On the Generalized Witsenhausen Counterexample

Pulkit Grover, Se Yong Park and Anant Sahai
Wireless Foundations, Department of EECS
University of California at Berkeley, CA-94720, USA
{pulkit, separk, sahai} @eecs.berkeley.edu

***Abstract*— Witsenhausen's counterexample is a deceptively simple distributed control problem that has remained unsolved for the last 40 years. Our recent work uses information-theoretic techniques that exploit implicit communication between the controllers to obtain characterizations of optimal costs for the Witsenhausen counterexample to within a constant factor uniformly over the problem parameters. To gain deeper insight into the nebulous concept of implicit communication, four modifications to the counterexample are considered — a zero-sum variant inspired by secrecy, two alternative orderings of the controllers, and a generalized Witsenhausen counterexample that includes quadratic costs on all states and inputs. For the first three modifications, implicit communication is either detrimental, impossible, or useless, and the optimal strategies are linear. However, nonlinear strategies outperform linear strategies by a substantial factor for the last, where implicit communication is central to strategy design. This is true even though linear strategies are no longer arbitrarily far from optimal with the inclusion of these additional cost terms.**

I. Introduction

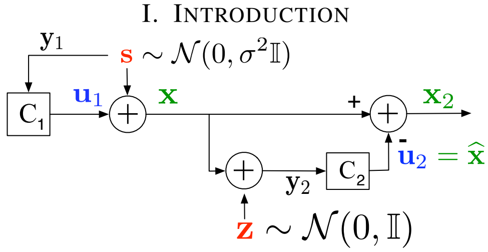

Fig. 1. Vector Witsenhausen's counterexample: the objective is to minimize the total average cost $\frac{1}{m}\mathbb{E}\left[k_w^2\|\mathbf{u}_w^m\|^2 + \|\mathbf{x}^m - \widehat{\mathbf{x}}^m\|^2\right]$. The first controller is called "weak" (w) since it has an input cost and the second is called "blurry" (b) because of its noisy observations.

Discovered in 1968 in [1], Witsenhausen's counterexample (Fig. 1) has become *the* canonical showpiece demonstrating that distributed control can be hard. Despite its simplicity, the optimal control law for the counterexample is still unknown and linear strategies can be arbitrarily bad [2]. The hardness is traditionally attributed partly to its nonconvexity, and partly to the NP-completeness of its natural[1] discrete counterpart [3].

Perhaps the larger conceptual difficulty lies in understanding the inherent possibility of implicit communication [2] between the two controllers — a feature that is ubiquitous in distributed control. Recent work [4] has shown that for an asymptotic vector version of the problem, quantization-based implicit communication strategies (generalized from [2]) attain within a factor of 4.45 of the optimal cost for all problem parameters. More sophisticated implicit communication strategies based on dirty-paper coding yield an improved factor[2] of 2. The results can be carried over to the scalar case and indeed to any finite vector length [6], [7] to obtain approximate optimality. For example, [7] shows that for the scalar case (the original counterexample), quantization-based strategies attain within a factor of 8 of the optimal cost.

This paper considers various extensions/modifications of Witsenhausen's counterexample directed towards gaining deeper insight into implicit communication and the difficulty of finding the optimal strategy. In Section II, we first consider a secrecy problem where the controllers are adversaries and so implicit communication is intuitively detrimental. We then consider two alternative orderings of the controllers: where there is no possibility (simultaneous ordering) of implicit communication or no incentive (reverse ordering). For each of these three problems, we show that linear strategies are optimal, and thereby strengthening the case for implicit communication as being the core difficulty.

In Section III we consider an extension of the counterexample that includes costs on all states and inputs. We show that even though linear strategies are no longer arbitrarily bad with inclusion of these additional costs, nonlinear strategies continue to outperform linear strategies by a substantial factor in many cases. This shows that Witsenhausen's counterexample is not an obscure corner case that exaggerates the importance of nonlinear strategies — rather it is one of many such problems in this large space.

II. Structural modifications

Bansal and Basar [8] considered modifications of Witsenhausen's counterexample with parameterized cost functions that contain Witsenhausen's counterexample as a special case. They show that whenever the cost function does not contain a product of two decision variables, affine control laws are optimal. Here we give additional interesting variations of the Witsenhausen's counterexample for which linear strategies are optimal.

A. A zero-sum variant

Inspired by [9], we convert Witsenhausen's counterexample into a zero-sum game-theoretic problem. The new problem turns out to be quite similar to the problem[3] considered

[1] Just as a DMC is the natural discrete counterpart of the Gaussian channel in information theory.

[2] The factor is further improved to 1.3 in [5] by improving the lower bound.

[3] The difference from the problem in [10] is that the net coefficient of $\mathbb{E}\left[U_w^2\right]$ can be negative in our case. Ho *et al.* state that the optimal strategies are linear for their problem as well.

in [10, Section IV].

The state transition function is the same as the original counterexample. However, the objective of the weak *jammer* is to minimize $k_w^2\mathbb{E}[U_w^2] - \mathbb{E}[X_2^2]$. Equipped with the knowledge of the weak jammer's policy, the blurry *eavesdropper* wants to minimize $E[X_2^2]$. The problem is to find $\inf_{U_w}\sup_{U_b} k_w^2\mathbb{E}[U_w^2] - \mathbb{E}[X_2^2]$. Since there is no cost constraint on U_b, the optimal policy for the blurry eavesdropper is $\mathbb{E}[X|Y_b]$. But what should the weak jammer do?

We call this problem the *Secrecy Witsenhausen Problem* since the jammer wants to hide X from the eavesdropper.

Theorem 1: Let $P^\star$ be the unique nonnegative solution t satisfying

$$k_w^2\sqrt{t}(\sigma^2+\sqrt{\sigma^2 t}+t+1)^2-\sqrt{t}-\sigma=0.$$

Then the optimal payoff of the secrecy Witsenhausen problem is

$$\begin{aligned}&\inf_{U_w}\sup_{U_b} k_w^2\mathbb{E}[U_w^2]-\mathbb{E}[X_2^2]\\ &\quad = k_w^2P^\star-\frac{\sigma^2+2\sqrt{\sigma^2P^\star}+P^\star}{\sigma^2+2\sqrt{\sigma^2P^\star}+P^\star+1},\end{aligned}$$

and is achieved by the linear strategies of state amplification and LLSE:

$$U_w=\frac{\sqrt{P^\star}}{\sigma}Y_w,\;\; U_b=\frac{\sigma^2+2\sqrt{\sigma^2P^\star}+P^\star}{\sigma^2+2\sqrt{\sigma^2P^\star}+P^\star+1}Y_b.$$

Proof: We can lower bound the payoff by forcing $U_b=\frac{\mathbb{E}[X^2]}{\mathbb{E}[X^2]+1}Y_b$.

$$\begin{aligned}&\inf_{U_w}\sup_{U_b} k_w^2\mathbb{E}[U_w^2]-\mathbb{E}[(X-U_b)^2]\\
&\geq \inf_{U_w} k_w^2\mathbb{E}[U_w^2]-\mathbb{E}\left[\left(X-\frac{\mathbb{E}[X^2]}{\mathbb{E}[X^2]+1}Y_b\right)^2\right]\\
&\geq \min_{\mathbb{E}[U_w^2]} k_w^2\mathbb{E}[U_w^2]-1\\
&\quad+\frac{1}{\mathbb{E}[S^2]+2\sqrt{\mathbb{E}[S^2]\mathbb{E}[U_w^2]}+\mathbb{E}[U_w^2]+1} \qquad (1)\\
&= \min_{\mathbb{E}[U_w^2]} k_w^2\mathbb{E}[U_w^2]-1\\
&\quad+\frac{1}{\sigma^2+2\sqrt{\sigma^2\mathbb{E}[U_w^2]}+\mathbb{E}[U_w^2]+1} \qquad (2)\\
&= k_w^2P^\star-\frac{\sigma^2+2\sqrt{\sigma^2P^\star}+P^\star}{\sigma^2+2\sqrt{\sigma^2P^\star}+P^\star+1} \qquad (3)\end{aligned}$$

(1) follows from the Cauchy-Schwartz inequality. The optimization problem (2) is convex. To see this, consider the function $k_w^2t+\frac{1}{\sigma^2+2\sqrt{\sigma^2t}+t+1}$ for $t\geq 0$. The first term is linear in t and the second term is convex-∪ since it is a composition of a nonincreasing convex-∪ function $\frac{1}{t}(t>0)$ and a concave-∩ function $\sigma^2+2\sqrt{\sigma^2t}+t+1$ [11, Pg. 84]. The minimum is achieved by the unique $P^\star(\geq 0)$ satisfying

$$\begin{aligned}0&=\frac{d}{dt}\left(k_w^2t+\frac{1}{\sigma^2+2\sqrt{\sigma^2t}+t+1}\right)\\
&=k_w^2-\frac{\frac{\sigma}{\sqrt{t}}+1}{\left(\sigma^2+2\sqrt{\sigma^2t}+t+1\right)^2}\\
&\Leftrightarrow k_w^2\sqrt{t}(\sigma^2+\sqrt{\sigma^2t}+t+1)^2-\sqrt{t}-\sigma=0. \qquad (4)\end{aligned}$$

We can upper bound the payoff by forcing $U_w=\frac{\sqrt{P^\star}}{\sigma}S$.

$$\begin{aligned}&\inf_{U_w}\sup_{U_b} k_w^2\mathbb{E}[U_w^2]-\mathbb{E}[X_2^2]\\
&\leq \sup_{U_b} k_w^2\mathbb{E}\left[\left(\frac{\sqrt{P^\star}}{\sigma}S\right)^2\right]-\mathbb{E}[X_2^2]\\
&= k_w^2P^\star-\inf_{U_b}\mathbb{E}\left[\left(\left(1+\frac{\sqrt{P^\star}}{\sigma}\right)S-U_b\right)^2\right]\\
&= k_w^2P^\star-\frac{\sigma^2+2\sqrt{\sigma^2P^\star}+P^\star}{\sigma^2+2\sqrt{\sigma^2P^\star}+P^\star+1} \qquad (5)\end{aligned}$$

(5) comes from MMSE $\left[\left(1+\frac{\sqrt{P^\star}}{\sigma}\right)S \mid \left(1+\frac{\sqrt{P^\star}}{\sigma}\right)S+Z\right]=\frac{\sigma^2+2\sqrt{\sigma^2P^\star}+P^\star}{\sigma^2+2\sqrt{\sigma^2P^\star}+P^\star+1}$.

By (3) and (5) the theorem is proved. ■

B. Reordering the controllers in Witsenhausen's counterexample

1) Simultaneous Control: Consider the case of both controllers operating simultaneously. Formally, let the underlying random variables be $S\sim\mathcal{N}(0,\sigma^2)$ and $Z\sim\mathcal{N}(0,1)$. The state transition equation is $X_2=S+U_w-U_b$. The controllers' observations are $Y_w=S$ and $Y_b=S+Z$, so $U_w=\gamma_w(Y_w)$ and $U_b=\gamma_b(Y_b)$. The object is to minimize the same objective function: $k_w^2\mathbb{E}[U_w^2]+\mathbb{E}[X_2^2]$. We call this the *Simultaneous Control Problem*.

R. Radner solved this in a generalized form [12, Theorem 5]: If every primitive random variable is jointly Gaussian and the objective function is a convex quadratic function in the product space of measurable functions of each controller's observation, the optimal strategy is linear.

Corollary 1: [12, Theorem 5] The optimal cost of the simultaneous control problem is

$$\inf_{U_w,U_b} k_w^2\mathbb{E}[U_w^2]+\mathbb{E}[X_2^2]=\frac{k_w^2\sigma^2}{k_w^2+1+k_w^2\sigma^2}$$

and is achieved by the linear control strategies, $U_w=\frac{-1}{k_w^2+1+k_w^2\sigma^2}Y_w$ and $U_b=\frac{k_w^2\sigma^2}{k_w^2+1+k_w^2\sigma^2}Y_b$.

Proof: S, Z, Y_w, and Y_b are jointly Gaussian. Moreover,

$$\begin{aligned}&\min_{U_w,U_b} k_w^2\mathbb{E}[U_w^2]+\mathbb{E}[(S+U_w-U_b)^2]\\
&=\min_{U_w,U_b}\mathbb{E}\Bigg[\begin{bmatrix}U_w & U_b\end{bmatrix}\begin{bmatrix}1+k_w^2 & -1\\ -1 & 1\end{bmatrix}\begin{bmatrix}U_w\\ U_b\end{bmatrix}\\
&\quad -2\begin{bmatrix}U_w & U_b\end{bmatrix}\begin{bmatrix}-S\\ S\end{bmatrix}+S^2\Bigg].\end{aligned}$$

Since $\begin{bmatrix} 1+k_w^2 & -1 \\ -1 & 1 \end{bmatrix}$ is positive definite (both the determinant and the trace are strictly positive except in the trivial case $k = 0$) the optimal strategy is linear [12, Theorem 5]. Let $U_w = b_w Y_w$ and $U_b = b_b Y_b$.

$$\begin{aligned}
&\min_{b_w, b_b} k_w^2 \mathbb{E}\left[U_w^2\right] + \mathbb{E}\left[X_2^2\right] \\
&= \min_{b_w, b_b} \begin{bmatrix} b_w & b_b \end{bmatrix} \begin{bmatrix} k_w^2\sigma^2 + \sigma^2 & -\sigma^2 \\ -\sigma^2 & 1+\sigma^2 \end{bmatrix} \begin{bmatrix} b_w \\ b_b \end{bmatrix} \\
&\quad - 2 \begin{bmatrix} b_w & b_b \end{bmatrix} \begin{bmatrix} -\sigma^2 \\ \sigma^2 \end{bmatrix} + \sigma^2 \\
&= \frac{k_w^2\sigma^2}{k_w^2 + 1 + k_w^2\sigma^2}
\end{aligned}$$

where the minimum is achieved by

$$b_w = \frac{-1}{k_w^2 + 1 + k_w^2\sigma^2}, \quad b_b = \frac{k_w^2\sigma^2}{k_w^2 + 1 + k_w^2\sigma^2}.$$

■

Notice that there is no possibility for implicit communication since neither controller sees anything that depends directly on the other's actions.

2) Reverse Witsenhausen: Another natural variation of Witsenhausen's counterexample is reversing the order of the two controllers. We call this problem the *Reverse Witsenhausen Problem*. As before, the underlying random variables are $S \sim \mathcal{N}(0, \sigma^2)$ and $Z \sim \mathcal{N}(0, 1)$. The state transition equations are $X = S - U_b$ and $X_2 = X + U_w$. The controllers' observations are $Y_b = S + Z$ and $Y_w = X$, so $U_b = \gamma_b(Y_b)$ and $U_w = \gamma_w(Y_w)$. The objective is minimizing the same objective function: $k_w^2 \mathbb{E}[U_w^2] + \mathbb{E}[X_2^2]$.

Theorem 2: The optimal cost for the reverse Witsenhausen problem is

$$\inf_{U_w, U_b} k_w^2 \mathbb{E}[U_w^2] + \mathbb{E}[X_2^2] = \left(\frac{k_w^2}{k_w^2 + 1}\right) \frac{\sigma^2}{\sigma^2 + 1}$$

and is achieved by the linear control strategies: $U_w = -\frac{1}{k_w^2+1} Y_w$ and MMSE, $U_b = \frac{\sigma^2}{\sigma^2+1} Y_b$.

Proof:

$$\begin{aligned}
&\inf_{U_w, U_b} k_w^2 \mathbb{E}[U_w^2] + \mathbb{E}[X_2^2] \\
&= \inf_{U_w, U_b} k_w^2 \mathbb{E}[U_w^2] + \mathbb{E}\left[(S - U_b)^2\right] \\
&\quad + 2\mathbb{E}\left[U_w (S - U_b)\right] + \mathbb{E}[U_w^2] \\
&\geq \inf_{U_w, U_b} k_w^2 \mathbb{E}[U_w^2] + \mathbb{E}\left[(S - U_b)^2\right] \\
&\quad - 2\sqrt{\mathbb{E}\left[U_w^2\right]} \sqrt{\mathbb{E}\left[(S - U_b)^2\right]} + \mathbb{E}[U_w^2] \qquad (6) \\
&= \inf_{P > 0, U_b} k_w^2 P + \left(\sqrt{\mathbb{E}\left[(S - U_b)^2\right]} - \sqrt{P}\right)^2 \qquad (7) \\
&\geq \min_{P > 0} k_w^2 P + \left(\sqrt{\frac{\sigma^2}{\sigma^2 + 1}} - \sqrt{P}\right)^2 \qquad (8) \\
&= \left(\frac{k_w^2}{k_w^2 + 1}\right) \frac{\sigma^2}{\sigma^2 + 1}
\end{aligned}$$

(6) follows from the Cauchy-Schwartz inequality. We can get (7) by denoting $\mathbb{E}[U_w^2]$ as P. (8) follows from $MMSE[S|Y_b] = \frac{\sigma^2}{\sigma^2+1}$ and its minimum is achieved by $P = \left(\frac{1}{(k_w^2+1)^2}\right) \frac{\sigma^2}{\sigma^2+1}$.

Moreover, the inequalities (6) and (8) are tight when U_w and U_b are chosen as $-\frac{1}{k_w^2+1} Y_w$ and $\frac{\sigma^2}{\sigma^2+1} Y_b$ respectively. Therefore, the lower bound is achievable. ■

Notice that although the blurry controller could conceivably communicate to the weak one, there is nothing that the weak one wants to learn. It already has a clear view of everything that impacts the cost function.

3) Comparing the various controller orderings: The freedom in placing the distributed controllers can be thought of as another design parameter. To understand this effect, we compare the performance of the three problems: Witsenhausen's counterexample, the simultaneous control problem and the reverse Witsenhausen problem. To distinguish the three problems, we use superscripts (W), (S) and (R) respectively.

We can easily notice that the reverse Witsenhausen problem has lower costs than the simultaneous control problem.

Corollary 2: $J^{(R)}(k, \sigma) \leq J^{(S)}(k, \sigma), \quad \forall k, \sigma > 0$

Proof: Direct calculation from Corollary 1 and Theorem 2. ■

This result can be intuitively understood. In the simultaneous control problem, the weak controller C_w does not know anything about the blurry controller C_b's observation noise Z, and so the uncertainty of Z shows up in X_2. In the reverse Witsenhausen problem, however, C_w can compensate for the noise Z that was added into X since it acts later with a perfect observation.

This result may seem obvious since the reverse Witsenhausen problem uses two time slots while the simultaneous control problem uses only one time slot. However, this intuition fails in the comparison between Witsenhausen's counterexample and the simultaneous control problem. The preference between the two architectures varies with the parameters (σ, k_w).

When $k = 0.1$ and $\sigma = 10$,

$$J^{(S)}(0.1, 10) = 0.4975 > 0.1468 \geq J^{(W)}(0.1, 10)$$

where the upper bound of the Witsenhausen's problem is the evaluation of [6, Theorem 2] (with quantization bin-size $B = 7.074$).

When $k = 1, \sigma = 0.5$,

$$J^{(S)}(1, 0.5) = 0.1111 < 0.1153 \leq J^{(W)}(1, 0.5)$$

where the lower bound of the Witsenhausen's problem is the evaluation of [5, Corollary 1].

To gain some insight, consider the case of small k_w and large σ. In Witsenhausen's counterexample, C_w can reject the noise Z using a quantization strategy [2]. The required power for this "non-causal" disturbance rejection is not significantly greater than the variance of the noise, 1. However, in the simultaneous control problem C_w does not know about the observation noise Z and there is no second stage either. Thus, to make X_2 very small, C_w has to directly reduce the initial state S and that costs on the order of σ^2.

The situation is changed when σ is small. In Witsenhausen's

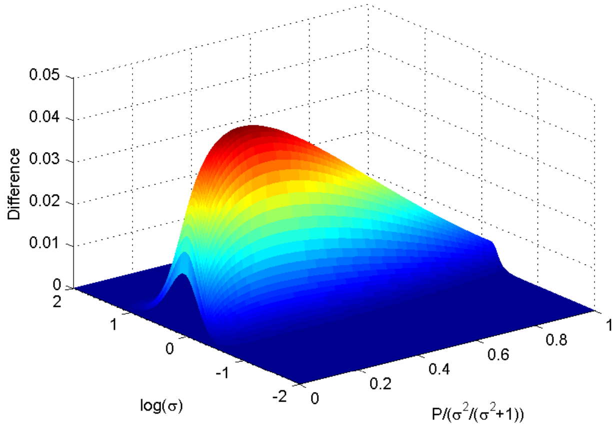

Fig. 2. The difference between a lower bound on the cost for Witsenhausen's problem and the optimal cost for the reversed Witsenhausen problem.

counterexample, C_w would choose to decrease S rather than to reject the noise Z that has a larger variance. This will result in decreasing the Signal-to-Noise (SNR) ratio at C_b. However, in the simultaneous problem any u_w does not effect the observations of C_b. Thus, the SNR at C_b remains the same and the cost of the simultaneous control problem is smaller than that of Witsenhausen's counterexample.
In Fig. 2, numerical calculations show that the lower bound of Witsenhausen's problem [5, Corollary 1] is always larger than the cost of the reverse Witsenhausen problem. Unfortunately, we do not yet have an analytical proof that gives any more insight into why this should be is true.

III. Costs on various states and inputs

In this section, we explore extensions of Witsenhausen's counterexample by adding costs on the remaining terms $\mathbf{s}^m$, $\mathbf{x}^m$ and $\widehat{\mathbf{x}}^m$. The objective for the controllers is to minimize

$$J(k_s, k_w, k_x, k_b, \sigma) = \tfrac{1}{m}\mathbb{E}\big[k_s^2\|\mathbf{S}^m\|^2 + k_w^2\|\mathbf{U}_w^m\|^2 + k_x^2\|\mathbf{X}^m\|^2 + k_b^2\|\widehat{\mathbf{X}}^m\|^2 + \|\mathbf{X}^m - \widehat{\mathbf{X}}^m\|^2\big], \quad (9)$$

over the choice of control laws. The problem is denoted by $W(k_s, k_w, k_x, k_b, \sigma)$. The asymptotically (as $m \to \infty$) optimal cost for given problem parameters will be denoted by $J^\star(k_s, k_w, k_x, k_b, \sigma)$. Note that the original counterexample corresponds to $k_s = k_x = k_b = 0$ and $m = 1$. By deriving lower and upper bounds on the total cost, we will show that vector-quantization-based nonlinear strategies achieve within a constant factor of the optimal cost even with nonzero k_s, k_x or k_b. The lower bounds will be expressed in variational form as an optimization over some variables, e.g. P. $P^\star$ will be used to denote the optimizing value of the lower bound (and a similar notation will be followed for other variables).

We first need the following simple lemma.

Lemma 1: For any positive real numbers a, b, c, d,

$$\frac{a+c}{b+d} \leq \max\left\{\frac{a}{b}, \frac{c}{d}\right\}. \quad (10)$$

Proof: Without loss of generality, assume that $\frac{c}{d} \leq \frac{a}{b}$. Then, $\frac{a+c}{b+d} = \frac{\frac{a}{b}+\frac{c}{b}}{1+\frac{d}{b}} = \frac{\frac{a}{b}+\frac{c}{d}\times\frac{d}{b}}{1+\frac{d}{b}} \overset{\frac{c}{d}\leq\frac{a}{b}}{\leq} \frac{\frac{a}{b}+\frac{a}{b}\times\frac{d}{b}}{1+\frac{d}{b}} = \frac{a}{b}$. ■

A. Cost on the initial state $\mathbf{s}^m$

Theorem 3: For $W(k_s, k_w, 0, 0, \sigma)$,

$$\min_P k_s^2\sigma^2 + k_w^2 P + \nu_w(P) \leq J^\star(k_s, k_w, 0, 0, \sigma) \leq 11\left(\min_P k_s^2\sigma^2 + k_w^2 P + \nu_w(P)\right),$$

where $\nu_w(P) = \left(\left(\sqrt{\kappa(P)} - \sqrt{P}\right)^+\right)^2$, and $\kappa(P) = \frac{\sigma^2}{(\sigma+\sqrt{P})^2+1}$.

Proof: Since $\mathbb{E}\left[\|\mathbf{S}^m\|^2\right] = m\sigma^2$, and $\mathbf{S}^m$ is unchanged by the control strategy, this constant term $k_s^2\sigma^2$ is added to both the upper bound (obtained using either simple linear or vector-quantization-based strategies) and the lower bound (of [4, Theorem 3]).

It was shown in [4, Theorem 1] that the ratio of the upper and lower bounds is smaller than 11 for $k_s = 0$. In Lemma 1, denoting by a and b the upper and lower bounds *without* the cost on the initial state, and using $c = d = k_s^2\sigma^2$, the new ratio $\frac{a+c}{b+d}$ is bounded by $\max\left\{\frac{a}{b}, \frac{c}{d}\right\} \leq \max\{11, 1\} = 11$. ■

B. Cost on the state $\mathbf{x}^m$

Theorem 4: For $W(0, k_w, k_x, 0, \sigma)$,

$$J^\star(0, k_w, k_x, 0, \sigma) \geq \min\ k_w^2 P + k_x^2 P_x + \nu_b(P, P_x),$$

$\nu_b(P, P_x) = \left(\left(\sqrt{\frac{\sigma^2}{P_x+1}} - \sqrt{P}\right)^+\right)^2$, where the minimum is over $P \geq 0$ and P_x satisfying $\left((\sigma - \sqrt{P})^+\right)^2 \leq P_x \leq (\sigma + \sqrt{P})^2$.

Proof: Let $\frac{1}{m}\mathbb{E}\left[\|\mathbf{U}_w^m\|^2\right] \leq P$ and $P_x := \frac{1}{m}\mathbb{E}\left[\|\mathbf{X}^m\|^2\right]$. For given P and P_x, we will obtain a lower bound on $\frac{1}{m}\mathbb{E}\left[\|\mathbf{X}^m - \widehat{\mathbf{X}}^m\|^2\right]$. We will use the following [4, Lemma 3],

$$\frac{1}{m}\mathbb{E}\left[\|\mathbf{X}^m - \widehat{\mathbf{X}}^m\|^2\right] \geq \left(\left(\sqrt{\mathbb{E}\left[\frac{1}{m}\|\mathbf{S}^m - \widehat{\mathbf{X}}^m\|^2\right]} - \sqrt{P}\right)^+\right)^2. \quad (11)$$

Viewing $\widehat{\mathbf{X}}^m$ as an estimate for $\mathbf{S}^m$, and following the proof of [4, Theorem 1], $\mathbb{E}\left[\|\mathbf{S}^m - \widehat{\mathbf{X}}^m\|^2\right] \geq \frac{\sigma^2}{P_x+1}$. The term $\nu_b(P, P_x)$ in Theorem 4 is now obtained using (11). It remains to show that P_x lies within the specified limits. This follows from the observation that the maximum value of P_x is obtained when the input $\mathbf{u}_w^m$ is aligned with the initial state, and minimum when it is aligned opposite to the initial state (and if $P \geq \sigma^2$, P_x can be forced to zero, its minimum value). ■

Theorem 5: For $W(0, k_w, k_x, 0, \sigma)$,

$$\left(\min_{P,P_x} k_w^2 P + k_x^2 P_x + \nu_b(P, P_x)\right) \leq J^\star(0, k_w, k_x, 0, \sigma) \leq 11\left(\min_{P,P_x} k_w^2 P + k_x^2 P_x + \nu_b(P, P_x)\right),$$

for $P \geq 0$ and P_x satisfying $\left((\sigma-\sqrt{P})^+\right)^2 \leq P_x \leq (\sigma+\sqrt{P})^2$, and where $\nu_b(P,P_x) = \left(\left(\sqrt{\frac{\sigma^2}{P_x+1}}-\sqrt{P}\right)^+\right)^2$.

Proof: Consider the cost $\frac{1}{m}k_x^2\mathbb{E}\left[\|\mathbf{x}^m\|\right]^2$, that corresponds to the middle term in (12).

For $P^\star > \frac{\sigma^2}{9}$, the lower bound is larger than $\frac{k_w^2\sigma^2}{9}$. Thus the ratio of the costs attained using the zero-forcing upper bound, $k_w^2\sigma^2$ and the lower bound is smaller than 9.

For $P^\star \leq \frac{\sigma^2}{9}$, we first further lower bound the expression in Theorem 4 by

$$J^\star(0,k_w,k_x,0,\sigma) \geq \min_{P\geq 0}\ k_w^2 P + k_x^2((\sigma-\sqrt{P})^+)^2 + \left(\left(\sqrt{\frac{\sigma^2}{(\sigma+\sqrt{P})^2+1}}-\sqrt{P}\right)^+\right)^2, \tag{12}$$

since $((\sigma-\sqrt{P})^+)^2 \leq P_x \leq (\sigma+\sqrt{P})^2$.

Since $P^\star \leq \frac{\sigma^2}{9}$, $(\sigma-\sqrt{P^\star})^2 \geq \frac{4}{9}\sigma^2 = d$. Further, for any power $P(<\sigma^2)$ in the upper bound, the corresponding cost term in the upper bound $(\sigma+\sqrt{P})^2 \leq 4\sigma^2 = c$. Using Lemma 1, this term can increase the ratio to a maximum of $\frac{c}{d} = 9$. However, without the middle term in (12), the ratio is bounded by $\frac{a}{b} = 11$ [4, Thm. 1]. Thus the overall ratio is still bounded by 11. ■

C. Cost on the input $\mathbf{u}_b^m = \widehat{\mathbf{x}}^m$

Theorem 6: For $W(0,k_w,0,k_b,\sigma)$,

$$J^\star(0,k_w,0,k_b,\sigma) \geq \inf_{P\geq 0,\zeta\geq 0} k_w^2 P + k_b^2\zeta + \left(\left(\sqrt{\eta(P,\zeta)}-\sqrt{P}\right)^+\right)^2,$$

where,

$$\eta(P,\zeta) = \begin{cases} \left(1-\sqrt{\frac{\bar{P}}{\bar{P}+1}}\frac{\sqrt{\zeta}}{\sigma}\right)^2\sigma^2 + \frac{\zeta}{\bar{P}+1} & \text{for } \zeta < \frac{\bar{P}\sigma^2}{\bar{P}+1} \\ \frac{\sigma^2}{\bar{P}+1} & \text{otherwise.} \end{cases} \tag{13}$$

where $\bar{P} = \left(\sigma+\sqrt{P}\right)^2$.

Proof: As in Theorem 5, we again use [4, Lemma 3] to lower bound $\frac{1}{m}\mathbb{E}\left[\|\mathbf{S}^m-\widehat{\mathbf{X}}^m\|^2\right]$. Let $\zeta := \frac{1}{m}\mathbb{E}\left[\|\widehat{\mathbf{X}}^m\|^2\right]$. By an application of the data-processing inequality [13, Pg. 32], $I(\mathbf{S}^m,\widehat{\mathbf{X}}^m) \leq I(\mathbf{X}^m,\mathbf{Y}_b^m)$. Let $\bar{C} = \frac{1}{2}\log_b\left(1+\bar{P}\right)$ denote the capacity of an AWGN channel with input power constraint $\bar{P}$ and noise variance 1. Then, $I(\mathbf{X}^m,\mathbf{Y}_b^m) \leq m\bar{C}$, and thus, $I(\mathbf{S}^m,\widehat{\mathbf{X}}^m) \leq m\bar{C}$. Imposing this constraint along with the power constraint on $\widehat{\mathbf{X}}^m$, for any control strategy,

$$\frac{1}{m}\mathbb{E}\left[\|\mathbf{S}^m-\widehat{\mathbf{X}}^m\|^2\right] \geq \inf_{\substack{p(\widehat{\mathbf{X}}^m|\mathbf{S}^m):\frac{1}{m}I(\mathbf{S}^m;\widehat{\mathbf{X}}^m)\leq\bar{C} \\ \frac{1}{m}\mathbb{E}\left[\|\widehat{\mathbf{X}}^m\|^2\right]\leq\zeta}} \mathbb{E}\left[d(\mathbf{S}^m,\widehat{\mathbf{X}}^m)\right]$$

Define Q as a discrete random variable (playing the role of time-sharing random variable) distributed uniformly on the dimensions $\{1,2,\ldots,m\}$. Define $S = S_Q$, $X = X_Q$, $Y_b = Y_{2,Q}$, and $\widehat{X} = \widehat{X}_Q$. Then,

$$\begin{aligned}
&\frac{1}{m}\mathbb{E}\left[\|\mathbf{S}^m-\widehat{\mathbf{X}}^m\|^2\right] \\
&\overset{(a)}{\geq} \inf_{\substack{p(\widehat{\mathbf{X}}^m|\mathbf{S}^m):I(S;\widehat{X})\leq\bar{C} \\ \frac{1}{m}\mathbb{E}\left[\|\widehat{\mathbf{X}}^m\|^2\right]\leq\zeta}} \mathbb{E}\left[d(\mathbf{S}^m,\widehat{\mathbf{X}}^m)\right] \\
&\overset{(b)}{=} \inf_{\substack{p(\widehat{\mathbf{X}}^m|\mathbf{S}^m): \\ I(S;\widehat{X})\leq\bar{C} \\ \frac{1}{m}\sum_{i=1}^m\mathbb{E}\left[\widehat{X}_i^2\right]\leq\zeta}} \frac{1}{m}\sum_{i=1}^m\mathbb{E}\left[d(S_i,\widehat{X}_i)\right] \\
&\overset{(c)}{=} \inf_{\substack{p(\widehat{\mathbf{X}}^m|\mathbf{S}^m):I(S;\widehat{X})\leq\bar{C} \\ \mathbb{E}\left[\widehat{X}^2\right]\leq\zeta}} \mathbb{E}\left[d(S,\widehat{X})\right] \\
&\overset{(d)}{=} \inf_{\substack{p(\widehat{X}|S):I(S;\widehat{X})\leq\bar{C} \\ \mathbb{E}\left[\widehat{X}^2\right]\leq\zeta}} \mathbb{E}\left[d(S,\widehat{X})\right]
\end{aligned} \tag{14}$$

Here (a) follows from the fact that $I(S;\widehat{X}) \leq \frac{1}{m}I(\mathbf{S}^m;\widehat{\mathbf{X}}^m)$ (proven below), (b) follows from the definition of the vector distortion and second norm, (c) follows from the the definition of the time-sharing random variable Q, and (d) uses the fact that the expressions in the objective and the constraints depend on $p(\widehat{\mathbf{X}}^m|\mathbf{S}^m)$ only through $p(\widehat{X}|S)$. To see the inequality used in (a),

$$\begin{aligned}
I(\mathbf{S}^m;\widehat{\mathbf{X}}^m) &= h(\mathbf{S}^m) - h(\mathbf{S}^m|\widehat{\mathbf{X}}^m) \\
&= \sum_{i=1}^m\left(h(S_i) - h(S_i|S^{i-1},\widehat{\mathbf{X}}^m)\right) \\
&\overset{(e)}{\geq} \sum_{i=1}^m\left(h(S_i) - h(S_i|\widehat{X}_i)\right) \\
&= \sum_{i=1}^m I(S_i;\widehat{X}_i) = mI(S;\widehat{X}|Q) \\
&= m\left(h(S|Q) - h(S|\widehat{X},Q)\right) \\
&= m\left(h(S) - h(S|\widehat{X},Q)\right) \\
&\overset{(f)}{\geq} m\left(h(S) - h(S|\widehat{X})\right) = mI(S;\widehat{X}),
\end{aligned}$$

where (e) and (f) use the fact that conditioning reduces differential entropy. From (14),

$$\begin{aligned}
&\frac{1}{m}\mathbb{E}\left[\|\mathbf{S}^m-\widehat{\mathbf{X}}^m\|^2\right] \\
&\geq \inf_{\substack{p(\widehat{X}|S):I(S;\widehat{X})\leq\bar{C} \\ \mathbb{E}\left[\widehat{X}^2\right]\leq\zeta}} \mathbb{E}\left[d(S,\widehat{X})\right] \\
&= \inf_{\substack{p(\widehat{X}|S): \\ I(S;\widehat{X})\leq\bar{C}}} \sup_{\lambda\geq 0} \mathbb{E}\left[(S-\widehat{X})^2\right] + \lambda\left(\mathbb{E}\left[\widehat{X}^2\right]-\zeta\right) \\
&\overset{(g)}{=} \sup_{\lambda\geq 0} \inf_{\substack{p(\widehat{X}|S): \\ I(S;\widehat{X})\leq\bar{C}}} \mathbb{E}\left[(S-\widehat{X})^2\right] + \lambda\left(\mathbb{E}\left[\widehat{X}^2\right]-\zeta\right)
\end{aligned}$$

$$= \sup_{\lambda\geq 0}\left(\inf_{p(\widehat{X}|S):I(S;\widehat{X})\leq\bar{C}}(\lambda+1)\mathbb{E}\left[\widehat{X}^2\right]-2\mathbb{E}\left[S\widehat{X}\right]\right)$$
$$-\lambda\zeta+\sigma^2$$
$$= \sup_{\lambda\geq 0}(\lambda+1)\inf_{p(\widehat{X}|S):I(S;\widehat{X})\leq\bar{C}}\mathbb{E}\left[\left(\widehat{X}-\frac{S}{\lambda+1}\right)^2\right]$$
$$-\lambda\zeta+\sigma^2\left(1-\frac{1}{\lambda+1}\right)$$
$$\overset{(h)}{\geq} \sup_{\lambda\geq 0}(\lambda+1)\left(\frac{\frac{\sigma^2}{(\lambda+1)^2}}{\bar{P}+1}\right)-\lambda\zeta+\frac{\lambda\sigma^2}{\lambda+1}$$
$$= \sup_{\lambda\geq 0}\frac{\sigma^2}{(\lambda+1)(\bar{P}+1)}-\lambda\zeta+\frac{\lambda\sigma^2}{\lambda+1}, \quad (15)$$

where (g) follows from strong duality in convex optimization [11, Pg. 226] (the optimization problem in (15) is convex [11, Pg. 1–2] because for fixed $p(S)$, $I(S;\widehat{X})$ is a convex-∪ function of $p(\widehat{X}|S)$, and the other constraint and the objective function are linear functions of $p(\widehat{X}|S)$), (h) follows from the fact that "uncoded" transmission is optimal for the optimization problem inside the brackets (it is the problem of minimizing the distortion of the Gaussian source $\frac{\mathbf{S}^m}{1+\lambda}$ across an AWGN channel subject to a power constraint of $\bar{P}$ on the channel input).

To optimize over λ, we differentiate w.r.t. λ and find where the derivative is zero.

$$\frac{\partial}{\partial\lambda}\left(\frac{\sigma^2}{(\lambda+1)(\bar{P}+1)}-\lambda\zeta+\left(1-\frac{1}{\lambda+1}\right)\sigma^2\right)=0$$
$$\text{i.e. } -\frac{\sigma^2}{(\lambda+1)^2(\bar{P}+1)}-\zeta+\frac{\sigma^2}{(\lambda+1)^2}=0$$
$$\text{i.e. } \frac{\sigma^2}{(\lambda+1)^2}\left(1-\frac{1}{\bar{P}+1}\right)=\zeta$$
$$\text{i.e.} \frac{\bar{P}\sigma^2}{(\bar{P}+1)(\lambda+1)^2}=\zeta\Rightarrow(\lambda+1)^2=\frac{\sigma^2\bar{P}}{(\bar{P}+1)\zeta}.$$

Double differentiate in order to verify its concavity:

$$\frac{\partial^2}{\partial\lambda^2}\left(\frac{\sigma^2}{(\lambda+1)(\bar{P}+1)}-\lambda\zeta+\left(1-\frac{1}{\lambda+1}\right)\sigma^2\right)$$
$$= 2\frac{\sigma^2}{(\lambda+1)^3(\bar{P}+1)}-2\frac{\sigma^2}{(\lambda+1)^3}$$
$$= -\frac{2\sigma^2\bar{P}}{(\lambda+1)^3(\bar{P}+1)}<0 \text{ because } \lambda\geq 0.$$

The maximizing value of λ is thus $\lambda^\star=\left(\sqrt{\frac{\bar{P}\sigma^2}{(\bar{P}+1)\zeta}}-1\right)^+$. Thus, for $\lambda^\star>0$ (which corresponds to $\zeta<\frac{\bar{P}\sigma^2}{\bar{P}+1}$), the optimizing $\widehat{X}$ is given by

$$\widehat{X}^\star = \frac{\sigma\sqrt{\bar{P}}}{(\lambda^\star+1)(\bar{P}+1)}Y_b=\frac{\sqrt{\zeta}}{\sqrt{\bar{P}+1}}Y_b. \quad (16)$$

The power of the estimate $\widehat{X}$ is $\frac{\zeta}{\bar{P}+1}\left(\bar{P}+1\right)=\zeta$. Thus, when $\zeta<\frac{\bar{P}\sigma^2}{\bar{P}+1}$, the optimizing $\widehat{X}$ merely scales Y_b so that the average power in the estimate is ζ. For $\zeta<\frac{\bar{P}\sigma^2}{\bar{P}+1}$, the following lower bound can therefore be obtained from (15)

$$\frac{1}{m}\mathbb{E}\left[\|\mathbf{S}^m-\widehat{\mathbf{X}}^m\|^2\right]$$
$$\geq \frac{\sigma^2}{\sqrt{\frac{\bar{P}\sigma^2}{(\bar{P}+1)\zeta}}(\bar{P}+1)}-\left(\sqrt{\frac{\bar{P}\sigma^2}{(\bar{P}+1)\zeta}}-1\right)\zeta$$
$$+\left(1-\sqrt{\frac{(\bar{P}+1)\zeta}{\bar{P}\sigma^2}}\right)\sigma^2$$
$$= \sigma^2+\zeta-\frac{2\sigma^2\sqrt{\zeta}\bar{P}}{\sqrt{\bar{P}\sigma^2}\sqrt{\bar{P}+1}}$$
$$= \left(1-\sqrt{\frac{\bar{P}}{\bar{P}+1}}\frac{\sqrt{\zeta}}{\sigma}\right)^2\sigma^2+\frac{\zeta}{\bar{P}+1},$$

where the last equality is obtained by completing the square.

Now consider the case when $\zeta<\frac{\bar{P}\sigma^2}{\bar{P}+1}$. Not surprisingly, the power of the constraint-free $MMSE$ estimate of S, $\widehat{S}_{mmse}=\frac{\sqrt{\bar{P}\sigma^2}}{\bar{P}+1}Y_b$ is also $\frac{\bar{P}\sigma^2}{\bar{P}+1}$. Thus $X=\widehat{S}_{mmse}$ if $\zeta\geq\frac{\bar{P}\sigma^2}{\bar{P}+1}$, obtaining a lower bound of $\frac{\sigma^2}{\bar{P}+1}$. Observe that this lower bound is also valid for all ζ from [4, Theorem 3].

This proves the theorem. ■

Theorem 7: For $W(0,k_w,0,k_b,\sigma)$,

$$\inf_{P\geq 0,\zeta\geq 0}k_w^2P+k_b^2\zeta+\left(\left(\sqrt{\eta(P,\zeta)}-\sqrt{P}\right)^+\right)^2$$
$$\leq J^\star(0,k_w,0,k_b,\sigma)$$
$$\leq 16\left(\inf_{P\geq 0,\zeta\geq 0}k_w^2P+k_b^2\zeta+\left(\left(\sqrt{\eta(P,\zeta)}-\sqrt{P}\right)^+\right)^2\right).$$

Proof: The upper bound is the minimum of that achieved by four schemes : Zero-forcing (ZF), Zero-Input-MMSE-Estimation (ZIME), Zero-Input-Zero-Estimate (ZIZE), and the Vector-Quantization (VQ) scheme of [4]. We now provide upper bounds on the asymptotic average costs by bounding those achieved by these four schemes.

In Zero-Forcing, the input $\mathbf{u}_w^m=-\mathbf{s}^m$ forces the state from $\mathbf{s}^m$ to $\mathbf{x}^m=0$. The cost for Zero-Forcing is $k_w^2\times\sigma^2+k_b^2\times 0+0$, since the estimate $\widehat{\mathbf{x}}^m=0=\mathbf{x}^m=k_w^2\sigma^2$.

In Zero-Input-MMSE-Estimation, the first controller has input $\mathbf{u}_w^m=0$, and thus $\mathbf{x}^m=\mathbf{s}^m$. The second controller uses $\widehat{\mathbf{x}}^m=\frac{\sigma^2}{\sigma^2+1}\times\mathbf{Y}_b^m$. The $MMSE$ is $\frac{\sigma^2}{\sigma^2+1}<1$. Thus the total cost is bounded above by $k_w^2\times 0+k_b^2\frac{\sigma^2}{\sigma^2+1}+1<k_b^2\sigma^2+1$.

For Zero-Input-Zero-Estimation, the first and the second controller both have inputs zero, and the average $MMSE$ is σ^2. Thus the total cost is σ^2.

For the Vector-Quantization scheme [4] (applicable only for $\sigma^2>1$), $P=1$, $\zeta=\sigma^2-1$, and the $MMSE$ cost is asymptotically zero. Thus the total cost is bounded above by $k_w^2+k_b^2(\sigma^2-1)<k_w^2+k_b^2\sigma^2$.

For any value of $(P^\star,\zeta^\star)$ in the lower bound, we now show that the ratio of the upper and lower bounds is uniformly bounded by 16.

Case 1: $P^\star\geq\frac{\sigma^2}{16}$. The lower bound then is no smaller than $k_w^2\frac{\sigma^2}{16}$. Using the zero-forcing strategy, the upper bound is no larger than $k_w^2\sigma^2$. Thus the ratio of the upper and the lower bounds is at most 16.

In the other two cases, $P^\star < \frac{\sigma^2}{16}$, and hence $\sqrt{P^\star} < \frac{\sigma}{4}$, $\bar{P^\star} = (\sigma + \sqrt{P^\star})^2 < \frac{25}{16}\sigma^2$.

Case 2: $P^\star < \frac{\sigma^2}{16}$, $\sigma^2 \leq 1$.

$$\begin{aligned} MMSE &\geq \left(\left(\sqrt{\frac{\sigma^2}{\bar{P^\star}+1}} - \sqrt{P^\star}\right)^+\right)^2 \\ &\stackrel{(a)}{\geq} \left(\sqrt{\frac{\sigma^2}{\frac{25}{16}+1}} - \sqrt{\frac{\sigma^2}{16}}\right)^2 \\ &= \left(\sqrt{\frac{16}{41}}\sigma - \frac{\sigma}{4}\right)^2 \geq 0.14\sigma^2 > \frac{\sigma^2}{7}, \end{aligned}$$

where (a) uses the fact that $\bar{P} < \frac{25}{16}\sigma^2 \leq \frac{25}{16}$. Using the ZIZE upper bound of σ^2, the ratio of the upper and lower bound is smaller than 7.

Case 3: $\zeta^\star \leq \frac{\sigma^2}{16}$, $P^\star < \frac{\sigma^2}{16}$, $\sigma^2 > 1$. For $\zeta^\star \leq \frac{\sigma^2}{16}$, clearly $\zeta^\star < \frac{\sigma^2}{2} \leq \frac{\bar{P^\star}}{\bar{P^\star}+1}\sigma^2$ (since $\bar{P^\star} > \sigma^2 \geq 1$, $\frac{\bar{P^\star}}{\bar{P^\star}+1} \geq \frac{1}{2}$), and thus

$$\begin{aligned} MMSE &\geq \left(\left(\sqrt{\left(1 - \sqrt{\frac{\bar{P^\star}}{\bar{P^\star}+1}}\frac{\sqrt{\zeta}}{\sigma}\right)^2 \sigma^2 + \frac{\zeta}{\bar{P^\star}+1}} \right.\right. \\ &\quad \left.\left. -\sqrt{P^\star}\right)^+\right)^2 \\ &\stackrel{(a)}{\geq} \left(\left(1 - \frac{1}{4}\right)\sigma - \sqrt{\frac{\sigma^2}{16}}\right)^2 = \frac{\sigma^2}{4}, \end{aligned}$$

where (a) uses $\frac{\bar{P^\star}}{\bar{P^\star}+1} < 1$, $\zeta \leq \frac{\sigma^2}{16}$, and $P^\star < \frac{\sigma^2}{16}$, and lower bounds the term $\frac{\zeta^\star}{\bar{P^\star}+1}$ by 0. Using the ZIZE upper bound of σ^2 again, the ratio is smaller than 4.

Case 4: $\zeta^\star > \frac{\sigma^2}{16}$, $P^\star < \frac{\sigma^2}{16}$, $\sigma^2 > 1$

If $P^\star > \frac{1}{16}$, the lower bound is larger than $\frac{k_w^2}{16} + k_b^2\frac{\sigma^2}{16}$. The VQ upper bound is smaller than $k_w^2 + k_b^2\sigma^2$. The ratio of the upper and lower bounds is therefore smaller than 16.

If $P^\star \leq \frac{1}{16}$,

$$\begin{aligned} MMSE &\geq \left(\left(\sqrt{\frac{\sigma^2}{\bar{P^\star}+1}} - \sqrt{P^\star}\right)^+\right)^2 \\ &\geq \left(\sqrt{\frac{1}{(1+\frac{1}{4})^2+1}} - \sqrt{\frac{1}{16}}\right)^2 \\ &= \left(\sqrt{\frac{16}{41}} - \frac{1}{4}\right)^2 \geq 0.14 \geq \frac{1}{7}. \end{aligned}$$

The lower bound is thus $k_b^2\frac{\sigma^2}{16} + \frac{1}{7}$. The ZIME upper bound is $k_b^2\sigma^2+1$. Using Lemma 1, the ratio of the upper and lower bounds is smaller than $\max\{16, 7\} = 16$. ■

We now argue that for the generic problem of (9), the ratio of upper and lower bounds is also smaller than 16. For $P^\star > \frac{\sigma^2}{9}$, the lower bound on the generic cost is larger than $k_s^2\sigma^2 + k_w^2\frac{\sigma^2}{9}$, whereas the zero-forcing strategy attains a cost of $k_s^2\sigma^2 + k_w^2\sigma^2$ (because $\mathbf{x}^m = \widehat{\mathbf{x}}^m = 0$). Thus we only need to consider $P^\star \leq \frac{\sigma^2}{9}$. As shown in Theorem 5, for $P^\star \leq \frac{\sigma^2}{9}$, the ratio of the additional costs due to $\mathbf{x}^m$ is at most 9. Using Lemma 1 and noting that cost due to $\mathbf{s}^m$ is the same in upper and lower bounds, the twin costs on $\mathbf{s}^m$ and $\mathbf{x}^m$ can raise the ratio to at most 9.

The lower bound on $MMSE$ for given ζ, the power of estimate $\widehat{\mathbf{X}}^m$, continues to be valid. Thus following the steps of Theorem 7, the ratio of the upper and lower bounds for the generic problem (9) is also smaller than 16.

IV. Numerical results

For the Witsenhausen counterexample, Fig. 3 shows that while the optimal linear strategy yields costs that are arbitrarily worse than the true optimum, nonlinear quantization-based strategies (used in conjunction with linear strategies) succeed in obtaining a bounded ratio.

How well do purely linear strategies perform for $W(k_s, k_w, k_x, k_b, \sigma)$? Fig. 4 shows a slice each in the three-parameter spaces of $W(0, k_w, k_x, 0, \sigma)$ and $W(0, k_b, k_x, 0, \sigma)$ — we plot the ratio of the costs achieved by the optimal linear strategy and the respective lower bound of Theorem 4 and Theorem 6 for fixed $k_x = 0.01, k_s = k_b = 0$ and for $k_b = 0.01, k_s = k_x = 0$. It might seem surprising that for these slices, linear strategies are not unboundedly bad (although the constant factor is large). The reason is that their associated additional costs grow linearly in σ^2 and dominate the other two original costs in the small-k_w large-σ^2 regime.

This raises a natural question — are nonlinear strategies at all useful for nonzero k_s, k_x or k_b? The answer is an emphatic yes — Fig. 5, 6, and 7 show that for nonzero k_s, k_x or k_b, nonlinear strategies achieve a maximum ratio of 4.13 which is significantly smaller than the ≈ 45 (or more) achieved by linear strategies for the same parameter value 0.01 (for $k_x = 0.01$, or $k_b = 0.01$, this is shown in Fig. 4. For $k_s = 0.01$, the linear penalty ratio is even larger – about 60 – but the figure is omitted due to space constraints).

An application of Lemma 1 shows that the ratio of the upper and lower bounds for $k_s \neq 0$ and $k_x = k_b = 0$ is a decreasing function of k_s. Numerical results (see Fig. 5, 6 and 7) suggest that this is true for the *maximum* ratio for the other two parameters as well, indicating that Witsenhausen's counterexample is the correct corner point in this space in the sense that the approximation results seem to be the hardest to obtain here.

Acknowledgments

We gratefully acknowledge the support of the National Science Foundation (CCF-0729122, CCF-0917212 and CNS-0932410), Sumitomo Electric and Samsung Corporation. We thank Hari Palaiyanur for helpful discussions, and Aaron Wagner for help in developing the tighter lower bound of [5].

References

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

[2] S. K. Mitter and A. Sahai, "Information and control: Witsenhausen revisited," in *Learning, Control and Hybrid Systems: Lecture Notes in Control and Information Sciences 241*, Y. Yamamoto and S. Hara, Eds. New York, NY: Springer, 1999, pp. 281–293.

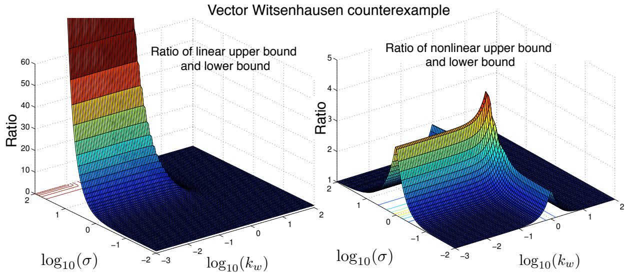

Fig. 3. The left figure shows the ratio of costs attained by optimal linear strategies and the lower bounds of [4, Theorem 3] for $k_s = k_x = k_b = 0$ (the original counterexample).The ratio diverges to infinity in the small-k large-σ regime. The ratio is bounded using the optimal linear strategy combined with the vector-quantization-based nonlinear strategy.

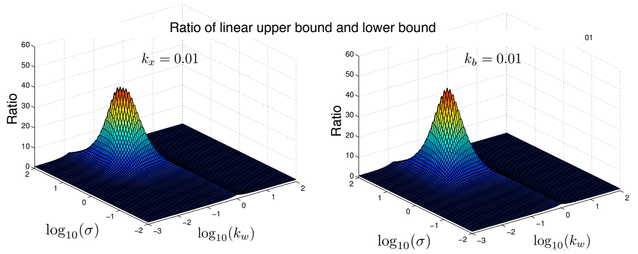

Fig. 4. Ratio of costs attained by optimal linear strategies and the respective lower bounds for $k_x = 0.01, k_s = k_b = 0$ (left) and $k_b = 0.01, k_s = 0, k_x = 0$ (right). The ratio is bounded uniformly by 45 and 50 respectively over all possible choices of (k_w, σ) for fixed k_x or k_b.

[3] 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.

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

[5] P. Grover, A. B. Wagner, and A. Sahai, "Information embedding meets distributed control," in *Submitted to IEEE Information Theory Workshop*, 2009.

[6] S. Y. Park, P. Grover, and A. Sahai, "A constant-factor approximately optimal solution to the Witsenhausen counterexample," in *Proceedings of the 48th IEEE Conference on Decision and Control*, Dec. 2009.

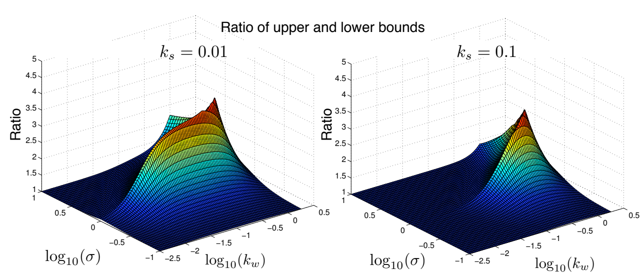

Fig. 5. Ratio of costs attained by vector quantization (and linear) strategies and the lower bounds for $k_s = 0.01$ (left) and $k_s = 0.1$ (right). The maximum ratio for $k_s = 0.1$ is slightly smaller than that for $k_s = 0.01$.

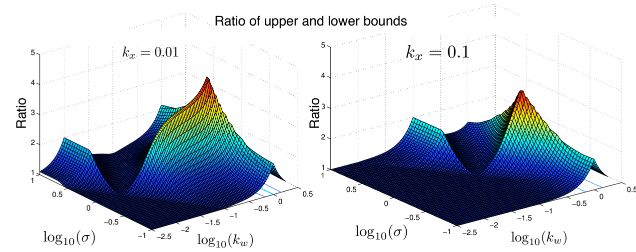

Fig. 6. Ratio of costs attained by vector quantization (and linear) strategies and the lower bounds of Theorem 5 for $k_x = 0.01$ (left) and $k_x = 0.1$ (right). The maximum ratio of 4.13 for $k_x = 0.01$ is significantly smaller than ≈ 40 that is attained by linear strategies.

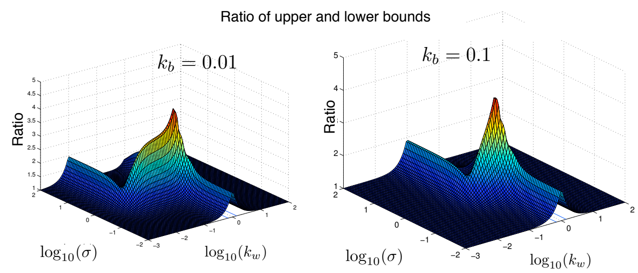

Fig. 7. Ratio of costs attained by vector quantization (and linear) strategies and the lower bounds for $k_b = 0.01$ (left) and $k_b = 0.1$ (right). The maximum ratio for $k_b = 0.1$ is almost same that of $k_b = 0.01$, but is observed to decrease for higher values of k_b.

[7] P. Grover, A. Sahai, and S. Y. Park, "The finite-dimensional Witsenhausen counterexample," in *7th International Symposium on Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks (WiOpt), Workshop on Control over Communication Channels (ConCom)*, 2009. [Online]. Available: http://www.eecs.berkeley.edu/∼pulkit/WitsenhausenConCom.pdf

[8] R. Bansal and T. Basar, "Stochastic teams with nonclassical information revisited: When is an affine control optimal?" *IEEE Trans. Automat. Contr.*, vol. 32, pp. 554–559, Jun. 1987.

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

[10] Y. C. Ho, M. P. Kastner, and E. Wong, "Teams, signaling, and information theory," *IEEE Trans. Automat. Contr.*, vol. 23, no. 2, pp. 305–312, Apr. 1978.

[11] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, 2004.

[12] R. Radner, "Team decision problems," *The Annals of Mathematical Statistics*, 1962.

[13] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 1st ed. New York: Wiley, 1991.