case, where the encoders observe only noisy versions of the sources. For the quadratic Gaussian case, we establish a separation result to the effect that multiterminal coding aimed at reconstructing the noisy sources subject to the rate constraints, followed by estimation of the remote sources from these reconstructions, is optimal under certain regularity conditions on the structure of the coding scheme.**

***Index Terms*— Direct and remote rate-distortion, high-resolution quantization, multiterminal source coding, Shannon lower bound.**

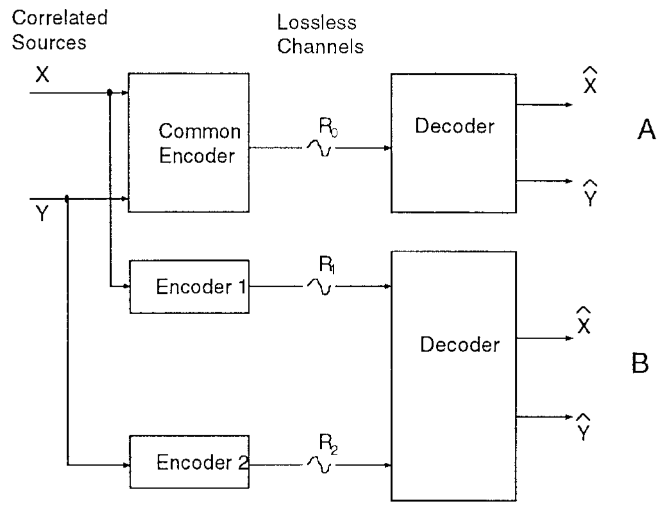

Fig. 1. Direct coding of correlated sources: (*A*) joint encoder and decoder and (*B*) separate encoders and common decoder.

I. Introduction

IN their celebrated paper [18], Slepian and Wolf showed that it is possible to encode correlated information sources separately and decode them jointly with a vanishing error, without paying in rate relative to jointly encoding the sources. That is, in spite of the fact that the encoders operate independently of one another, they can reduce the rate sum to the *joint* entropy of the dependent sources. Specifically, suppose X and Y are discrete memoryless correlated sources, encoded separately at rates R_1 and R_2, respectively; see Fig. 1—Case B. Then, there exist independent X- and Y-codes from which the sources can be decoded jointly with arbitrarily small error frequency if and only if

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

where $H(\cdot)$ denotes entropy [5].

A natural lossy extension of the multiterminal (or distributed) data compression problem is to allow some distortion, say, at levels D_x and D_y, in the reconstruction of the sources. A general solution to this problem in terms of single letter information quantities has not been found.[1] Yet, we do know the solution for some important special cases [2]–[4], [11], [16], [27]. These partial solutions tell us, for example, that unlike the lossless coding case treated by Slepian and Wolf, in multiterminal *lossy* source coding there is a cost in rate due to the separation of the encoders; see, e.g., [27] and [28]. However, many questions are still open in the general setting. One is whether or not the optimal solution can be separated into a "simple" quantization stage for each terminal followed by a Slepian–Wolf lossless coding stage.

In this work we address the problem of multiterminal lossy coding of continuous sources in the extreme of *high resolution*. As in the single-terminal case [12], the high-resolution limit allows us to get an explicit "single-letter" solution for the problem. Our solution is based on extending the Shannon lower bound (SLB) [1], and then showing that this extension becomes tight as the distortion levels D_x and D_y go to zero. For the squared-error distortion case, the extended bound has the following simple structure, highly reminiscent of the Slepian–Wolf rate region (1):

$$R_1 \geq h(X|Y) - \tfrac{1}{2}\log 2\pi e D_x \tag{2a}$$

$$R_2 \geq h(Y|X) - \tfrac{1}{2}\log 2\pi e D_y \tag{2b}$$

$$R_1 + R_2 \geq h(X, Y) - \tfrac{1}{2}\log(2\pi e)^2 D_x D_y \tag{2c}$$

where $h(\cdot)$ denotes differential entropy; see Fig. 2.

Manuscript received April 12, 1997; revised April 29, 1998. This work was supported in part by the Wolfson Research Awards, administered by the Israel Academy of Science and Humanities, and by the NSF under Grants NCR-9216975 and IRI-9310670. This work was presented in part at the International Symposium on Information Theory, Whistler, BC, Canada, September 1995. This work was performed while R. Zamir was at Cornell University.

R. Zamir is with the Department of Elelctrical Engineering–Systems, Tel-Aviv University, 69978 Tel-Aviv, Israel (e-mail: zamir@eng.tau.ac.il).

T. Berger is with the School of Electrical Engineering, Cornell University, Ithaca, NY 14853 USA (e-mail: berger@ee.cornell.edu).

Communicated by K. Zeger, Associate Editor At Large.
Publisher Item Identifier S 0018-9448(99)00061-9.

[1] Haroutunian [10] alleged to have found a simple solution to the general multiterminal source-coding problem, but the arguments were invalid and the results have been retracted.

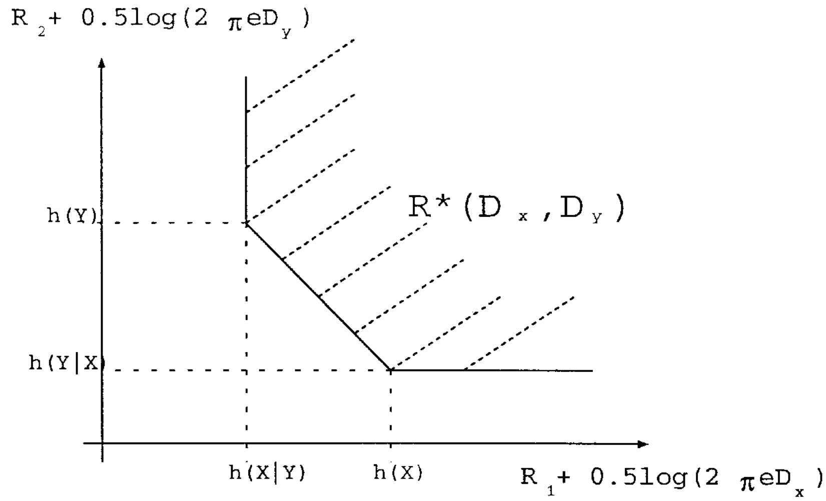

Fig. 2. The rate region $\mathcal{R}^*(D_x, D_y)$.

Note that by the SLB, the joint rate-distortion function of (X, Y) is also lower-bounded by the right-hand side of (2c). Thus our result implies that at high resolution there is no loss in rate sum relative to jointly encoding the sources. Furthermore, the forward theorem implies that optimal rate-distortion performance at high-resolution conditions can be achieved *systematically*, by first quantizing each source to its allowed distortion (ignoring the presence of the other source), and then employing Slepian–Wolf lossless encoding to the quantizers' outputs. For example, the first stage can be implemented by lattice or tessellating quantizers [21], [29].

In the setting discussed above, the encoders observe *directly* the sources which need to be reconstructed at the joint decoder. In this work we also consider *indirect (or remote)* multiterminal source coding, where the sources X and Y are noisy versions of a source Θ (referred to as the "clean" or "remote" source) which needs to be reconstructed at the joint decoder. Let

$$D^{\min} = E[\Theta - E(\Theta|X, Y)]^2 \triangleq \mathrm{COV}(\Theta|X, Y)$$

be the minimum mean-squared error in estimating Θ from (X, Y), and suppose that the desired distortion after encoding and decoding with a finite rate is some $D > D^{\min}$. If the noisy versions of the source are encoded *jointly* (see Fig. 3—Case 1), then it is well known, [7], [8], [17], [25], that to minimize the rate we should encode the optimal estimator

$$\hat{\Theta} = E\{\Theta|X, Y\}$$

with distortion $D - D^{\min}$. Thus the indirect rate-distortion function of the source is given by the ordinary rate-distortion function of the random variable $\hat{\Theta}$ at distortion level $D - D^{\min}$.

However, when the encoder of X is separated from the encoder of Y, as shown in Fig. 3—Case 2, it is generally impossible to calculate the conditional mean $E\{\Theta|X, Y\} = \mathrm{func}(X, Y)$ at the encoders. What can we do then? A few special cases of this more general multiterminal source coding problem are considered in [15], [20], [22], and [24], but its general solution is unknown.

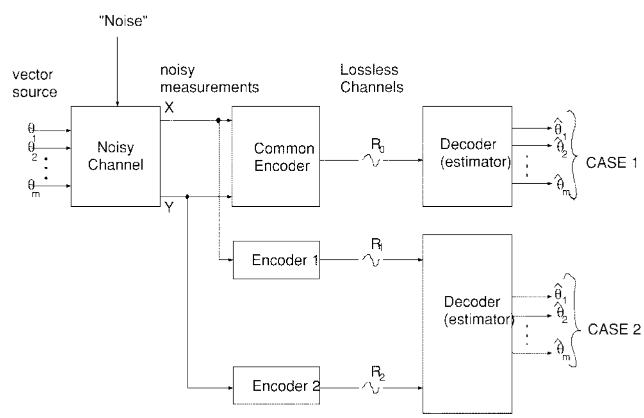

Fig. 3. Remote coding of correlated sources: (1) joint coding and (2) separate coding.

Using our results for the direct coding case, we investigate the rate region of multiterminal remote source coding in the extreme of high resolution, i.e., when $D - D^{\min}$ is small. Restricting the discussion to the Gaussian source/Gaussian measurements case, we derive asymptotic inner and outer bounds to the rate region in terms of the "Shannon outer bound" (2).

Unfortunately, unlike in the direct-coding case, the inner and outer bounds do not coincide asymptotically. Interestingly, the gap between these bounds depends on the correlation coefficient ρ between the quantization errors of the *noisy* sources X and Y, a factor which does not play a role in the direct coding case. The outer bound coincides with the inner bound if one limits attention to the class of encoders for which ρ goes to zero as $D \to D_{\min}$. We show that this zero correlation performance can be obtained by first applying the *direct* coding scheme to the noisy sources X and Y at appropriate distortion levels D_x and D_y (thus achieving a rate pair in (2)), and then estimating the remote source from $\hat{X}$ and $\hat{Y}$.

In Sections II and III, we present our main results for the direct- and the remote-coding problems, respectively. Section IV contains comparisons of the asymptotic rate sum in the multiterminal problem both to the rate of a joint encoder and to the rate sum in the closely related CEO problem [22]. In Section V, we suggest a systematic multiterminal coding scheme, composed of dithered lattice quantizers and Slepian–Wolf encoding, which approaches the asymptotic rate region for large lattice dimension.

For the sake of clarity we assume throughout the paper a configuration with two encoders and a mean-squared error distortion measure. However, the results extend straightforwardly to any number of encoders, and for the direct coding case to general difference distortion measures.

II. Direct-Coding Case

In this section we define the direct multiterminal source-coding scheme and prove an asymptotically tight outer bound (called the "Shannon outer bound") for the rate-distortion region.

Let $\{(X_i, Y_i)\}_{i=1}^{\infty}$ be a sequence of independent drawings of a pair of dependent, real-valued random variables (X, Y). We wish to block-encode $\underline{X} = X_1 \cdots X_n$ and $\underline{Y} = Y_1 \cdots Y_n$ *separately*, and decode them *jointly*, such that

$$\frac{1}{n} E \sum_{i=1}^{n} (\hat{X}_i - X_i)^2 \le D_x$$

and

$$\frac{1}{n} E \sum_{i=1}^{n} (\hat{Y}_i - Y_i)^2 \le D_y$$

where $\underline{\hat{X}} = X_1 \cdots X_n$ and $\underline{\hat{Y}} = Y_1 \cdots Y_n$ are the reconstructed sources. The encoding and decoding functions have the form

$$\begin{aligned} &f_1\colon \mathcal{R}^n \to \{1, \cdots, 2^{nR_1}\} \qquad f_2\colon \mathcal{R}^n \to \{1, \cdots, 2^{nR_2}\} \\ &g\colon \{1, \cdots, 2^{nR_1}\} \times \{1, \cdots, 2^{nR_2}\} \to \mathcal{R}^n \times \mathcal{R}^n \end{aligned} \tag{3}$$

where $(\underline{\hat{X}}, \underline{\hat{Y}}) = g(f_1(\underline{X}), f_2(\underline{Y}))$. The (D_x, D_y)-admissible rate region $\mathcal{R}(D_x, D_y) = \{(R_1, R_2)\}$ is defined in the usual way; see, e.g., [2] and [5]. Fig. 1, part (*B*) shows this coding system.

Assume that the joint differential entropy $h(X, Y)$ exists and is finite, and let

$$\mathcal{R}^*(D_x, D_y) = \left\{ (R_1, R_2)\colon \begin{aligned} R_1 &\ge h(X|Y) - \tfrac{1}{2}\log 2\pi e D_x \\ R_2 &\ge h(Y|X) - \tfrac{1}{2}\log 2\pi e D_y \\ R_1 + R_2 &\ge h(X, Y) - \tfrac{1}{2}\log (2\pi e)^2 D_x D_y. \end{aligned} \right\} \tag{4}$$

Theorem 1 below states that $\mathcal{R}^*(D_x, D_y)$ outer bounds $\mathcal{R}(D_x, D_y)$, while Theorem 2 states that this outer bound becomes tight as the distortions approach zero.

Theorem 1 (Shannon Outer Bound): For any mean-squared error levels D_x and D_y

$$\mathcal{R}(D_x, D_y) \subseteq \mathcal{R}^*(D_x, D_y). \tag{5}$$

Proof: Let [1], [9]

$$R_{X|Y}(D_x) = \inf_{\{\hat{X}\colon E\{(\hat{X}-X)^2\} \le D_x\}} I(X; \hat{X}|Y) \tag{6}$$

denote the conditional rate-distortion function of X given Y, associated with the case when Y is available at both the encoder and the decoder, and let

$$R_{X,Y}(D_x, D_y) = \inf_{\{\hat{X}, \hat{Y}\colon E\{(\hat{X}-X)^2\} \le D_x,\, E\{(\hat{Y}-Y)^2\} \le D_y\}} I(XY; \hat{X}\hat{Y}) \tag{7}$$

denote the joint rate-distortion function of (X, Y) under a double distortion constraint, associated with *joint* encoding of the two sources. By the conditional Shannon lower bound [9]

$$R_{X|Y}(D_x) \ge h(X|Y) - \frac{1}{2}\log 2\pi e D_x \tag{8}$$

and by the regular Shannon lower bound

$$R_{X,Y}(D_x, D_y) \ge h(X, Y) - \frac{1}{2}\log (2\pi e)^2 D_x D_y. \tag{9}$$

Since additional information (Y is available to the X-encoder and X is available to the Y encoder) and fewer constraints (encoders do not need to help one another) may only reduce the coding rate, we have for every $(R_1, R_2) \in \mathcal{R}(D_x, D_y)$

$$\begin{aligned} R_1 &\ge R_{X|Y}(D_x) \\ R_2 &\ge R_{Y|X}(D_y) \\ R_1 + R_2 &\ge R_{X,Y}(D_x, D_y). \end{aligned} \tag{10}$$

Combining the lower bounds (8) and (9) with (10), we obtain the proof of the theorem. □

As in the single-terminal case, if the joint distribution of (X, Y) is "smooth," then as D_x and D_y go to zero, R_1 and R_2 must go to infinity. We say that two rate regions $\mathcal{R}_1(D_x, D_y)$ and $\mathcal{R}_2(D_x, D_y)$ coincide asymptotically as $D_x, D_y \to 0$, and denote that $\mathcal{R}_1(D_x, D_y) \cong \mathcal{R}_2(D_x, D_y)$, if for any $\epsilon > 0$, and D_x, D_y small enough,

$$\mathcal{R}_2^{(-\epsilon)}(D_x, D_y) \subseteq \mathcal{R}_1(D_x, D_y) \subseteq \mathcal{R}_2^{(+\epsilon)}(D_x, D_y) \tag{11}$$

where

$$\begin{aligned} \mathcal{R}^{(+\epsilon)} &\triangleq \{(R_1, R_2)\colon (R_1 + \epsilon, R_2 + \epsilon) \in \mathcal{R}\} \\ \mathcal{R}^{(-\epsilon)} &\triangleq \{(R_1, R_2)\colon (R_1 - \epsilon, R_2 - \epsilon) \in \mathcal{R}\}. \end{aligned} \tag{12}$$

Theorem 2 (Asymptotic Characterization): If

$$E\{X^2\}, E\{Y^2\} < \infty$$

and

$$h(X, Y) > -\infty$$

then as $D_x, D_y \to 0$ the outer bound (5) becomes achievable, i.e.,

$$\mathcal{R}(D_x, D_y) \cong \mathcal{R}^*(D_x, D_y). \tag{13}$$

Proof: We start by constructing an inner bound to $\mathcal{R}(D_x, D_y)$, which holds for any D_x, D_y. For that, we use [2, Theorem 6.1] which provides achievable rates for the lossy multiterminal problem. Although the derivation in [2] assumes sources with a finite alphabet, the extension to real-valued sources satisfying a second-moment condition is straightforward following the method of [3] and [26]. Let $N_1 \sim \mathcal{N}(0, D_x)$ and $N_2 \sim \mathcal{N}(0, D_y)$ be Gaussian variables, independent of (X, Y) and of one another. It is easy to verify that $(X+N_1) \leftrightarrow X \leftrightarrow Y \leftrightarrow (Y+N_2)$ form a "long" Markov chain, i.e., $X+N_1$ and Y are conditionally independent given X, etc. Thus by [2, Theorem 6.1], the rate region

$$\mathcal{R}^{\text{add}}(D_x, D_y) = \left\{ (R_1, R_2): \begin{array}{r} R_1 > I(X; X+N_1|Y+N_2) \\ R_2 > I(Y; Y+N_2|X+N_1) \\ R_1+R_2 > I(XY; X+N_1, Y+N_2) \end{array} \right\} \tag{14}$$

is achievable (for $n \to \infty$). Combining that with the Shannon outer bound of Theorem 1, we have for any D_x and D_y

$$\mathcal{R}^{\text{add}}(D_x, D_y) \subseteq \mathcal{R}(D_x, D_y) \subseteq \mathcal{R}^*(D_x, D_y). \tag{15}$$

We now show that the inner and outer bounds in (15) coincide asymptotically. Note first that

$$\begin{aligned} I(X; X+N_1|Y+N_2) &= h(X+N_1|Y+N_2) - h(N_1) \\ &= h(X+N_1|Y+N_2) - \frac{1}{2}\log 2\pi e D_x. \end{aligned} \tag{16}$$

Using the proof technique of [12, Theorem 1], it follows that $h(X+N_1, Y+N_2) \to h(X, Y)$ as D_x and D_y go to zero, provided that the variances and the joint differential entropy of X and Y are finite. Similarly, $h(Y+N_2) \to h(Y)$ as D_y goes to zero. Thus

$$h(X+N_1|Y+N_2) = h(X+N_1, Y+N_2) - h(Y+N_2)$$

goes to

$$h(X, Y) - h(Y) = h(X|Y)$$

as D_x and D_y go to zero. By substituting in (16), we obtain

$$I(X; X+N_1|Y+N_2) = h(X|Y) - \frac{1}{2}\log 2\pi e D_x + o(1) \tag{17}$$

where $o(1) \to 0$ as $D_x, D_y \to 0$. Similarly

$$I(Y; Y+N_2|X+N_1) = h(Y|X) - \frac{1}{2}\log 2\pi e D_y + o(1) \tag{18}$$

and

$$\begin{aligned} I(XY; X+N_1, Y+N_2) \\ = h(X, Y) - \frac{1}{2}\log(2\pi e)^2 D_x D_y + o(1). \end{aligned} \tag{19}$$

By comparing (17)–(19) with the definitions (4) and (14) of $\mathcal{R}^*(D_x, D_y)$ and $\mathcal{R}^{\text{add}}(D_x, D_y)$, we see that for any $\epsilon > 0$

$$\begin{aligned} (R_1, R_2) \in \mathcal{R}^*(D_x, D_y) \\ \Rightarrow (R_1+\epsilon, R_2+\epsilon) \in \mathcal{R}^{\text{add}}(D_x, D_y) \end{aligned} \tag{20}$$

if D_x and D_y are small enough. Thus by definition (12)

$$\mathcal{R}^*(D_x, D_y) \subseteq \mathcal{R}^{\text{add}(+\epsilon)}(D_x, D_y). \tag{21}$$

Combining with (15), we have

$$\begin{aligned} \mathcal{R}^{\text{add}}(D_x, D_y) &\cong \mathcal{R}(D_x, D_y) \\ &\cong \mathcal{R}^*(D_x, D_y), \qquad \text{as } D_x, D_y \to 0 \end{aligned} \tag{22}$$

as desired. □

Remarks:

1) Using the general form of the SLB [1], it is easy to verify that Theorem 1 applies also to a general difference distortion measure $d(\hat{x} - x)$, where the "Shannon outer bound" is given in this case by

$$\mathcal{R}^*(D_x, D_y) = \left\{ (R_1, R_2): \begin{array}{r} R_1 \geq h(X|Y) - h_{\max}(d, D_x) \\ R_2 \geq h(Y|X) - h_{\max}(d, D_y) \\ R_1 + R_2 \geq h(X, Y) - h_{\max}(d, D_x) \\ - h_{\max}(d, D_y) \end{array} \right\} \tag{23}$$

where

$$h_{\max}(d, D) \triangleq \max_{\{Z:\, Ed(Z) \leq D\}} h(Z). \tag{24}$$

If in addition d satisfies the conditions of [26] and [12, Sec. III] (e.g., d is an rth-power distortion measure $d(\hat{x}-x) = |\hat{x}-x|^r$), and $Ed(X), Ed(Y) < \infty$, then also Theorem 2 holds, i.e., $\mathcal{R}^*(D_x, D_y)$ of (23) is asymptotically tight. See the proof technique of [12].

2) We can further characterize the rate at which $\mathcal{R}(D_x, D_y)$ converges to $\mathcal{R}^*(D_x, D_y)$ as D_x and D_y go to zero in the squared-error distortion case. As shown in the Appendix, under a certain smoothness condition on the joint density of (X, Y), the distance between $\mathcal{R}(D_x, D_y)$ and $\mathcal{R}^*(D_x, D_y)$, as measured by the "ϵ" in the definition of "$\cong$" in (11), goes to zero as $O(D_x) + O(D_y)$.

3) For (X, Y) Gaussian with variances σ_x^2 and σ_y^2 and correlation coefficient ρ, the rate region (4) becomes

$$\mathcal{R}^*(D_x, D_y) = \left\{ (R_1, R_2): \begin{array}{r} R_1 \geq \frac{1}{2}\log\left(\frac{(1-\rho^2)\sigma_x^2}{D_x}\right) \\ R_2 \geq \frac{1}{2}\log\left(\frac{(1-\rho^2)\sigma_y^2}{D_y}\right) \\ R_1 + R_2 \geq \frac{1}{2}\log\left(\frac{(1-\rho^2)\sigma_x^2\sigma_y^2}{D_x D_y}\right) \end{array} \right\} \tag{25}$$

thus showing that the inner bound $\mathcal{R}^*_{in}(D_x, D_y)$ of [2] and [3] is asymptotically tight as $D_x, D_y \to 0$.

4) In Section V we show that for small distortions, lattice quantizers followed by Slepian–Wolf encoding can approach the rate region $\mathcal{R}^*(D_x, D_y)$, by letting the lattice dimension go to infinity.

III. Remote-Coding Case

A. Definition and Results

Next we restrict our attention to the quadratic Gaussian case but consider the following more general, *indirect* coding problem [1, pp. 78, 124]. The decoder needs to reconstruct a vector source $\underline{\Theta} = (\Theta_1 \cdots \Theta_m)$ with average squared errors $D_{\theta 1}, \cdots, D_{\theta m}$, from encodings of X and Y. The sources $\underline{\Theta}$, X and Y are zero-mean, jointly Gaussian, and memoryless. We refer to $\underline{\Theta}$ as the *remote* (or "clean") source, and to X and Y as its *noisy versions*.

We consider *separate* encoding of X and Y, as shown in Fig. 3—Case 2.[2] Let $\underline{D}_\theta = (D_{\theta 1}, \cdots, D_{\theta m})$ denote the distortion vector, and denote by $\mathcal{R}(\underline{D}_\theta) = \{(R_1, R_2)\}$ the set of $\underline{D}_\theta$-admissible rate pairs (R_1, R_2) of the X- and Y-encoders. To state our results, we will sometimes assume that the encoders are restricted to a certain class $\mathcal{C}$ of coding systems, e.g., lattice quantizers followed by Slepian–Wolf coding. In this case we denote the set of $\underline{D}_\theta$-admissible rate pairs by $\mathcal{R}(\underline{D}_\theta, \mathcal{C})$.

When X and Y are available to the decoder with infinite resolution, the optimal reconstruction of $\underline{\Theta}$ is its conditional mean given (X, Y) [19], which due to the Gaussian statistics has a linear form

$$E(\underline{\Theta}|X, Y) = T \cdot (X, Y)^t \tag{26}$$

where T is some $m \times 2$ matrix and $(\cdot)^t$ denotes transpose. The mean squared errors are then the diagonal elements $D_{\theta 1}^{\min} \cdots D_{\theta m}^{\min}$ of the conditional covariance matrix

$$\mathrm{COV}(\underline{\Theta}|X, Y) \stackrel{\Delta}{=} E\{[\underline{\Theta} - E(\underline{\Theta}|X, Y)] \cdot [\underline{\Theta} - E(\underline{\Theta}|X, Y)]^t\} \tag{27}$$

where the outer expectation in (27) is taken over the joint distribution of $(\underline{\Theta}, X, Y)$.

Clearly, for any coding rates the decoder can achieve only distortions satisfying $D_{\theta j} \geq D_{\theta j}^{\min}$, and usually when

$$\underline{D}_\theta \to \underline{D}_\theta^{\min} \stackrel{\Delta}{=} (D_{\theta 1}^{\min}, \cdots, D_{\theta m}^{\min})$$

the coding rates go to infinity. We are interested in the asymptotic behavior of the rate region in this limit.

In order to state our main results for the remote coding problem, we need to introduce some definitions. For T of (26) and a positive vector $\underline{\Delta} = (\Delta_1, \cdots, \Delta_m)$, define a set of 2×2 nonnegative definite matrices

$$\mathcal{K}(\underline{\Delta}, T) = \{K: (TKT^t)_{j,j} \leq \Delta_j, \qquad \text{for } j = 1 \cdots m\}. \tag{28}$$

[2] This formulation can be regarded as a special case of *direct* multiterminal coding of the vector $(X, Y, \Theta_1, \cdots, \Theta_m)$, assuming zero rate is allocated to the encoders of $\Theta_1, \cdots, \Theta_m$, and no distortion constraint is imposed on X and Y.

The set $\mathcal{K}(\underline{\Delta}, T)$ may be interpreted as follows (see Lemma 2 below). The remote source $\underline{\Theta}$ can be decoded with distortion vector $\underline{D}_\theta^{\min} + \underline{\Delta}$ *if and only if* the noisy source (X, Y) can be decoded from the decoder's input with error covariance K belonging to the set $\mathcal{K}(\underline{\Delta}, T)$.

For a positive vector $\underline{\Delta}$, define

$$\mathcal{R}^{**}(\underline{\Delta}) = \bigcup_{\{a, b: \mathrm{diag}[a, b] \in \mathcal{K}(\underline{\Delta}, T)\}} \mathcal{R}^*(a, b) \tag{29}$$

where $\mathrm{diag}[a, b]$ denotes a diagonal matrix whose diagonal elements are a and b, and where the rate region $\mathcal{R}^*(\cdot)$ is defined in (4). Theorem 3 states that the region $\mathcal{R}^{**}(\underline{D}_\theta - \underline{D}_\theta^{\min})$ constitutes an asymptotic inner bound for the remote coding rate region $\mathcal{R}(\underline{D}_\theta)$.

Theorem 3 (Asymptotic Inner Bound—Remote Case): Assume that $\underline{\Theta}$, X and Y are jointly Gaussian variables with finite second moments, such that the covariance matrix $\mathrm{COV}(X, Y)$ is not singular, and the optimal estimation matrix T defined in (26) has no all-zero columns. Then

$$\mathcal{R}(\underline{D}_\theta) \stackrel{\sim}{\supset} \mathcal{R}^{**}(\underline{D}_\theta - \underline{D}_\theta^{\min}), \qquad \text{as } \underline{D}_\theta \to \underline{D}_\theta^{\min} \tag{30}$$

where the notation $\stackrel{\sim}{\supset}$ means that for any $\epsilon > 0$ and $\underline{D}_\theta - \underline{D}_\theta^{\min}$ small enough

$$\mathcal{R}(\underline{D}_\theta) \supseteq \mathcal{R}^{**(-\epsilon)}(\underline{D}_\theta - \underline{D}_\theta^{\min}). \tag{31}$$

The proof is given in the second part of this section. We proceed to develop an asymptotic outer bound for $\mathcal{R}(\underline{D}_\theta)$.

First we need a few more definitions. Let $W = (f_1(\underline{X}), f_2(\underline{Y}))$ denote the information received by the common decoder. Let $\underline{\Theta}_1 \cdots \underline{\Theta}_n$ denote the corresponding samples of the remote vector source. A minimum mean-squared error decoder estimates $\underline{\Theta}_i$ as $E(\underline{\Theta}_i|W)$, so the time average of the error covariance in $\underline{\Theta}$ is given by

$$\overline{K}_{\underline{\Theta}} = \frac{1}{n} \sum_{i=1}^{n} \mathrm{COV}(\underline{\Theta}_i|W) \tag{32}$$

where $\mathrm{COV}(\cdot)$ is defined in (27). Similarly, the time average error covariance of the conditional mean estimator $E(X_i, Y_i|W)$ is given by

$$\overline{K}_{x,y} = \frac{1}{n} \sum_{i=1}^{n} \mathrm{COV}(X_i, Y_i|W). \tag{33}$$

Note that $\overline{K}_{x,y}$ is well defined regardless of whether or not the decoder (for $\underline{\Theta}$) actually estimates (X, Y).

For some given coding scheme, let ρ be the correlation coefficient associated with the covariance matrix $\overline{K}_{x,y}$ defined above. Given a certain class of encoders $\mathcal{C}$ (of possibly different dimensions), define

$$\rho(\underline{D}_\theta, \mathcal{C}) = \sup|\rho| \tag{34}$$

where the supremum is taken over all encoders in the class $\mathcal{C}$ which achieve distortion $\underline{D}_\theta$ (assuming there are such), and let

$$\rho_0(\mathcal{C}) = \limsup_{\underline{D}_\theta \to \underline{D}_\theta^{\min}} \rho(\underline{D}_\theta, \mathcal{C}). \tag{35}$$

Note that for unrestricted encoders, $\rho_0(\mathcal{C})$ may take any value between zero and one for some sources [23]. The following theorem provides an asymptotic outer bound for the region $\mathcal{R}(\underline{D}_\theta, \mathcal{C})$, of $\underline{D}_\theta$-admissible rate pairs of encoders from $\mathcal{C}$, in terms of the quantities defined above, as the distortion vector $\underline{D}_\theta$ approaches its minimal value:

Theorem 4 (Asymptotic Outer Bound—Remote Case): Assume the conditions of Theorem 3 hold. Let $\mathcal{C}$ be some arbitrary class of encoders. Then, the rate region of encoders from $\mathcal{C}$ satisfies

$$\mathcal{R}(\underline{D}_\theta, \mathcal{C}) \tilde{\subset} \mathcal{R}^{**}\left(\frac{\underline{D}_\theta - \underline{D}_\theta^{\min}}{1-\rho_0(\mathcal{C})}\right), \qquad \text{as } \underline{D}_\theta \to \underline{D}_\theta^{\min}. \quad (36)$$

Remarks:

1) The maximization in (34) which defines the asymptotic correlation coefficient $\rho_0(\mathcal{C})$ is taken over encoders of arbitrary dimension. Thus the outer bound in (36) is not of the desirable "single-letter" variety. Furthermore, the outer bound is useless if $\rho_0(\mathcal{C}) = 1$.
2) In spite of the above, a high correlation coefficient between the X- and Y-quantization errors is unlikely if these errors must vanish in the mean-squared sense. In fact, as shown in [23], under natural conditions on the structure of the encoders and on the "smoothness" of the joint distribution of (X, Y), if D_x and D_y go to zero then the correlation coefficient between the X- and Y-quantization errors must vanish, i.e.,
$$\rho_0(\mathcal{C}) = 0. \quad (37)$$
In particular, lattice (or tessellating) quantizers satisfy the asymptotic zero correlation property of (37); see [23].
3) If the rank of the $m \times 2$ matrix T is equal to two, then $\underline{D}_\theta \to \underline{D}_\theta^{\min}$ implies that $D_x, D_y \to 0$, i.e., high-resolution encoding of the remote source requires high-resolution encoding of the noisy sources.[3] In such a case, as discussed in Remark 2 above, a "well-behaved" class of coding schemes would have $\rho_0(\mathcal{C}) = 0$, and, therefore, $\mathcal{R}(\underline{D}_\theta, \mathcal{C}) \tilde{\subset} \mathcal{R}^{**}(\underline{D}_\theta - \underline{D}_\theta^{\min})$.
4) Some examples of sources and separate encoders for which $\rho_0(\mathcal{C}) \neq 0$ are given in [23], showing that this property comes only at the expense of the quantization efficiency. We conjecture that if $\text{rank}(T) = 2$, any "good" multiterminal coding scheme satisfies (37), whereupon Theorems 3 and 4 imply that the *un*restricted rate region satisfies
$$\mathcal{R}(\underline{D}_\theta) \cong \mathcal{R}^{**}(\underline{D}_\theta - \underline{D}_\theta^{\min}), \qquad \text{as } \underline{D}_\theta \to \underline{D}_\theta^{\min}. \quad (38)$$
5) Section V shows that the right-hand side of (38) can be realized by a "systematic" coding scheme, composed of lattice (or tessellating) quantizers followed by Slepian–Wolf-encoding.
6) If rank $T < 2$, then $\underline{D}_\theta \to \underline{D}_\theta^{\min}$ does not imply that $D_x, D_y \to 0$. In such a case, a coding scheme with $\rho \neq 0$ may exceed the performance of $\mathcal{R}^{**}(\underline{D}_\theta - \underline{D}_\theta^{\min})$. For example, suppose $m = 1$ and $E(\Theta|X, Y) = X + Y$, and assume that (X, Y) are uniformly distributed over a narrow off-diagonal strip. Letting each quantization cell be a union of two sufficiently separated small intervals results in high ambiguity in $X - Y$ but high accuracy in $X + Y$, i.e., small distortion between $\hat{\Theta}$ and Θ. This multiterminal quantizing scheme saves 1 bit (or in general, $\log$(number of intervals in a single cell)) relative to a scheme that has the same distortion but uses connected cells. Note that in the irregular scheme, D_x and D_y do not vanish as $D_\theta \to 0$, and the correlation coefficient ρ between the X and Y quantization errors goes to -1.
7) In Fig. 4 we show the asymptotic rate-distortion region $\mathcal{R}^{**}(D_\theta - D_\theta^{\min})$ for encoding a single remote source with distortion $D_\theta > D_\theta^{\min} = (1/\sigma_\theta^2 + 1/\sigma_1^2 + 1/\sigma_2^2)^{-1}$, where $X = \Theta + N_1$, $Y = \Theta + N_2$, $\Theta \sim \mathcal{N}(0, \sigma_\theta^2)$, and $N_j \sim \mathcal{N}(0, \sigma_j^2)$, and where Θ, N_1, and N_2 are independent. (The rate region is the area bounded by the convex hull of the curves which are drawn in the figure.) This is an example of a quadratic Gaussian CEO problem [22] with two nonidentical "agents."

[3] In the more general setting of L noisy sources this condition becomes rank $T = L$.

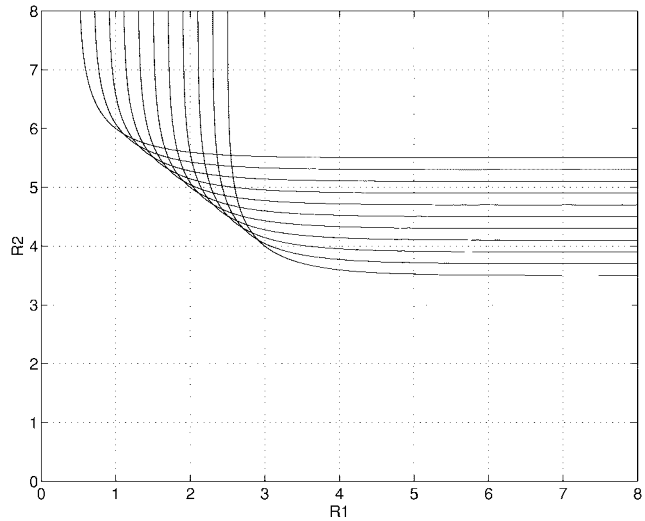

Fig. 4. The asymptotic rate region for two encoders and one remote source.

B. Derivation of Results

We now state two basic lemmas, relating the estimation functions and the second-order moments of the errors in $\underline{\Theta}$ and in (X, Y) at the decoder; recall the definitions of W, $\overline{K}_{\underline{\Theta}}$, and $\overline{K}_{x,y}$ above.

Lemma 1 below implies that we can view the optimal decoder for $\underline{\Theta}$ as an optimal decoder for (X, Y), followed by the linear transformation T, i.e., by the optimal estimate of $\underline{\Theta}$ from (X, Y) (see (26)). Lemma 2 below shows how the second-order moments of the errors in coding (X, Y) determine the achievable distortions in coding $\underline{\Theta}$. Lemma 2 also justifies the interpretation of the set $\mathcal{K}$ given after (28).

Lemma 1 (Separation Principle):

$$E(\underline{\Theta}_i|W) = T \cdot E(X_i, Y_i|W)^t \quad (39)$$

where T is defined in (26).

Proof: Since the encoder does not have direct access to $\underline{\Theta}$, the codewords W depend on $\underline{\Theta}_1 \cdots \underline{\Theta}_n$ only through $X_1 \cdots X_n, Y_1 \cdots Y_n$. Moreover, since $(\underline{\Theta}_i, X_i, Y_i)$ are drawn independently for $i = 1 \cdots n$

$$\underline{\Theta}_i \leftrightarrow (X_i, Y_i) \leftrightarrow W \text{ form a Markov chain.} \tag{40}$$

Iterating the expectation in (39), we now write

$$\begin{aligned} E(\underline{\Theta}_i|W) &= E\{E(\underline{\Theta}_i|W, X_i, Y_i)|W\} &(41)\\ &= E\{E(\underline{\Theta}_i|X_i, Y_i)|W\} &(42)\\ &= E\{T \cdot (X_i, Y_i)^t|W\} &(43)\end{aligned}$$

where (42) follows from the Markov chain (40), and (43) follows from (26). □

Lemma 2 (Pythagoras Theorem):

$$\overline{K}_{\underline{\Theta}} = \mathrm{COV}\,(\underline{\Theta}|X, Y) + T\overline{K}_{x,y}T^t. \tag{44}$$

The relatively simple proof of Lemma 2 is deferred to the Appendix. Note that Lemma 2 holds if the encoding is done either separately or jointly, i.e., for both cases of Fig. 3.

Proof of Theorem 3: We need to show that the rate region $\mathcal{R}^{**}(\underline{D}_\theta - \underline{D}_\theta^{\min})$ defined in (29) is asymptotically achievable in the limit as $n \to \infty$ and then $\underline{D}_\theta \to \underline{D}_\theta^{\min}$. We first construct the inner bound

$$\mathcal{R}(\underline{D}_\theta) \supseteq \bigcup_{\{D_x, D_y:\, \mathrm{diag}[D_x, D_y] \in \mathcal{K}(\underline{D}_\theta - \underline{D}_\theta^{\min}, T)\}} \mathcal{R}^{\mathrm{add}}(D_x, D_y) \tag{45}$$

that holds for *any* distortion vector $\underline{D}_\theta$, where $\mathcal{K}$ is defined in (28), and $\mathcal{R}^{\mathrm{add}}(D_x, D_y)$ is defined in (14).

Pick arbitrary D_x and D_y, and let $N_1 \sim \mathcal{N}(0, D_x)$ and $N_2 \sim \mathcal{N}(0, D_y)$ be independent of $(\underline{\Theta}, X, Y)$ and of one another. As discussed in the previous subsection, the "long" Markov chain $(X + N_1) \leftrightarrow X \leftrightarrow Y \leftrightarrow (Y + N_2)$ induces a direct coding system of two encoders and a common decoder that, for any $(R_1, R_2) \in \mathcal{R}^{\mathrm{add}}(D_x, D_y)$, achieves distortions approaching D_x and D_y in X and Y as the coding blocklength n goes to infinity. The separation principle of Lemma 1 then tells us how to estimate $\underline{\Theta}$ optimally from $\underline{\hat{X}}$ and $\underline{\hat{Y}}$. As we show below, the Pythagoras Theorem (Lemma 2) implies that if $\mathrm{diag}\,[D_x, D_y]$ belongs to the set $\mathcal{K}(\underline{D}_\theta - \underline{D}_\theta^{\min}, T)$, then this estimate achieves average distortion $\underline{D}_\theta$ or less, and (45) will follow.

We first need to assess the covariance matrix $\overline{K}_{x,y}$ associated with the X and Y quantization errors in the system above. Assume for a moment that both the encoding and the decoding are done via a *joint-distortion-typicality rule*, associated with the distribution of the "long" Markov chain $(X+N_1) \leftrightarrow X \leftrightarrow Y \leftrightarrow (Y+N_2)$ (see, e.g., [5, pp. 351–362]). The time-average covariance of the errors in coding X and Y would be close to that induced by the chain, i.e., to $\mathrm{COV}\,(N_1, N_2) = \mathrm{diag}\,[D_x, D_y]$. More precisely

$$\begin{aligned}\mathrm{COV}\,(N_1, N_2) - \epsilon I &\le \frac{1}{n}\sum_{i=1}^{n} \mathrm{COV}\,(\hat{X}_i - X_i, \hat{Y}_i - Y_i)\\ &\le \mathrm{COV}\,(N_1, N_2) + \epsilon I \end{aligned} \tag{46}$$

where $\hat{X}_i = \hat{X}_i(W)$ and $\hat{Y}_i = \hat{Y}_i(W)$ are the outputs at time i of the joint-distortion-typical decoder, W is the decoder input, I is the 2×2 identity matrix, and $\epsilon > 0$ can be made as small as desired by picking n large enough. Throughout the paper an inequality between matrices is in the sense that the difference matrix is nonnegative-definite.[4] We now replace the joint-distortion-typical decoder by an optimal, conditional expectation decoder, while keeping the same encoder. By doing so the error covariance can only decrease,[5] i.e., $\mathrm{COV}\,(X_i, Y_i|W) \le \mathrm{COV}\,(\hat{X}_i - X_i, \hat{Y}_i - Y_i)$. Recalling the definition of the average error covariance $\overline{K}_{x,y}$ from (33), and combining with the right inequality in (46), we thus obtain the upper bound (in a matrix sense)

$$\overline{K}_{x,y} \le \mathrm{COV}\,(N_1, N_2) + \epsilon I = \mathrm{diag}\,[D_x + \epsilon, D_y + \epsilon]. \tag{47}$$

We are now in a position to use Lemma 2, and substitute (47) into (44) to obtain the upper bound

$$\overline{K}_{\underline{\Theta}} \le \mathrm{COV}\,(\underline{\Theta}|X, Y) + T \cdot \mathrm{diag}\,[D_x + \epsilon, D_y + \epsilon] \cdot T^t \tag{48}$$

on the average error in $\underline{\Theta}$ of the coding system described above. (Note that the matrix inequality is preserved when (44) and (47) are combined.) Since $\mathrm{diag}\,[D_x, D_y]$ belongs to $\mathcal{K}(\underline{D}_\theta - \underline{D}_\theta^{\min}, T)$, (48) implies that $\overline{K}_{\underline{\Theta}_{j,j}} \le D_{\theta j} + \epsilon (TT^t)_{j,j}$, for $j = 1 \cdots m$. Thus the distortion constraints $D_{\theta 1} \cdots D_{\theta m}$ are approached as closely as desired by taking n large enough, and (45) is proved.

To complete the proof of Theorem 3, we will show that as $\underline{D}_\theta \to \underline{D}_\theta^{\min}$, the inner bound in (45) approaches $\mathcal{R}^{**}(\underline{D}_\theta - \underline{D}_\theta^{\min})$. It is easy to verify from (28) that since T has no all-zero columns, if a pair (D_x, D_y) belongs to the union in (45) (i.e., $\mathrm{diag}\,[D_x, D_y] \in \mathcal{K}(\underline{D}_\theta - \underline{D}_\theta^{\min}, T)$), then

$$\max\{D_x, D_y\} \le k \cdot \max_j (D_{\theta j} - D_{\theta j}^{\min})$$

for some finite k. Thus all pairs (D_x, D_y) in the union (45) go to zero *uniformly* as $\underline{D}_\theta \to \underline{D}_\theta^{\min}$. By (22) this implies that each term in the union approaches $\mathcal{R}^*(D_x, D_y)$ uniformly. Thus also the entire union approaches the corresponding union over $\mathcal{R}^*(D_x, D_y)$, which is $\mathcal{R}^{**}(\underline{D}_\theta - \underline{D}_\theta^{\min})$. □

Proof of Theorem 4: We construct below the following outer bound:

$$\mathcal{R}(\underline{D}_\theta, \mathcal{C}) \subseteq \mathcal{R}^{**}\left(\frac{\underline{D}_\theta - \underline{D}_\theta^{\min}}{1 - \rho_0(\mathcal{C}) - \epsilon(\underline{D}_\theta)}\right) \tag{49}$$

where $\epsilon(\underline{D}_\theta) \to 0$ as $\underline{D}_\theta \to \underline{D}_\theta^{\min}$. Since $\mathcal{R}^{**}(t\underline{\Delta})$ is continuous in t for all $t > 0$,[6] the right-hand side of (49) is asymptotically equal ($\cong$) to $\mathcal{R}^{**}((\underline{D}_\theta - \underline{D}_\theta^{\min})/1 - \rho_0(\mathcal{C}))$, which proves the desired asymptotic outer bound in (36).

[4] This definition induces a partial ordering between positive-definite matrices. Although not every pair of positive-definite matrices can be compared, it is sometimes possible to compare a nondiagonal matrix with a diagonal matrix, as done in (46).

[5] The distortion-typical decoding rule induced by the chain $(X + N_1) \leftrightarrow X \leftrightarrow Y \leftrightarrow (Y + N_2)$ is suboptimal since $X + N_1$ is in general not the conditional expectation of X given $X + N_1$.

[6] We have

$$\begin{aligned}\mathcal{R}^{**}(t\underline{\Delta}) &= \{(R_1, R_2) : (R_1 + \log(t), R_2 + \log(t)) \in \mathcal{R}^{**}(\underline{\Delta})\}\\ &= \mathcal{R}^{**(+\log(t))}(\underline{\Delta}).\end{aligned}$$

Let $\epsilon > 0$ be arbitrary. By definition (35), it is possible to find $\delta > 0$ small enough such that $\rho(\underline{D}_\theta, \mathcal{C}) \le \rho_0(\mathcal{C}) + \epsilon$ for all $\underline{D}_\theta \le \underline{D}_\theta^{\min} + \delta \cdot \underline{1}$, where $\underline{1}$ is a vector of ones. Now, assume an arbitrary coding scheme in $\mathcal{C}$ with $\underline{D}_\theta \le \underline{D}_\theta^{\min} + \delta \cdot \underline{1}$. Let

$$\overline{K}_{x,y} = \begin{pmatrix} \alpha & \rho\sqrt{\alpha\beta} \\ \rho\sqrt{\alpha\beta} & \beta \end{pmatrix} \tag{50}$$

be the X- and Y-error covariance (33) associated with this coding scheme (i.e., $D_x = \alpha$, $D_y = \beta$ and the correlation coefficient is ρ). By the above $|\rho| \le \rho_0(\mathcal{C}) + \epsilon$. This implies that the matrix

$$\begin{aligned} &\overline{K}_{x,y} - (1 - \rho_0(\mathcal{C}) - \epsilon) \cdot \mathrm{diag}[\alpha, \beta] \\ &\qquad = \begin{pmatrix} (\rho_0(\mathcal{C}) + \epsilon)\alpha & \rho\sqrt{\alpha\beta} \\ \rho\sqrt{\alpha\beta} & (\rho_0(\mathcal{C}) + \epsilon)\beta \end{pmatrix} \end{aligned} \tag{51}$$

is nonnegative-definite, since it is symmetric with positive diagonal elements and positive determinant. Thus

$$\overline{K}_{x,y} \ge (1 - \rho_0(\mathcal{C}) - \epsilon) \cdot \mathrm{diag}\,[\alpha, \beta].$$

Combining with Lemma 2, we obtain

$$\begin{aligned} &\overline{K}_{\underline{\Theta}} - \mathrm{COV}\,(\underline{\Theta}|X, Y) \\ &\quad = T\overline{K}_{x,y}T^t \ge T \cdot (1 - \rho_0(\mathcal{C}) - \epsilon) \cdot \mathrm{diag}\,[\alpha, \beta] \cdot T^t. \end{aligned} \tag{52}$$

(Note that the matrix inequality is preserved when (44) and (51) are combined.) Since the diagonal elements of a nonnegative-definite matrix are nonnegative, we deduce from (52) that

$$D_{\theta j} - D_{\theta j}^{\min} \ge (T \cdot (1 - \rho_0(\mathcal{C}) - \epsilon) \cdot \mathrm{diag}\,[\alpha, \beta] \cdot T^t)_{j,j}, \qquad \text{for } j = 1 \cdots m \tag{53}$$

which by the definition of the set $\mathcal{K}$ in (28) implies

$$\mathrm{diag}\,[\alpha, \beta] \in \mathcal{K}\left(\frac{\underline{D}_\theta - \underline{D}_\theta^{\min}}{1 - \rho_0(\mathcal{C}) - \epsilon}, T \right). \tag{54}$$

On the other hand, the Shannon outer bound (Theorem 1) and (50) imply that the rates (R_1, R_2) of this coding scheme belong to $\mathcal{R}^*(\alpha, \beta)$. Thus the definition of $\mathcal{R}^{**}(\cdot)$ and (54) imply

$$(R_1, R_2) \in \mathcal{R}^{**}\left(\frac{\underline{D}_\theta - \underline{D}_\theta^{\min}}{1 - \rho_0(\mathcal{C}) - \epsilon} \right). \tag{55}$$

Since, in the limit as $\underline{D}_\theta \to \underline{D}_\theta^{\min}$, ϵ can be chosen arbitrarily small, this proves (49) and, hence, Theorem 4. □

IV. The Minimum Rate Sum at High Resolution

To enhance insight into the results of Sections II and III, and to compare them to related problems, it is convenient to consider the minimum possible sum of rates over all admissible rate pairs. In light of Theorems 2–4, we restrict our attention to the asymptotic rate regions $\mathcal{R}^*(\cdot)$ for the direct case and $\mathcal{R}^{**}(\cdot)$ for the remote case.

We start with the direct coding case. As in the case of the Slepian–Wolf region (1), the minimum rate sum of the region $\mathcal{R}^*$ is limited only by the last line in (4). Thus in direct coding it is given asymptotically (as $D_x, D_y \to 0$) by

$$\min_{(R_1, R_2) \in \mathcal{R}^*(D_x, D_y)} R_1 + R_2 = h(X, Y) - \tfrac{1}{2} \log(2\pi e)^2 D_x D_y \tag{56}$$

provided that $h(X, Y)$ exists and is finite, and $EX^2, EY^2 < \infty$.

In the remote-coding case, since the region $\mathcal{R}^{**}$ is expressed as a union of regions of the form of $\mathcal{R}^*$, its minimum possible rate sum is given by the minimum of (56) over the admissible values of D_x and D_y. Thus from (29) and (36)

$$\begin{aligned} &\min_{(R_1, R_2) \in \mathcal{R}^{**}(\underline{D}_\theta - \underline{D}_\theta^{\min})} R_1 + R_2 \\ &\qquad = h(X, Y) - \frac{1}{2} \log(2\pi e)^2 \det{}^{**} \end{aligned} \tag{57}$$

where

$$\det{}^{**} = \max D_x D_y = \max \left\{ \det \mathrm{diag}\,[D_x, D_y] \right\} \tag{58}$$

where the maximization is taken over all (D_x, D_y) such that $\mathrm{diag}\,[D_x, D_y] \in \mathcal{K}(\underline{D}_\theta - \underline{D}_\theta^{\min}, T)$, and "det" denotes determinant. Note that if T satisfies the conditions in Theorem 4, $\det^{**}$ is finite for all $\underline{D}_\theta > \underline{D}_\theta^{\min}$.

A. The Loss Relative to Joint Encoding

Suppose X and Y are encoded jointly, and denote the joint coding rate R_0; see Case A of Fig. 1 and Fig. 3 for the direct and the remote case, respectively. As shown in [28], for the quadratic case the minimum rate sum in direct multiterminal source coding is at most 1 bit above R_0. In view of our results above, it is interesting to assess this rate loss in the extreme of high resolution for both the direct- and the remote-coding cases.

By the Shannon lower bound for jointly encoding (X, Y) with distortions (D_x, D_y) (see (9)), R_0 is bounded from below by the right-hand side of (56). Thus in the direct-coding case the rate loss for separating the encoders is asymptotically zero!

As for the remote-coding case, note that Lemma 2 holds also for joint encoding, implying that a joint encoder must satisfy

$$\overline{K}_{x,y} \in \mathcal{K}(\underline{D}_\theta - \underline{D}_\theta^{\min}, T). \tag{59}$$

Applying the Shannon lower bound for the joint rate distortion function of (X, Y) under the distortion constraint in (59), we obtain

$$R_0 \ge h(X, Y) - \frac{1}{2} \log 2\pi e \det{}^{**}_{\text{joint}} \tag{60}$$

where

$$\det{}^{**}_{\text{joint}} = \max\{\det K\} \tag{61}$$

where the maximization is taken over all $K \in \mathcal{K}(\underline{D}_\theta - \underline{D}_\theta^{\min}, T)$.[7] In fact, if $\det^{**}_{\text{joint}} < \infty$, then the joint rate-distortion function of the Gaussian pair (X, Y), relative to constraint (59), is *equal* to the lower bound in (60) for

[7] The constraint of (59) has the flavor of the *error spectrum* constraint of [14]. Specifically, for T square and orthonormal [14] implies that (61) is maximized by $K = T^t \mathrm{diag}\,[\Delta_1, \Delta_2] T$.

$\underline{D}_\theta - \underline{D}_\theta^{\min}$ below a certain critical value; see, e.g., [1, pp. 133, 250]. Hence, for $\underline{D}_\theta - \underline{D}_\theta^{\min}$ small the lower bound (60) becomes tight. Comparing with (57), we conclude that the rate sum loss for separating the encoders in the remote case satisfies

$$\min_{(R_1, R_2) \in \mathcal{R}(\underline{D}_\theta, \mathcal{C})} R_1 + R_2 - R_0 = \frac{1}{2} \log \left(\frac{\det^{**}_{\text{joint}}}{\det^{**}} \right) + o(1) \tag{62}$$

where $o(1) \to 0$ as $\underline{D}_\theta \to \underline{D}_\theta^{\min}$, provided that (37) (i.e., $\rho_0(\mathcal{C}) = 0$) holds.

This rate loss comes from the fact that a zero correlation coefficient between the X, Y-quantization errors, as condition (37) implies, is not optimal for estimating $\underline{\Theta}$. For example, suppose a bivariate remote source (Θ_1, Θ_2), and measurements $X = \Theta_1 + \Theta_2 + N_1$ and $Y = \Theta_1 - \Theta_2 + N_2$, where $\Theta_1, \Theta_2, N_1, N_2$ are independent zero-mean Gaussian variables, $\text{Var}(\Theta_1) = \text{Var}(\Theta_2) = \sigma_\theta^2$ and $\text{Var}(N_1) = \text{Var}(N_2) = \sigma_N^2$. For convenience we will assume that $\sigma_N^2 \ll \sigma_\theta^2$, implying that the estimation matrix and the minimum distortions are approximately

$$T \approx \frac{1}{2} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$$

and $D_{\theta_1}^{\min} = D_{\theta_2}^{\min} \approx \sigma_N^2/2$, respectively. Suppose now that the remote sources need to be reconstructed with highly nonsymmetric distortions $D_{\theta_1} = D_{\theta_1}^{\min} + \delta$ and $D_{\theta_2} = D_{\theta_2}^{\min} + 99\delta$ for some $\delta > 0$. Then, (58) and (61) are maximized by (see preceding footnote)

$$K = 2\delta \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \quad \text{and} \quad K = 100\delta \begin{pmatrix} 1 & -0.98 \\ -0.98 & 1 \end{pmatrix} \tag{63}$$

respectively, so $\det^{**} = 4\delta^2$ while $\det^{**}_{\text{joint}} = 400\delta^2$. By (62), the rate sum loss of the multiterminal coding scheme is $\frac{1}{2}\log(100) \approx 3.3$ bit per sample in the limit as $\delta \to 0$. To gain these ≈ 3.3 bits, the joint encoder quantizes X and Y with highly negatively correlated errors (the correlation coefficient should be -0.98). Intuitively, such correlated errors may be achieved using quantization cells which are narrow in the 45° direction and wide in the $-45°$ direction in the X, Y-plain; see Fig. 5. This structure gives higher resolution in coding $X + Y$ than in coding $X - Y$, which in return allows the better estimate of Θ_1 relative to Θ_2. In contrast to this, when the encoders are separated, i.e., independent, the quantization cells are *rectangular* in the X, Y-plain. Thus to achieve small distortion in estimating Θ_1 they must be narrow in both directions, implying a necessary excess rate.

When $\det^{**}_{\text{joint}} = \infty$, the rate loss in (62) equals infinity. To better understand this situation, consider the more general case where there are L (rather than two) noisy measurements of $\underline{\Theta} = \Theta_1, \cdots, \Theta_m$. In this case, the matrices T and K are $m \times L$ and $L \times L$, respectively. Let $m' = \text{rank}\, T$. The case $\det^{**}_{\text{joint}} = \infty$ happens when $m' < L$, and in particular when $m < L$, in which case it is enough for the joint encoder to encode only m' linear combinations of the L noisy measurements in order to satisfy the m distortion constraints on $\underline{\Theta}$. This fact implies that the set $\mathcal{K}$ is unbounded, and that $\det^{**}_{\text{joint}} = \infty$. As opposed to that, the separate encoders must

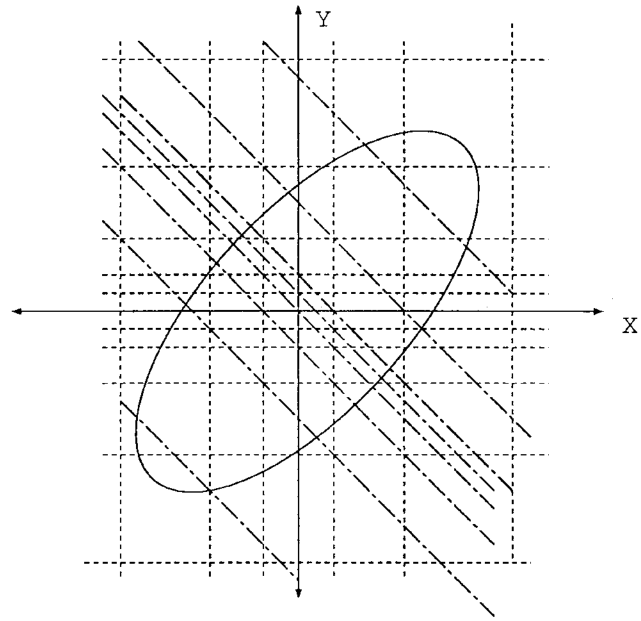

Fig. 5. Remote encoding of (Θ_1, Θ_2) via the noisy sources $X = \Theta_1 + \Theta_2 + N_1$ and $Y = \Theta_1 - \Theta_2 + N_2$. Doted lines show the effective two-dimensional cells of separate quantizers; "dash–dot" lines show the cells of an efficient joint quantizer.

encode *all* the L noisy measurements. Thus we will typically have for $\underline{D}_\theta - \underline{D}_\theta^{\min}$ small

$$\frac{R_1 + \cdots + R_L}{R_0} \approx \frac{L}{m'} \tag{64}$$

implying that $R_1 + \cdots + R_L - R_0 \to \infty$ as $\underline{D}_\theta \to \underline{D}_\theta^{\min}$. See a specific example below (the CEO problem).

B. Comparison with the Quadratic Gaussian CEO Problem

The following multiterminal remote source-coding problem was considered in [22]. There is a single remote source Θ and L noisy measurements $\underline{X} = X_1 \cdots X_L$, $X_i = \Theta + N_i$, where $\Theta, N_1, \cdots, N_L$ are independent zero-mean Gaussian variables, the variance of Θ is σ_θ^2, and the common variance of $N_1 \cdots N_L$ is σ_N^2. The estimation matrix and the minimum distortion are given for large L by $T \approx 1/L(1, \cdots, 1)$ and $D_\theta^{\min} \approx \sigma_N^2/L$. The rate of a *joint* encoder which observes $X_1 \cdots X_L$ is given by the Wolf–Ziv formula [25], which for large L becomes $R_0 \approx h(\Theta) - \frac{1}{2} \log 2\pi e(D_\theta - \sigma_N^2/L)$, so

$$D_\theta \approx \frac{\sigma_N^2}{L} + \sigma_\theta^2 2^{-2R_0}. \tag{65}$$

For *separated* encoders and large $R_{\text{sum}} \triangleq R_1 + \cdots + R_L$, we have on the one hand, [15], [22],

$$\lim_{L \to \infty} D_\theta \approx \frac{\sigma_N^2}{2R_{\text{sum}}}. \tag{66}$$

(This result, shown in [22] to be an upper bound and conjectured there to be tight, recently has been established by Oohama [15].) On the other hand, for a fixed L and in the limit as $R_{\text{sum}} \to \infty$, we have from (57) (generalized to arbitrary L)

$R_{\rm sum} = h(\underline{X}) - (L/2)\log 2\pi e L(D_\theta - D_\theta^{\min})$, thus[8]

$$D_\theta \approx \frac{\sigma_N^2}{L} + \frac{\sigma_N^2}{L} 2^{-(2/L)R_{\rm sum}}. \tag{67}$$

First note from (65) and (67) that indeed, as discussed in the previous subsection, for fixed L and $D_\theta - D_\theta^{\min} \to 0$ the rate sum in separate encoding needs to be asymptotically (almost) L times larger than the rate in joint encoding to achieve the same distortion D_θ. Second, observe that while the high-resolution encoding distortion in (67) decays exponentially with $R_{\rm sum}$, the distortion in (66) decays only linearly with $R_{\rm sum}$. Formally, this difference follows since the high-resolution analysis assumes $R_{\rm sum} \to \infty$ for a fixed L, i.e., high rate per *each* terminal, while the analysis in [22] assumes $L \to \infty$ for a fixed $R_{\rm sum}$, i.e., vanishing average rate per terminal.[9]

V. Multiterminal Dithered Lattice Quantization

Theorem 2 and (22) suggest that understanding the rate-region $\mathcal{R}^{\rm add}(D_x, D_y)$, associated with additive noise test channels, provides insight into how to achieve optimal high-resolution performance in the direct multiterminal coding case. By the separation principle of Lemma 1 and by the Pythagoras Theorem (Lemma 2), this understanding also provides the key to asymptotic optimality in the remote coding case. In this section we show that entropy coded (lattice) dithered quantizers (ECDQ) [29] can realize $\mathcal{R}^{\rm add}(D_x, D_y)$ in the limit of large lattice dimension. A similar result was shown in [21] for tessellating quantizers. We briefly review ECDQ definitions and results and refer the interested reader to [29]–[31] for more details.

Let Q_1 and Q_2 be K-dimensional unbounded lattice quantizers, and let $\underline{Z}_1$ and $\underline{Z}_2$ (the dither vectors) be independent random vectors, uniformly distributed over the basic cell of Q_1 and Q_2, respectively. Assume that the lattice quantizers are scaled so that $\frac{1}{K}E\|\underline{Z}_1^2\| = D_x$ and $\frac{1}{K}E\|\underline{Z}_2^2\| = D_y$. Following [29], we have for any source pair $\underline{X}, \underline{Y} \in \mathcal{R}^K$

$$\frac{1}{K}E\|Q_1(\underline{X}+\underline{Z}_1) - \underline{Z}_1 - \underline{X}\|^2 = \frac{1}{K}E\|Z_1\|^2 = D_x$$
$$\frac{1}{K}E\|Q_2(\underline{Y}+\underline{Z}_2) - \underline{Z}_2 - \underline{Y}\|^2 = \frac{1}{K}E\|Z_2\|^2 = D_y \tag{68}$$

where the expectation is taken with respect to $(\underline{X}, \underline{Y}, \underline{Z}_1, \underline{Z}_2)$. Also

$$H(Q_1(\underline{X}+\underline{Z}_1)|\underline{Z}_1) = I(\underline{X};\, \underline{X} - \underline{Z}_1)$$

and

$$H(Q_2(\underline{Y}+\underline{Z}_2)|\underline{Z}_2) = I(\underline{Y};\, \underline{Y} - \underline{Z}_2). \tag{69}$$

By extension, it is easy to show also that[10]

$$\begin{aligned} &H(Q_1(\underline{X}+\underline{Z}_1)|\underline{Z}_1,\, Q_2(\underline{Y}+\underline{Z}_2),\, \underline{Z}_2) \\ &\quad = I(\underline{X};\, \underline{X}-\underline{Z}_1|\underline{Y}-\underline{Z}_2) \\ &H(Q_2(\underline{Y}+\underline{Z}_2)|\underline{Z}_2,\, Q_1(\underline{X}+\underline{Z}_1),\, \underline{Z}_1) \\ &\quad = I(\underline{Y};\, \underline{Y}-\underline{Z}_2|\underline{X}-\underline{Z}_1) \\ &H(Q_1(\underline{X}+\underline{Z}_1),\, Q_2(\underline{Y}+\underline{Z}_2)|\underline{Z}_1,\, \underline{Z}_2) \\ &\quad = I(\underline{XY};\, \underline{X}-\underline{Z}_1,\, \underline{Y}-\underline{Z}_2). \end{aligned} \tag{70}$$

Note the resemblance between the quantities on the right-hand side of (70) and the characterization of $\mathcal{R}^{\rm add}$ in (14).

The proposed multiterminal-ECDQ works as follows. Assume that the (pseudo-) random dither vectors $\underline{Z}_1$ and $\underline{Z}_2$ are available to both the encoders and the decoder, and are drawn independently at each quantization stage. After adding the dither vectors to the sources, Q_1 and Q_2 quantize $\underline{X}+\underline{Z}_1$ and $\underline{Y}+\underline{Z}_2$, respectively. Then, a block of m such successive outputs of the quantizers is Slepian–Wolf-encoded, conditioned on the dither vectors.[11] This is possible provided that m is large enough, and the coding rates (R_1, R_2) satisfy (1) relative to the joint and the conditional entropies of Q_1 and Q_2 in (70). At the decoder, the Slepian–Wolf-code is decoded (correctly with high probability), and the dither is subtracted to obtain the reconstructions $\hat{\underline{X}} = Q_1(\underline{X}+\underline{Z}_1) - \underline{Z}_1$ and $\hat{\underline{Y}} = Q_2(\underline{Y}+\underline{Z}_2) - \underline{Z}_2$. Equation (68) shows that $\hat{\underline{X}}$ and $\hat{\underline{Y}}$ satisfy the desired distortions, while the entropies in (70) characterize the region of admissible rates.

However, since $\underline{Z}_1$ and $\underline{Z}_2$ are not Gaussian, this rate region does not coincide with $\mathcal{R}^{add}(D_x, D_y)$. Nevertheless, it follows from [13] and [31] that

$$\begin{aligned} &\frac{1}{K} I(\underline{X}, \underline{Y};\, \underline{X}-\underline{Z}_1, \underline{Y}-\underline{Z}_2) \\ &\quad = I(X, Y;\, X+N_1, Y+N_2) + \log(2\pi e G_K) + o(1) \end{aligned} \tag{71}$$

for any (X, Y) satisfying the conditions of Theorem 2, where N_1 and N_2 are the Gaussian variables associated with $\mathcal{R}^{\rm add}(D_x, D_y)$, G_K is the normalized second moment of Q_1 and Q_2 (assuming both have the same structure), and $o(1) \to 0$ as $D_x, D_y \to 0$. Thus at high resolution, multiterminal-ECDQ has a combined rate redundancy of $\log(2\pi e G_K)$ bits per sample over $\mathcal{R}^{\rm add}$. Since for optimal lattice quantizers, [30], $\log(2\pi e G_K) \to 0$ as $K \to \infty$, we conclude that multiterminal-ECDQ approaches optimal high-resolution performance for large lattice dimension.

The performance of *low*-resolution ECDQ can be enhanced by incorporating linear *pre-* and *post-filtering*, corresponding in the memoryless case above to pre- and post-2×2 matrices operating on X and Y [31]. In the multiterminal case, however, the pre-matrix must be diagonal due to the separation

[8] Here $\det^{**}$ is achieved by $D_{x_1} = \cdots = D_{x_L} = L(D_\theta - D_\theta^{\min})$. Note that $(1/L)h(\underline{X}) \to h(N)$ as $L \to \infty$.

[9] It is tempting to suggest that in [22] the total rate is effectively allocated to a *finite* number $L^{\rm eff} = 2R_{\rm sum}$ of "high-rate terminals," so (66) roughly equals $\approx \sigma_N^2/L_{\rm eff}$, like the dominant term in (65) and (67).

[10] The expression for the joint entropy follows straightforwardly, while the expressions for the conditional entropies follow by subtracting the entropies in (69) from the joint entropy, and using the independence of $(\underline{Z}_1, \underline{Z}_2)$ and $(\underline{X}, \underline{Y})$.

[11] Since Slepian–Wolf encoding works for a finite-source alphabet, we assume here that the sources have bounded supports. Otherwise, the sources should be first limited to a bounded set, such that the probability and the second moment outside this set are negligible with respect to 1 and D_x, D_y, respectively.

of the terminals, leading to some degradation in performance. Yet, for jointly Gaussian sources, such a scheme with optimal filters will achieve, in the limit as $K \to \infty$, the inner bound $\mathcal{R}^*_{\text{in}}(D_x, D_y)$ of [2] for *any* D_x and D_y.

APPENDIX

A. Speed of Convergence to $\mathcal{R}^*(D_x, D_y)$ (Section II, Remark 2)

In the case of squared-error distortion measure (N_1 and N_2 in (14) are Gaussian with variances D_x and D_y, respectively), we have from the vector version of De-Bruijn's identity [5], [6]

$$\begin{aligned} h(X+N_1, Y+N_2) = h(X, Y) &+ \frac{J_x(X, Y)}{2} D_x + \frac{J_y(X, Y)}{2} D_y \\ &+ O(D_x^2) + O(D_y^2) \end{aligned} \tag{72}$$

provided that $h(X+\sqrt{t}Z_1, Y+\sqrt{\tau}Z_2)$ is twice differentiable with respect to (t, τ) at $(t, \tau) = (0, 0)$. Here Z_1 and Z_2 are independent Gaussians with unit variance, and

$$J_x(X, Y) \triangleq \int \left(\frac{\partial f(x, y)}{\partial x}\right)^2 \frac{dx\,dy}{f(x, y)}$$

and

$$J_y(X, Y) \triangleq \int \left(\frac{\partial f(x, y)}{\partial y}\right)^2 \frac{dx\,dy}{f(x, y)}$$

are the marginal Fisher informations of $f(x, y)$, the joint density of the random variables (X, Y). Assuming the differentiability condition above holds, we can use (72) to express the "$o(1)$" terms in (17)–(19). Then, choosing $\epsilon = O(D_x) + O(D_y)$ in (20) and (21) shows that the rate region $\mathcal{R}(D_x, D_y)$ converges to $\mathcal{R}^*(D_x, D_y)$ as $O(D_x) + O(D_y)$.

B. Proof of Lemma 2

We expand $\overline{K}_{\underline{\Theta}}$ in (32) as follows:

$$\begin{aligned} \overline{K}_{\underline{\Theta}} &= \frac{1}{n} \sum_{i=1}^{n} \text{COV}\{\underline{\Theta}_i - E(\underline{\Theta}_i|W)\} \\ &= \frac{1}{n} \sum_{i=1}^{n} \text{COV}\{\underline{\Theta}_i - E(\underline{\Theta}_i|X_i, Y_i) \\ &\quad + E(\underline{\Theta}_i|X_i, Y_i) - E(\underline{\Theta}_i|W)\} \\ &= \frac{1}{n} \sum_{i=1}^{n} \text{COV}\{\underline{\Theta}_i - E(\underline{\Theta}_i|X_i, Y_i)\} \\ &\quad + \text{COV}\{E(\underline{\Theta}_i|X_i, Y_i) - E(\underline{\Theta}_i|W)\} \end{aligned} \tag{73}$$

where in the last step we dropped the cross terms (the crosscorrelation matrices)

$$\begin{aligned} \text{cross}_i &= E\{[\underline{\Theta}_i - E(\underline{\Theta}_i|X_i, Y_i)] \\ &\quad \cdot [E(\underline{\Theta}_i|X_i, Y_i) - E(\underline{\Theta}_i|W)]^t\} \\ &= E\{[\underline{\Theta}_i - T(X_i, Y_i)^t] \cdot [(X_i, Y_i)T^t - E(\underline{\Theta}_i|W)]^t\}, \\ &\qquad i = 1 \cdots n \end{aligned} \tag{74}$$

which are equal to zero. To see that $\text{cross}_i = 0$, we first iterate the expectation in (74) (with inner expectation conditioned on (X_i, Y_i)), to obtain

$$\begin{aligned} \text{cross}_i = E\{E\{[\underline{\Theta}_i - T(X_i, Y_i)^t] \cdot [(X_i, Y_i)T^t \\ - E(\underline{\Theta}_i|W)^t]|X_i, Y_i\}\}. \end{aligned} \tag{75}$$

Next we observe that by (40) $\underline{\Theta}_i$ and W are conditionally independent given (X_i, Y_i), thus the inner expectation in (75) factors, i.e.,

$$\begin{aligned} \text{cross}_i = E\{E\{\underline{\Theta}_i - T(X_i, Y_i)^t|X_i, Y_i\} \\ \cdot E\{(X_i, Y_i)T^t - E(\underline{\Theta}_i|W)^t|X_i, Y_i\}\}. \end{aligned} \tag{76}$$

Finally, by (26), the first multiplicand in the inner expectation in (76) is zero, thus $\text{cross}_i = 0$ as desired. We now proceed from (73). Since $(\underline{\Theta}_i, X_i, Y_i)$ are drawn independent and identically distributed (i.i.d.), the first term in the sum on the right-hand side of (73) is independent of i and is equal to $\text{COV}(\underline{\Theta}|X, Y)$. Using (26) and (39) we write the second term in the sum in the right-hand side of (73) as

$$\begin{aligned} &\text{COV}\{T \cdot (X_i, Y_i)^t - T \cdot E\{X_i, Y_i|W\}^t\} \\ &= T \cdot \text{COV}(X_i, Y_i|W) \cdot T^t. \end{aligned}$$

Thus by the definition of $\overline{K}_{x,y}$ in (33) we obtain

$$\overline{K}_{\underline{\Theta}} = \text{COV}(\underline{\Theta}|X, Y) + T\overline{K}_{x,y}T^t$$

as desired. □

ACKNOWLEDGMENT

H. Viswanathan showed that multiterminal quantization errors may be correlated even at high resolution, and derived sufficient conditions for the errors to be asymptotically uncorrelated. We gratefully acknowledge many helpful discussions with him and with R. Yeung.

REFERENCES

[1] T. Berger, *Rate Distortion Theory: A Mathematical Basis for Data Compression*. Englewood Cliffs, NJ: Prentice-Hall, 1971.
[2] ——, "Multiterminal source coding," in *The Information Theory Approach to Communications*, G. Longo, Ed. New York: Springer-Verlag, 1977.
[3] T. Berger and S. Y. Tung, "Encoding of correlated analog sources," in *Proc. IEEE–USSR Joint Workshop on Information Theory*, 1975, pp. 7–10.
[4] T. Berger and R. Yeung, "Multiterminal source encoding with one distortion criterion," *IEEE Trans. Inform. Theory*, vol. 35, pp. 228–236, Mar. 1989.
[5] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York: Wiley, 1991.
[6] A. Dembo, "Information inequalities and uncertainty principles," Tech. Rep. 75, Dept. Statist., Stanford Univ., July 1990.
[7] R. L. Dobrushin and B. S. Tsybakov, "Information transmission with additional noise," *IRE Trans. Inform. Theory*, vol. IT-8, pp. 293–304, 1962.
[8] T. Fine, "Optimum mean square quantization of a noisy input," *IEEE Trans. Inform. Theory*, vol. IT-11, pp. 293–294, 1965.
[9] R. M. Gray, "A new class of lower bounds to information rates of stationary sources via conditional rate-distortion functions," *IEEE Trans. Inform. Theory*, vol. IT-19, pp. 480–489, July 1973.
[10] E. A. Haroutunian, "Multiterminal source coding achievable rates and reliability," *IEEE Trans. Inform. Theory*, vol. 42, pp. 2094–2101, Nov. 1996.

[11] A. Kaspi and T. Berger, "Rate-distortion for correlated sources with partially separated encoders," *IEEE Trans. Inform. Theory*, vol. IT-28, pp. 828–840, Nov. 1982.
[12] T. Linder and R. Zamir, "On the asymptotic tightness of the Shannon lower bound," *IEEE Trans. Inform. Theory*, vol. 40, pp. 2026–2031, Nov. 1994.
[13] T. Linder and K. Zeger, "Asymptotic entropy constrained performance of tessellating and universal randomized lattice quantization," *IEEE Trans. Inform. Theory*, vol. 40, pp. 575–579, Mar. 1994.
[14] R. A. McDonald and P. M. Schultheiss, "Information rates of Gaussian signals under criteria constraining the error spectrum," *Proc. IEEE*, vol. 52, pp. 415–416, 1964.
[15] Y. Oohama, "The rate-distortion function for the quadratic Gaussian CEO problem," in *School of Information Science and Electrical Engineering*, Kyushu University, Fukuoka, Japan, Jan. 1997, preprint.
[16] ——, "Gaussian multiterminal source coding," *IEEE Trans. Inform. Theory*, vol. 43, pp. 1912–1923, Nov. 1997.
[17] D. J. Sakrison, "Source encoding in the presence of random disturbance," *IEEE Trans. Inform. Theory*, vol. IT-14, pp. 165–166, Jan. 1968.
[18] D. Slepian and J. K. Wolf, "Noiseless coding of correlated information sources," *IEEE Trans. Inform. Theory*, vol. IT-19, pp. 471–480, July 1973.
[19] H. L. Van Trees, *Detection Estimation and Modulation Theory*. New York: Wiley, 1968.
[20] J. N. Tsitsiklis, "Decentralized detection by a large number of sensors," *Math. Contr. Signals, and Syst.*, vol. 1, pp. 167–182, 1988.
[21] H. Viswanathan, "Entropy coded tessellating quantization of correlated sources is asymptotically optimal," preprint.
[22] H. Viswanathan and T. Berger, "The quadratic Gaussian CEO problem," *IEEE Trans. Inform. Theory*, vol. 43, pp. 1549–1559, Sept. 1997.
[23] H. Viswanathan and R. Zamir, "On the asymptotic whiteness of high resolution quantization errors," in preparation.
[24] H. S. Witsenhausen, "Indirect rate distortion problems," *IEEE Trans. Inform. Theory*, vol. IT-26, pp. 518–521, Sept. 1980.
[25] J. K. Wolf and J. Ziv, "Transmission of noisy information to a noisy receiver with minimum distortion," *IEEE Trans. Inform. Theory*, vol. IT-16, pp. 406–411, July 1970.
[26] 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.
[27] A. D. Wyner and J. Ziv, "The rate-distortion function for source coding with side information at the decoder," *IEEE Trans. Inform. Theory*, vol. IT-22, pp. 1–10, Jan. 1976.
[28] R. Zamir, "The rate loss in the Wyner–Ziv problem," *IEEE Trans. Inform. Theory*, vol. 42, pp. 2073–2084, Nov. 1996.
[29] R. Zamir and M. Feder, "On universal quantization by randomized uniform/lattice quantizer," *IEEE Trans. Inform. Theory*, vol. 38, pp. 428–436, Mar. 1992.
[30] ——, "On lattice quantization noise," *IEEE Trans. Inform. Theory*, vol. 42, pp. 1152–1159, July 1996.
[31] ——, "Information rates of pre/post filtered dithered quantizers," *IEEE Trans. Inform. Theory*, vol. 42. pp. 1340–1353, Sept. 1996.