Information Embedding meets Distributed Control

Pulkit Grover[†], Aaron B. Wagner[‡] and Anant Sahai[†]

***Abstract*— We consider the problem of information embedding where the encoder modifies a white Gaussian host signal in a power-constrained manner, and the decoder recovers both the embedded message and the *modified* host signal. This extends the recent work of Sumszyk and Steinberg to the continuous-alphabet Gaussian setting. We show that a dirty-paper coding based strategy achieves the optimal rate for perfect recovery of the modified host and the message. We also provide bounds for the extension wherein the modified host signal is recovered only to within a specified distortion. Our results specialized to the zero-rate case provide the tightest known lower bounds on the asymptotic costs for the vector version of a famous open problem in distributed control — the Witsenhausen counterexample. Using this bound, we characterize the asymptotically optimal costs for the vector Witsenhausen problem to within a factor of 1.3 for all problem parameters, improving on the earlier best known bound of 2.**

I. Introduction

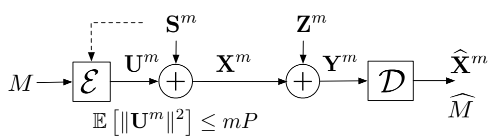


Fig. 1. The host signal $\mathbf{S}^m$ is first modified by the encoder using a power constrained input $\mathbf{U}^m$. The modified host signal $\mathbf{X}^m$ and the message M are then reconstructed at the decoder. The problem is to find the minimum distortion in reconstruction of $\mathbf{X}^m$ given P, the power constraint, and R, the rate of reliable message transmission.

The problem of interest in this paper (see Fig. 1) derives its motivation from an information-theoretic standpoint, as well as from a distributed-control perspective. Information-theoretically, the problem is an extension of an information embedding problem recently addressed by Sumszyk and Steinberg [1] — the encoder ensures that the decoder recovers the *modified* host signal $\mathbf{X}^m$ perfectly, along with the message. Philosophically, the work in [1] naturally follows the lead of Steinberg's earlier work in [2] where distributed source-coding is explored with the added constraint of having the eventual reconstructions known at the encoder. There, this added constraint made the distributed lossy-coding problem solveable in general even though in its traditional form, the problem famously remains open except for the two-user Gaussian case [3].

In [1], the authors assume that the host signal $\mathbf{S}^m$, the modified host signal (the channel input) $\mathbf{X}^m$ and the channel output $\mathbf{Y}^m$ are all *finite-alphabet*. In this paper, we consider the Gaussian (and hence infinite-alphabet) version of the problem and derive tight upper and lower bounds. The extension is non-trivial [4] because Fano's inequality-based techniques do not work for the infinite-alphabet problem. Instead, we have to consider allowig the decoder to reconstruct $\mathbf{X}^m$ to within a specified average distortion. We derive upper and lower bounds on the rate for nonzero distortion, and show their tightness in the limit of zero distortion.

Within information theory, our problem is also closely related to those in [5]–[7] (these connections are described in detail in [8]). Relative to the state amplification problem of Kim, Sutivong and Cover [5], the crucial difference is that their decoder reconstructs the initial state $\mathbf{S}^m$ whereas ours must recover $\mathbf{X}^m$. In [6], the encoder wants to instead reveal as little information about the initial state as possible. In [7], Kotagiri and Laneman consider two ("informed" and "uninformed") encoders that transmit over a Multiple Access Channel with the channel interference $\mathbf{S}^m$ known only to the informed encoder. The informed encoder helps the uninformed encoder communicate its message by modifying the interference $\mathbf{S}^m$. Intuitively, an improved reconstruction of the *modified* interference would lead to an improved decoding of the uninformed user's message.

We ourselves have a different motivation for addressing this problem — that of addressing a long-standing open problem in distributed control called the Witsenhausen counterexample [9]. It was well known that *centralized* Linear-Quadratic-Gaussian (LQG) stochastic control systems are easy to analyze — the optimal control law is linear in the observation [10]. For general distributed systems, however, the inherent difference in "who knows what" brings out a double-role of the control actions — while minimizing the local cost terms, they also *communicate* implicitly to the other controllers to reduce the knowledge differential. That this double nature can make life difficult was demonstrated by Witsenhausen's 1968 counterexample — a deceptively simple two-stage distributed LQG system. Witsenhausen showed in the scalar case that a simple nonlinear two-point quantization strategy can outperform the best linear strategies, despite everything being LQG with no bandwidth-mismatch. However, the problem of finding the optimal control law has proved to be quite challenging and the problem remains open. The non-convexity of the problem makes the search for an optimal strategy hard [11], [12]. In fact, suitably relaxed discrete analogs of the problem are even NP complete [13].

The connection between our problem and the Witsenhausen counterexample is that the special case of zero-rate communication in our formulation is precisely the asymptotic vector version of the counterexample with Witsenhausen's

†Wireless Foundations, Department of EECS, University of California at Berkeley. Email: {pulkit, sahai} @ eecs.berkeley.edu. ‡ School of Electrical and Computer Engineering, Cornell University. Email: wagner @ ece.cornell.edu.

two distributed controllers playing the roles of $\mathcal{E}$ and $\mathcal{D}$ in Fig. 1. This connection was made by Grover and Sahai in [8] where upper and lower bounds were developed on the minimum distortion. Using these bounds, they show that a combination of a linear and a DPC-based strategy attains within a factor of 2 of the optimal for any weighted sum of the power and distortion costs. These were the first results showing any-sort of optimality for this type of problem. In a follow-up work, Grover, Sahai and Park [14] used generalized sphere-packing ideas to obtain tighter lower bounds for non-asymptotic vector lengths. Lattice-based quantization strategies were given that provably get to within a constant factor of the optimal cost for each vector length, including Witsenhausen's original scalar counterexample.

In this context, our contribution here is a better lower-bound that improves the constant-factor approximation for the asymptotic optimal costs from 2 in [8] to about 1.3. Techniques similar to those developed in [14] can likely be used to obtain tighter nonasymptotic finite-length results. Further, with the new lower bound, the ratio of the upper and lower bound on the distortion (a formulation of more information-theoretic flavor) is bounded by 1.5 for all values of power P and all problem parameters. This is a vast improvement over [8] where this factor was infinity.

This constant-factor spirit is inspired from constant-gap approximation results for the capacity[1] of some notoriously hard multiuser wireless communication problems [15], [16]. A parallel hope thus arises in distributed control, since the issue of implicit communication between controllers is as ubiquitous there as interference is in wireless.

II. Problem Statement

The host signal $\mathbf{S}^m$ is distributed $\mathcal{N}(0,\sigma^2\mathbb{I})$, and the message M is independent of $\mathbf{S}^m$ and distributed uniformly over $\{1,2,\ldots,2^{mR}\}$. The encoder $\mathcal{E}_m$ maps $(M,\mathbf{S}^m)$ to $\mathbf{X}^m$ by distorting $\mathbf{S}^m$ using input $\mathbf{U}^m$ of average power (for each message) at most P, i.e. $\mathbb{E}\left[\|\mathbf{S}^m-\mathbf{X}^m\|^2\right]\leq mP$. Additive White Gaussian noise $\mathbf{Z}^m\sim\mathcal{N}(0,\sigma_z^2\mathbb{I})$, where $\sigma_z^2=1$, is added to $\mathbf{X}^m$ by the channel. The decoder $\mathcal{D}_m$ maps the channel outputs $\mathbf{Y}^m$ to estimates $\widehat{\mathbf{X}}^m$ of the modified host signal $\mathbf{X}^m$, and an estimate $\widehat{M}$ of the message.

Define the error probability $\epsilon_m(\mathcal{E}_m,\mathcal{D}_m)=\Pr(M\neq\widehat{M})$. For the encoder-decoder sequence $\{\mathcal{E}_m,\mathcal{D}_m\}_{m=1}^{\infty}$, define the minimum asymptotic distortion $MMSE(P,R)$ as

$$\inf_{\{\mathcal{E}_m,\mathcal{D}_m\}_{m=1}^{\infty}:\epsilon_m(\mathcal{E}_m,\mathcal{D}_m)\to 0}\ \limsup_{m\to\infty}\frac{1}{m}\mathbb{E}\left[\|\mathbf{X}^m-\widehat{\mathbf{X}}^m\|^2\right].$$

We are interested in the tradeoff between the rate R, the power P, and $MMSE(P,R)$.

The conventional control-theoretic weighted cost formulation [9] defines the total cost to be

$$J=\frac{1}{m}k^2\|\mathbf{U}^m\|^2+\frac{1}{m}\|\mathbf{X}^m-\widehat{\mathbf{X}}^m\|^2. \tag{1}$$

The objective is to minimize the average cost, $\mathbb{E}\left[J\right]$ at rate R. The average is taken over the realizations of the host signal, the channel noise, and the message. At $R=0$, the problem is the vector Witsenhausen counterexample [8].

[1]In this context, a factor of 2 approximation in transmit power would be a slightly stronger result than a one-bit approximation in the capacity.

III. Main Results

A. Lower bounds on $MMSE$ for rate R and power P

Theorem 1: For the problem as stated in Section II, for communicating reliably at rate R with input power P, the asymptotic average mean-square error in recovering $\mathbf{X}^m$ is lower bounded as follows. For $P\geq 2^{2R}-1$,

$$\begin{aligned}&MMSE(P,R)\\ \geq\ &\inf_{\sigma_{SU}}\sup_{\gamma>0}\frac{1}{\gamma^2}\Bigg(\Bigg(\sqrt{\frac{\sigma^2 2^{2R}}{1+\sigma^2+P+2\sigma_{SU}}}\\ &-\sqrt{(1-\gamma)^2\sigma^2+\gamma^2P-2\gamma(1-\gamma)\sigma_{SU}}\Bigg)^+\Bigg)^2,\end{aligned}$$

where $\max\left\{-\sigma\sqrt{P},\frac{2^{2R}-1-P-\sigma^2}{2}\right\}\leq\sigma_{SU}\leq\sigma\sqrt{P}$. For $P<2^{2R}-1$, reliable communication is not possible.

Corollary 1: For the vector Witsenhausen problem ($R=0$) with $\mathbb{E}\left[\|\mathbf{U}^m\|^2\right]\leq mP$, the following is a lower bound on the $MMSE$ in the estimation of $\mathbf{X}^m$.

$$\begin{aligned}&MMSE(P,0)\\ \geq\ &\inf_{\sigma_{SU}}\sup_{\gamma>0}\frac{1}{\gamma^2}\Bigg(\Bigg(\sqrt{\frac{\sigma^2}{1+\sigma^2+P+2\sigma_{SU}}}\\ &-\sqrt{(1-\gamma)^2\sigma^2+\gamma^2P-2\gamma(1-\gamma)\sigma_{SU}}\Bigg)^+\Bigg)^2.\end{aligned}$$

where $\sigma_{SU}\in[-\sigma\sqrt{P},\sigma\sqrt{P}]$. Further, this bound holds for all values of m (and not merely in the limit $m\to\infty$).
For conceptual clarity and conciseness, we only consider the case $R=0$ (Corollary 1). Tools developed herein extend directly to the case $R>0$, and are detailed in [17].

Proof: [Of Corollary 1]

For any chosen encoding map $\mathcal{E}$ and decoding map $\mathcal{D}$, there is a Markov chain $\mathbf{S}^m\to\mathbf{X}^m\to\mathbf{Y}^m\to\widehat{\mathbf{X}}^m$. Using the data-processing inequality

$$I(\mathbf{S}^m;\widehat{\mathbf{X}}^m)\leq I(\mathbf{X}^m;\mathbf{Y}^m). \tag{2}$$

The terms in the inequality can be bounded by single letter expressions as follows. Define Q as a random variable uniformly distributed over $\{1,2,\ldots,m\}$. Define $S=S_Q$, $U=U_Q$, $X=X_Q$, $Z=Z_Q$, $Y=Y_Q$ and $\widehat{X}=\widehat{X}_Q$. Using the memorylessness of the channel (see [17] for details),

$$I(\mathbf{X}^m;\mathbf{Y}^m)\leq mI(X;Y), \tag{3}$$

and using the fact that the source is iid,

$$mI(S;\widehat{X})\leq I(\mathbf{S}^m;\widehat{\mathbf{X}}^m). \tag{4}$$

Using (2), (3) and (4),

$$I(S;\widehat{X})\leq I(X;Y). \tag{5}$$

Also observe that from the definitions of S, X, $\widehat{X}$ and Y, $\mathbb{E}\left[\frac{1}{m}\|\mathbf{S}^m-\mathbf{X}^m\|^2\right]=\mathbb{E}\left[(S-X)^2\right]$, and $\mathbb{E}\left[\frac{1}{m}\|\mathbf{X}^m-\widehat{\mathbf{X}}^m\|^2\right]=\mathbb{E}\left[(X-\widehat{X})^2\right]$.

Using the Cauchy-Schwartz inequality, the correlation $\sigma_{SU} = \mathbb{E}[SU]$ must satisfy the following constraint,

$$|\sigma_{SU}| = |\mathbb{E}[SU]| \leq \sqrt{\mathbb{E}[S^2]}\sqrt{\mathbb{E}[U^2]} = \sigma\sqrt{P}. \quad (6)$$

Also,

$$\mathbb{E}\left[X^2\right] = \mathbb{E}\left[(S+U)^2\right] = \sigma^2 + P + 2\sigma_{SU}. \quad (7)$$

Since $Z = Y - X \perp\!\!\!\perp X$, and because a Gaussian input distribution maximizes the mutual information across an average-power constrained AWGN channel, we can upper bound the RHS of (5) as follows

$$I(X;Y) \leq \frac{1}{2}\log_2\left(1 + \frac{\sigma^2 + P + 2\sigma_{SU}}{1}\right). \quad (8)$$

Now, looking at the LHS in (5),

$$\begin{aligned} I(S;\widehat{X}) &= h(S) - h(S|\widehat{X}) \\ &= h(S) - h(S - \gamma\widehat{X}|\widehat{X})\ \forall\gamma \\ &\stackrel{(a)}{\geq} h(S) - h(S - \gamma\widehat{X}) \\ &= \frac{1}{2}\log_2\left(2\pi e\sigma^2\right) - h(S - \gamma\widehat{X}), \quad (9) \end{aligned}$$

where (a) follows from the fact that conditioning reduces entropy. Note here that the result holds for any γ, and in particular, γ can depend on σ_{SU}. Now,

$$\begin{aligned} & h(S - \gamma\widehat{X}) \\ = \ & h\left((1-\gamma)S - \gamma U - \gamma(\widehat{X} - X)\right). \quad (10) \end{aligned}$$

Using the Cauchy-Schwartz inequality, we can bound the second moment for the random variable in (10),

$$\begin{aligned} & \mathbb{E}\left[\left((1-\gamma)S - \gamma U - \gamma(\widehat{X} - X)\right)^2\right] \\ \leq & \left(\sqrt{(1-\gamma)^2\sigma^2 + \gamma^2 P - 2\gamma(1-\gamma)\sigma_{SU}}\right. \\ & \left. + \gamma\sqrt{\mathbb{E}\left[(\widehat{X} - X)^2\right]}\right)^2, \quad (11) \end{aligned}$$

that is obtained by aligning[2] $X - \widehat{X}$ with $(1-\gamma)S - \gamma U$. Thus,

$$\begin{aligned} & I(S;\widehat{X}) \\ \geq & \frac{1}{2}\log_2\left(2\pi e\sigma^2\right) - h(S - \gamma\widehat{X}) \\ \geq & \frac{1}{2}\log_2\left(\sigma^2\right) \\ & - \log_2\left(\sqrt{(1-\gamma)^2\sigma^2 + \gamma^2 P - 2\gamma(1-\gamma)\sigma_{SU}}\right. \\ & \left. + \gamma\sqrt{\mathbb{E}\left[(\widehat{X} - X)^2\right]}\right). \quad (12) \end{aligned}$$

Using (5), (8) and (12), followed by algebraic manipulations gives us the expression for the lower bound, where the $\sup_{\gamma>0}$ is obtained because any choice of $\gamma > 0$ is allowed, and the $\inf_{\sigma_{SU}}$ is obtained by allowing for all encoders and decoders (the encoding and decoding strategy influence the bound only through σ_{SU}, and σ_{SU} has to satisfy (6)). The details are in [17] since space here is at a premium. ■

[2] In general, since $\widehat{\mathbf{X}}^m$ is a function of $\mathbf{Y}^m$, this alignment is not actually possible unless the recovery of $\mathbf{X}^m$ is exact. The derived bound is therefore loose except in the zero-distortion case.

B. The tightness at $MMSE(P,R) = 0$

For an achievable strategy, we use the combination of linear and dirty-paper coding strategies of [8], except that here it is used to communicate a message at rate R as well.

Corollary 2: Given power P, a combination of linear and DPC-based strategies achieves the maximum rate $C(P)$ in the limit $MMSE(P,R) = 0$, where $C(P)$ is given by

$$\begin{aligned} & C(P) \\ = & \sup_{\sigma_{SU}} \frac{1}{2}\log_2\left(\frac{(P\sigma^2 - \sigma_{SU}^2)(1 + \sigma^2 + P + 2\sigma_{SU})}{\sigma^2(\sigma^2 + P + 2\sigma_{SU})}\right), \end{aligned}$$

where $\sigma_{SU} \in [-\sigma\sqrt{P}, 0]$.

Proof: We only outline the proof, see [17] for details.

The achievability:

We first show that the combination of linear and DPC strategies [8] achieves the expression in Corollary 2. A DPC strategy with the DPC parameter $\alpha = 1$ (ensuring perfect recovery of $\mathbf{X}^m$) achieves the following rate [18, Eq. (6)]

$$R = \frac{1}{2}\log_2\left(\frac{P(P + \sigma^2 + 1)}{P + \sigma^2}\right). \quad (13)$$

Now, the combination strategy divides P into P_{lin} and P_{dpc}. The linear part scales $\mathbf{S}^m$, modifying the initial state to variance $\widetilde{\sigma}^2 = (\sigma - \sqrt{P_{lin}})^2$. The now achievable rate is

$$\widetilde{R} = \sup_{P_{lin}, P_{dpc}: P = P_{lin} + P_{dpc}} \frac{1}{2}\log_2\left(\frac{P_{dpc}(P_{dpc} + \widetilde{\sigma}^2 + 1)}{P_{dpc} + \widetilde{\sigma}^2}\right). \quad (14)$$

Let $\sigma_{SU} = -\sigma\sqrt{P_{lin}}$ (note that as P_{lin} varies from 0 to P, σ_{SU} varies from 0 to $-\sigma\sqrt{P}$). Then, $P_{dpc} = P - \frac{\sigma_{SU}^2}{\sigma^2}$, and $P_{dpc} + \widetilde{\sigma}^2 = P_{dpc} + \sigma^2 + P_{lin} - 2\sigma\sqrt{P_{lin}} = P + \sigma^2 + 2\sigma_{SU}$. The resulting expression matches the expression in Corollary 2.

The converse:

Since we are free to choose γ, let $\gamma = \gamma^* = \frac{\sigma^2 + \sigma_{SU}}{\sigma^2 + P + 2\sigma_{SU}}$. Then, $1 - \gamma^* = \frac{P + \sigma_{SU}}{\sigma^2 + P + 2\sigma_{SU}}$. Since $MMSE = 0$, it has to be the case that the term inside $(\cdot)^+$ in the lower bound is non-positive for some value of σ_{SU}. This immediately yields

$$\begin{aligned} 2^{2R} \leq & \sup_{\sigma_{SU}}\left\{\frac{1}{\sigma^2}\left((1-\gamma^*)^2\sigma^2 + {\gamma^*}^2 P - 2\gamma(1-\gamma^*)\sigma_{SU}\right)\right. \\ & \left. \times(1 + \sigma^2 + P + 2\sigma_{SU})\right\} \\ = & \sup_{\sigma_{SU}} \frac{(P\sigma^2 - \sigma_{SU}^2)(1 + \sigma^2 + P + 2\sigma_{SU})}{\sigma^2(\sigma^2 + P + 2\sigma_{SU})}, \end{aligned}$$

where $\sigma_{SU} \in \{-\sigma\sqrt{P}, \sigma\sqrt{P}\}$. Observe that the term $(P\sigma^2 - \sigma_{SU}^2)$ is oblivious to the sign of σ_{SU}. Further,

$$\frac{1 + \sigma^2 + P + 2\sigma_{SU}}{\sigma^2 + P + 2\sigma_{SU}} = 1 + \frac{1}{\sigma^2 + P + 2\sigma_{SU}} \quad (15)$$

is clearly larger for $\sigma_{SU} < 0$ if we fix $|\sigma_{SU}|$. Thus the supremum in the upper bound is attained at some $\sigma_{SU} < 0$, and we get that the expression in the Corollary is an upper bound on the rate as well. ■

IV. Numerical Results

Witsenhausen's control theoretic formulation seeks to minimize the sum of weighted costs $k^2P + MMSE$. Fig. 2(b) shows that asymptotically, the ratio of upper and lower bounds on the weighted cost is bounded by 1.3, an improvement over the ratio of 2 in [8] and shown in Fig. 2(a). The upper bound is obtained using the combination of linear and DPC-based strategy of [8]. The new lower bound is the one derived in Corollary 1. A ridge of ratio 2 along $\sigma^2 = \frac{\sqrt{5}-1}{2}$ present in Fig. 2(a) does not exist anymore with the new lower bound since this small-k regime corresponds to target $MMSE$s close to zero – where the new lower bound is tight. This is illustrated in Fig. 3 which also explains why the ridge along $k \approx 1.67$ still survives in the new ratio plot.

Fig. 4 shows the ratio of upper and lower bounds on $MMSE(P, 0)$ versus P and σ. While the ratio with the bound of [8] was unbounded (Fig. 4, top), the new ratio is bounded by a factor of 1.5 (Fig. 4, bottom). Fig. 5 shows that the the price of increasing rate is an increased average distortion in $\mathbf{X}^m$ estimation at high rates.

The MATLAB code for these figures can be found in [19].

Acknowledgments

P. Grover and A. Sahai acknowledge the support of the National Science Foundation (CNS-403427 and CCF-729122) and Sumitomo Electric. A.B. Wagner acknowledges the support of NSF 06-42925 (CAREER) grant. We thank Hari Palaiyanur and Se Yong Park for suggesting improvements on an earlier draft of the paper.

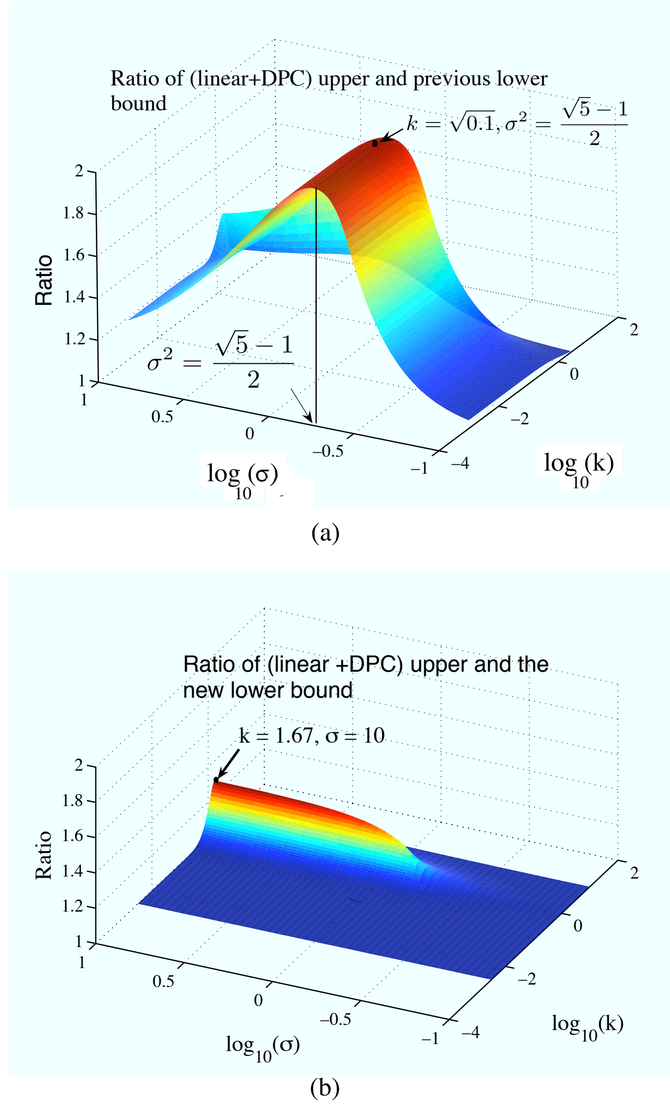


Fig. 2. The ratio of upper and lower bounds on the total asymptotic cost for the vector Witsenhausen counterexample with the lower bound taken from [8] in (a) and from Corollary 1 in (b). As compared to the previous best known ratio of 2 [8], the ratio here is smaller than 1.3. Further, an infinite ridge along $\sigma^2 = \frac{\sqrt{5}-1}{2}$ and small k that is present in lower bounds of [8] is no longer present here. This is a consequence of the tightness lower bound at $MMSE = 0$, and hence for small k. A ridge remains along $k \approx 1.67$ ($\log_{10}(k) \approx 0.22$) and large σ, and this can be understood by observing Fig. 3 for $\sigma = 10$.

References

[1] O. Sumszyk and Y. Steinberg, "Information embedding with reversible stegotext," in *Proceedings of the 2009 IEEE Symposium on Information Theory*, Seoul, Korea, Jun. 2009.

[2] Y. Steinberg, "Simultaneous transmission of data and state with common knowledge," in *Proceedings of the 2008 IEEE Symposium on Information Theory*, Toronto, Canada, Jun. 2008, pp. 935–939.

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

[4] Y. Steinberg, personal communication, Jun. 2009.

[5] Y.-H. Kim, A. Sutivong, and T. M. Cover, "State amplification," *IEEE Trans. Inform. Theory*, vol. 54, no. 5, pp. 1850–1859, May 2008.

[6] N. Merhav and S. Shamai, "Information rates subject to state masking," *IEEE Trans. Inform. Theory*, vol. 53, no. 6, pp. 2254–2261, Jun. 2007.

[7] S. P. Kotagiri and J. N. Laneman, "Multiaccess channels with state known to some encoders and independent messages," *EURASIP Journal on Wireless Communications and Networking*, no. 450680, 2008.

[8] P. Grover and A. Sahai, "Vector Witsenhausen counterexample as assisted interference suppression," *Special issue on Information Processing and Decision Making in Distributed Control Systems of the International Journal on Systems, Control and Communications (IJSCC), to appear*, Sep. 2009. [Online]. Available: http://www.eecs.berkeley.edu/~sahai/

[9] H. S. Witsenhausen, "A counterexample in stochastic optimum control," *SIAM Journal on Control*, vol. 6, no. 1, pp. 131–147, Jan. 1968.

[10] D. Bertsekas, *Dynamic Programming*. Belmont, MA: Athena Scientific, 1995.

[11] R. Bansal and T. Basar, "Stochastic teams with nonclassical information revisited: When is an affine control optimal?" *IEEE Transactions on Automatic Control*, vol. 32, pp. 554–559, Jun. 1987.

[12] J. T. Lee, E. Lau, and Y.-C. L. Ho, "The Witsenhausen counterexample: A hierarchical search approach for nonconvex optimization problems," *IEEE Trans. Automat. Contr.*, vol. 46, no. 3, 2001.

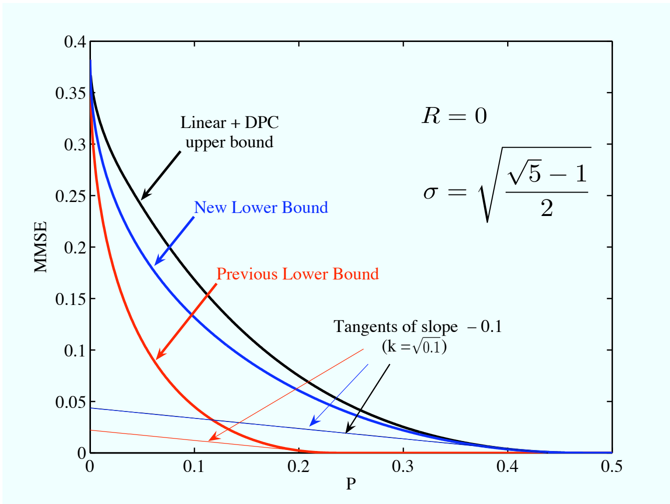
[13] C. H. Papadimitriou and J. N. Tsitsiklis, "Intractable problems in control theory," *SIAM Journal on Control and Optimization*, vol. 24, no. 4, pp. 639–654, 1986.

[14] P. Grover, A. Sahai, and S. Y. Park, "The finite-dimensional witsenhausen counterexample," *Proceedings of the Workshop on Control over Communication Channels (ConCom)*, Jul. 2009.

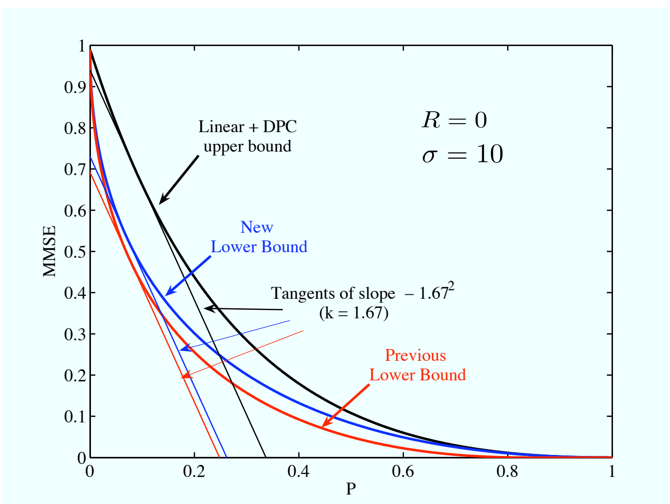
[15] R. Etkin, D. Tse, and H. Wang, "Gaussian interference channel capacity to within one bit," *IEEE Trans. Inform. Theory*, pp. 5534–5562, Dec. 2008.

[16] A. Avestimehr, S. Diggavi, and D. Tse, "A deterministic approach to wireless relay networks," in *Proc. of the Allerton Conference on Communications, Control and Computing*, October 2007.

[17] P. Grover, A. B. Wagner, and A. Sahai, "Information



(a)



(b)

Fig. 3. Upper and lower bounds on P vs asymptotic $MMSE$ for $\sigma = \sqrt{\frac{\sqrt{5}-1}{2}}$ (square-root of the Golden ratio; Fig. (a)) and $\sigma = 10$ (b) for zero-rate (the vector Witsenhausen counterexample). Tangents are drawn to evaluate the total cost for $k = \sqrt{0.1}$ for $\sigma = \sqrt{\frac{\sqrt{5}-1}{2}}$, and for $k = 1.67$ for $\sigma = 10$ (slope $= -k^2$). The intercept on the MMSE axis of the tangent provides the respective bound on the total cost. The tangents to the upper bound and the new lower bound almost coincide for small values of k. At $k \approx 1.67$ and $\sigma = 10$, however, our bound is not significantly better than that in [8] and hence the ridge along $k \approx 1.67$ remains in the new ratio plot in Fig. 2.

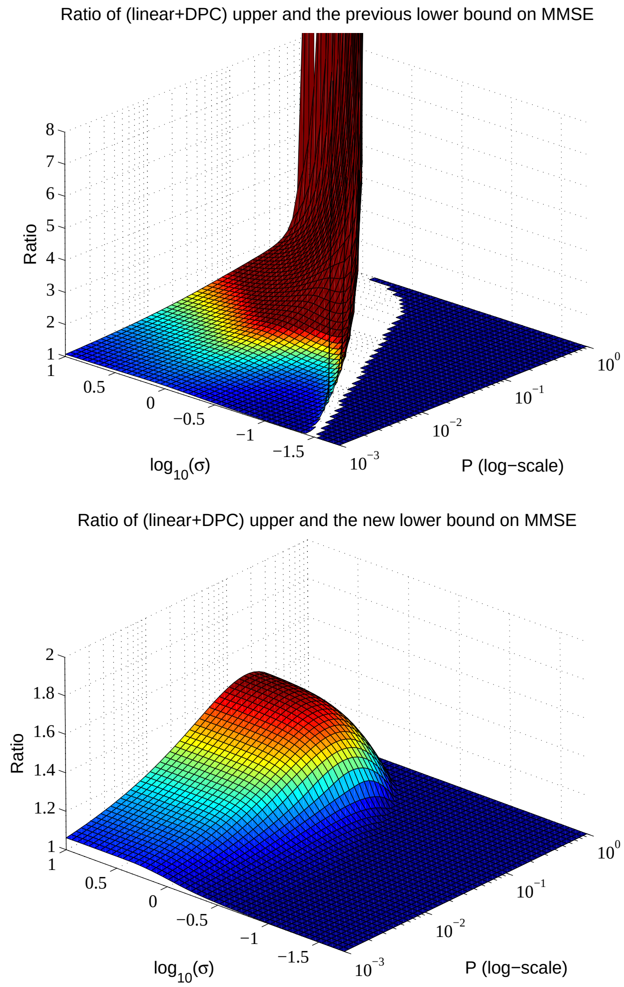


Fig. 4. Ratio of upper and lower bounds on MMSE vs P and σ at $R = 0$. Whereas the ratio diverges to infinity with the old lower bound of [8] (top), it is bounded by 1.5 for the new bound (bottom). This is a consequence of the improved tightness of the new bound at small $MMSE$.

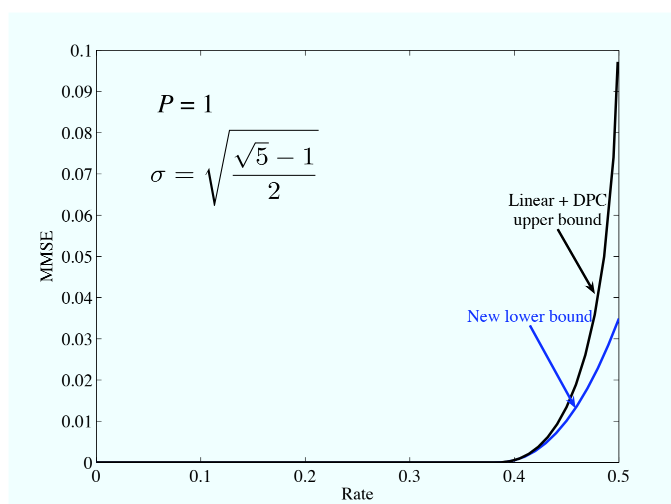


Fig. 5. Plot of upper and lower bounds on $MMSE$ vs rate for fixed power $P = 1$ and $\sigma = \sqrt{\frac{\sqrt{5}-1}{2}}$. Higher rates require higher average distortion in the reconstruction of $\mathbf{X}^m$.

embedding meets distributed control," *Submitted to IEEE Transactions on Information Theory*, Aug. 2009. [Online]. Available: http://www.eecs.berkeley.edu/∼pulkit/InfoEmbedding.pdf

[18] M. H. M. Costa, "Writing on dirty paper," *IEEE Trans. Inform. Theory*, vol. 29, no. 3, pp. 439–441, May 1983.

[19] "Code for 'Information embedding meets distributed control'." [Online]. Available: http://www.eecs.berkeley.edu/∼pulkit/InformationEmbedding.htm