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 to encode the message, 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. When specialized to the zero-rate case, our results 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 numerically to within a factor of 1.3 for all problem parameters, improving on the earlier best known bound of 2.

I. Introduction

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] is directed towards understanding how a communication problem changes when an additional requirement, that of the encoder being able to produce a copy of the reconstruction

†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. An abridged version of this paper will be presented at the 2010 Information Theory Workshop (ITW), Cairo, Egypt.

at the decoder, is imposed on the system (in source coding context, the issue was explored by Steinberg in [2]). The problem is also closely connected to other information theory problems [3]–[6]. We refer the interested reader to [7], where these connections are discussed in detail.

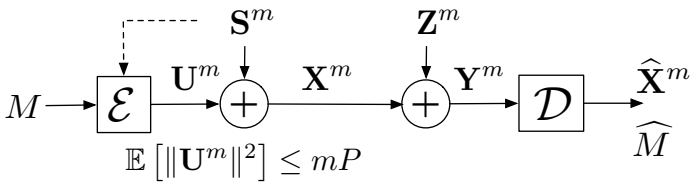


Fig. 1. The host signal $\mathbf{S}^m$ is first modified by the encoder using a power constrained input $\mathbf{U}^m$. The modified host signal $\mathbf{X}^m$ and the message M are then reconstructed at the decoder. The problem is to find the minimum distortion in reconstruction of $\mathbf{X}^m$ given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

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 version of their problem. The extension is non-trivial [8] because simple Fano's inequality-based techniques do not work for the infinite-alphabet formulation. Experience in infinite-alphabet problems might even suggest that (asymptotic) perfect reconstruction may be impossible because the problem is set in continuous space. Intriguingly, asymptotic perfect reconstruction is possible in our problem because the encoder can ensure that the modified host signal takes only finitely many values. We provide tight results characterizing the tradeoff between rate and power for perfect reconstruction. Further, we relax the assumption of perfect recovery of the host signal by considering recovery within a specified nonzero distortion, and for this problem we provide upper and lower bounds on the tradeoff between rate, power and average distortion.

The nonzero distortion problem is closely related to the vector version of a famous distributed control problem called the Witsenhausen counterexample [9] — at zero communication rate, the two problems are the same [7]. The scalar counterexample is believed to be quite challenging (see [7] for a survey of prior results showing why it is believed to be so). As a conceptual simplification, Grover and Sahai [7] considered the long-blocklength limit of the counterexample. Further, they relaxed the requirement of

obtaining a provably optimal strategy to a weaker objective of obtaining strategies that attain within a constant factor of the optimal cost. For the weighted sum cost of power and average distortion costs (see Section II), they then show that random binning based dirty-paper coding techniques attain within a factor of 2 of the optimal cost for all problem parameters (*i.e.* the weights and the variances of the random variables). Backing off from the infinite blocklength limit, Grover, Sahai and Park [10] then showed that similar constant factor results can also be obtained for finite vector lengths, including the scalar case. The achievable strategy, which yields the upper bounds, now uses lattices instead of random codebooks. The lower bound is obtained by applying sphere-packing ideas from information theory to the bound of [7].

The lower bound in this paper specialized to rate zero provides an improved lower bound to the costs of the vector Witsenhausen counterexample in the long-blocklength limit. Using this improved bound, we show that the ratio of upper and lower bounds is smaller than 1.3 regardless of the choice of the weights and the problem parameters. This is an improvement over the previously best known maximum ratio of 2 [7].

Control theory has long wrestled with the Witsenhausen counterexample. Because it is a canonical problem, a comprehensive distributed-control theory would necessarily include a good understanding of the counterexample. Information-theory has had long-standing canonical problems of its own. In a line of investigation started by Gupta and Kumar [11], the question of the capacity of a large wireless network with the number of nodes is studied. Restricting attention to obtaining just the scaling of the total capacity, it turns out that obtaining precise solutions to canonical problems in information theory is not necessary. The bar for what might constitute a reasonable information-theoretic solution was thus lowered. More recently, calculation of channel capacity within a finite number of bits[1] for canonical information-theory problems (e.g. the interference channel [12]) has led to significant advances in approximating capacity for larger network communication problems [13], [14]. The recent results on Witsenhausen's counterexample

[1]The constant-factor results on control costs are closely related to results on bounded gap from capacity in information-theory. A factor of 2 approximation in power would be a slightly stronger result than a one-bit approximation in the capacity.

thus raise a parallel hope in distributed control.

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 an estimate $\widehat{\mathbf{X}}^m$ 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 follows

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

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

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

where $k \in \mathbb{R}^+$. 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 [7].

III. Main Results

A. Lower bounds on $MMSE(P, R)$

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 $P \geq 2^{2R} - 1$,

$$MMSE(P, R) \geq \inf_{\sigma_{SU}} \sup_{\gamma > 0} \frac{1}{\gamma^2} \left(\left(\sqrt{\frac{\sigma^2 2^{2R}}{1 + \sigma^2 + P + 2\sigma_{SU}}} - \sqrt{(1-\gamma)^2\sigma^2 + \gamma^2 P - 2\gamma(1-\gamma)\sigma_{SU}} \right)^+ \right)^2, \tag{2}$$

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 at rate R is not possible.

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

$$MMSE(P,0) \geq \inf_{\sigma_{SU}} \sup_{\gamma>0} \frac{1}{\gamma^2} \left(\left(\sqrt{\frac{\sigma^2}{1+\sigma^2+P+2\sigma_{SU}}} - \sqrt{(1-\gamma)^2\sigma^2 + \gamma^2 P - 2\gamma(1-\gamma)\sigma_{SU}} \right)^+ \right)^2. \tag{3}$$

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

Proof: [Of Theorem 1] For conceptual clarity, we first derive the result for the case $R = 0$ (Corollary 1). The tools developed are then used to derive the lower bound for $R > 0$.

Proof: [Of Corollary 1]

For any chosen pair of encoding map $\mathcal{E}_m$ and decoding map $\mathcal{D}_m$, 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

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

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 $\widehat{X} = \widehat{X}_Q$. Then,

$$\begin{aligned}
I(\mathbf{X}^m; \mathbf{Y}^m) &= h(\mathbf{Y}^m) - h(\mathbf{Y}^m|\mathbf{X}^m) \\
&\overset{(a)}{\leq} \sum_i h(Y_i) - h(\mathbf{Y}^m|\mathbf{X}^m) \\
&= \sum_i h(Y_i) - h(Y_i|X_i) \\
&= \sum_i I(X_i; Y_i) \\
&= mI(X;Y|Q) \\
&= m\left(h(Y|Q) - h(Y|X,Q)\right) \\
&\leq m\left(h(Y) - h(Y|X,Q)\right) \\
&\overset{(b)}{=} m\left(h(Y) - h(Y|X)\right) = mI(X;Y),
\end{aligned} \tag{5}$$

where (a) follows from an application of the chain-rule for entropy followed by using the fact that conditioning reduces entropy, and (b) follows from the observation that the additive noise Z_i is iid across time, and independent of the input X_i (thus $Y \perp\!\!\!\perp Q|X$). Also,

$$\begin{aligned} I(\mathbf{S}^m; \widehat{\mathbf{X}}^m) &= h(\mathbf{S}^m) - h(\mathbf{S}^m|\widehat{\mathbf{X}}^m) \\ &= \sum_i h(S_i) - h(\mathbf{S}^m|\widehat{\mathbf{X}}^m) \\ &\overset{(a)}{\geq} \sum_i \Big(h(S_i) - h(S_i|\widehat{X}_i)\Big) \\ &= \sum_i I(S_i; \widehat{X}_i) = mI(S; \widehat{X}|Q) \\ &= m\Big(h(S|Q) - h(S|\widehat{X}, Q)\Big) \\ &\overset{(b)}{\geq} m\Big(h(S) - h(S|\widehat{X})\Big) = mI(S; \widehat{X}), \end{aligned} \tag{6}$$

where (a) and (b) again follow from the fact that conditioning reduces entropy, and (b) also uses the observation that since S_i are iid, S, S_i, and $S|Q = q$ are distributed identically.

Now, using (4), (5) and (6),

$$mI(S; \widehat{X}) \leq I(\mathbf{S}^m; \widehat{\mathbf{X}}^m) \leq I(\mathbf{X}^m; \mathbf{Y}^m) \leq mI(X; Y). \tag{7}$$

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

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

$$|\sigma_{SU}| = |\mathbb{E}\left[SU\right]| \leq \sqrt{\mathbb{E}\left[S^2\right]}\sqrt{\mathbb{E}\left[U^2\right]} \leq \sigma\sqrt{P}. \tag{8}$$

Also,

$$\mathbb{E}\left[X^2\right] = \mathbb{E}\left[(S+U)^2\right] = \sigma^2 + P + 2\sigma_{SU}. \tag{9}$$

Since $Z = Y - X \perp\!\!\!\perp X$, and Gaussian input distribution maximizes the mutual information across an average power constrained AWGN channel,

$$I(X; Y) \leq \frac{1}{2}\log_2\left(1 + \frac{P + \sigma^2 + 2\sigma_{SU}}{1}\right). \tag{10}$$

$$\begin{aligned} I(S;\widehat{X}) &= h(S) - h(S|\widehat{X}) \\ &= h(S) - h(S - \gamma\widehat{X}|\widehat{X}) \ \forall\gamma \\ &\overset{(a)}{\geq} h(S) - h(S - \gamma\widehat{X}) \\ &= \frac{1}{2}\log_2\left(2\pi e\sigma^2\right) - h(S - \gamma\widehat{X}), \end{aligned} \tag{11}$$

where (a) follows from the fact that conditioning reduces entropy. Also note here that the result holds for any $\gamma > 0$, and in particular, γ can depend on σ_{SU}. Now,

$$\begin{aligned} h(S - \gamma\widehat{X}) &= h(S - \gamma(\widehat{X} - X) - \gamma X) \\ &= h\left(S - \gamma(\widehat{X} - X) - \gamma S - \gamma U\right) \\ &= h\left((1-\gamma)S - \gamma U - \gamma(\widehat{X} - X)\right). \end{aligned} \tag{12}$$

The second moment of a sum of two random variables A and B can be bounded as follows

$$\begin{aligned} \mathbb{E}\left[(A+B)^2\right] &= \mathbb{E}\left[A^2\right] + \mathbb{E}\left[B^2\right] + 2\mathbb{E}\left[AB\right] \\ &\overset{\text{Cauchy-Schwartz ineq.}}{\leq} \mathbb{E}\left[A^2\right] + \mathbb{E}\left[B^2\right] + 2\sqrt{\mathbb{E}\left[A^2\right]}\sqrt{\mathbb{E}\left[B^2\right]} \\ &= (\sqrt{\mathbb{E}\left[A^2\right]} + \sqrt{\mathbb{E}\left[B^2\right]})^2, \end{aligned} \tag{13}$$

with equality when A and B are aligned, i.e. $A = \lambda B$ for some $\lambda \in \mathbb{R}$. For the random variable under consideration in (12), choosing $A = (1-\gamma)S - \gamma U$, and $B = -\gamma(\widehat{X} - X)$ in (13)

$$\mathbb{E}\left[\left((1-\gamma)S - \gamma U - \gamma(\widehat{X} - X)\right)^2\right] \leq \left(\sqrt{(1-\gamma)^2\sigma^2 + \gamma^2 P - 2\gamma(1-\gamma)\sigma_{SU}} + \gamma\sqrt{\mathbb{E}\left[(\widehat{X} - X)^2\right]}\right)^2. \tag{14}$$

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

$$\begin{aligned} I(S;\widehat{X}) &\geq \frac{1}{2}\log_2\left(2\pi e\sigma^2\right) - h(S - \gamma\widehat{X}) \\ &\geq \frac{1}{2}\log_2\left(\frac{\sigma^2}{\left(\sqrt{(1-\gamma)^2\sigma^2 + \gamma^2 P - 2\gamma(1-\gamma)\sigma_{SU}} + \gamma\sqrt{\mathbb{E}\left[(\widehat{X} - X)^2\right]}\right)^2}\right). \end{aligned} \tag{15}$$

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

Using (7), $I(S;\widehat{X}) \leq I(X;Y)$. Using the lower bound on $I(S;\widehat{X})$ from (15) and the upper bound on $I(X;Y)$ from (10), we get

$$\frac{1}{2}\log_2\left(\frac{\sigma^2}{\left(\sqrt{(1-\gamma)^2\sigma^2+\gamma^2P-2\gamma(1-\gamma)\sigma_{SU}}+\gamma\sqrt{\mathbb{E}\left[(\widehat{X}-X)^2\right]}\right)^2}\right) \leq \frac{1}{2}\log_2\left(1+\frac{P+\sigma^2+2\sigma_{SU}}{1}\right),$$

for the choice of $\mathcal{E}_m$ and $\mathcal{D}_m$. Since $\log_2(\cdot)$ is a monotonically increasing function,

$$\frac{\sigma^2}{\left(\sqrt{(1-\gamma)^2\sigma^2+\gamma^2P-2\gamma(1-\gamma)\sigma_{SU}}+\gamma\sqrt{\mathbb{E}\left[(\widehat{X}-X)^2\right]}\right)^2} \leq 1+P+\sigma^2+2\sigma_{SU}$$

$$\text{i.e.} \quad \left(\sqrt{(1-\gamma)^2\sigma^2+\gamma^2P-2\gamma(1-\gamma)\sigma_{SU}}+\gamma\sqrt{\mathbb{E}\left[(\widehat{X}-X)^2\right]}\right)^2 \geq \frac{\sigma^2}{1+P+\sigma^2+2\sigma_{SU}},$$

$$\text{Since } \gamma>0,\ \gamma\sqrt{\mathbb{E}\left[(\widehat{X}-X)^2\right]} \geq \sqrt{\frac{\sigma^2}{1+P+\sigma^2+2\sigma_{SU}}}-\sqrt{(1-\gamma)^2\sigma^2+\gamma^2P-2\gamma(1-\gamma)\sigma_{SU}}.$$

Because the RHS may not be positive, we take the maximum of zero and the RHS and obtain the following lower bound for $\mathcal{E}_m$ and $\mathcal{D}_m$.

$$\mathbb{E}\left[(\widehat{X}-X)^2\right] \geq \frac{1}{\gamma^2}\left(\left(\sqrt{\frac{\sigma^2}{1+P+\sigma^2+2\sigma_{SU}}}-\sqrt{(1-\gamma)^2\sigma^2+\gamma^2P-2\gamma(1-\gamma)\sigma_{SU}}\right)^+\right)^2. \quad (16)$$

Because the bound holds for every $\gamma>0$,

$$\mathbb{E}\left[(\widehat{X}-X)^2\right] \geq \sup_{\gamma>0}\frac{1}{\gamma^2}\left(\left(\sqrt{\frac{\sigma^2}{1+P+\sigma^2+2\sigma_{SU}}}-\sqrt{(1-\gamma)^2\sigma^2+\gamma^2P-2\gamma(1-\gamma)\sigma_{SU}}\right)^+\right)^2, \quad (17)$$

for the chosen $\mathcal{E}_m$ and $\mathcal{D}_m$. Now, from (8), σ_{SU} can take values in $[-\sigma\sqrt{P},\sigma\sqrt{P}]$. Because the lower bound depends on $\mathcal{E}_m$ and $\mathcal{D}_m$ only through σ_{SU}, we obtain the following lower bound for all $\mathcal{E}_m$ and $\mathcal{D}_m$,

$$\mathbb{E}\left[(\widehat{X}-X)^2\right] \geq \inf_{|\sigma_{SU}|\leq\sigma\sqrt{P}}\sup_{\gamma>0}\frac{1}{\gamma^2}\left(\left(\sqrt{\frac{\sigma^2}{1+P+\sigma^2+2\sigma_{SU}}}-\sqrt{(1-\gamma)^2\sigma^2+\gamma^2P-2\gamma(1-\gamma)\sigma_{SU}}\right)^+\right)^2, \quad (18)$$

which proves Corollary 1. Notice that we did not take limits in m everywhere, and hence the lower bound holds for all values of m. ■

The case of nonzero rate

Continuing to the proof of Theorem 1, consider now the problem when the encoder wants to also communicate a message M reliably to the decoder at rate R.

Using Fano's inequality, since $\Pr(M \neq \widehat{M}) = \epsilon_m \to 0$ as $m \to \infty$, $H(M|\widehat{M}) \leq m\delta_m$ where $\delta_m \to 0$. Thus,

$$\begin{aligned} I(M;\widehat{M}) &= H(M) - H(M|\widehat{M}) \\ &= mR - H(M|\widehat{M}) \\ &\geq mR - m\delta_m = m(R-\delta_m). \end{aligned} \tag{19}$$

As before, we consider a mutual information inequality that follows directly from the Markov chain $(M, \mathbf{S}^m) \to \mathbf{X}^m \to \mathbf{Y}^m \to (\widehat{\mathbf{X}}^m, \widehat{M})$:

$$I(M, \mathbf{S}^m; \widehat{M}, \widehat{\mathbf{X}}^m) \leq I(\mathbf{X}^m; \mathbf{Y}^m). \tag{20}$$

The RHS can be bounded above as in (5). For the LHS,

$$\begin{aligned} I(M, \mathbf{S}^m; \widehat{M}, \widehat{\mathbf{X}}^m) &= I(M; \widehat{M}, \widehat{\mathbf{X}}^m) + I(\mathbf{S}^m; \widehat{M}, \widehat{\mathbf{X}}^m | M) \\ &\geq I(M; \widehat{M}) + I(\mathbf{S}^m; \widehat{M}, \widehat{\mathbf{X}}^m | M) \\ &= I(M; \widehat{M}) + h(\mathbf{S}^m|M) - h(\mathbf{S}^m|\widehat{M}, \widehat{\mathbf{X}}^m, M) \\ &\overset{\mathbf{S}^m \perp\!\!\!\perp M}{=} I(M; \widehat{M}) + h(\mathbf{S}^m) - h(\mathbf{S}^m|\widehat{M}, \widehat{\mathbf{X}}^m, M) \\ &\geq I(M; \widehat{M}) + h(\mathbf{S}^m) - h(\mathbf{S}^m|\widehat{\mathbf{X}}^m) \\ &\geq I(M; \widehat{M}) + I(\mathbf{S}^m; \widehat{\mathbf{X}}^m) \\ &\overset{\text{using (6)}}{\geq} I(M; \widehat{M}) + mI(S; \widehat{X}). \end{aligned} \tag{21}$$

From (19), (20) and (21), we obtain

$$\begin{aligned} m(R-\delta_m) + mI(S;\widehat{X}) &\overset{\text{using (19)}}{\leq} I(M;\widehat{M}) + mI(S;\widehat{X}) \\ &\overset{\text{using (21)}}{\leq} I(M, \mathbf{S}^m; \widehat{M}, \widehat{\mathbf{X}}^m) \\ &\overset{\text{using (20)}}{\leq} I(\mathbf{X}^m; \mathbf{Y}^m) \overset{\text{using (5)}}{\leq} mI(X;Y). \end{aligned} \tag{22}$$

$I(X;Y)$ and $I(S;\widehat{X})$ can be bounded as before in (10) and (15). Observing that as $m \to \infty$, $\delta_m \to 0$, we get the following lower bound on the $MMSE$ for nonzero rate,

$$MMSE(P,R) \geq \inf_{\sigma_{SU}} \sup_{\gamma>0} \frac{1}{\gamma^2} \left(\left(\sqrt{\frac{\sigma^2 2^{2R}}{1+\sigma^2+P+2\sigma_{SU}}} - \sqrt{(1-\gamma)^2\sigma^2 + \gamma^2 P - 2\gamma(1-\gamma)\sigma_{SU}} \right)^+ \right)^2 . \tag{23}$$

In the limit $\delta_m \to 0$, we require from (22) that $I(X;Y) \geq R$. This gives the following constraint on σ_{SU},

$$\begin{aligned} \frac{1}{2}\log_2\left(1+P+\sigma^2+2\sigma_{SU}\right) &\geq R \\ \text{i.e. } \sigma_{SU} &\geq \frac{2^{2R}-1-P-\sigma^2}{2}, \end{aligned} \tag{24}$$

yielding (in conjunction with (8)) the constraint on σ_{SU} in Theorem 1. The constraint on P in the Theorem follows from Costa's result [3], because the rate R must be smaller than the capacity over a power constrained AWGN channel with known interference, $\frac{1}{2}\log_2(1+P)$. ■

It is insightful to see how the lower bound in Corollary 1 is an improvement over that in [7]. The lower bound in [7] is given by

$$MMSE(P,0) \geq \left(\left(\sqrt{\frac{\sigma^2}{\sigma^2+P+2\sigma\sqrt{P}+1}} - \sqrt{P} \right)^+ \right)^2 , \tag{25}$$

which again holds for all m. Because any γ provides a valid lower bound in Corollary 1, choosing $\gamma = 1$ in Corollary 1 provides the following (loosened) bound,

$$MMSE(P,0) \geq \inf_{|\sigma_{SU}| \leq \sigma\sqrt{P}} \left(\left(\sqrt{\frac{\sigma^2}{\sigma^2+P+2\sigma_{SU}+1}} - \sqrt{P} \right)^+ \right)^2 , \tag{26}$$

which is minimized for $\sigma_{SU} = \sigma\sqrt{P}$. This immediately yields the lower bound (25) of [7].

B. The upper bound and the tightness at $MMSE = 0$

We use the combination of linear and dirty-paper coding strategies of [7], except that we communicate a message at rate R as well. We summarize the strategy briefly, and refer the interested reader to [7] for a detailed description and analysis of the achievability.

The encoder divides its input into two parts $\mathbf{U}_{lin}^m$ and $\mathbf{U}_{dpc}^m$ of powers P_{lin} and P_{dpc} respectively, such that $P = P_{lin} + P_{dpc}$ (by construction, $\mathbf{U}_{lin}^m$ and $\mathbf{U}_{dpc}^m$ turn out to be orthogonal in the limit). We refer to

P_{lin} as the *linear* part of the power, and P_{dpc} the *dirty-paper coding* part of the power. The linear part is used to scale the host signal down by a factor β (using $\mathbf{U}_{lin}^m = -\beta \mathbf{S}^m$) so that the scaled down host signal has variance $\widetilde{\sigma}^2 = \sigma^2(1-\beta)^2$, where $\beta^2\sigma^2 = P_{lin}$. Using the remaining P_{dpc} power, the transmitter dirty-paper codes against the scaled-down host signal $(1-\beta)\mathbf{S}^m$ with the DPC parameter α [3] allowed to be arbitrary (unlike in [3], where it is eventually chosen to be the MMSE parameter).

A plain DPC strategy achieves the following rate [3, Eq. (6)]

$$R = \frac{1}{2}\log_2\left(\frac{P(P+\sigma^2+1)}{P\sigma^2(1-\alpha)^2 + P + \alpha^2\sigma^2}\right), \tag{27}$$

The strategy recovers $\mathbf{U}^m + \alpha\mathbf{S}^m$ at the decoder with high probability. Because we also have a linear part here, the achieved rate is

$$R = \frac{1}{2}\log_2\left(\frac{P_{dpc}(P_{dpc}+\widetilde{\sigma}^2+1)}{P_{dpc}\widetilde{\sigma}^2(1-\alpha)^2 + P_{dpc} + \alpha^2\widetilde{\sigma}^2}\right). \tag{28}$$

The decoder now decodes the codeword $\mathbf{U}_{dpc}^m + \alpha(1-\beta)\mathbf{S}^m$. It then performs an MMSE estimation for estimating $\mathbf{X}^m = \mathbf{S}^m + \mathbf{U}^m = (1-\beta)\mathbf{S}^m + \mathbf{U}_{dpc}^m$ using the channel output $\mathbf{Y}^m = (1-\beta)\mathbf{S}^m + \mathbf{U}_{dpc}^m + \mathbf{Z}^m$ and the decoded codeword $\alpha(1-\beta)\mathbf{S}^m + \mathbf{U}_{dpc}^m$. The obtained $MMSE$ can now be minimized over the choice of α and β under the constraint (28).

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

$$C(P) = \sup_{\sigma_{SU}\in[-\sigma\sqrt{P},0]} \frac{1}{2}\log_2\left(\frac{(P\sigma^2 - \sigma_{SU}^2)(1+\sigma^2+P+2\sigma_{SU})}{\sigma^2(\sigma^2+P+2\sigma_{SU})}\right). \tag{29}$$

Proof:

The achievability

The combination of linear and DPC-based strategies of [7] recovers $\mathbf{U}_{dpc}^m + \alpha(1-\beta)\mathbf{S}^m$ at the decoder with high probability. In order to perfectly recover $\mathbf{X}^m = (1-\beta)\mathbf{S}^m + \mathbf{U}_{dpc}^m$, we can use $\alpha = 1$, and hence the strategy would achieve a rate of

$$R_{ach} = \sup_{P_{lin},P_{dpc}:P=P_{lin}+P_{dpc}} \frac{1}{2}\log_2\left(\frac{P_{dpc}(P_{dpc}+\widetilde{\sigma}^2+1)}{P_{dpc}+\widetilde{\sigma}^2}\right), \tag{30}$$

where we take a supremum over P_{lin}, P_{dpc} such that they sum up to P. 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 $P_{dpc} + \widetilde{\sigma}^2 = P_{dpc} + \sigma^2 + P_{lin} - 2\sigma\sqrt{P_{lin}} = P + \sigma^2 + 2\sigma_{SU}$. Thus,

$$R_{ach} = \sup_{\sigma_{SU} \in [-\sigma\sqrt{P}, 0]} \frac{1}{2} \log_2 \left(\frac{\left(P - \frac{\sigma_{SU}^2}{\sigma^2}\right)(P + \sigma^2 + 2\sigma_{SU} + 1)}{P + \sigma^2 + 2\sigma_{SU}} \right). \tag{31}$$

Simple algebra shows that this expression matches that 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, $1 - \gamma^* = \frac{P + \sigma_{SU}}{\sigma^2 + P + 2\sigma_{SU}}$. Thus, we get

$$0 \geq \inf_{\sigma_{SU}} \frac{1}{\gamma^{*^2}} \left(\left(\sqrt{\frac{\sigma^2 2^{2R}}{1 + \sigma^2 + P + 2\sigma_{SU}}} - \sqrt{(1-\gamma^*)^2\sigma^2 + \gamma^{*^2} P - 2\gamma^*(1-\gamma^*)\sigma_{SU}} \right)^+ \right)^2. \tag{32}$$

It has to be the case that the term inside $(\cdot)^+$ is non-positive for some value of σ_{SU}. This immediately yields

$$\begin{aligned}
2^{2R} &\leq \sup_{\sigma_{SU}} \frac{1}{\sigma^2} \left((1-\gamma^*)^2\sigma^2 + \gamma^{*^2} P - 2\gamma^*(1-\gamma^*)\sigma_{SU} \right) (1 + \sigma^2 + P + 2\sigma_{SU}) \\
&= \sup_{\sigma_{SU}} \frac{1}{\sigma^2} \frac{((P+\sigma_{SU})^2\sigma^2 + (\sigma^2 + \sigma_{SU})^2 P - 2(P+\sigma_{SU})(\sigma^2+\sigma_{SU})\sigma_{SU})}{(\sigma^2 + P + 2\sigma_{SU})^2} (1 + \sigma^2 + P + 2\sigma_{SU}) \\
&= \sup_{\sigma_{SU}} \frac{1}{\sigma^2} \frac{(P^2\sigma^2 - \sigma_{SU}^2\sigma^2 + 2P\sigma_{SU}\sigma^2 + P\sigma^4 - P\sigma_{SU}^2 - 2\sigma_{SU}^3)}{(\sigma^2 + P + 2\sigma_{SU})^2} (1 + \sigma^2 + P + 2\sigma_{SU}) \\
&= \sup_{\sigma_{SU}} \frac{1}{\sigma^2} \frac{((P\sigma^2 - \sigma_{SU}^2)(P + \sigma^2 + 2\sigma_{SU}))}{(\sigma^2 + P + 2\sigma_{SU})^2} (1 + \sigma^2 + P + 2\sigma_{SU}) \\
&= \sup_{\sigma_{SU}} \frac{(P\sigma^2 - \sigma_{SU}^2)(1 + \sigma^2 + P + 2\sigma_{SU})}{\sigma^2(\sigma^2 + P + 2\sigma_{SU})}
\end{aligned}$$

Thus, we get the following upper bound on $C(P)$,

$$C(P) \leq \sup_{\sigma_{SU} \in [-\sigma\sqrt{P}, \sigma\sqrt{P}]} \frac{1}{2} \log_2 \left(\frac{(P\sigma^2 - \sigma_{SU}^2)(1 + \sigma^2 + P + 2\sigma_{SU})}{\sigma^2(\sigma^2 + P + 2\sigma_{SU})} \right). \tag{33}$$

The term $(P\sigma^2 - \sigma_{SU}^2)$ is oblivious to the sign of σ_{SU}. However, the term

$$\frac{1 + \sigma^2 + P + 2\sigma_{SU}}{\sigma^2 + P + 2\sigma_{SU}} = 1 + \frac{1}{\sigma^2 + P + 2\sigma_{SU}} \tag{34}$$

is clearly larger for $\sigma_{SU} < 0$ if we fix $|\sigma_{SU}|$. Thus the supremum in (33) is attained at some $\sigma_{SU} < 0$, and we get

$$C(P) \leq \sup_{\sigma_{SU} \in [-\sigma\sqrt{P}, 0]} \frac{1}{2} \log_2 \left(\frac{(P\sigma^2 - \sigma_{SU}^2)(1 + \sigma^2 + P + 2\sigma_{SU})}{\sigma^2(\sigma^2 + P + 2\sigma_{SU})} \right), \tag{35}$$

which matches the expression in (31). Thus for perfect reconstruction ($MMSE = 0$), the combination of linear and DPC strategy proposed in [7] is optimal. ■

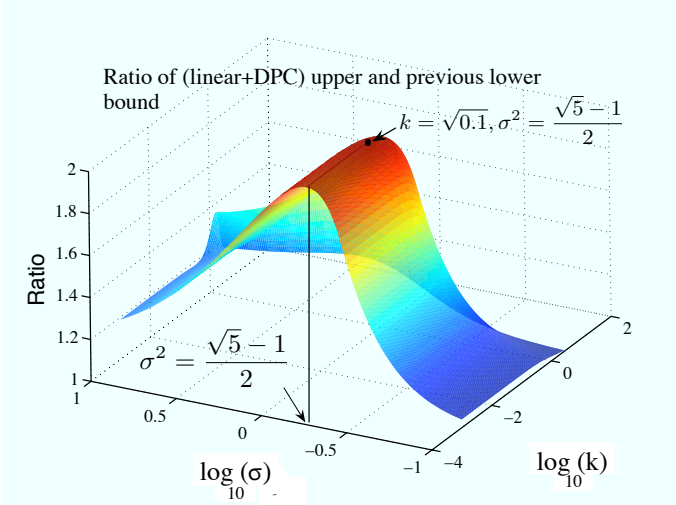
IV. Numerical results

Witsenhausen's original 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 new lower bounds (from Corollary 1) on the weighted cost is bounded by 1.3, an improvement over the ratio of 2 in [7]. A ridge of ratio 2 along $\sigma^2 = \frac{\sqrt{5}-1}{2}$ present in Fig. 2(a) (obtained using the old bound from [7]) 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 (top). Also shown in Fig. 3 (bottom) is the lack of tightness in the bounds at small P. The figure explains how this looseness results in the the ridge along $k \approx 1.67$ still surviving in the new ratio plot.

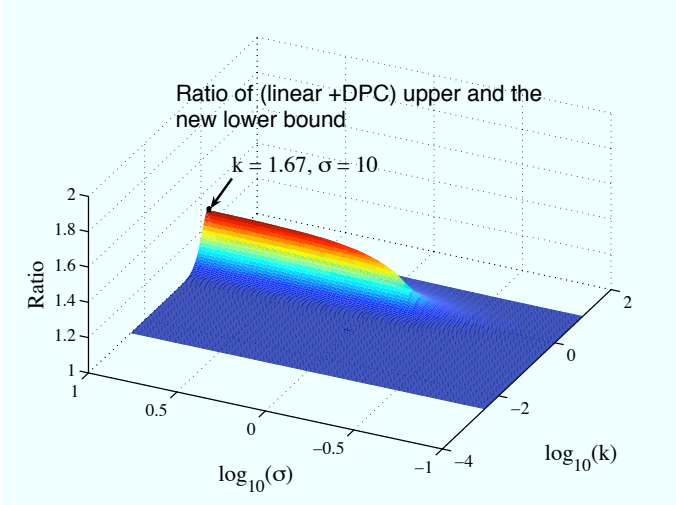
Fig. 4 shows the ratio of upper and lower bounds on $MMSE(P, 0)$ versus P and σ. While the ratio with the bound of [7] was unbounded (Fig. 4, top), the new ratio is bounded by a factor of 1.5 (Fig. 4, bottom). This is again a reflection of the tightness of the bound at small $MMSE$. A flipped perspective is shown in Fig. 5, where we compute the ratio of upper and lower bounds on required power to attain a specified $MMSE$. As a further evidence of the lack of tightness in the small-P regime, the ratio of upper and lower bounds on required power diverges to infinity along the path $MMSE = \frac{\sigma^2}{\sigma^2+1}$.

Fig. 6 shows the upper and the lower bounds for $R = 0.5$. Again, the bounds are not tight in the small-P regime — now the looseness is at the lowest power $P = 1$ at which communication at $R = 0.5$ is possible. As shown in Corollary 2, the bounds are still tight at $MMSE = 0$. Fig. 7 shows the upper and lower bounds on $MMSE$ as a function of the rate R for fixed power $P = 1$ and σ^2 equal to the Golden ratio. The figure demonstrates that beyond the maximum rate with zero distortion, the price of increasing rate is an increased distortion in the estimation of $\mathbf{X}^m$.

The MATLAB code for these figures can be found in [15].



(a)



(b)

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 [7] in (a) and from Corollary 1 in (b). As compared to the previous best known ratio of 2 [7], the ratio here is smaller than 1.3. Further, an infinitely long ridge along $\sigma^2 = \frac{\sqrt{5}-1}{2}$ and small k that is present in lower bounds of [7] 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 $\sigma = 10$.

ACKNOWLEDGMENTS

P. Grover and A. Sahai acknowledge the support of the National Science Foundation (CNS-403427, CNS-093240, CCF-0917212 and CCF-729122) and Sumitomo Electric. A. B. Wagner acknowledges the support of NSF CSF-06-42925 (CAREER) grant. We thank Hari Palaiyanur, Se Yong Park and Gireeja

Ranade for helpful discussions.

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

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

[5] 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.

[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] 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*, 2010. [Online]. Available: http://www.eecs.berkeley.edu/∼sahai/

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

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

[10] 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.

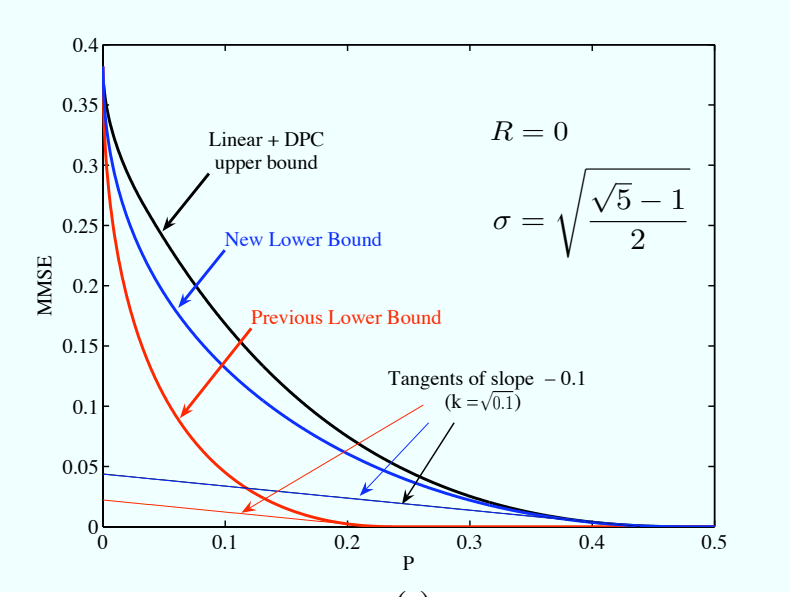
[11] P. Gupta and P. Kumar, "The capacity of wireless networks," *Information Theory, IEEE Transactions on*, vol. 46, no. 2, pp. 388–404, Mar 2000.

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

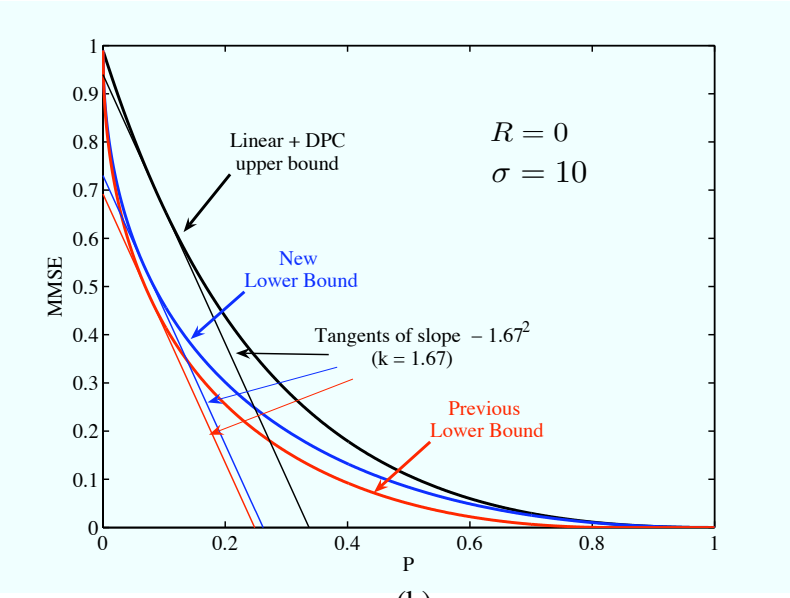
[13] A. Avestimehr, S. Diggavi, and D. Tse, "A deterministic approach to wireless relay networks," in *Proc. of the Allerton Conference on Communications, Control and Computing*, October 2007.

[14] A. S. Avestimehr, "Wireless network information flow: A deterministic approach," Ph.D. dissertation, UC Berkeley, Berkeley, CA, 2008.

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



(a)



(b)

Fig. 3. Upper and lower bounds on asymptotic $MMSE$ vs P 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 [7] and hence the ridge along $k \approx 1.67$ remains in the new ratio plot in Fig. 2.

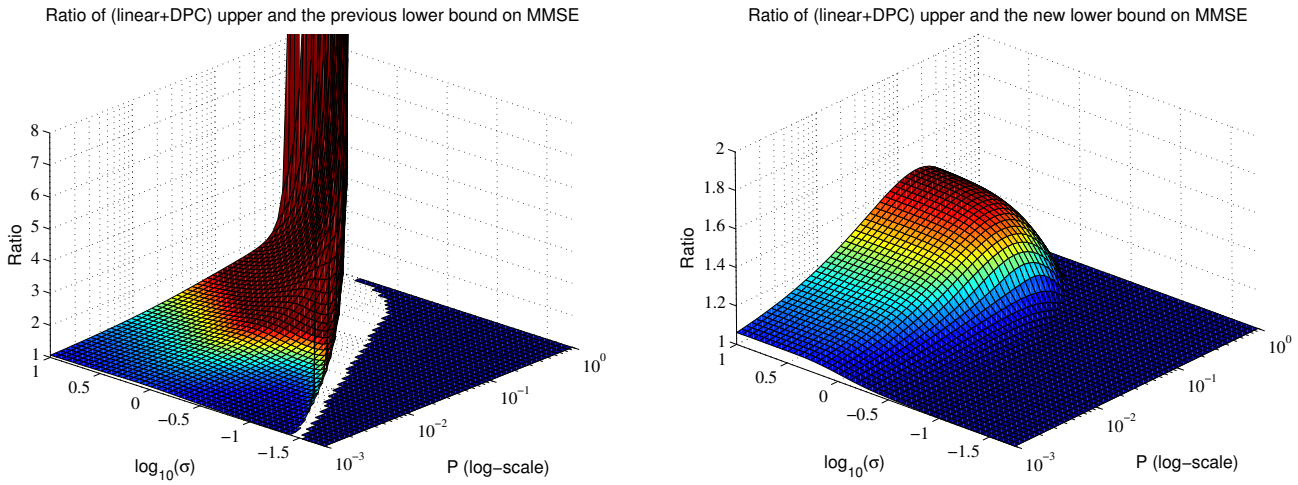


Fig. 4. Ratio of upper and lower bounds on $MMSE$ vs P and σ at $R = 0$. Whereas the ratio diverges to infinity with the old lower bound of [7] (top), it is bounded by 1.5 for the new bound (bottom). This is a consequence of the improved tightness of the new bound at small $MMSE$.

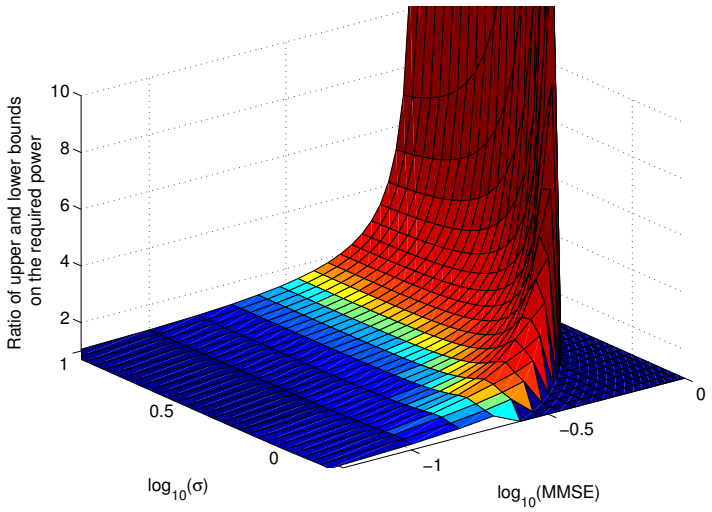


Fig. 5. Ratio of upper and lower bounds on P vs $MMSE$ and σ at $R = 0$. Interestingly, the ratio increases to infinity as $\sigma \to \infty$ along the path where P is close to zero (corresponding to $MMSE = \frac{\sigma^2}{\sigma^2+1}$.)

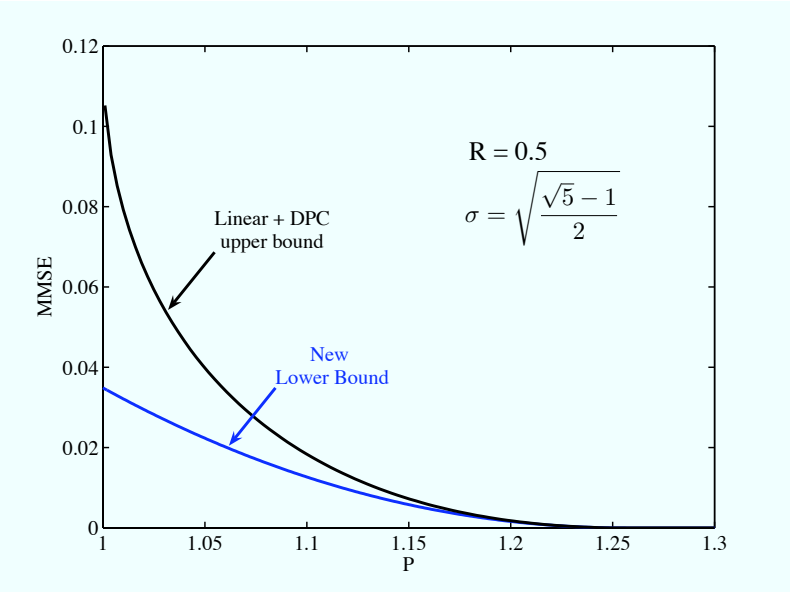


Fig. 6. Upper and lower bounds on P vs $MMSE$ for $\sigma = \sqrt{\frac{\sqrt{5}-1}{2}}$ for $R = 0.5$. Though the bounds match at $MMSE = 0$ (by Corollary 2), the bounds do not match at the minimum power ($P = 1$ here) for nonzero rates. Below $P = 1$, communication at $R = 0.5$ is not possible.

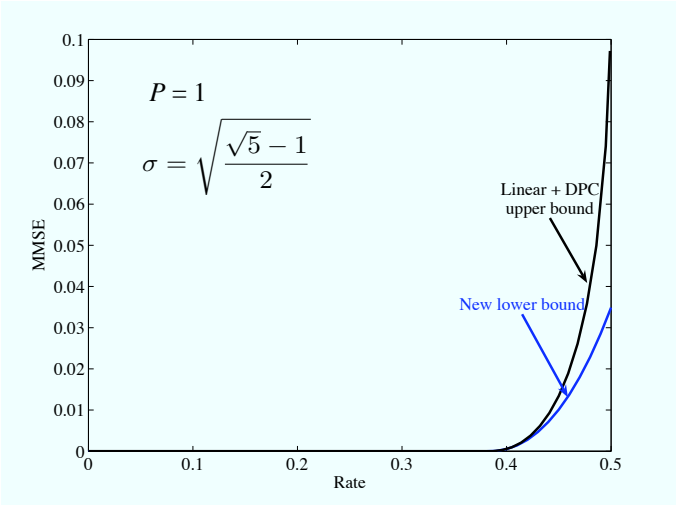


Fig. 7. Plot of upper and lower bounds on $MMSE$ vs rate for fixed power $P = 1$ and $\sigma = \sqrt{\frac{\sqrt{5}-1}{2}}$. Higher rates require higher average distortion in the reconstruction of $\mathbf{X}^m$.