A LOWER BOUND TO THE AWGN REMOTE RATE-DISTORTION FUNCTION

Michael Gastpar

University of California, Berkeley
Department of EECS
Berkeley, CA 94720-1770
gastpar@eecs.berkeley.edu

ABSTRACT

In the remote source coding problem, an underlying source is observed in noise. The noisy observations must be encoded into a bit stream in such a way as to enable the decoder to produce a good approximation to the original source sequence. The trade-off between the rate of the bit stream and the fidelity of the reconstructed source sequence is sometimes referred to as the remote rate-distortion function. This paper focuses on a special case of the remote source coding problem: The encoder obtains M noisy versions of each underlying source sample. The probability density function of the underlying source is arbitrary, but the observation noise is assumed to be Gaussian (hence the name "AWGN remote rate-distortion function"). The goal is to reconstruct the underlying source sequence to within mean-squared error. For this scenario, a new lower bound to the rate-distortion function is presented.

The investigations are motivated by a study of the fundamental performance trade-offs in certain sensor network scenarios. The presented lower bound on the remote rate-distortion function is one of the building blocks for a cut-set argument that leads to an upper bound to the performance achievable in these sensor networks.

1. INTRODUCTION

In this paper, we derive a lower bound to the remote rate-distortion function when the observation noise is Gaussian. More precisely, as illustrated in Figure 1, the encoder gets to observe M noisy versions of one and the same underlying source. In each time instant n (in discrete time), the source output is $S[n]$, which is sampled in an independent and identically distributed (iid) fashion from a fixed and known (but arbitrary) source distribution $p_S(s)$. The encoder obtains M noisy observations of this source, according to

$$U_m[n] \quad = \quad S[n] + W_m[n], \text{ for } m = 1, 2, \ldots, M, \quad (1)$$

where, for the scope of this paper, $W_m[n]$ are assumed to be iid Gaussian random variables (both over m and over n)

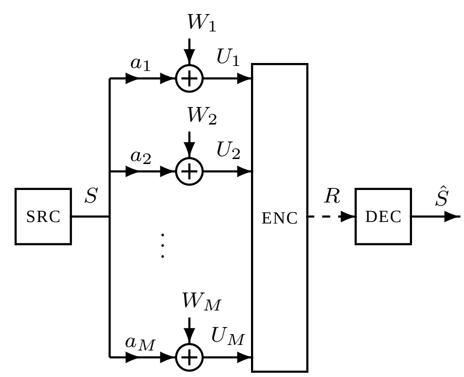

Fig. 1. The AWGN remote source coding problem: An underlying source S, not necessarily Gaussian, is observed M-fold in Gaussian noise.

of mean zero and variance σ_W^2. We denote the vector of M observations by $U[n] = (U_1[n], \ldots, U_M[n])^T$.

The encoder encodes long blocks of observation vectors, $\{U[j]\}_{j=1}^n$, into a bit stream that appears at a rate of R bits per observed vector at the decoder. The decoder attempts to reconstruct the underlying source sequence $S[n]$ with respect to mean-squared error,

$$D \quad \stackrel{\text{def}}{=} \quad \lim_{n\to\infty} E\left[\frac{1}{n}\sum_{j=1}^{n}\left|S[j]-\hat{S}[j]\right|^2\right]. \qquad (2)$$

The corresponding rate-distortion question is then, as usual, that of optimally trading off the coding rate R and the resulting overall distortion D. Our problem has one more parameter of interest, namely the number of observations M. The focus of the present paper is on the case when M is "large".

If the underlying source sequence $\{S[j]\}_{j=1}^{\infty}$ is a sequence of iid Gaussian random variables (of mean zero and variance σ_S^2), the solution is well known. For $a_1 = a_2 =$

$\ldots = a_M = 1$, it can be expressed as

$$D_{remote}(R, M) = \frac{\sigma_S^2}{1 + \frac{M\sigma_S^2}{\sigma_W^2}} + \frac{\sigma_S^2 2^{-2R}}{1 + \frac{\sigma_W^2}{M\sigma_S^2}}. \quad (3)$$

This permits various interpretations, such as: How should the total rate R grow with M? Clearly, the first summand decreases like M^{-1}, and hence, in a scaling sense, there is no point in making the second term decay faster; to make it decay like M^{-1}, the total rate should be increased like $R \sim \log_2 M$.

In this paper, we show that the rate-distortion behavior of the form of Equation (3) continues to hold for a class of distributions $p_S(s)$ of the underlying source *beyond* the Gaussian case. First, it is easy to see that Equation (3) is an *upper* bound to $D_{remote}(R, M)$ for *any* iid source S with distribution $p_S(s)$ with variance σ_S^2. For completeness, an explicit argument is included in Appendix A. This shows that the essential behavior given in Equation (3) *can* be achieved for arbitrary distribution of the underlying source S, but leaves open the question whether a significantly better performance is achievable for specific source distributions $p_S(s)$. In this paper, we therefore focus on the derivation of a *lower* bound to $D_{remote}(R, M)$. For a class of source distributions $p_S(s)$, our lower bound confirms that the essential behavior of Equation (3) continues to hold.

Our investigations are motivated by sensor network scenarios such as considered in [1]: Suppose that M sensors independently encode the observations $U_m[n]$, where $m = 1, 2, \ldots, M$, for transmission across a multiple-access channel towards a central data acquisition center. Then, the best currently known lower bounds for the achievable end-to-end distortion result from a cut-set argument involving the idealized system of Figure 1 (idealized in the sense that the M sensors can collaboratively compress the M observation sequences), where the rate R is determined by a separate argument. In [1], it is shown that this simple bound, in fact, characterizes the optimum scaling behavior for the sensor network when the underlying source S is Gaussian. The results of the present paper permit to extend this insight to a more general class of sources.

2. THE REMOTE RATE-DISTORTION FUNCTION

The remote source coding problem was first studied in a joint source-channel coding context by Dobrushin and Tsybakov in [2]. Wolf and Ziv [3] studied the single-shot version of this problem (single source output, single channel use), a result that is also implicitly present in [2], as argued in [4]. For the case of quadratic distortion, they show elegant decompositions of the remote distortion criterion. That such elegant decompositions are not restricted to quadratic distortion criteria was recognized in [5, p.78-81]. Namely, the original distortion measure $d(s, \hat{s})$ is converted into a distortion measure expressed in terms of the observation U as $d(u, \hat{s}) = E_{S|u}[d(S, \hat{s})]$. This point of view was extended in [6].

In the present paper, we focus on the case of *mean-squared error* distortion. Here, the remote rate-distortion function can be expressed as [5, p.78-81]

$$D_{remote}(R) = \min E\left[\left|S - \hat{S}\right|^2\right] \quad (4)$$

where the minimization is over all distributions $p(\hat{s}|u)$ for which

$$I(U; \hat{S}) \leq R, \quad (5)$$

where U denotes the (column) vector of noisy observations, i.e.,

$$U \stackrel{\text{def}}{=} \begin{pmatrix} U_1 \\ U_2 \\ \vdots \\ U_M \end{pmatrix} = \begin{pmatrix} a_1 S + W_1 \\ a_2 S + W_2 \\ \vdots \\ a_M S + W_M \end{pmatrix} \quad (6)$$

To apply the elegant decomposition of [3], let $f(U)$ denote the minimum mean-squared error (MMSE) estimate of S based on the vector of noisy observations, i.e., $f(U) = E[S|U]$. Then,

$$\begin{aligned} &D_{remote}(R) \\ &= \min E\left[|S - f(U) + f(U) - \hat{S}|^2\right] \\ &= E\left[|S - f(U)|^2\right] + \min E\left[|f(U) - \hat{S}|^2\right], \end{aligned} \quad (7)$$

where the minimum is over all $p(\hat{s}|u)$ that satisfy $I(U; \hat{S}) \leq R$, and where the second equality holds since for the MMSE estimate,

$$E\left[(S - f(U))(f(U) - \hat{S})^H\right] = 0. \quad (8)$$

The goal of this paper is to derive a lower bound to the expression (7).

Evaluation of (7) for the Gaussian source. Before proceeding, let us briefly outline how (3) is found from (7). For the special case where $a_m = 1$, for $m = 1, 2, \ldots, M$ (which is the case of interest in this short note), it is shown in Appendix B that $\sum_{m=1}^M U_m$ is a sufficient statistic for S given U, irrespective of the source distribution $p_S(s)$. For general a_m, it can be shown that the sufficient statistic is $\sum_{m=1}^M a_m^* U_m$. When the underlying source S is Gaussian,

$$f(U) = E[S|U] = E[SU^H]\left(E[UU^H]\right)^{-1} U, \quad (9)$$

from which we find that

$$E\left[|S-f(U)|^2\right] = \frac{\sigma_S^2\sigma_W^2}{\sum_{k=1}^M |a_k|^2\sigma_S^2+\sigma_W^2}. \qquad (10)$$

Evaluating this, and using $A=\sum_{k=1}^M |a_k|^2$, yields

$$D_{remote}(R) = \frac{\sigma_S^2}{1+\frac{A\sigma_S^2}{\sigma_W^2}} + \frac{\sigma_S^2 2^{-2R}}{1+\frac{\sigma_W^2}{A\sigma_S^2}}.$$

Formula (3) is recovered by noting that $A=M$ in the case $a_m=1$, for $m=1,2,\ldots,M$.

3. LOWER BOUND

The spirit of the more general lower bound to the behavior of $D_{remote}(R,M)$ that we derive in this paper is for the source distribution $p_S(s)$ to be relatively "close" to the Gaussian, i.e., a well-behaved distribution. The theorem below gives one way of making this intuitive notion precise. However, as we will briefly mention in the conclusions, there are source distributions $p_S(s)$ for which the behavior is fundamentally different.

Theorem 1. *For the AWGN remote rate-distortion problem with $a_m=1$, for $m=1,2,\ldots,M$,[1] consider the class of underlying sources S with probability density function $p_S(s)$ with differential entropy $h(S)>-\infty$ and define*

$$g(x) \stackrel{\text{def}}{=} \frac{1}{\frac{d}{dx}h\left(\sqrt{x}S+W\right)} \qquad (11)$$

where W is Gaussian with zero mean and unit variance. If there exists x_0 such that for all $x\geq x_0$, $g(x)$ is a convex function of x, then there exists M_0 such that for all $M\geq M_0$,

$$D_{remote}(R,M) \geq \frac{\sigma_S^2}{\frac{2\pi e}{2^{2h(\tilde{S})}}+\frac{M\sigma_S^2}{\sigma_W^2}} + \frac{2^{2h(E[S|U])}}{2\pi e}2^{-2R}, \qquad (12)$$

where $\tilde{S}=S/Var(S)$ and $h(X)=-\int dx p_X(x)\log_2 p_X(x)$ denotes the differential entropy function with the binary logarithm.

The proof of this theorem is given in Appendix B.

Remark. Clearly, Condition (11) is not readily verified for all distributions of the underlying source S. Note, however, that if we set $W\equiv 0$, it is easy to verify that $\frac{d}{dx}h\left(\sqrt{x}S+W\right)=\frac{1}{2x}$, and hence, $g(x)$ is trivially convex. Intuitively speaking, for large enough x, the effect of W will disappear, such that most distributions $p_S(s)$ will actually satisfy Condition (11).

[1] In this note, we restrict attention to this simple case; it is, however, straightforward to extend our results to the more general case.

Remark. Note that (for a fixed source variance) the first term in Equation (12) is *maximized* when the source S is Gaussian. To see this, simply note that among all sources $\tilde{S}$ of variance 1, the one that maximizes the differential entropy $h(\tilde{S})$ is the Gaussian. For the second term, we can also obtain a rough understanding by noting that as $M\to\infty$, we have $Var(E[S|U])\to\sigma_S^2$ (in an appropriate sense that depends on the underlying source distribution $p_S(s)$). At fixed $Var(E[S|U])$, the entropy $h(E[S|U])$ is maximized when the random variable $E[S|U]$ is Gaussian. This is the case when the source S is Gaussian. Therefore, in this limit, the second term is also *maximized* when the source S is Gaussian.

Remark. In the context of the analysis of performance bounds for sensor networks, along the lines of [1], the theorem can be used to establish that the insights and conclusions given for the case of Gaussian sources in [1] continue to hold for an interesting class of source distributions $p_S(s)$.

4. EXAMPLES AND APPLICATIONS

In this section, we briefly illustrate Theorem 1. Specifically, we first show that the bound (Equation (12)) reduces to the correct expression in the special case where the underlying source S is Gaussian. Then, we outline how to evaluate the bound in the case where S is distributed according to a Laplacian distribution.

A. Gaussian source S

Theorem 1 is tight for the case where the underlying source S is Gaussian. In this case, we find that

$$h(\tilde{S}) = \frac{1}{2}\log_2 2\pi e, \qquad (13)$$

which recovers the first term. For the second term, we need to determine $h(E[S|U])$. In the Gaussian case, it is well known that

$$E[S|U] = E[SU^H]\left(E[UU^H]\right)^{-1}U. \qquad (14)$$

In the special case at hand,

$$E[UU^H] = \begin{pmatrix} \sigma_S^2+\sigma_W^2 & \sigma_S^2 & \ldots & \sigma_S^2 \\ \sigma_S^2 & \sigma_S^2+\sigma_W^2 & \ldots & \sigma_S^2 \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_S^2 & \sigma_S^2 & \ldots & \sigma_S^2+\sigma_W^2 \end{pmatrix}.$$

This is a circulant matrix, and hence, easy to manipulate. For example, the determinant can be expressed in closed form as

$$\det(E[UU^H]) = (M\sigma_S^2+\sigma_W^2)(\sigma_W^2)^{M-1}. \qquad (15)$$

Moreover, the inverse of the matrix $E[UU^H]$ is easily found to be of the same structure, with elements

$$\frac{1}{M\sigma_S^2+\sigma_W^2}\left(1+(M-1)\frac{\sigma_S^2}{\sigma_W^2}\right) \qquad (16)$$

along the diagonal, and elements

$$-\frac{1}{M\sigma_S^2+\sigma_W^2}\frac{\sigma_S^2}{\sigma_W^2} \tag{17}$$

everywhere else. With this, it is easy to evaluate the expression in Equation (14) to yield

$$E[S|U] \quad = \quad \frac{\sigma_S^2}{M\sigma_S^2+\sigma_W^2}\sum_{m=1}^{M} U_m, \tag{18}$$

which is a Gaussian random variable with mean zero and variance

$$\begin{aligned} \sigma^2 &= \left(\frac{\sigma_S^2}{M\sigma_S^2+\sigma_W^2}\right)^2\left(M^2\sigma_S^2+M\sigma_W^2\right) \\ &= \frac{M\sigma_S^2}{M\sigma_S^2+\sigma_W^2} \\ &= \frac{\sigma_S^2}{1+\frac{\sigma_W^2}{M\sigma_S^2}}. \end{aligned} \tag{19}$$

The differential entropy of a Gaussian random variable of variance σ^2 is well known to be

$$h(E[S|U]) \quad = \quad \frac{1}{2}\log_2(2\pi e)\sigma^2. \tag{20}$$

Using this in the lower bound presented in Theorem 1 yields the exact formula for the remote AWGN rate-distortion function for the Gaussian source, which was given above in Equation (3).

B. Laplacian source S

Suppose S is distributed according to the Laplacian distribution and has variance σ_S^2, i.e.,

$$p_S(s) \quad = \quad \frac{1}{2b}\exp(-|s|/b), \tag{21}$$

where $\sigma_S^2 = 2b^2$. For the unit-variance Laplacian $\tilde{S}$, it is well known that

$$\begin{aligned} h(\tilde{S}) &= \frac{1+\ln(\sqrt{2})}{\ln 2} = \log_2(\sqrt{2}e) \\ &= \frac{1}{2}\log_2(2e^2). \end{aligned} \tag{22}$$

Hence, the lower bound takes the shape

$$\begin{aligned} & D_{remote}(R,M) \\ & \geq \quad \frac{\sigma_S^2}{\frac{\pi}{e}+\frac{M\sigma_S^2}{\sigma_W^2}} + \frac{2^{2h(E[S|U])}}{2\pi e}2^{-2R}, \end{aligned} \tag{23}$$

We refrain from attempting to evaluate $h(E[S|U])$ explicitly and simply note that as $M \to \infty$, this goes to a constant.

5. CONCLUSIONS AND EXTENSIONS

This paper provides a lower bound to the remote rate-distortion function when a source S is observed M-fold in additive white Gaussian noise. The lower bound can be used, e.g., to establish lower bounds to the scaling behavior of certain sensor network scenarios of interest, along the lines of [1].

As for extensions, an interesting case *not* covered by Theorem 1 concerns the scenario where the underlying source S is discrete. Here, the first term in (7) no longer decays harmonically in M, but *exponentially*. This has been studied in [7].

In another extension, the results of this paper can be applied to the CEO source coding problem, where, with reference to Figure 1, the M noisy observations $U_1, U_2, \ldots, U_M$ have to be encoded *separately*, see e.g. [8, 9, 10]. Such an application is outlined in [11].

ACKNOWLEDGEMENTS

Stimulating discussions with K. Eswaran (Berkeley) and Prof. Dr. M. Vetterli (EPFL) are gratefully acknowledged. The material in this paper was supported in part by the National Science Foundation under award CCF-0347298 (CAREER).

A. UPPER BOUND

For the sake of completeness, we also briefly outline a simple outer bound to the AWGN remote rate-distortion function. Specifically, we show that the case of a Gaussian source S serves as an upper bound: For any source S of variance σ_S^2, the rate needed to attain a distortion D is no larger than (3) (rewritten as a rate-distortion function).

To see this, it is convenient to rewrite the remote rate-distortion function in the following form:

$$R_{remote}(D) \quad = \quad \min I(U;\hat{S}) \tag{24}$$

where the minimization is over all distributions $p(\hat{s}|u)$ for which

$$E\left[\left|S-\hat{S}\right|^2\right] \quad \leq \quad D. \tag{25}$$

To obtain an upper bound, it suffices to select *some* conditional distribution $p(\hat{s}|u)$ for which $E[|S-\hat{S}|^2] \leq D$. Specifically, we select the conditional distribution that can be written as

$$\hat{S}_0 \quad = \quad \frac{\sigma_S^2}{M\sigma_S^2+\sigma_W^2+\sigma_Z^2/M}\left(\sum_{m=1}^{M} U_m + Z\right), \tag{26}$$

where Z is independent of all other random variables and is zero-mean Gaussian with variance σ_Z^2. Next, the variance

σ_Z^2 is selected such that $E[|S-\hat{S}|^2]=D$. The crux of the proof is that in the case where the underlying source is Gaussian, this $\hat{S}_0$ indeed achieves the remote rate-distortion function. To finish the proof, it suffices to write out

$$R_{remote}(D) \leq I(U;\hat{S}_0) = h(\hat{S}_0) - h(\hat{S}_0|U), \quad (27)$$

where $h(\hat{S}_0|U)$ is fixed, irrespective of the distribution of the underlying source S. For the term $h(\hat{S}_0)$, note that Equation (26) fixes the variance of $\hat{S}_0$, and hence, is maximized when $\hat{S}_0$ is Gaussian, which happens when the underlying source S is Gaussian.

B. PROOF OF THEOREM 1

Proof. Consider the remote rate-distortion function as in Equation (7). Let us first determine a lower bound to the first term. To this end, define

$$T_M = \frac{1}{\sqrt{M\sigma_W^2}} \sum_{m=1}^{M} U_m = \sqrt{\frac{M\sigma_S^2}{\sigma_W^2}}\tilde{S} + \tilde{W}, \quad (28)$$

where $\tilde{S}$ and $\tilde{W}$ are independent, $\tilde{S}$ is a renormalized version of S (with unit variance) and $\tilde{W}$ is zero-mean Gaussian with unit variance.

To see that T_M is a sufficient statistic for S given U, simply note that

$$p(s|u) = \frac{p(u|s)p(s)}{p(u)}, \quad (29)$$

where

$$\begin{aligned} p(u|s) &= \prod_{m=1}^{M} \frac{1}{\sqrt{2\pi}\sigma_W} \exp\left(-\frac{(u_m-s)^2}{2\sigma_W^2}\right) \\ &= \frac{1}{(2\pi\sigma_W^2)^{M/2}} \exp\left(-\sum_{m=1}^{M} \frac{u_m^2}{2\sigma_W^2}\right) \\ &\quad \cdot \exp\left(-\frac{Ms^2 - 2s\sum_{m=1}^{M} u_m}{2\sigma_W^2}\right) \end{aligned} \quad (30)$$

and hence,

$$\begin{aligned} p(u) &= \frac{1}{(2\pi\sigma_W^2)^{M/2}} \exp\left(-\sum_{m=1}^{M} \frac{u_m^2}{2\sigma_W^2}\right) \\ &\quad \cdot \int \exp\left(-\frac{M\tilde{s}^2 - 2\tilde{s}\sum_{m=1}^{M} u_m}{2\sigma_W^2}\right) p(\tilde{s})d\tilde{s}. \end{aligned}$$

Combining terms, we get

$$p(s|u) = \frac{\exp\left(-\frac{Ms^2-2s\sum_{m=1}^{M}u_m}{2\sigma_W^2}\right)p(s)}{\int \exp\left(-\frac{M\tilde{s}^2-2\tilde{s}\sum_{m=1}^{M}u_m}{2\sigma_W^2}\right)p(\tilde{s})d\tilde{s}}, \quad (31)$$

showing that the sum of the noisy observations is a sufficient statistic irrespective of the source distribution $p(s)$. Hence $E[S|U]=E[S|T_M]$. Furthermore,

$$E\left[|S-E[S|U]|^2\right] = \sigma_S^2 E\left[\left|\tilde{S}-E[\tilde{S}|T_M]\right|^2\right]. \quad (32)$$

There are various ways to proceed from here. In order to obtain the simple bound stated in the theorem, we exploit an elegant argument due to Guo *et al* [12, 13], noting that since the additive noise W in Equation (28) is Gaussian, it is true that

$$\begin{aligned} E\left[\left|\tilde{S}-E[\tilde{S}|T_M]\right|^2\right] &= 2\frac{d}{d\frac{M\sigma_S^2}{\sigma_W^2}} I(\tilde{S};T_M) \quad (33) \\ &= 2\frac{d}{d\frac{M\sigma_S^2}{\sigma_W^2}} \mathcal{H}(T_M), \quad (34) \end{aligned}$$

where we use $\mathcal{H}$ to denote the differential entropy with the natural logarithm, i.e., $\mathcal{H}(X) = -\int dx p_X(x)\ln p_X(x)$.

A simple lower bound to $\mathcal{H}(T_M)$ follows from the entropy power inequality, see e.g. [14],

$$e^{2\mathcal{H}(T_M)} \geq e^{2\mathcal{H}\left(\sqrt{\frac{M\sigma_S^2}{\sigma_W^2}}\tilde{S}\right)} + e^{2\mathcal{H}(\tilde{W})}. \quad (35)$$

In other words, the entropy $\mathcal{H}(T_M)$ is lower bounded by the following expression:

$$\mathcal{H}_{lower}(M) \stackrel{\text{def}}{=} \frac{1}{2}\ln\left(\frac{M\sigma_S^2}{\sigma_W^2}e^{2\mathcal{H}(\tilde{S})} + 2\pi e\right), \quad (36)$$

and we find moreover that

$$\begin{aligned} f_{lower}(M) &\stackrel{\text{def}}{=} 2\frac{d}{d\frac{M\sigma_S^2}{\sigma_W^2}} \mathcal{H}_{lower}(M) \\ &= \frac{1}{\frac{M\sigma_S^2}{\sigma_W^2} + \frac{2\pi e}{e^{2\mathcal{H}(\tilde{S})}}}. \end{aligned} \quad (37)$$

To complete the proof, we have to show that under the stated assumptions and for large enough M,

$$\mathcal{H}_{lower}(M) \leq \mathcal{H}(T_M) \quad (38)$$

implies that

$$f_{lower}(M) \leq 2\frac{d}{d\frac{M\sigma_S^2}{\sigma_W^2}} \mathcal{H}(T_M) \stackrel{\text{def}}{=} f(M). \quad (39)$$

By assumption, $1/f(M)$ is a convex function for all M for which $M\sigma_S^2/\sigma_W^2 \geq x_0$. But since $1/f_{lower}(M)$ is linear, there can be at most two intersections between $1/f(M)$ and $1/f_{lower}(M)$. Denote the largest M such that $1/f(M) = 1/f_{lower}(M)$ by M_0. If they do not intersect, set $M_0 = \lceil x_0 \sigma_W^2 / \sigma_S^2 \rceil$. For all $M \geq M_0$, we must have that $f_{lower}(M) \leq f(M)$. To see this, suppose by contradiction that $f_{lower}(M) > f(M)$ for all $M > M_0$. But then, there must be some M_1 such that for all $M \geq M_1$, $\mathcal{H}_{lower}(M) > \mathcal{H}(T_M)$, contradicting (35).

This implies that we can extend (34) to conclude that for all $M \geq M_0$,

$$E\left[\left|\tilde{S} - E[\tilde{S}|T_M]\right|^2\right] \quad \geq \quad f_{lower}(M). \qquad (40)$$

Changing the base of the logarithm back to 2 yields the claimed bound.

For the second term of Equation (7), we can use a Shannon lower bound. Note that $f(U) = E[S|U]$ is a sufficient statistic of S given U. Therefore,

$$\begin{aligned} &\min_{p(\hat{s}|u): I(U;\hat{S}) \leq R} E\left[|f(U) - \hat{S}|^2\right] \\ &= \min_{p(\hat{s}|f(u)): I(f(U);\hat{S}) \leq R} E\left[|f(U) - \hat{S}|^2\right]. \end{aligned} \qquad (41)$$

But the latter can be lower bounded by the standard Shannon lower bound, see e.g. [15, p.370], which says

$$h(E[S|U]) - \frac{1}{2}\log_2(2\pi e)D \quad \leq \quad R. \qquad (42)$$

This yields the claimed bound.

C. REFERENCES

[1] M. Gastpar and M. Vetterli, "Power, spatio-temporal bandwidth, and distortion in large sensor networks," *IEEE Journal on Selected Areas in Communications (Special Issue on Self-Organizing Distributive Collaborative Sensor Networks)*, vol. 23, no. 4, pp. 745–754, April 2005.

[2] R. L. Dobrushin and B. S. Tsybakov, "Information transmission with additional noise," *IRE Transactions on Information Theory*, vol. IT–18, pp. S293–S304, 1962.

[3] J. K. Wolf and J. Ziv, "Transmission of noisy information to a noisy receiver with minimum distortion," *IEEE Transactions on Information Theory*, vol. IT–16, pp. 406–411, July 1970.

[4] H. S. Witsenhausen, "Indirect rate distortion problems," *IEEE Transactions on Information Theory*, vol. IT–26, pp. 518 – 521, September 1980.

[5] T. Berger, *Rate Distortion Theory: A Mathematical Basis For Data Compression*, Prentice-Hall, Englewood Cliffs, NJ, 1971.

[6] M. Gastpar, "On remote sources and channels with feedback," in *Proc 38th Annual Conference on Information Sciences and Systems (CISS)*, Princeton, NJ, March 17-19 2004.

[7] K. Eswaran and M. Gastpar, "Rate loss in the CEO problem," in *Proc 39th Annual Conference on Information Sciences and Systems (CISS)*, Baltimore, MD, March 16-18 2005.

[8] T. Berger, Z. Zhang, and H. Viswanathan, "The CEO problem," *IEEE Transactions on Information Theory*, vol. IT–42, pp. 887–902, May 1996.

[9] Y. Oohama, "The rate-distortion function for the quadratic Gaussian CEO problem," *IEEE Transactions on Information Theory*, vol. IT–44, no. 3, pp. 1057–1070, May 1998.

[10] H. Viswanathan and T. Berger, "The quadratic Gaussian CEO problem," *IEEE Transactions on Information Theory*, vol. IT–43, no. 5, pp. 1549–1559, September 1997.

[11] K. Eswaran and M. Gastpar, "On the quadratic awgn ceo problem and non-gaussian sources," in *Proc IEEE Int Symp Info Theory*, Adelaide, Australia, September 4-9 2005.

[12] D. Guo, S. Shamai, and S. Verdú, "Mutual information and MMSE in Gaussian channels," in *Proc IEEE Int Symp Info Theory*, Chicago, IL, June 2004.

[13] D. Guo, S. Shamai, and S. Verdú, "Mutual information and minimum mean-square error in gaussian channels," *IEEE Transactions on Information Theory*, vol. IT–51, no. 4, pp. 1261–1282, April 2005.

[14] N. M. Blachman, "The convolution inequality for entropy powers," *IEEE Transactions on Information Theory*, vol. IT–11, pp. 267–271, April 1965.

[15] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, Wiley, New York, 1991.