Remote Vector Gaussian Source Coding With Decoder Side Information Under Mutual Information and Distortion Constraints

Chao Tian, *Member, IEEE*, Jun Chen, *Member, IEEE*

Abstract—Let X, Y, Z be zero-mean, jointly Gaussian random vectors of dimensions n_x, n_y and n_z, respectively. Let $\mathcal{P}$ be the the set of random variables of W such that $W \leftrightarrow Y \leftrightarrow (X, Z)$ is a Markov string. We consider the following optimization problem

$$\min_{W \in \mathcal{P}} I(Y; W|Z),$$

subject to one of the following two possible constraints: (1) $I(X; W|Z) \geq R_I$, (2) the mean squared error between X and $\hat{X} = \mathbb{E}(X|W, Z)$ is less than d. The problem under the first kind of constraint is motivated by MIMO relay channels with an oblivious transmitter and a relay connected to the receiver through a dedicated link, while for the second case it is motivated by source coding with decoder side information where the sensor observation is noisy. In both cases, we show that the joint Gaussian solution is optimal. Moreover, explicit water filling interpretations are given for both cases, which suggest transform coding approaches performed in different transform domains, and that the optimal solution for one problem is in general suboptimal for the other.

Index Terms—Gaussian vector, source coding with side information, relay channel, transform coding.

I. INTRODUCTION

In a relay channel [1], the relay wishes to help the transmitter to convey information reliably to the end receiver. In certain cases, the transmitter does not actively cooperate with the relay, which falls into the oblivious cooperation framework [2]; it can also be understood as a special case of the nomadic transmitter setting where two relays send summary information regarding their received signals to a final central processing center [3], but in this special case the link between one relay and the processing center is taken to have an almost infinite capacity. In both settings, the transmitter may send its signal without awareness of the existence of the relay, and the relay does not necessarily know exactly the codebooks the transmitter is using[1].

We are interested in the case that the relay is connected to the receiver through a dedicated link (e.g. a separate frequency band or a land line) of a maximum rate R. We assume the channel noises are additive and Gaussian, and moreover the

Chao Tian is with AT&T Labs-Research, Florham Park, NJ, USA.

Jun Chen is with the Department of Electrical and Computer Engineering, McMaster University, Hamilton, ON L8S 4K1, Canada (email: junchen@ece.mcmaster.ca).

[1]If the relay knows the codebooks of the transmitter, the relay may be able to decode the messages, or when it is not able to decode, may use any coding strategy to help the transmitter. The restricted problem setting we consider here essentially requires the relay to always use the compress-and-forward strategy given in [1].

transmitter which has n_x transmit antennas is using Gaussian signaling X; this is a reasonable and popular assumption [3]. The relay which has n_y receive antennas observes Y, and it is a noisy version of X. The relay now has the task of source coding this noisy observation for the end receiver. The end receiver, which is the decoder in this source coding setting, has n_z antennas and thus also has side information Z through its own observation of the channel. Assuming channel state information and the probability distribution of X are known at the relay and end receiver, and using well-known property of typical sequences, it can be shown that the system can reliably transmit at rate arbitrarily close to R'_I, if there exists a random variable $W \leftrightarrow Y \leftrightarrow (X, Z)$, such that

$$I(X; Z, W) \geq R'_I, \quad I(Y; W|Z) \leq R. \quad (1)$$

This can be rewritten as the optimization problem

$$R(R_I) \triangleq \min_{W \in \mathcal{P}} I(Y; W|Z), \quad (2)$$

subject to

$$I(X; W|Z) = I(X; Z, W) - I(X; Z)$$
$$\geq R'_I - I(X; Z) \triangleq R_I, \quad (3)$$

which is exactly the optimization problem under the first kind of constraint given in the abstract. Roughly speaking, the relay utilizes a source coding scheme based on the Wyner-Ziv coding technique [4], and the quantity $I(Y; W|Z)$ is exactly this coding rate, however the receiver places the mutual information constraint (3), instead of attempting to reconstruct the source X to satisfy certain distortion constraint as in [4].

Let us turn to the more conventional rate-distortion setting where the encoder observes a noisy version of the source, and the end receiver, which has certain side information, is interested in recovering the source within certain distortion. Wyner and Ziv's solution [4] generalizes naturally to this remote source setting (see [5], [6]). The single letter optimization posed in the abstract under the second constraint is thus exactly the rate distortion function under a mean squared error (MSE) distortion criterion, i.e., the quantity $I(Y; W|Z)$ is again the source coding rate. We shall denote this function as $R(d)$.

It is now clear that the essential task is to determine the optimal random variable W, since in both problems it provides the optimal quantization codebook, and perhaps more importantly, the optimal resource allocation for these vector source coding problems. One key question in this context is

whether the optimal solutions of the random variables W are the same under these two constraints.

In this correspondence, we show that for both cases a jointly Gaussian vector solution is optimal for W, which also naturally leads to transform coding and water-fill interpretations. However, the water filling is performed in different spaces for the two problems, which implies that in general the optimal solutions in the rate distortion setting are sub-optimal solutions in the relay setting, and vice versa. It is thus important to distinguish between these two constraints, and use the appropriate one depending on the particular system requirement.

One may also ask the question whether if giving the decoder side information $\boldsymbol{Z}$ to the encoder helps or not. Indeed our result suggests that encoder side information does not help for both cases, however it does play a role in determining the transform domains.

The solution of the vector problem under the mutual information constraint is largely inspired by the work in [8] for the case without side information at the decoder. However, our proof is significantly simpler than theirs, despite the fact that side information is introduced. Our work is also relevant to [7]. However in [7] only distortion-constrained problem with direct observation of the source is considered, which does not include the problems considered here.

In the rest of the paper, the following notation is adopted. Random vectors are written in bold capital letters, and random variables in capital letters. $\boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{Z}$ are jointly Gaussian, zero-mean, real-valued random vectors. We write their variance matrices as $\Sigma_x, \Sigma_y, \Sigma_z$, and covariance matrices as Σ_{xy}, Σ_{yz} and Σ_{xz}, respectively. $\boldsymbol{X}$ is the underlying source of interest, $\boldsymbol{Y}$ is the noisy observation at the encoder, and $\boldsymbol{Z}$ is the decoder side information. We do not assume that $\boldsymbol{Y}$ and $\boldsymbol{Z}$ are conditionally independent given $\boldsymbol{X}$, i.e., the channel noises are not necessarily independent. All the logarithms are base e.

Before proceeding, we shall rewrite the relation between $\boldsymbol{X}$, $\boldsymbol{Y}$ and $\boldsymbol{Z}$ to simplify the proof given in the sequel. It is clear that without loss of generality we can write

$$
\begin{aligned}
\boldsymbol{Y} &= K_{yz}\boldsymbol{Z} + \boldsymbol{N_1}, \\
\boldsymbol{X} &= K_{xy}\boldsymbol{Y} + K_{xz}\boldsymbol{Z} + \boldsymbol{N_2} \\
&= (K_{xy}K_{yz} + K_{xz})\boldsymbol{Z} + K_{xy}\boldsymbol{N_1} + \boldsymbol{N_2},
\end{aligned} \tag{4}
$$

where $\boldsymbol{N_1}$, $\boldsymbol{N_2}$ are zero-mean jointly Gaussian random vectors independent of each other; $\boldsymbol{N_1}$ is independent of $\boldsymbol{Z}$ and $\boldsymbol{N_2}$ is independent of $(\boldsymbol{Y}, \boldsymbol{Z})$. More precisely, we have by applying well-known linear estimation calculation [9] that

$$
\begin{aligned}
K_{yz} &= \Sigma_{yz}\Sigma_z^{-1}, \\
K_{xy} &= \Sigma_{xy}\Sigma_y^{-1} + \Sigma_{xy}\Sigma_y^{-1}\Sigma_{yz}\Delta^{-1}\Sigma_{zy}\Sigma_y^{-1} \\
&\quad - \Sigma_{xz}\Delta^{-1}\Sigma_{zy}\Sigma_y^{-1}, \\
K_{xz} &= \Sigma_{xz}\Delta^{-1} - \Sigma_{xy}\Sigma_y^{-1}\Sigma_{yz}\Delta^{-1},
\end{aligned} \tag{5}
$$

where $\Delta = \Sigma_z - \Sigma_{zy}\Sigma_y^{-1}\Sigma_{yz}$. We shall assume the inverses of the above matrices are well defined, which is usually satisfied except for some degenerate cases. The following covariance

matrices are also important,

$$
\begin{aligned}
\Sigma_{N_1} &= \Sigma_{y|z} = \Sigma_y - K_{yz}\Sigma_{zy} \\
\Sigma_{N_2} &= \Sigma_{x|yz} = \Sigma_x - K_{xy}\Sigma_{yx} - K_{xz}\Sigma_{xz} \\
\Sigma_{x|z} &= \Sigma_x - \Sigma_{xz}\Sigma_z^{-1}\Sigma_{zx}.
\end{aligned} \tag{6}
$$

II. SCALAR RANDOM VARIABLE PROBLEMS

For the scalar case, denote $k_{zy} \triangleq K_{zy}$, $k_{xy} \triangleq K_{xy}$ and $k_{xz} \triangleq K_{xz}$, respectively. We consider the scalar Gaussian case first. Since we are not aware a formal proof for the case under the mutual information constraint, such a proof is given; for the other case, we simply restate the known result.

A. The mutual information constraint problem

Theorem 1: For the scalar Gaussian source and $R_I \geq 0$,

$$
R(R_I) = \infty, \quad \text{if} \quad R_I \geq \frac{1}{2}\log\frac{\Sigma_{x|z}}{\Sigma_{x|yz}}, \tag{7}
$$

and otherwise

$$
R(R_I) = \frac{1}{2}\log\frac{\Sigma_{x|z} - \Sigma_{x|yz}}{\Sigma_{x|z}\exp(-2R_I) - \Sigma_{x|yz}}. \tag{8}
$$

Proof: We start the converse proof by writing

$$
I(X; W|Z) = h(X|Z) - h(X|W, Z). \tag{9}
$$

To bound the second term, we have

$$
\begin{aligned}
h(X|W, Z) &= h(k_{xy}Y + k_{xz}Z + N_2|W, Z) \\
&= h(k_{xy}Y + N_2|W, Z).
\end{aligned} \tag{10}
$$

Then it follows

$$
\begin{aligned}
&\exp[2h(k_{xy}Y + N_2|W, Z)] \\
&\overset{(a)}{\geq} \exp[2h(k_{xy}Y|W, Z)] + \exp[2h(N_2)] \\
&= k_{xy}^2\exp[2h(Y|W, Z)] + \exp[2h(N_2)] \\
&= k_{xy}^2\exp[-2I(Y; W|Z) + 2h(Y|Z)] + \exp[2h(N_2)] \\
&= k_{xy}^2\exp[-2I(Y; W|Z) + 2h(N_1)] + \exp[2h(N_2)],
\end{aligned} \tag{11}
$$

where (a) is by applying the conditional version of the entropy power inequality [10]. Thus we have

$$
\begin{aligned}
I(Y; W|Z) &\geq \frac{1}{2}\log[k_{xy}^2\exp(2h(N_1))] \\
&\quad - \frac{1}{2}\log[\exp(2h(X|Z) - 2I(W; X|Z)) - \exp(2h(N_2))].
\end{aligned}
$$

Because the function $-\log(r\exp(-2x) - q)$ is monotonically increasing in x when $r > 0$, and $I(W; X|Z) \geq R_I$, we have

$$
\begin{aligned}
I(Y; W|Z) &\geq \frac{1}{2}\log[k_{xy}^2\exp(2h(N_1))] \\
&\quad - \frac{1}{2}\log[\exp(2h(X|Z) - 2R_I) - \exp(2h(N_2))]. \tag{12}
\end{aligned}
$$

Clearly when we take $W = Y + N_3$ where N_3 is a zero-mean Gaussian independent of everything else such that $I(W; X|Z) = R_I$, all the inequalities become equality.

Since $I(W; X|Z) \leq I(W, Y; X|Z)$, the Markov string $W \leftrightarrow Y \leftrightarrow (X, Z)$ implies that $I(W, Y; X|Z) = I(Y; X|Z)$. Thus when $R_I \geq I(Y; X|Z)$, the optimization problem is not feasible. Substitute appropriate values into the quantities, and it is easy to see $\Sigma_{x|z} - \Sigma_{x|yz} = k_{xy}^2\Sigma_{y|z}$ because it is the variance of $k_{xy}N_1$. The proof is thus complete. ∎

B. The rate-distortion problem

Theorem 2 ([5] and [6]): For the scalar Gaussian source, we have for $\Sigma_{x|z} \geq d \geq \Sigma_{x|yz}$,

$$R(d) = \frac{1}{2} \log \frac{\Sigma_{x|z} - \Sigma_{x|yz}}{d - \Sigma_{x|yz}}; \qquad (13)$$

and $R(d) = 0$ if $d > \Sigma_{x|z}$; $R(d) = \infty$ if $0 \leq d < \Sigma_{x|yz}$.

We omit the proof here; see [5] and [6]. It can be shown that in both cases encoder side information does not help, i.e., even when the Markov string is relaxed to $W \leftrightarrow (Z, Y) \leftrightarrow X$ the solutions do not change. Further notice the similarity between the functions $R(R_I)$ and $R(d)$. Indeed, for the scalar problem, for both cases the solution is given by the same Gaussian solution $W = Y + N_3$ with the variance of N_3 properly chosen. However, as we shall illustrate next, for the vector problem, their solutions no longer coincide.

III. RANDOM VECTOR PROBLEMS

In this section we discuss the vector version of the problems. We provide a proof by decomposing the problem into a set of component-wise problems. It is conceivable that the enhancement technique given in [11], [12] may be of use, however our proof directly yields the water filling interpretation.

A. The mutual information constraint problem

The following notation is needed. For a symmetric positive definite matrix S, let its eigenvalue decomposition be $U^T \Lambda U$, where Λ is the diagonal matrix with eigenvalues on its diagonal. Define $S^{1/2} = U^T \Lambda^{1/2} U$, where $\Lambda^{1/2}$ is the diagonal matrix where the squared root of each element of Λ is taken. Similarly, $S^{-1/2}$ is also defined for a symmetric positive definite matrix, however with an inverse on each non-zero element of Λ. The main result in this subsection is the following theorem.

Theorem 3: Let the eigenvalue of $\Sigma_{y|z}^{1/2} K_{xy}^T \Sigma_{x|z}^{-1} K_{xy} \Sigma_{y|z}^{1/2}$ be given as μ_i. Then for $0 \leq R_I \leq \frac{1}{2} \log |\Sigma_{x|z}| - \frac{1}{2} \log |\Sigma_{x|yz}| \triangleq R_I^*$,

$$R(R_I) = \sum_{i=1}^{n_x} \frac{1}{2} \log \left(\mu_i \left[\left(\frac{1 - \mu_i}{1 - \gamma} \right)^- - (1 - \mu_i) \right]^{-1} \right),$$

where $(x)^- = \min(x, 1)$, $\log \frac{0}{0} \triangleq 0$, and $\gamma \in [0, 1)$ is chosen such that

$$-\frac{1}{2} \sum_{i=1}^{n_x} \log \left(\frac{1 - \mu_i}{1 - \gamma} \right)^- = R_I.$$

If $R_I > R_I^*$, then $R(R_I) = \infty$.

We shall consider a specific transform; this transform is usually referred to as the canonical correlation analysis in statistics. First find the singular value decomposition of $\Sigma_{x|z}^{-1/2} K_{xy} \Sigma_{y|z}^{1/2} = U^T \Lambda^I V$, where U and V are two orthogonal matrices and Λ^I is a diagonal matrix with singular values on the diagonal. Define $T_x = U \Sigma_{x|z}^{-1/2}$ and $T_y = V \Sigma_{y|z}^{-1/2}$, which in fact gives the transforms $\mathbf{X'} = T_x \mathbf{X}$ and $\mathbf{Y'} = T_y \mathbf{Y}$,

respectively. Notice that the transformation matrices are full rank, and under these transforms,

$$\Sigma_{x'|z} = T_x \Sigma_{x|z} T_x^T = U \Sigma_{x|z}^{-1/2} \Sigma_{x|z} \Sigma_{x|z}^{-1/2} U^T = I_{n_x}, \quad (14)$$

where I_{n_x} is the identity matrix of dimension n_x. Similarly,

$$\Sigma_{y'|z} = T_y \Sigma_{y|z} T_y^T = V \Sigma_{y|z}^{-1/2} \Sigma_{y|z} \Sigma_{y|z}^{-1/2} V^T = I_{n_y}. \quad (15)$$

Moreover, we have

$$\begin{aligned} \Sigma_{x'y'|z} &= T_x K_{xy} \Sigma_{y|z} T_y^T \\ &= U \Sigma_{x|z}^{-1/2} K_{xy} \Sigma_{y|z} \Sigma_{y|z}^{-1/2} V^T = \Lambda^I. \end{aligned} \quad (16)$$

These facts imply that after the transform, the components of $\mathbf{X'}$, respectively the components of $\mathbf{Y'}$, are independent and identically distributed with unit variance given $\mathbf{Z}$. The vectors of $\mathbf{X'}$ and $\mathbf{Y'}$ are only component-wise dependent given $\mathbf{Z}$. Moreover, since the element of $\Sigma_{x'y'|z} = \Lambda^I$ is a covariance of two random variables of unit variance, its diagonal values satisfy $0 \leq |\lambda_i| \leq 1$.

Proof: Since the transform T_x and T_y are full rank, and thus information lossless, we have

$$I(\mathbf{X}; W|\mathbf{Z}) = I(\mathbf{X'}; W|\mathbf{Z}). \quad (17)$$

Thus we have by the conditional independence and the chain rule

$$\begin{aligned} I(\mathbf{X}; W|\mathbf{Z}) &= \sum_{i=1}^{n_x} [h(X_i'|\mathbf{Z}) - h(X_i'|W, \mathbf{Z}, \mathbf{X}_{i-}')] \\ &\leq \sum_{i=1}^{n_x} [h(X_i'|\mathbf{Z}) - h(X_i'|W, \mathbf{Z}, \mathbf{X}_{i-}', \mathbf{Y}_{i-}')] \\ &\overset{(a)}{=} \sum_{i=1}^{n_x} [h(X_i'|\mathbf{Z}) - h(X_i'|W, \mathbf{Z}, \mathbf{Y}_{i-}')] \\ &= \sum_{i=1}^{n_x} I(X_i'; W, \mathbf{Y}_{i-}'|\mathbf{Z}), \end{aligned} \quad (18)$$

where (a) is due to the Markov string $X_i' \leftrightarrow (\mathbf{Y}_{i-}', W, \mathbf{Z}) \leftrightarrow \mathbf{X}_{i-}'$. On the other hand we have

$$R \geq I(\mathbf{Y}; W|\mathbf{Z}) = \sum_{i=1}^{n_y} h(Y_i'|\mathbf{Z}) - h(Y_i'|W, \mathbf{Z}, \mathbf{Y}_{i-}')]$$

$$= \sum_{i=1}^{n_y} I(Y_i'; W, \mathbf{Y}_{i-}'|\mathbf{Z}). \quad (19)$$

Now define $(W, \mathbf{Y}_{i-}') = W_i$, and it is clear that $X_i' \leftrightarrow (\mathbf{Z}, Y_i') \leftrightarrow W_i$ is a Markov string. Thus we have relaxed the problem into

$$\min \sum_{i=1}^{n_y} I(Y_i'; W_i|\mathbf{Z}), \quad (20)$$

$$\text{subject to: } \sum_{i=1}^{n_x} I(X_i'; W_i|\mathbf{Z}) \geq R_I, \quad (21)$$

where $X_i' \leftrightarrow (\mathbf{Z}, Y_i') \leftrightarrow W_i$ for $i \leq \max(n_x, n_y)$.

Using the same outer bounding derivation as the scalar problem we can further write the relaxed problem as[2]

$$\min \sum_{i=1}^{n_I} -\frac{1}{2} \log[\exp(-2R_i) - (1 - \lambda_i^2)] + \frac{1}{2} \log[\lambda_i^2] \quad (22)$$

$$\text{subject to: } \sum_{i=1}^{n_I} R_i \geq R_I, \ -\frac{1}{2} \log(1 - \lambda_i^2) \geq R_i \geq 0, \quad (23)$$

where n_I is the number of non-zero λ_i values.

A water-filling solution can be derived from the above optimization problem. More precisely, by the Kuhn-Tucker condition we can find the optimizing solution as

$$R_i = \frac{1}{2} \log \left(\frac{\lambda - 1}{\lambda(1 - \lambda_i^2)} \right)^+,$$

where we use $(x)^+$ to denote $\max(x, 1)$, and λ is the "water level" such that the sum-rate constraint is satisfied with equality. Now defining $\gamma = \lambda^{-1}$ and noticing $\lambda_i^2 = \mu_i$ give the formula in the theorem.

It remains to show that the solution of the relaxed problem is indeed achievable, i.e., there exists a random variable (vector) W that matches this lower bound, but this is obvious by letting $W = T_y Y + N_3$, where N_3 is a Gaussian random vector with independent components such that the equality $I(X_i'; Y_i' + N_{3,i}|Z) = R_i$ holds. This indeed is a Gaussian solution, and the theorem is proved. ∎

B. The rate-distortion problem

Assuming the MSE distortion is taken to be averaged over each component, we have the following theorem.

Theorem 4: Let the eigenvalues of $K_{xy}\Sigma_{y|z}K_{xy}^T$ be given as λ_i^D, $i = 1, 2, ..., n_x$. Then for $\text{Tr}(\Sigma_{x|z}) \geq n_x d \geq \text{Tr}(\Sigma_{x|yz})$,

$$R(d) = \sum_{i=1}^{n_x} \frac{1}{2} \log \left[\frac{\lambda_i^D}{\lambda} \right]^+, \quad (24)$$

where $\text{Tr}(\cdot)$ is the trace of a matrix, and $\lambda > 0$ is chosen such that

$$\sum_{i=1}^{n_x} \min(\lambda, \lambda_i^D) = n_x d - \text{Tr}\left[\Sigma_{x|yz}\right]. \quad (25)$$

If $n_x d < \text{Tr}(\Sigma_{x|yz})$, then $R(d) = \infty$; if $n_x d > \text{Tr}(\Sigma_{x|z})$, then $R(d) = 0$.

The proof of this theorem is rather standard, however we include it below for completeness.

Proof: To prove the converse, consider the fact that

$$\text{Tr}[\mathbb{E}(X - \hat{X})(X - \hat{X})^T] \leq n_x d. \quad (26)$$

It is clear that

$$\mathbb{E}(X - \hat{X})(X - \hat{X})^T$$
$$= \mathbb{E}(K_{xy}Y + K_{xz}Z - \hat{X})(K_{xy}Y + K_{xz}Z - \hat{X})^T + \Sigma_{x|yz}$$

because N_2 is independent of Y and Z, thus independent of W and Z. Now define $\bar{X} = \hat{X} + (K_{xy}K_{yz} + K_{xz})Z$, from which we see that

$$\text{Tr}\left[\mathbb{E}(X - \hat{X})(X - \hat{X})^T\right]$$
$$= \text{Tr}\left[\mathbb{E}(K_{xy}N_1 - \bar{X})(K_{xy}N_1 - \bar{X})^T\right]$$
$$+ \text{Tr}\left[\Sigma_{x|yz}\right] \leq n_x d. \quad (27)$$

Thus there exists a function $\bar{X} = f'(W, Z)$ such that

$$\text{Tr}\left[\mathbb{E}(K_{xy}N_1 - \bar{X})(K_{xy}N_1 - \bar{X})^T\right]$$
$$\leq n_x d - \text{Tr}\left[\Sigma_{x|yz}\right] \triangleq n_x d'. \quad (28)$$

Denote the eigenvalue decomposition of $K_{xy}\Sigma_{y|z}K_{xy}^T$ as $U^T \Lambda^D U$. Denote $N' = U K_{xy} N_1$, which clearly has independent components. We have the following chain of inequalities.

$$I(Y; W|Z) \geq I(U K_{xy} Y; W|Z)$$
$$= h(U K_{xy} Y|Z) - h(U K_{xy} Y|W, Z)$$
$$= h(U K_{xy} N_1) - h(U K_{xy} N_1|W, Z)$$
$$= \sum_{i=1}^{n_x} [h(N_i') - h(N_i'|W, Z, N_i'^-)]$$
$$\geq \sum_{i=1}^{n_x} [h(N_i') - h(N_i'|W, Z)]$$
$$= \sum_{i=1}^{n_x} I(N_i'; W, Z)$$
$$\geq \sum_{i=1}^{n_x} I(N_i'; \mathbb{E}(N_i'|W, Z)) = \sum_{i=1}^{n_x} I(N_i'; \hat{N}_i'), \quad (29)$$

where $\hat{N}_i' \triangleq \mathbb{E}(N_i'|W, Z)$. Notice that U is an orthogonal transform, thus preserves the MSE distortion. Define $d_i = \mathbb{E}(N_i' - \hat{N}_i')^2$, then it is clear that we have $\sum_{i=1}^{n} d_i \leq n_x d'$. This is a conventional rate-distortion problem, and thus

$$R(d) \geq \sum \frac{1}{2} \log \left[\frac{\lambda_i^D}{\lambda} \right]^+, \quad (30)$$

where λ is the water level such that

$$\sum_{i=1}^{n_x} \min(\lambda, \lambda_i^D) = n_x d - \text{Tr}\left[\Sigma_{x|yz}\right]. \quad (31)$$

This establish the lower bound for the rate distortion function. The fact that this lower bound is achievable is straightforward by letting $W = U K_{xy} Y + N_3$, where N_3 is a Gaussian vector having independent components, with their variances determined by

$$\sigma_{3,i}^2 = \frac{\min(\lambda, \lambda_i^D)}{\lambda_i^D - \min(\lambda, \lambda_i^D)}. \quad (32)$$

The proof is complete. ∎

IV. CONCLUDING REMARKS

From the proofs of Theorem 3 and Theorem 4, it is seen that the two transforms in which water filling is performed are different under the two kinds of constraints. For the mutual information constraint, the transform is $T_y = V\Sigma_{y|z}^{-1/2}$,

where V is determined by the eigenvalue decomposition of $\Sigma_{y|z}^{1/2} K_{xy}^T \Sigma_{x|z}^{-1} K_{xy} \Sigma_{y|z}^{1/2}$, and the water filling is along these eigenvalues. For the MSE distortion constraint, the transform is determined by the eigenvalue decomposition of $K_{xy} \Sigma_{y|z} K_{xy}^T$, and the water filling is along these eigenvalues. Moreover, the transform T_y may not be orthogonal, thus does not preserve trace and the MSE distortion. This result implies that the optimal solution for one problem is in general suboptimal for the other. Thus when MIMO system is considered for the relay problem, it is important to recognize that the Wyner-Ziv source coding component should be used under the mutual information constraint, instead of distortion constraint. It should also be noted that the problem under mutual information constraint considered here includes the hypothesis testing problem in [13] as a special case when the source is vector Gaussian, and thus the given result provides an explicit solution for it.

It is theoretically interesting to consider the same optimization problem when the two kinds of constraints are placed simultaneously. The extremal inequality in [11] indicates that the optimal value is attained by timesharing Gaussian solutions. However, it is unclear whether a single Gaussian solution suffices and whether there exists an explicit water filling interpretation.

REFERENCES

[1] T. Cover and A. El Gamal, "Capacity theorems for the relay channel," *IEEE Trans. Information Theory,* vol. 25, no. 5pp. 572-584, Sep. 1979.
[2] M. Katz and S. Shamai, "Oblivious cooperation in collocated wireless networks," in proc. *2006 IEEE International Symposium on Information Theory,* Jul. 2006, pp. 2062-2066.
[3] A. Sanderovich, S. Shamai, Y. Steinberg, and G. Kramer, "Communication via decentralized processing," *IEEE Trans. Information Theory*, vol. 54, no. 7, pp. 3008–3023, Jul. 2008.
[4] A. D. Wyner and J. Ziv, "The rate-distortion function for source coding with side information at the decoder," *IEEE Trans. Information Theory*, vol. 22, no. 1, pp. 1–10, Jan. 1976.
[5] H. Yamamoto and K. Itoh, "Source coding theory for multi-terminal communication systems with a remote source," *Trans. of the IECE of Japan*, vol. E63, no. 10, pp. 700-706, 1980.
[6] S. Draper and G. Wornell, "Side information aware coding strategies for estimation under communication constraints," *IEEE Journal of Selected Areas in Communication*, Vol. 22, No. 6, Aug. 2004.
[7] M. Gastpar, P. L. Dragotti and M. Vetterli, "The distributed Karhunen-Loeve transform," *IEEE Trans. Information Theory*, vol. 52, no. 12, pp. 5177–5196, Dec. 2006.
[8] A. Globerson and N. Tishby, "On the optimality of the Gaussian information bottleneck curve," Hebrew University Technical Report, 2004.
[9] T. Kailath, A. Sayed and B. Hassibi, *Linear estimation,* Prentice Hall, New Jersey, 2000.
[10] T. M. Cover and J. A. Thomas, *Elements of information theory*, New York: Wiley, 1991.
[11] T. Liu and P. Viswanath, "An extremal inequality motivated by multiterminal information-theoretic problems," *IEEE Transactions on Information Theory*, vol. 53, no. 5, pp. 1839-1851, May 2007.
[12] H. Weingarten, Y. Steinberg, and S. Shamai, "The capacity region of the Gaussian multiple-input multiple-output broadcast channel," *IEEE Trans. Information Theory*, vol. 52, no. 9, pp. 3936–3964, Sep. 2006.
[13] R. Ahlswede and I. Csiszár, "Hypothesis testing with communication constraints," *IEEE Trans. Information Theory*, vol. 32, no. 4, pp. 533-542, Jul. 1986.

Chao Tian (S'00, M'05) received the B.E. degree in Electronic Engineering from Tsinghua University, Beijing, China, in 2000 and the M.S. and Ph.D. degrees in Electrical and Computer Engineering from Cornell University, Ithaca, NY in 2003 and 2005, respectively.

Dr. Tian was a postdoctoral researcher at Ecole Polytechnique Federale de Lausanne (EPFL) from 2005 to 2007. He joined AT&T Labs–Research, Florham Park, New Jersey in 2007, where he is now a Senior Member of Technical Staff. His research interests include multi-user information theory, joint source-channel coding, quantization design and analysis, as well as image/video coding and processing.

Jun Chen (S'03M'06) received the B.E. degree with honors in communication engineering from Shanghai Jiao Tong University, Shanghai, China, in 2001 and the M.S. and Ph.D. degrees in electrical and computer engineering from Cornell University, Ithaca, NY, in 2003 and 2006, respectively.

He was a Postdoctoral Research Associate in the Coordinated Science Laboratory at the University of Illinois at Urbana-Champaign, Urbana, IL, from 2005 to 2006, and a Josef Raviv Memorial Postdoctoral Fellow at the IBM Thomas J. Watson Research Center, Yorktown Heights, NY, from 2006 to 2007. He is currently an Assistant Professor of Electrical and Computer Engineering at Mc- Master University, Hamilton, ON, Canada. He holds the Barber-Gennum Chair in Information Technology. His research interests include information theory, wireless communications, and signal processing.