2011 50th IEEE Conference on Decision and Control and
European Control Conference (CDC-ECC)
Orlando, FL, USA, December 12-15, 2011

On Optimal Policies for Control and Estimation Over a Gaussian Relay Channel

Ali A. Zaidi*, Serdar Yüksel†, Tobias J. Oechtering*, and Mikael Skoglund*

Abstract— **The problem of causal transmission of a memoryless Gaussian source over a two-hop memoryless Gaussian relay channel is considered. The source and the relay encoders have average transmit power constraints, and the performance criterion is mean squared distortion. The main contribution of this paper is to show that unlike in the case of a point-to-point channel, linear encoding schemes are not optimal over a two-hop relay channel in general, extending the sub-optimality results which are known for more than two hops. In some cases, simple three level quantization policies employed at the source and at the relay can outperform the best linear policies. Further a lower bound on the distortion is derived and it is shown that the distortion bounds derived using cut-set arguments are not tight in general for sensor networks.**

I. Introduction

Consider a physical phenomenon characterized by a sequence of independent and identically distributed real valued Gaussian random variables $\{X_n\}_{n\in\mathbb{Z}_+}$ having zero mean and variance σ_x^2, where n denotes a discrete time index. We wish to instantly communicate this physical phenomenon to a remote destination over a two-hop relay channel with as high fidelity as possible. The system model is illustrated in Fig. 1. According to the figure, at a discrete time $n \in \mathbb{Z}_+$ the source encoder $\mathcal{E}$ observes X_n and produces an output signal $S_{e,n} = f_{1,n}(\{X_i\}_{i=1}^n)$ suitable for transmission, where $f_{1,n} : \mathbb{R}^n \mapsto \mathbb{R}$ is a causal mapping. The encoder mapping $f_{1,n}$ has to satisfy the following average power constraint,

$$\mathbb{E}[S_{e,n}^2] \le P_S. \tag{1}$$

The transmitted signal $S_{e,n}$ is then observed in noise by the relay node $\mathcal{R}$ as $Y_n = S_{e,n} + Z_{r,n}$, where $\{Z_{r,n}\}_{n\in\mathbb{Z}_+}$ is a zero mean white Gaussian noise sequence of variance N_r. The relay node applies a causal mapping on the received signal $f_{2,n} : \mathbb{R}^n \mapsto \mathbb{R}$ to produce $S_{r,n} = f_{2,n}(\{Y_i\}_{i=1}^n)$ under the following average relay power constraint,

$$\mathbb{E}[S_{r,n}^2] \le P_R. \tag{2}$$

The signal $S_{r,n}$ is then transmitted over a Gaussian channel. Accordingly the destination node $\mathcal{D}$ receives $R_n = S_{r,n} + Z_{d,n}$, where $\{Z_{d,n}\}_{n\in\mathbb{Z}_+}$ is a zero mean white Gaussian noise sequence of variance N_d. Upon receiving R_n the decoder wishes to reconstruct the transmitted variable X_n by applying a mapping $g_n : \mathbb{R}^n \mapsto \mathbb{R}$ to produce $\hat{X}_n = g_n(\{R_i\}_{i=1}^n)$. The encoder, the relay, and the decoder are all causal and delay-free (zero delay). The objective is to choose the encoder, relay, and decoder mappings such that following distortion

$$D = \lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} \mathbb{E}[(X_n - \hat{X}_n)^2] \tag{3}$$

is minimized subject to the constraints in (1) and (2).

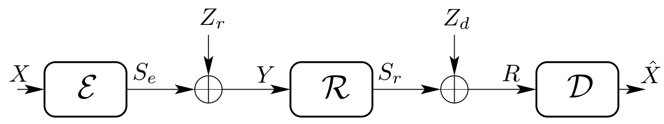

Fig. 1. A memoryless Gaussian source is transmitted in a causal fashion to a remote destination via a relay node. The destination wishes to reconstruct the source with minimum expected squared-error and zero-delay.

It is well-known that linear encoding is optimal for transmission of a Gaussian source over a point-to-point Gaussian channel when the distortion measure is mean squared error [1, 2]. From [3–6] we know that linear policies are also optimal if the encoder observes a noisy version of a Gaussian source. Moreover in [7] Gastpar has shown that linear (uncoded) scheme is even optimal in a simple Gaussian sensor network setting where each sensor node observes a noisy version of a Gaussian source and all the sensor nodes simultaneously transmit over a multiple-access Gaussian channel.

Lipsa and Martins studied a multi-stage decision (encoding) problem in [8, 9] and provided counter-examples (based on functions whose output can take on only two values) to show that linear policies are not optimal when the number of stages are sufficiently large. However their counter-example does not hold when the number of stages is either three or four. Therefore we highlight some unanswered questions pertaining to the transmission of a Gaussian source over the two-hop relay channel under discussion: i) Are linear policies optimal? ii) If not, then under what circumstances linear scheme can be optimal and what is the greatest lower bound on the distortion? iii) What are the optimal policies? In this paper we address the first two questions and demonstrate that non-linear policies based on simple three-level quantizer functions can beat the best linear policies in some cases, thus the linear encoding is not optimal in general. Moreover we discuss that linear encoding policies are person-by-person optimal, however they do not guarantee global optimality

This work was supported in part by the European Commission through the FP7 project FeedNetBack (co-design for networked control systems).

*Ali A. Zaidi, Tobias J. Oechtering, and Mikael Skoglund are with the School of Electrical Engineering and the ACCESS Linnaeus Center, Royal Institute of Technology (KTH), Stockholm, Sweden. {zaidi, oech, skoglund}@kth.se

†Serdar Yüksel is with the Mathematics and Engineering Program, Department of Mathematics and Statistics, Queen's University, Kingston, Ontario, Canada. yuksel@mast.queensu.ca

978-1-61284-799-3/11/$26.00 ©2011 IEEE

as the given team problem is non-convex in the encoding policies. We also derive a lower bound on distortion which is however not tight in general. The final question in the direction of finding the optimal policies is challenging and will be a subject of our future work.

This is a team decision problem under non-classical information structure [10]. One popular example of such problems is the well-known Witsenhausen's counterexample, which looks deceptively simple and remains unsolved till today [11].

The problem of causal transmission transmission over a two-hop relay channel is motivated by control applications, where the sensor measurements of a dynamical system are transmitted via a relay node to a remote decoder which has to control the system in real time. Control of a linear time invariant system over various types of relay channels has been been studied in [12–14], where sub-optimal linear schemes are used to derive conditions on mean-square stability. In [15], memoryless non-linear relay mappings are shown to outperform linear mappings for instantaneous transmission of a Gaussian source over an orthogonal three-node relay channel. A similar observation has been made in the control context in [16] for two parallel Gaussian relay channels. However the problem we are studying in this paper is fundamentally different from the problems addressed in [15, 16] due to the absence of a direct link (or parallel channels) from the source to the decoder.

II. Distortion Lower Bound

We derive a lower bound on the distortion using Bansal and Başar's approach [17] and the data processing inequality [18]. Consider the following series of inequalities:

$$\begin{aligned}
\frac{1}{N}\sum_{n=1}^{N} I(X_n;\hat{X}_n) &\overset{(a)}{=} \frac{1}{N}\left(\sum_{n=1}^{N} H(X_n) - \sum_{n=1}^{N} H(X_n|\hat{X}_n)\right) \\
&\overset{(b)}{=} \frac{1}{N}\left(H(X^N) - \sum_{n=1}^{N} H(X_n|\hat{X}_n)\right) \\
&\overset{(c)}{\le} \frac{1}{N}\left(H(X^N) - \sum_{n=1}^{N} H(X_n|\hat{X}^N, X^{n-1})\right) \\
&= \frac{1}{N}\Big(H(X^N) - H(X^N|\hat{X}^N)\Big) = \frac{1}{N} I(X^N;\hat{X}^N) \\
&\overset{(d)}{\le} \frac{1}{N} I(S_e^N;R^N) \overset{(e)}{\le} \frac{1}{N}\min\{I(S_e^N;Y^N), I(S_r^N;R^N)\} \\
&\overset{(f)}{\le} \frac{1}{N}\min\left\{\sum_{n=1}^{N} I(S_{e,n};Y_n), \sum_{n=1}^{N} I(S_{r,n};R_n)\right\} \\
&\overset{(g)}{\le} \frac{1}{2}\min\left\{\log\left(1+\frac{P_S}{N_r}\right), \log\left(1+\frac{P_R}{N_d}\right)\right\}, \qquad (4)
\end{aligned}$$

where (a) follows from the definition of mutual information; (b) follows from independence of the sequence $\{X_n\}_{n=1}^N$ and by defining $X^N \triangleq \{X_n\}_{n=1}^N$; (c) follows from conditioning reduces entropy; (d) and (e) follow from the data processing inequality with Markov chain [18]; (f) follows from the fact that the channels are memoryless and conditioning reduces entropy; and (g) follows from the fact that mutual information is maximized by Gaussian distribution. Further consider the following inequalities:

$$\begin{aligned}
\frac{1}{N}\sum_{n=1}^{N} I(X_n;\hat{X}_n) &\overset{(a)}{\ge} \frac{1}{2N}\sum_{n=1}^{N}\log\left(\frac{\sigma_x^2}{\mathbb{E}[(X_n-\hat{X}_n)^2]}\right) \\
&= \frac{1}{2N}\sum_{n=1}^{N}\log\left(\sigma_x^2\right) - \frac{1}{2N}\sum_{n=1}^{N}\log\left(\mathbb{E}[(X_n-\hat{X}_n)^2]\right) \\
&\overset{(b)}{\ge} \frac{1}{2}\log\left(\sigma_x^2\right) - \frac{1}{2}\log\left(\frac{1}{N}\sum_{n=1}^{N}\mathbb{E}\left[(X_n-\hat{X}_n)^2\right]\right), \qquad (5)
\end{aligned}$$

where (a) follows from the rate distortion theorem for an i.i.d. Gaussian source [18]; and (b) follows from the concavity of the logarithm function. Now from (4) and (5), we obtain the following lower bound on the distortion by simple algebraic manipulation.

$$\begin{aligned}
D &= \lim_{N\to\infty}\frac{1}{N}\sum_{n=1}^{N}\mathbb{E}\left[(X_n-\hat{X}_n)^2\right] \\
&\ge \sigma_x^2\max\left\{\frac{N_r}{P_S+N_r}, \frac{N_d}{P_R+N_d}\right\}. \qquad (6)
\end{aligned}$$

Remark 1: It is observed in [19, Theorem 3.5] that due to the presence of the two channel noise components (Z_r and Z_d), we have $I(S_e;R) < \min\{I(S_e;Y), I(S_r;R)\}$. Therefore the bound in (6) is not tight. However it becomes tight when variance of any of the two channel noise components approaches zero. In [19, Theorem 3.5] the authors discussed that $I(S_e;R)$ is strictly lower than the capacity of a two-hop relay channel which follows from block coding arguments and cut-set bound. This tells us that the distortion bounds obtained using cut set arguments are not tight in general for sensor networks due to the zero-delay reconstruction.

III. Linear Policies

In this section we find the optimal linear encoding policies and the distortion obtained under these policies. Since the source is memoryless and the encoders are causal, the optimal encoders are memoryless [20]. This can be easily verified by showing that if we transmit a linear combination of the current and the previous source observations, then the previous observations will only contribute to noise as the source is memoryless. We therefore restrict our study to memoryless linear policies, in the sense that the encoders merely transmit a scaled version of the received signal. That is, the source and the relay encoders transmit the following:

$$S_{e,n} = \sqrt{\frac{a_n}{\sigma_x^2}}X_n, \quad S_{r,n} = \sqrt{\frac{b_n}{a_n+N_r}}Y_n,$$

where $a_n, b_n \in R_+$ are time varying gain coefficients which are chosen such that the transmit power constraints in (1) and (2) are satisfied, i.e., $a_n \le P_S$ and $b_n \le P_R$. The decoder accordingly receives

$$R_n = \sqrt{\frac{a_n b_n}{\sigma_x^2(a_n+N_r)}}X_n + \sqrt{\frac{b_n}{a_n+N_r}}Z_{r,n} + Z_{d,n},$$

and computes the minimum mean squared-error (MMSE) estimate according to $\hat{X}_n = \mathbb{E}[X_n|R^n] = \mathbb{E}[X_n|R_n]$, where we have used the notation $R^n = \{R_i\}_{i=1}^n$ and the fact that the $\{R_n, R_{n-k}\}$ are mutually independent for all $k \neq n$. Since X_n is Gaussian, the distortion per time instant follows from a straightforward computation [21],

$$\mathbb{E}[(X_n - \hat{X}_n)^2] = \sigma_x^2 \left(1 - \frac{a_n b_n}{(a_n + N_r)(b_n + N_d)}\right)$$

which leads to

$$D_L = \lim_{N\to\infty} \frac{\sigma_x^2}{N} \sum_{n=1}^{N} \left(1 - \frac{a_n b_n}{(a_n + N_r)(b_n + N_d)}\right) \quad (7)$$

The optimal choice of the gain coefficients $0 < a_n \leq P_S$, $0 < b_n \leq P_R$, which minimizes (7) is $\{a_n^\star = P_S, b_n^\star = P_R\}$. This choice of the gain coefficients leads to the following lowest distortion that is obtained under the best linear encoding scheme.

$$D_L^\star = \sigma_x^2 \left(1 - \frac{P_S P_R}{(P_S + N_r)(P_R + N_d)}\right). \quad (8)$$

We have so far found a strict lower bound on distortion in (6) and an upper bound in (8) using the best linear scheme. However we still do not know how good linear policies are and under what circumstances they are optimal? In the following we show that the linear policies are person-by-personal optimal, however they do not guarantee team optimality.

Person-by-person Optimality of Linear Policies and Concavity of the Team Problem

Let us fix the source encoder to be linear. Given a linear and memoryless policy at the source encoder, we now find an optimal relaying policy which minimizes the per time instant distortion $\mathbb{E}[(X_n - E[X_n|R^n])^2]$, where $R^n \triangleq \{R_i\}_{i=1}^n$. We can rewrite the per stage distortion as

$$\begin{aligned}&\mathbb{E}[(X_n - E[X_n|R^n])^2] \overset{(a)}{=} \mathbb{E}\left[(X_n - E[X_n|Y^n])^2\right] \\ &+ \mathbb{E}\left[(E[X_n|Y^n] - E[X_n|R^n])^2\right] \\ &\overset{(b)}{=} \mathbb{E}\left[(X_n - c_n Y_n)^2\right] + \mathbb{E}\left[(c_n Y_n - E[X_n|R^n])^2\right], \quad (9)\end{aligned}$$

where (a) follows from $\mathbb{E}[(X_n - E[X_n|Y^n])(E[X_n|Y^n] - E[X_n|R^n])] = 0$ (by the orthogonality principle of MMSE estimation); and (b) follows from the fact that the source encoder is linear and memoryless and the MMSE estimation of a Gaussian variable is linear, i.e. $E[X_n|Y^n] = c_n Y_n$, where c_n is a scalar. According to (9), an optimal relaying policy is the one which minimizes $E[(c_n Y_n - E[X_n|R^n])^2]$, since the remaining term in the per time instant distortion function is independent of the relaying policy. This problem was studied by Bansal and Başar in [6], from which it follows that an optimal relay encoding policy is linear and memoryless if we fix the source encoder to be linear memoryless. This observation can also be made from [3–5, 7, 22, 23]. Now if we fix the relay encoder policy to be linear and memoryless, one can observe that the problem becomes equivalent to the transmission of a Gaussian source over a point to point Gaussian channel subject to an average power constraint, for which it is well-known that linear (memoryless) encoding is optimal in the sense of minimizing mean squared distortion [1, 2]. Hence if we fix either the source encoder or the relay encoder to be linear, then the greatest lower bound on the distortion is given by (8). That is linear policies are person-by-person optimal.

We know that in a decentralized team optimization problem person-by-person optimal solutions are globally optimal if the cost function is convex in the policies of the decision makers and the cost function satisfies differentiability conditions in the policies [24]. Let us now investigate convexity of the distortion given in (3). Let P be an observation channel from the input variable X at source encoder to the channel output variable R such that $P(\cdot|x)$ is a probability measure on the Borel σ-algebra $\mathcal{B}(\mathbb{R})$ on $\mathbb{R}$ for every $x \in \mathbb{R}$, and $P(A|\cdot) : \mathbb{R} \mapsto [0 : 1]$ is a Borel measurable function for every $A \in \mathcal{B}(\mathbb{R})$. Similarly we define P_1 as an observation channel from the variable X to the variable Y, and P_2 as an observation channel from the variable Y to the variable R. From [25, Theorem 4.1] it follows that the distortion in (3) is concave in the joint observation channel $P(A|x) = \int_{\mathbb{R}} P_2(A|y)P_1(dy|x)$ for every $A \in \mathcal{B}(\mathbb{R})$, where the individual channels P_1 and P_2 are induced by the source and the relay encoding policies. Thus the distortion in (3) is non-convex in the encoding policies. This implies that the person-by-person optimal encoding policies do not guarantee team optimality.

IV. Counter Example: Non-linear Policies

In this section we provide a simple counter example to show that linear policies are not optimal for causal transmission of a Gaussian source over the given two-hop relay channel. Consider the following time invariant policies at the source encoder and the relay encoder respectively:

$$f_1(x) = \left\{ \begin{array}{cc} a, & \text{for} \quad x > m_1 \\ 0, & \text{for} \quad |x| \leq m_1 \\ -a, & \text{for} \quad x < -m_1 \end{array} \right\}, \quad (10)$$

$$f_2(y) = \left\{ \begin{array}{cc} b, & \text{for} \quad y > m_2 \\ 0, & \text{for} \quad |y| \leq m_2 \\ -b, & \text{for} \quad y < -m_2 \end{array} \right\}, \quad (11)$$

where the scalars $a, b \in \mathbb{R}_+$. In (10) and (11) we have dropped the index n for the sake of simplicity without any loss as we are considering time invariant policies. According to these policies, the signals observed at the relay and the destination are respectively given by

$$Y = \left\{ \begin{array}{cc} a + Z_r, & \text{for} \quad X > m_1 \\ Z_r, & \text{for} \quad |X| \leq m_1 \\ -a + Z_r, & \text{for} \quad X < -m_1 \end{array} \right\}, \quad (12)$$

$$R = \left\{ \begin{array}{cc} b + Z_d, & \text{for} \quad Y > m_2 \\ Z_d, & \text{for} \quad |Y| \leq m_2 \\ -b + Z_d, & \text{for} \quad Y < -m_2 \end{array} \right\}. \quad (13)$$

$$\hat{X} = g(r) = \mathbb{E}[X|R=r] = \frac{1}{p(r)}\left(l_3(r) - l_1(r)\right)\sqrt{\frac{\sigma_x^2}{2\pi}}\exp\left(-\frac{{m_1}^2}{2\sigma_x^2}\right), \tag{15}$$

$$D_{NL} := \mathbb{E}\left[\left(X-\hat{X}\right)^2\right] = \int_{-\infty}^{\infty}\left(l_1(r)\int_{-\infty}^{-m_1}(x-g(r))^2 p(x)dx + l_2(r)\int_{-m_1}^{m_1}(x-g(r))^2 p(x)dx + l_3(r)\int_{m_1}^{\infty}(x-g(r))^2 p(x)dx\right)dr \tag{16}$$

where $p(r) = \left(l_1(r)+l_3(r)\right)Q\left(\frac{m_1}{\sigma_x}\right) + l_2(r)\left(1-2Q\left(\frac{m_1}{\sigma_x}\right)\right)$,

$$l_1(r) = \frac{1}{\sqrt{2\pi N_d}}\left[e^{\frac{-(r+b)^2}{2N_d}}Q\left(\frac{m_2-a}{\sqrt{N_r}}\right) + e^{\frac{-r^2}{2N_d}}\left(1-Q\left(\frac{m_2-a}{\sqrt{N_r}}\right)-Q\left(\frac{m_2+a}{\sqrt{N_r}}\right)\right) + e^{\frac{-(r-b)^2}{2N_d}}Q\left(\frac{m_2+a}{\sqrt{N_r}}\right)\right],$$

$$l_2(r) = \frac{1}{\sqrt{2\pi N_d}}\left[e^{\frac{-r^2}{2N_d}} + \left(e^{\frac{-(r+b)^2}{2N_d}} + e^{\frac{-(r-b)^2}{2N_d}} - 2e^{\frac{-r^2}{2N_d}}\right)Q\left(\frac{m_2}{\sqrt{N_r}}\right)\right],$$

$$l_3(r) = \frac{1}{\sqrt{2\pi N_d}}\left[e^{\frac{-(r+b)^2}{2N_d}}Q\left(\frac{m_2+a}{\sqrt{N_r}}\right) + e^{\frac{-r^2}{2N_d}}\left(1-Q\left(\frac{m_2-a}{\sqrt{N_r}}\right)-Q\left(\frac{m_2+a}{\sqrt{N_r}}\right)\right) + e^{\frac{-(r-b)^2}{2N_d}}Q\left(\frac{m_2-a}{\sqrt{N_r}}\right)\right],$$

The non-linear policies in (10) and (11) have to satisfy the average transmit power constraints. In Appendix I we have obtained conditions on $a, b \in \mathbb{R}_+$ to ensure the power constraints in (1) and (2) are satisfied. These conditions are:

$$a \leq \sqrt{\frac{P_S}{2Q\left(\frac{m_1}{\sigma_x}\right)}}, b \leq \sqrt{\frac{P_R}{2\kappa(m_1, m_2, a, \sigma_x, N_r)}}, \tag{14}$$

where $\kappa(m_1, m_2, a, \sigma_x, N_r) = \left(1-2Q\left(\frac{m_1}{\sigma_x}\right)\right)Q\left(\frac{m_2}{\sqrt{N_r}}\right) + Q\left(\frac{m_1}{\sigma_x}\right)\left(Q\left(\frac{m_2-a}{\sqrt{N_r}}\right)+Q\left(\frac{m_2+a}{\sqrt{N_r}}\right)\right)$ and $Q(x) \triangleq \frac{1}{\sqrt{2\pi}}\int_x^{\infty} e^{-\frac{\tau^2}{2}}d\tau$. For these non-linear encoding policies, the expressions for the MMSE decoder $g(R)$ and the corresponding distortion D_{NL} are derived in Appendix II and are reproduced in (15) and (16). The distortion D_{NL} can be computed numerically using (15), (16), and (14) for any fixed values of the system parameters $\{\sigma_x^2, P_S, P_R, N_d, N_r, m_1, m_2\}$. We now give two examples to demonstrate that the proposed simple non-linear scheme can outperform the best linear scheme. In the following examples we fix the values of the system parameters and then numerically compute the distortion for non-linear and linear policies according to (16) and (8) respectively. We also evaluate the lower bound in (6), however the reader should keep in mind that the bound is not tight in general as we have used cut-set arguments for deriving this bound.

Example 1: Fixing $\sigma_x^2 = P_S = P_R = 1$, $N_r = N_d = 4$, $m_1 = 2.45$, and $m_2 = 6.84$, we get: $D_{NL} = 0.926$, $D_L^\star = 0.96$, and $D = 0.8$.

Example 2: Fixing $\sigma_x^2 = P_S = P_R = 1$, $N_r = N_d = 10$, $m_1 = 2.85$, and $m_2 = 12.05$, we get: $D_{NL} = 0.964$, $D_L^\star = 0.992$, and $D = 0.909$.

The above examples validate the fact that linear policies are not optimal in general for the given two-hop relay channel when the source and the relay node have individual power constraints. Let us now consider a total power constraint on the source and the relay, i.e., $\mathbb{E}[S_{e,n}^2] + \mathbb{E}[S_{r,n}^2] = P$. According to Appendix III, the distortion is minimized for the linear policies by an equal power allocation $\mathbb{E}[S_{e,n}^2] = \mathbb{E}[S_{r,n}^2] = \frac{P}{2}$ if the two channel noises have equal variance i.e., $N_r = N_d$. In the above two counter-examples we have only considered the cases with equal source and relay transmit powers and equal noise variances, thus the linear policies are also not optimal when a total transmit power constraint is imposed on source and relay.

V. Conclusion

We studied the problem of mean square estimation of a Gaussian source over a two-hop Gaussian relay channel with average source and relay transmit power constraints. A strict lower bound on mean square distortion was derived. We observed that the distortion bounds obtained using cut-set arguments are not tight in general for sensor networks due to the zero-delay nature of the problem. Further it was shown that linear policies are person-by-person optimal for causal transmission and estimation of a Gaussian source over the given two-hop relay channel. However person-by-person optimality of the linear policies do not guarantee global optimality due to concavity property of the distortion function in the observation channel. A simple three level function was shown to outperform the best linear scheme in some cases, thus validating the fact that linear policies are not optimal in general. This observation is in accordance with the already known results for non-classical information structures [10]. We wish to identify necessary and sufficient conditions for optimal schemes for this problem using variational methods in future work. Some recent related results on functional properties of MMSE can be found in [26].

References

[1] T. Berger, *Rate Distortion Theory: A mathematical basis for data compression.* Englewood Cliffs, NJ: Prentice-Hall, 1971.

[2] M. Gastpar, B. Rimoldi, and M. Vetterli, "To code or not to code," in *IEEE ISIT*, 2000, p. 236.

[3] E. Ayanoglu and R. M. Gray, "The design of joint source and channel trellis waveform coders," *IEEE Trans. Inform. Theory*, vol. 33, no. 6, pp. 855–865, 1987.

[4] R. L. Dobrushin and B. S. Tsybakov, "Information transmission with additional nosie," *IRE Trans. Inf. Theory*, vol. 18, pp. 293–304, 1962.

[5] T. Fine, "Optimum mean-square quantization of a noisy input," *IEEE Trans. Inf. Theory*, vol. 11, pp. 293–294, April 1965.

[6] R. Bansal and T. Başar, "Solutions to a class of linear-quadratic-gaussian LQG stochastic team problems with nonclassical information," *Systems Control Letters*, vol. 9, no. 2, pp. 125–130, 1987.

[7] M. Gastpar, "Uncoded transmission is exactly optimal for a simple Gaussian sensor network," *IEEE Trans. Inf. Theory*, vol. 54, pp. 5247–5251, 2008.

[8] G. M. Lipsa and N. C. Martins, "Optimal memoryless control in Gaussian noise: A simple counterexample," *Automatica*, 2011, doi:10.1016/j.automatica.2010.12.001.

[9] ——, "Finite horizon optimal memoryless control of a delay in Gaussian noise: A simple counterexample." in *IEEE CDC*, December 2008, pp. 1628–1635.

[10] Y. C. Ho, "Team decision theory and information structures," *Proceedings of the IEEE*, vol. 68, no. 6, pp. 644–654, 1980.

[11] H. S. Witsenhausen, "A counterexample in stochastic optimum control," *SIAM J. Control*, vol. 6, pp. 131–147, 1968.

[12] A. A. Zaidi, T. J. Oechtering, and M. Skoglund, "Rate sufficient conditions for closed–loop control over AWGN relay channels," in *IEEE ICCA*, June 2010, pp. 602–607.

[13] A. A. Zaidi, T. J. Oechtering, S. Yüksel, and M. Skoglund, "Sufficient conditions for closed-loop control over a general half-duplex white Gaussian relay channel," in *IEEE ACC*, 2011, To appear.

[14] U. Kumar, V. Gupta, and J. N. Laneman, "Sufficient conditions for stabilizability over Gaussian relay channel and cascade channels," in *IEEE CDC*, December 2010, pp. 4765–4770.

[15] J. Karlsson and M. Skoglund, "Optimized low-delay source-channel-relay mapping," *IEEE Trans. Commun.*, vol. 58, no. 5, pp. 1397–1404, 2010.

[16] S. Yüksel and S. Tatikonda, "A counterexample in distributed optimal sensing and control," *IEEE Trans. Automat. Control*, vol. 54, no. 4, 2009.

[17] R. Bansal and T. Başar, "Simultaneous design of measurement and control strategies for stochastic systems with feedback," *Automatica*, vol. 25, no. 5, pp. 679–694, 1989.

[18] T. Cover and J. Thomas, *Elements of information theory*. John Wiley Sons, Inc., 2006.

[19] S. Yüksel and S. Tatikonda, "Distributed sensing and control over communication channels and a counterexample," in *Allerton Conference on Communication, Control, and Computing*, 2007, pp. 260–267.

[20] J. Walrand and P. Varaiya, "Optimal causal coding-decoding problems," *IEEE Trans. Inf. Theory*, vol. 29, pp. 814–820, 1983.

[21] M. Hayes, *Statistical digital signal processing and modelling*. John Wiley Sons, Inc., 1996.

[22] K. S. Gomadam and S. A. Jafar, "Optimal relay functionality for SNR maximization in memoryless relay networks," *IEEE Journal on Selected Areas in Communications 25*, vol. 25, no. 2, pp. 390–401, 2007.

[23] S. Yüksel, "On optimal causal coding of partially observed markov sources under classical and non-classical information structures," in *IEEE ISIT*, 2010, pp. 81–85.

[24] R. Radner, "Team decision problems," *Annals of Mathematical Statistics*, vol. 33, pp. 857–881, 1962.

[25] S. Yüksel and T. Linder, "Optimization and convergence of observation channels in stochastic control," *SIAM J. Control Optim.*, Submitted 2010. [Online]. Available: http://arxiv.org/abs/1009.3824

[26] Y. Wu and S. Verdú, "Functional properties of MMSE," in *IEEE ISIT*, 2010, pp. 1453 – 1457.

APPENDIX I
TRANSMIT POWER CONSTRAINTS

The parameter $a \in \mathbb{R}_+$ of the source mapping $f_1(.)$ in (10) is chosen such that

$$P_S \geq \mathbb{E}\left[S_e^2\right] = \mathbb{E}\left[f_1^2(X)\right] = \int_{-\infty}^{\infty} f_1^2(x)p(x)dx$$
$$= 2a^2 Q\left(\frac{m_1}{\sigma_x}\right) \Rightarrow a \leq \sqrt{\frac{P_S}{2Q\left(\frac{m_1}{\sigma_x}\right)}}, \tag{17}$$

which follows from (10), $p(x) = \frac{1}{\sqrt{2\pi\sigma_x^2}} e^{-\frac{x^2}{2\sigma_x^2}}$, $Q(x) \triangleq \frac{1}{\sqrt{2\pi}}\int_x^\infty e^{-\frac{\tau^2}{2}}d\tau$, and $Q(x) = Q(-x)$. From (12) we have

$$p(y|x) = \left\{ \begin{array}{ll} \frac{1}{\sqrt{2\pi N_r}} e^{-\frac{(y-a)^2}{2N_r}}, & \text{if} \quad x > m_1 \\ \frac{1}{\sqrt{2\pi N_r}} e^{-\frac{y^2}{2N_r}}, & \text{if} \quad |x| \leq m_1 \\ \frac{1}{\sqrt{2\pi N_r}} e^{-\frac{(y+a)^2}{2N_r}}, & \text{if} \quad x < -m_1 \end{array} \right\}. \tag{18}$$

The marginal pdf $p(y)$ can now be computed as

$$p(y) = \int_{\mathbb{R}} p(y|x)p(x)dx = \frac{1}{\sqrt{2\pi N_r}}\left[\left(e^{\frac{-(y+a)^2}{2N_r}} + e^{\frac{-(y-a)^2}{2N_r}}\right)\right.$$
$$\left. \times Q\left(\frac{m_1}{\sigma_x}\right) + e^{\frac{-y^2}{2N_r}}\left(1 - 2Q\left(\frac{m_1}{\sigma_x}\right)\right)\right], \tag{19}$$

The condition on the parameter b which ensures the average transmit power constraint at the relay node is obtained as

$$P_R \geq \mathbb{E}\left[S_r^2\right] = \mathbb{E}\left[f_2^2(Y)\right] = \int_{-\infty}^{\infty} f_2^2(y)p(y)dy$$
$$= 2b^2\left[Q\left(\frac{m_1}{\sigma_x}\right)\left(Q\left(\frac{m_2 - a}{\sqrt{N_r}}\right) + Q\left(\frac{m_2 + a}{\sqrt{N_r}}\right)\right) + \right.$$
$$\left.\left(1 - 2Q\left(\frac{m_1}{\sigma_x}\right)\right)Q\left(\frac{m_2}{\sqrt{N_r}}\right)\right] =: 2b^2\kappa(m_1, m_2, a, \sigma_x, N_r)$$
$$\Rightarrow b \leq \sqrt{\frac{P_R}{2\kappa(m_1, m_2, a, \sigma_x, N_r)}}, \tag{20}$$

which follows from (11),(19), and by defining $\kappa(m_1, m_2, a, \sigma_x, N_r)$.

APPENDIX II
DISTORTION CALCULATION

We first find the joint pdf $p(x, r)$ and the marginal pdf $p(r)$ in order to compute the MMSE estimator. Since we have the following Markov chain $X \to Y \to R$,

$$p(x, r) = \int_{\mathbb{R}} p(r|y)p(y|x)p(x)dy, \tag{21}$$

where $p(y|x)$ is given in (18), and from (13) we have

$$p(r|y) = \left\{ \begin{array}{ll} \frac{1}{\sqrt{2\pi N_d}} e^{-\frac{(r-b)^2}{2N_d}}, & \text{if} \quad y > m_2 \\ \frac{1}{\sqrt{2\pi N_d}} e^{-\frac{r^2}{2N_d}}, & \text{if} \quad |y| \leq m_2 \\ \frac{1}{\sqrt{2\pi N_d}} e^{-\frac{(r+b)^2}{2N_d}}, & \text{if} \quad y < -m_2 \end{array} \right\}. \tag{22}$$

For the interval $x < -m_1$, $p(x,r)$ is computed as follows:

$$p(x,r) = p(x)\int_{\mathbb{R}} p(r|y)p(y|x)dy$$
$$\overset{(a)}{=} p(x)\left[\frac{1}{\sqrt{2\pi N_r}}\left\{\int_{-\infty}^{-m_2} p(r|y)e^{\frac{-(y+a)^2}{2N_r}}dy\right.\right.$$
$$\left.\left.+\int_{-m_2}^{m_2} p(r|y)e^{\frac{-(y+a)^2}{2N_r}}dy + \int_{m_2}^{\infty} p(r|y)e^{\frac{-(y+a)^2}{2N_r}}dy\right\}\right]$$
$$\overset{(b)}{=} \frac{p(x)}{\sqrt{2\pi N_d}}\left[\frac{e^{\frac{-(r+b)^2}{2N_d}}}{\sqrt{2\pi N_r}}\int_{-\infty}^{-m_2} e^{\frac{-(y+a)^2}{2N_r}}dy\right.$$
$$\left.+\frac{e^{\frac{-r^2}{2N_d}}}{\sqrt{2\pi N_r}}\int_{-m_2}^{m_2} e^{\frac{-(y+a)^2}{2N_r}}dy + \frac{e^{\frac{-(r-b)^2}{2N_d}}}{\sqrt{2\pi N_r}}\int_{m_2}^{\infty} e^{\frac{-(y+a)^2}{2N_r}}dy\right]$$
$$\overset{(c)}{=} \frac{p(x)}{\sqrt{2\pi N_d}}\left[e^{\frac{-(r+b)^2}{2N_d}}Q\left(\frac{m_2-a}{\sqrt{N_r}}\right)+e^{\frac{-(r-b)^2}{2N_d}}Q\left(\frac{m_2+a}{\sqrt{N_r}}\right)\right.$$
$$\left.+e^{\frac{-r^2}{2N_d}}\left\{1-Q\left(\frac{m_2-a}{\sqrt{N_r}}\right)-Q\left(\frac{m_2+a}{\sqrt{N_r}}\right)\right\}\right]\overset{(d)}{=:}p(x)l_1(r) \quad (23)$$

where (a) follows from (18); (b) follows from (22); (c) follows from the definition of $Q(.)$; and (d) follows by defining $l_1(r)$. By following the same steps as above, the joint pdf $p(x,r)$ for $|x| \le m_1$ is given by

$$p(x,r) = \frac{p(x)}{\sqrt{2\pi N_d}}\left[e^{\frac{-r^2}{2N_d}}+\right.$$
$$\left.+\left\{e^{\frac{-(r+b)^2}{2N_d}}+e^{\frac{-(r-b)^2}{2N_d}}-2e^{\frac{-r^2}{2N_d}}\right\}Q\left(\frac{m_2}{\sqrt{N_r}}\right)\right]$$
$$=: p(x)l_2(r), \quad (24)$$

and for $x > m_1$,

$$p(x,r) = \frac{p(x)}{\sqrt{2\pi N_d}}\left[e^{\frac{-(r+b)^2}{2N_d}}Q\left(\frac{m_2+a}{\sqrt{N_r}}\right)\right.$$
$$+e^{\frac{-r^2}{2N_d}}\left\{1-Q\left(\frac{m_2-a}{\sqrt{N_r}}\right)-Q\left(\frac{m_2+a}{\sqrt{N_r}}\right)\right\}$$
$$\left.+e^{\frac{-(r-b)^2}{2N_d}}Q\left(\frac{m_2-a}{\sqrt{N_r}}\right)\right] =: p(x)l_3(r). \quad (25)$$

From (23), (24), and (25), we compute the following marginal pdf,

$$p(r) = \int_{\mathbb{R}} p(x,r)dx$$
$$= \int_{-\infty}^{-m_1} p(x,r)dx + \int_{-m_1}^{m_1} p(x,r)dx + \int_{m_1}^{\infty} p(x,r)dx$$
$$= l_1(r)\int_{-\infty}^{-m_1} p(x)dx + l_2(r)\int_{-m_1}^{m_1} p(x)dx$$
$$+ l_3(r)\int_{m_1}^{\infty} p(x)dx$$
$$= (l_1(r)+l_3(r))\,Q\left(\frac{m_1}{\sigma_x}\right) + l_2(r)\left(1-2Q\left(\frac{m_1}{\sigma_x}\right)\right). \quad (26)$$

The MMSE estimator can now be computed using (23), (24), (25), and (26) as follows.

$$\mathbb{E}[X|R=r] = \int_{\mathbb{R}} xp(x|r)dx = \frac{1}{p(r)}\int_{\mathbb{R}} xp(x,r)dx$$
$$= \frac{1}{p(r)}\left(\int_{-\infty}^{-m_1} xp(x,r)dx + \int_{-m_1}^{m_1} xp(x,r)dx\right.$$
$$\left.+\int_{m_1}^{\infty} xp(x,r)dx\right) = \frac{1}{p(r)}\left(l_1(r)\int_{-\infty}^{-m_1} xp(x)dx\right.$$
$$\left.+ l_2(r)\int_{-m_1}^{m_1} xp(x)dx + l_3(r)\int_{m_1}^{\infty} xp(x)dx\right)$$
$$\overset{(a)}{=} \frac{1}{p(r)}\left(l_3(r)-l_1(r)\right)\int_{m_1}^{\infty} xp(x)dx$$
$$= \frac{1}{p(r)}\left(l_3(r)-l_1(r)\right)\sqrt{\frac{\sigma_x^2}{2\pi}}\exp\left(-\frac{{m_1}^2}{2\sigma_x^2}\right) =: g(r), \quad (27)$$

where (a) follows from $\int_{-m_1}^{m_1} xp(x)dx = 0$. The associated mean squared error is given by

$$\mathbb{E}[(X-\mathbb{E}[X|R])^2] = \int_{\mathbb{R}^2} (x-g(r))^2\, p(x,r)d(x,r)$$
$$= \int_{-\infty}^{\infty}\left(l_1(r)\int_{-\infty}^{-m_1} (x-g(r))^2\, p(x)dx\right.$$
$$+ l_2(r)\int_{-m_1}^{m_1} (x-g(r))^2\, p(x)dx$$
$$\left.+ l_3(r)\int_{m_1}^{\infty} (x-g(r))^2\, p(x)dx\right)dr. \quad (28)$$

Appendix III
Optimal Power Allocation

According to (7), minimizing D_L is equivalent to minimizing $\frac{(a_n+N_r)(b_n+N_d)}{a_n b_n}$ subject to $a_n+b_n \le P$, and $a_n, b_n \in \mathbb{R}_+$. In order to solve this constraint optimization problem, we can use Lagrange multiplier method. However before applying this method we verify convexity of the function that we want to minimize. We can rewrite

$$\frac{(a_n+N_r)(b_n+N_d)}{a_n b_n} = \left(1+\frac{N_r}{a_n}\right)\left(1+\frac{N_d}{b_n}\right),$$

where the right hand side is clearly convex in (a_n, b_n).

The Lagrangian function is given by

$$J = \left(\frac{a_n b_n}{(a_n+N_r)(b_n+N_d)}\right) + \lambda\,(a_n+b_n-P).$$

From $\frac{\partial J}{\partial a_n} = 0$, $\frac{\partial J}{\partial b_n} = 0$, we get

$$a_n = \frac{-N_r+\sqrt{N_r^2-\frac{4N_r}{\lambda}}}{2}, b_n = \frac{-N_d+\sqrt{N_d^2-\frac{4N_d}{\lambda}}}{2},$$

where λ is chosen to ensure $a_n+b_n = P$ and $a_n, b_n \in \mathbb{R}_+$. We observe that for $N_r = N_d$, the optimal choice is $a_n^\star = b_n^\star = \frac{P}{2}$.