

A constant-factor approximately optimal solution to the Witsenhausen Counterexample

Se Yong Park, Pulkit Grover, and Anant Sahai {seypark,pulkit,sahai}@eecs.berkeley.edu

Abstract—Despite its simplicity (two controllers and otherwise LQG), Witsenhausen’s counterexample is one of the long-standing open problems in stochastic distributed control. Recently, it was proved that an asymptotic vector “relaxation” can be solved to within a constant factor of the optimal cost. A parallel result is shown here for the original scalar problem. Between linear strategies and explicit-signalling-based nonlinear strategies, the optimal performance can be obtained to within a constant factor that is uniformly bounded regardless of the problem parameters. The key contribution is a new lower bound that is much tighter than Witsenhausen’s bound for some parameter values.

I. INTRODUCTION

A special 40th anniversary session for the the Witsenhausen counterexample [1] was organized at the last CDC, so the papers there [2]–[7] give a good overview. The present paper builds on the line of investigation opened up in [6], [8], [9], and for space reasons, the reader is directed to [9], [10] for the full scholarly context [11]–[18]. Whereas in [9] a constant-factor-approximate optimal [19]–[21] result was obtained using information-theoretic tools for an asymptotically infinite-dimensional variant of the Witsenhausen counterexample, here the original scalar counterexample is attacked from first-principles¹.

The problem will be formally defined in the next section, but it is useful here to examine an illustrative particular path through the parameter-space of Witsenhausen problems (Figure 1). Witsenhausen’s lower bound from [1] is too optimistic when the initial state-variance σ_0^2 is large. This is because Witsenhausen effectively reveals the absolute value of the initial state to the second controller. This is too much information. Meanwhile, the bound from [9] is also too optimistic, but for a different reason — it assumes that the second stage cost can be zero using only a finite first-stage cost. As discussed in [9], this is coming from its information-theoretic assumption that noise *always* behaves typically.

As can be seen in Figure 1, the new bound does not suffer from either limitation. Its ratio with the cost of the nonlinear control strategy from [8] always stays bounded. The new bound is also based on a genie-aided argument, but instead of revealing the absolute-value of the initial state, the genie reveals a pair consisting of the initial state both with and without a constant displacement that depends on the average first-stage cost. The bound captures the key idea that the Gaussian observation noise always results in

This work was supported by a Samsung Fellowship and the National Science Foundation: 0403427, 0729122, 0917212, 0932410. All authors are with the EECS Department at UC Berkeley.

¹A more information-theoretic approach is taken in [22] and its journal version [10] by combining generalized sphere-packing ideas from [23] with the bound from [9]. While that approach gives better constants, it obscures the core reason for difficulty in the scalar case.

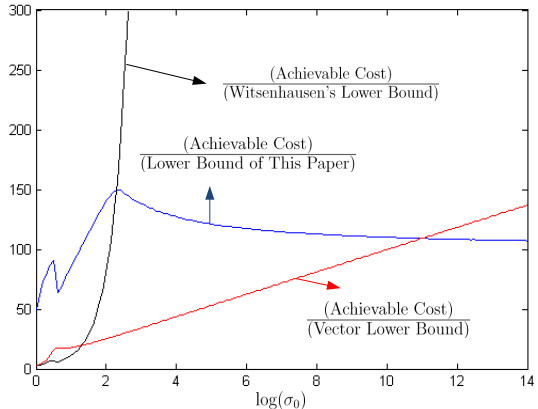


Fig. 1. The lower bounds vs the achievable cost as $k = \frac{1}{\sigma_0}$.

some probability of getting confused even between only two points. This is why the quantize-to-the-nearest-lattice-point strategy of [8] does essentially as well as can be expected.

Although the proof given here is for the traditional Gaussian distribution for the initial state, it is easily extended to any distribution that has a Reimann-integrable part to it.

II. PROBLEM STATEMENT

- The initial state \mathbf{X}_0 is distributed as $\mathcal{N}(0, \sigma_0^2)$.
- The state evolves as:

$$\begin{aligned} \mathbf{X}_1 &= f_1(\mathbf{X}_0, \mathbf{U}_1) = \mathbf{X}_0 + \mathbf{U}_1 \\ \mathbf{X}_2 &= f_2(\mathbf{X}_1, \mathbf{U}_2) = \mathbf{X}_1 - \mathbf{U}_2. \end{aligned}$$

- The outputs observed by the controllers:

$$\begin{aligned} \mathbf{Y}_1 &= \mathbf{X}_0 \\ \mathbf{Y}_2 &= \mathbf{X}_1 + \mathbf{W}, \end{aligned}$$

where $\mathbf{W} \sim \mathcal{N}(0, 1)$ is an additive Gaussian noise that is independent of \mathbf{X}_0 . The control signals \mathbf{U}_1 and \mathbf{U}_2 must be measurable functions of \mathbf{Y}_1 and \mathbf{Y}_2 respectively:

$$\mathbf{U}_1 = \gamma_1(\mathbf{Y}_1), \quad \mathbf{U}_2 = \gamma_2(\mathbf{Y}_2).$$

- The control objective is to minimize an expected quadratic cost, averaged over the random \mathbf{X}_0 and \mathbf{W} .

$$C(k, \mathbf{U}_1, \mathbf{X}_2; \gamma_1, \gamma_2) = C_1(k, \mathbf{U}_1) + C_2(\mathbf{X}_2).$$

where $C_1(k, \mathbf{U}_1) = k^2 \mathbf{U}_1^2$ and $C_2(\mathbf{X}_2) = \mathbf{X}_2^2$. Given a control strategy, the expected costs are a function of the problem parameters k and σ_0 . Let

$$\begin{aligned} \bar{C}(k, \sigma_0; \gamma_1, \gamma_2) &= \mathbb{E}[C(k, \mathbf{U}_1, \mathbf{X}_2; \gamma_1, \gamma_2)] \\ \bar{C}^*(k, \sigma_0) &= \inf_{\gamma_1(\cdot), \gamma_2(\cdot)} \bar{C}(k, \sigma_0; \gamma_1, \gamma_2). \end{aligned}$$

III. A NEW LOWER BOUND AND CONSTANT-FACTOR OPTIMALITY

This section proves the main result: linear strategies and the nonlinear quantization-based strategy of [8] can uniformly achieve a cost within a constant factor for all the problem parameters σ_0 and k .

In the quantization-based scheme of [8], the control \mathbf{U}_1 quantizes \mathbf{X}_0 to the nearest point on the uniform lattice $\{\dots, -2B, -B, 0, +B, +2B, \dots\}$, resulting in \mathbf{X}_1 that is quantized \mathbf{X}_0 . The second controller simply picks \mathbf{U}_2 as the lattice point closest to \mathbf{Y}_2 . The following function helps us bound the costs attained by this strategy.

Definition 1: For any k, a, b, c , all > 0

$$m_k(a, b, c) := \min_{B \geq \sqrt{\frac{3}{2c}}} ak^2 B^2 + bB \exp(-B^2 c).$$

Since $\frac{\partial^2}{\partial B^2} (ak^2 B^2 + bB \exp(-B^2 c)) = 2(ak^2 + 2bcB(cB^2 - \frac{3}{2}) \exp(-B^2 c))$, the function being minimized is a convex function of B in the interval $[\sqrt{\frac{3}{2c}}, \infty)$. Thus, the minimum exists and is achieved at a unique B .

Theorem 1: For the problem as stated in Section II with $\sigma_w^2 = 1$, the optimal expected cost $\bar{C}^*(k, \sigma_0)$ satisfies,

$$\begin{aligned} \frac{1}{8769} \min \left\{ k^2 \sigma_0^2, \frac{\sigma_0^2}{\sigma_0^2 + 1}, m_k \left(\frac{1}{4}, 1.6, \frac{1}{8} \right) \right\} &\leq \bar{C}^*(k, \sigma_0) \\ &\leq \min \left\{ k^2 \sigma_0^2, \frac{\sigma_0^2}{\sigma_0^2 + 1}, m_k \left(\frac{1}{4}, 1.6, \frac{1}{8} \right) \right\} \end{aligned}$$

We prove the theorem by providing upper and lower bounds and showing that their ratio is bounded. We need the following standard lemma from [24].

Lemma 1: Let $Q(x) := \int_x^\infty \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{u^2}{2}\right) du$. Then,

$$\frac{x^2 - 1}{\sqrt{2\pi} x^3} \exp\left(-\frac{x^2}{2}\right) \leq Q(x) \leq \frac{1}{\sqrt{2\pi} x} \exp\left(-\frac{x^2}{2}\right).$$

A. Upper bound on the optimal cost

Theorem 2: The total average cost for the Witsenhausen counterexample is bounded by

$$\bar{C}^*(k, \sigma_0) \leq \min \left\{ k^2 \sigma_0^2, \frac{\sigma_0^2}{\sigma_0^2 + 1}, m_k \left(\frac{1}{4}, 1.6, \frac{1}{8} \right) \right\}. \quad (1)$$

Proof: The first term is achievable using simple zero-forcing, i.e. $\gamma_1(\mathbf{Y}_1) = -\mathbf{Y}_1, \gamma_2(\mathbf{Y}_2) = 0$. The second term is achievable by zero-input and linear MMSE estimation, i.e. $\gamma_1(\mathbf{Y}_1) = 0, \gamma_2(\mathbf{Y}_2) = -\frac{\sigma_0^2}{\sigma_0^2 + 1} \mathbf{Y}_2$. The last term is achieved by the quantization-based strategy (see [8, Eq.(10), (11)]) with lattice spacing B :

$$\gamma_1^B(\mathbf{Y}_1) = -\mathbf{Y}_1 + B \lfloor \frac{\mathbf{Y}_1}{B} + \frac{1}{2} \rfloor, \quad \gamma_2^B(\mathbf{Y}_2) = B \lfloor \frac{\mathbf{Y}_2}{B} + \frac{1}{2} \rfloor.$$

To bound the second stage cost, denote the probability that the second controller makes an n -point quantization error by

$$\begin{aligned} P_{B_n} &:= \Pr(|\mathbf{X}_1 - \gamma_2^B(\mathbf{Y}_2)| = nB) \\ &= 2 \left(Q \left(\frac{(2n-1)B}{2} \right) - Q \left(\frac{(2n+1)B}{2} \right) \right). \quad (2) \end{aligned}$$

Since we are free to choose the quantization interval B , assume $\frac{B^2}{8} \geq \frac{3}{2}$. Now, for the second stage cost,

$$\begin{aligned} \mathbb{E}[C_2(\mathbf{X}_2; \gamma_1^B, \gamma_2^B)] &= \sum_{n=1}^{\infty} (nB)^2 P_{B_n} \\ &\stackrel{(a)}{=} \sum_{n=1}^{\infty} (nB)^2 2 \left(Q \left(\frac{(2n-1)B}{2} \right) - Q \left(\frac{(2n+1)B}{2} \right) \right) \\ &\stackrel{(b)}{=} \sum_{n=1}^{\infty} (n^2 - (n-1)^2) B^2 2Q \left(\frac{(2n-1)B}{2} \right) \\ &\stackrel{(c)}{\leq} \sum_{n=1}^{\infty} (2n-1) B^2 \frac{2}{\sqrt{2\pi} \frac{(2n-1)B}{2}} \exp \left(-\frac{(2n-1)^2 B^2}{8} \right) \\ &= \frac{4B}{\sqrt{2\pi}} \exp \left(-\frac{B^2}{8} \right) \sum_{n=1}^{\infty} \exp \left(-\frac{(2n-1)^2 B^2 - B^2}{8} \right) \\ &\stackrel{(d)}{\leq} \frac{4B}{\sqrt{2\pi}} \exp \left(-\frac{B^2}{8} \right) \sum_{n=1}^{\infty} \exp \left(-\frac{2(2n-2)B^2}{8} \right) \\ &\stackrel{(e)}{\leq} \frac{4B}{\sqrt{2\pi}} \exp \left(-\frac{B^2}{8} \right) \frac{1}{1 - \exp(-6)}. \end{aligned}$$

(a): By (2)

(b): Substituting $n+1$ for n in the second term

(c): By Lemma 1

(d): Since $\frac{(2n-1)^2 B^2 - B^2}{8} = \frac{2n(2n-2)B^2}{8} \geq \frac{2(2n-2)B^2}{8}$

(e): Since $\frac{B^2}{8} \geq \frac{3}{2}$, this gives a geometric series.

Together with the fact that the first stage cost is upper bounded by $k^2 \frac{B^2}{4}$ (since quantization only has to move the probability mass by at most $\frac{B}{2}$ in either direction), we have

$$\begin{aligned} \bar{C}^*(k, \sigma_0) &\leq \min_{\frac{B^2}{8} \geq \frac{3}{2}} k^2 \frac{B^2}{4} + \frac{4B}{\sqrt{2\pi}} \frac{\exp\left(-\frac{B^2}{8}\right)}{1 - \exp(-6)} \\ &\leq \min_{\frac{B^2}{8} \geq \frac{3}{2}} k^2 \frac{B^2}{4} + 1.6B \exp\left(-\frac{B^2}{8}\right) = m_k \left(\frac{1}{4}, 1.6, \frac{1}{8} \right). \quad \blacksquare \end{aligned}$$

B. Lower bound on the optimal cost

This section is the technical heart of the paper. But before we can prove a new lower bound for Witsenhausen's counterexample, we need the following technical lemma.

Lemma 2: For positive real numbers $d, w > 0$ and a real number c , let \mathcal{K} and \mathcal{L} be two sets in \mathbb{R} such that

$$\begin{aligned} \mathcal{K} &= \left\{ x \in \mathbb{R} \mid |x - c| \leq \frac{w}{2} \right\} \\ \mathcal{L} &= \left\{ x \in \mathbb{R} \mid |x - c + d| \leq \frac{w}{2} \right\}. \end{aligned}$$

Define a random variable \mathbf{X} such that $\Pr(\mathbf{X} \in \mathcal{K}) = \Pr(\mathbf{X} \in \mathcal{L}) = \frac{1}{2}$, and its noisy observation $\mathbf{Y} = \mathbf{X} + \mathbf{N}$ where \mathbf{N} is an independent $\mathcal{N}(0, 1)$ random variable.

Then, the MMSE of \mathbf{X} given \mathbf{Y} is lower bounded by $\mathbb{E}[(\mathbf{X} - \mathbb{E}[\mathbf{X}|\mathbf{Y}])^2] \geq s(d, w)$ where $s(d, w) = \frac{1}{\sqrt{2\pi}} \left(\frac{d-w}{2}\right)^2 \left(\frac{1}{d+w} - \frac{4}{(d+w)^3}\right) \exp\left(-\frac{(d+w)^2}{8}\right)$ and $(\cdot)^+$ denotes $\max\{\cdot, 0\}$.

Proof: For $d < w$, $s(d, w)$ is 0 and the lemma is trivial. Thus, assume that $d \geq w$. We first show that $\mathbb{E}[\mathbf{X}|\mathbf{Y} = y]$ is an increasing function of y . By Bayes' rule, $\frac{d}{dy}\mathbb{E}[\mathbf{X}|\mathbf{Y} = y]$

$$\begin{aligned} &= \frac{d}{dy} \frac{\int_{-\infty}^{\infty} x \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{(y-x)^2}{2}\right) dF(x)}{\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{(y-x)^2}{2}\right) dF(x)} \\ &= \frac{d}{dy} \frac{\int_{-\infty}^{\infty} x \exp\left(yx - \frac{x^2}{2}\right) dF(x)}{\int_{-\infty}^{\infty} \exp\left(yx - \frac{x^2}{2}\right) dF(x)} \\ &= \left(\int_{-\infty}^{\infty} x^2 \exp\left(yx - \frac{x^2}{2}\right) dF(x) \right. \\ &\quad \times \int_{-\infty}^{\infty} \exp\left(yx - \frac{x^2}{2}\right) dF(x) \\ &\quad \left. - \left(\int_{-\infty}^{\infty} x \exp\left(yx - \frac{x^2}{2}\right) dF(x) \right)^2 \right) \\ &\quad / \left(\int_{-\infty}^{\infty} \exp\left(yx - \frac{x^2}{2}\right) dF(x) \right)^2 \\ &\stackrel{(a)}{\geq} 0 \end{aligned}$$

$$(a): \quad \mathbb{E} \left[\mathbf{X}^2 \exp\left(y\mathbf{X} - \frac{\mathbf{X}^2}{2}\right) \right] \mathbb{E} \left[\exp\left(y\mathbf{X} - \frac{\mathbf{X}^2}{2}\right) \right] \geq \mathbb{E} \left[\mathbf{X} \exp\left(y\mathbf{X} - \frac{\mathbf{X}^2}{2}\right) \right]^2 \text{ using Cauchy-Schwartz.}$$

There are two cases : $\mathbb{E}[\mathbf{X}|\mathbf{Y} = c - \frac{d}{2}] \leq c - \frac{d}{2}$ or $\mathbb{E}[\mathbf{X}|\mathbf{Y} = c - \frac{d}{2}] \geq c - \frac{d}{2}$. Consider first the case when $\mathbb{E}[\mathbf{X}|\mathbf{Y} = c - \frac{d}{2}] \leq c - \frac{d}{2}$. Since $\mathbb{E}[\mathbf{X}|\mathbf{Y} = y]$ is an increasing function of y ,

$$\mathbb{E}[\mathbf{X}|\mathbf{Y} = y] \leq c - \frac{d}{2}, \quad \forall y \leq c - \frac{d}{2}. \quad (3)$$

$$\begin{aligned} &\text{Using this fact, the MMSE } \mathbb{E}[(\mathbf{X} - \mathbb{E}[\mathbf{X}|\mathbf{Y}])^2] \\ &\geq \mathbb{E} \left[(\mathbf{X} - \mathbb{E}[\mathbf{X}|\mathbf{Y}])^2 \mid \mathbf{X} \in \mathcal{K}, \mathbf{Y} \leq c - \frac{d}{2} \right] \\ &\quad \cdot \Pr \left(\mathbf{X} \in \mathcal{K}, \mathbf{Y} \leq c - \frac{d}{2} \right) \\ &\stackrel{(b)}{\geq} \left(\frac{d-w}{2} \right)^2 \Pr \left(\mathbf{X} \in \mathcal{K}, \mathbf{Y} \leq c - \frac{d}{2} \right) \\ &\stackrel{(c)}{\geq} \left(\frac{d-w}{2} \right)^2 \frac{1}{2} \Pr \left(\mathbf{N} \geq \frac{d+w}{2} \right) \\ &= \left(\frac{d-w}{2} \right)^2 \frac{1}{2} Q \left(\frac{d+w}{2} \right) \end{aligned} \quad (4)$$

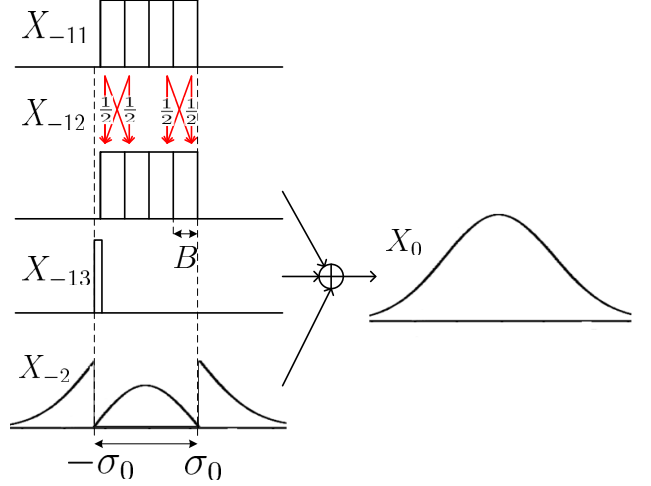


Fig. 2. The key decomposition of the Gaussian prior for the state X_0 used for proving Theorem 3.

$$\begin{aligned} &\stackrel{(d)}{\geq} \left(\frac{d-w}{2} \right)^2 \frac{1}{\sqrt{2\pi}} \left(\frac{1}{d+w} - \frac{4}{(d+w)^3} \right) \\ &\quad \cdot \exp\left(-\frac{(d+w)^2}{8}\right) \end{aligned}$$

- (b): By (3) and the definition of \mathcal{K} , $\inf_{x \in \mathcal{K}, y \leq c - \frac{d}{2}} (x - \mathbb{E}[\mathbf{X}|\mathbf{Y} = y])^2 \geq \inf_{x \in \mathcal{K}} (x - (c - \frac{d}{2}))^2 \geq (c - \frac{w}{2} - c + \frac{d}{2})^2 = (\frac{d-w}{2})^2$
- (c): Since $\sup_{x \in \mathcal{K}} \{x - (c - \frac{d}{2})\} = \frac{d+w}{2}$
- (d): By Lemma 1.

The case of $\mathbb{E}[\mathbf{X}|\mathbf{Y} = c - \frac{d}{2}] \geq c - \frac{d}{2}$ can be proved similarly by considering $\mathbf{X} \in \mathcal{L}$ and $\mathbf{Y} \geq c - \frac{d}{2}$, in (4). ■

Now, we are ready to prove a lower bound on the optimal cost. However, to make it work, it is conceptually simpler to work with a uniform random variable for the initial state. The trick is to give an appropriate “backstory” to this Gaussian random variable that involves a uniform random variable in a critical way. A version of this same trick was used in [25], [26]. The backstory is illustrated in Figure 2.

We interpret \mathbf{X}_0 as a convex combination of three random variables \mathbf{X}_{-11} , \mathbf{X}_{-13} (a small annoying subset of the uniform part that represents the boundary), and \mathbf{X}_{-2} (the non-uniform part of \mathbf{X}_0). Our bound pretends the cost is zero on the total probability mass in \mathbf{X}_{-2} and \mathbf{X}_{-13} . \mathbf{X}_{-11} is the uniform part that is of interest. We partition the real line into two sets:

$$\begin{aligned} \mathcal{E}_1 &= \{x \in \mathbb{R} \mid x \in (\sigma_0 - 2lB - B, \sigma_0 - 2lB], l \in \mathbb{Z}\} \\ \mathcal{E}_2 &= \{x \in \mathbb{R} \mid x \in (\sigma_0 - 2lB - 2B, \sigma_0 - 2lB - B], l \in \mathbb{Z}\}. \end{aligned}$$

When the realization of \mathbf{X}_{-11} lies in \mathcal{E}_1 , it will be distorted to its counterpart that lies in \mathcal{E}_2 with probability $\frac{1}{2}$, and vice versa. This distorted version of \mathbf{X}_{-11} is given to the second controller as side information. Therefore, the

only randomness left at the second controller is \mathbf{X}_{-12} and Lemma 2 allows the MMSE error to be lower bounded.

Theorem 3: For $\sigma_0^2 \geq 1$,

$$\bar{C}^*(k, \sigma_0) \geq \min \left\{ \frac{k^2 \sigma_0^2}{100}, 0.06, m_k \left(1, \frac{1}{274}, 50 \right) \right\}. \quad (5)$$

Proof: We write the Gaussian random variable \mathbf{X}_0 as a function of three random variables $\mathbf{X}_{-1}, \mathbf{X}_{-2}, \mathbf{X}_{-3}$. The random variable \mathbf{X}_{-1} is again written as a function of $\mathbf{X}_{-11}, \mathbf{X}_{-12}, \mathbf{X}_{-13}, \mathbf{X}_{-14}$. Define a scalar p_u and a function $\beta(x)$ as

$$p_u = \int_{-\sigma_0}^{\sigma_0} \frac{1}{\sqrt{2\pi}\sigma_0} \exp\left(-\frac{1}{2}\right) dx = \frac{2}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}\right)$$

$$\beta(x) = \begin{cases} -B & \text{if } x \in \mathcal{E}_1 \\ B & \text{if } x \in \mathcal{E}_2 \end{cases}$$

and set \mathcal{A} and \mathcal{A}' are defined as

$$\mathcal{A} = \{x \in (-\sigma_0, \sigma_0] | x + \beta(x) \in (-\sigma_0, \sigma_0)\}$$

$$\mathcal{A}' = (-\sigma_0, \sigma_0] \setminus \mathcal{A}.$$

The dependence of $\beta(x), \mathcal{A}$, and \mathcal{A}' on B is suppressed.

Since the distance between x and $x + \beta(x)$ is B , the Lebesgue measure of \mathcal{A} is lower bounded by

$$|\mathcal{A}| \geq 2\sigma_0 - B. \quad (6)$$

The laws for independent random variables $\mathbf{X}_{-2}, \mathbf{X}_{-3}, \mathbf{X}_{-11}, \mathbf{X}_{-12}, \mathbf{X}_{-13}$, and \mathbf{X}_{-14} are given as: (f is used for density functions)

$$f(\mathbf{X}_{-11} = x) = \begin{cases} \frac{1}{|\mathcal{A}|} & \text{if } x \in \mathcal{A} \\ 0 & \text{otherwise} \end{cases} \quad (7)$$

$$\Pr(\mathbf{X}_{-12} = x) = \begin{cases} \frac{1}{2} & \text{if } x = 1 \\ \frac{1}{2} & \text{if } x = 0 \end{cases}$$

$$f(\mathbf{X}_{-13} = x) = \begin{cases} \frac{1}{|\mathcal{A}'|} & \text{if } x \in \mathcal{A}' \\ 0 & \text{otherwise} \end{cases}$$

$$\Pr(\mathbf{X}_{-14} = x) = \begin{cases} \frac{|\mathcal{A}|}{|\mathcal{A}| + |\mathcal{A}'|} & \text{if } x = 1 \\ \frac{|\mathcal{A}'|}{|\mathcal{A}| + |\mathcal{A}'|} & \text{if } x = 0 \end{cases} \quad (8)$$

$$\mathbf{X}_{-1} = \begin{cases} \mathbf{X}_{-11} + \mathbf{X}_{-12}\beta(\mathbf{X}_{-11}) & \text{if } \mathbf{X}_{-14} = 1 \\ \mathbf{X}_{-13} & \text{if } \mathbf{X}_{-14} = 0 \end{cases}$$

$$f(\mathbf{X}_{-2} = x) = \begin{cases} \frac{1}{p_u} \left(\frac{1}{\sqrt{2\pi}\sigma_0} \exp\left(-\frac{x^2}{2\sigma_0^2}\right) - \frac{1}{\sqrt{2\pi}\sigma_0} \exp\left(-\frac{1}{2}\right) \right) & \text{if } x \in (-\sigma_0, \sigma_0] \\ \frac{1}{p_u} \frac{1}{\sqrt{2\pi}\sigma_0} \exp\left(-\frac{x^2}{2\sigma_0^2}\right) & \text{otherwise} \end{cases}$$

$$\Pr(\mathbf{X}_{-3} = x) = \begin{cases} p_u & \text{if } x = 1 \\ 1 - p_u & \text{if } x = 0 \end{cases} \quad (9)$$

$$\mathbf{X}_0 = \begin{cases} \mathbf{X}_{-1} & \text{if } \mathbf{X}_{-3} = 1 \\ \mathbf{X}_{-2} & \text{if } \mathbf{X}_{-3} = 0 \end{cases}.$$

Then, \mathbf{X}_{-1} is distributed uniformly in $(-\sigma_0, \sigma_0]$, and \mathbf{X}_0 is still $\mathcal{N}(0, \sigma_0^2)$. Suppose $\mathbb{E}[\gamma_1(\mathbf{X}_0)^2] = P$. Consider the sets

$$\mathcal{B} = \{x \in (-\sigma_0, \sigma_0] | |\gamma_1(x)| \leq 4\sqrt{P}\}$$

$$\mathcal{B}' = \{x \in (-\sigma_0, \sigma_0] | |\gamma_1(x)| > 4\sqrt{P}\}$$

$$\mathcal{D} = \{x \in (-\sigma_0, \sigma_0] | x \in \mathcal{A} \cap \mathcal{B}, x + \beta(x) \in \mathcal{B}\}.$$

An upper bound on the Lebesgue measure of \mathcal{B}' can be derived as follows:

$$\begin{aligned} \frac{1}{16} &\stackrel{(a)}{\geq} \Pr\left(|\gamma_1(\mathbf{X}_0)| > 4\sqrt{P}\right) \\ &\stackrel{(b)}{\geq} \Pr(\mathbf{X}_0 \in \mathcal{B}') \\ &\stackrel{(c)}{=} \int_{x \in \mathcal{B}'} \frac{1}{\sqrt{2\pi}\sigma_0} \exp\left(-\frac{x^2}{2\sigma_0^2}\right) dx \\ &\stackrel{(d)}{\geq} \int_{x \in \mathcal{B}'} \frac{1}{\sqrt{2\pi}\sigma_0} \exp\left(-\frac{1}{2}\right) dx \\ &= \frac{1}{\sqrt{2\pi}\sigma_0} \exp\left(-\frac{1}{2}\right) |\mathcal{B}'| = \frac{p_u}{2\sigma_0} |\mathcal{B}'| \quad (10) \end{aligned}$$

(a): Markov's inequality with the fact $\mathbb{E}[\gamma_1(\mathbf{X}_0)^2] = P$

(b): Set inclusion

(c): \mathbf{X}_0 is a Gaussian random variable $\mathcal{N}(0, \sigma_0^2)$

(d): $\mathcal{B}' \subset (-\sigma_0, \sigma_0]$.

Also, $|\{x | x + \beta(x) \in \mathcal{B}'\}|$

$$\begin{aligned} &\stackrel{(e)}{=} |\{x | x + \beta(x) \in \mathcal{B}', x \in \mathcal{E}_1\}| \\ &\quad + |\{x | x + \beta(x) \in \mathcal{B}', x \in \mathcal{E}_2\}| \\ &\stackrel{(f)}{=} |\{x | x - B \in \mathcal{B}', x \in \mathcal{E}_1\}| \\ &\quad + |\{x | x + B \in \mathcal{B}', x \in \mathcal{E}_2\}| \\ &\stackrel{(g)}{=} |\{x' | x' \in \mathcal{B}'\}| \quad (11) \end{aligned}$$

(e): Since $\mathcal{E}_1, \mathcal{E}_2$ is a partition of \mathbb{R}

(f): By definition of $\beta(x)$

(g): Since $\mathcal{E}_1, \mathcal{E}_2$ is a partition of \mathbb{R} and $\mathcal{E}_1 = \mathcal{E}_2 \pm B$

Using the above facts, a lower bound on $|\mathcal{D}|$ is given by

$$\begin{aligned} |\mathcal{D}| &\stackrel{(h)}{=} |\mathcal{A} \setminus \mathcal{B}' \setminus \{x | x + \beta(x) \in \mathcal{B}'\}| \\ &\geq |\mathcal{A}| - |\mathcal{B}'| - |\{x | x + \beta(x) \in \mathcal{B}'\}| \\ &\stackrel{(i)}{\geq} |\mathcal{A}| - 2|\mathcal{B}'| \\ &\stackrel{(j)}{\geq} 2\sigma_0 - B - 2|\mathcal{B}'| \quad (12) \end{aligned}$$

(h): $\mathcal{X} \cap \mathcal{Y} \cap \mathcal{Z} = \mathcal{X} \setminus \mathcal{Y}^c \setminus \mathcal{Z}^c$ for any $\mathcal{X}, \mathcal{Y}, \mathcal{Z}$

(i): By (11), (j): By (6).

A genie provides $\mathbf{X}_{-11}, \mathbf{X}_{-13}, \mathbf{X}_{-14}, \mathbf{X}_{-2}, \mathbf{X}_{-3}$ to the second controller. For this genie aided control action $\gamma_2^{(g)}$,

$$\begin{aligned}
& \mathbb{E}[C_2(\mathbf{X}_2; \gamma_1, \gamma_2^{(g)})] \\
& \geq \mathbb{E}[C_2(\mathbf{X}_2; \gamma_1, \gamma_2^{(g)}) | \mathbf{X}_{-11} \in \mathcal{D}, \mathbf{X}_{-14} = 1, \mathbf{X}_{-3} = 1] \\
& \quad \cdot \Pr(\mathbf{X}_{-11} \in \mathcal{D}, \mathbf{X}_{-14} = 1, \mathbf{X}_{-3} = 1) \\
& \stackrel{(k)}{\geq} s(B, 8\sqrt{P}) \cdot \Pr(\mathbf{X}_{-11} \in \mathcal{D}, \mathbf{X}_{-14} = 1, \mathbf{X}_{-3} = 1) \\
& \stackrel{(l)}{=} s(B, 8\sqrt{P}) \cdot \Pr(\mathbf{X}_{-11} \in \mathcal{D}) \Pr(\mathbf{X}_{-14} = 1) \\
& \quad \cdot \Pr(\mathbf{X}_{-3} = 1) \\
& \stackrel{(m)}{\geq} \frac{|\mathcal{D}|}{|\mathcal{A}|} \frac{|\mathcal{A}|}{|\mathcal{A}| + |\mathcal{A}'|} p_u s(B, 8\sqrt{P}) \\
& \stackrel{(n)}{\geq} \frac{1}{2\sigma_0} (2\sigma_0 - B - 2|B'|) p_u s(B, 8\sqrt{P}) \\
& \stackrel{(o)}{\geq} \left(p_u - p_u \frac{B}{2\sigma_0} - \frac{2}{16} \right) s(B, 8\sqrt{P}) \tag{13}
\end{aligned}$$

(k): Conditioned on the given event \mathbf{X}_0 is either \mathbf{X}_{-11} or $\mathbf{X}_{-11} + \beta(\mathbf{X}_{-11})$ with probability $\frac{1}{2}$ each, and $\mathbf{X}_{-11} \in \mathcal{D}$ tells that $|\gamma_1(\cdot)|$ at both points is less than $4\sqrt{P}$. Thus, Lemma 2 justifies the inequality.

(l): By independence of $\mathbf{X}_{-11}, \mathbf{X}_{-14}$ and \mathbf{X}_{-3}

(m): By definitions (7), (8) and (9), and $\mathcal{C} \subseteq \mathcal{A}$

(n): By (12) and since $\{\mathcal{A}, \mathcal{A}'\}$ is a partition of $(-\sigma_0, \sigma_0]$

(o): By (10).

Consider the case $P \geq \frac{\sigma_0^2}{100}$. Then,

$$\bar{C}^*(k, \sigma_0) \geq k^2 E[\gamma_1(\mathbf{X}_0)^2] = k^2 \frac{\sigma_0^2}{100}. \tag{14}$$

When $\frac{3}{100} \leq P \leq \frac{\sigma_0^2}{100}$, substituting $B = 12\sqrt{P}$ in (13) gives

$$\begin{aligned}
& \mathbb{E}[C_2(\mathbf{X}_2; \gamma_1, \gamma_2^{(g)})] \\
& \stackrel{(p)}{\geq} \left(p_u - p_u \frac{12\sqrt{P}}{2\sigma_0} - \frac{2}{16} \right) s(12\sqrt{P}, 8\sqrt{P}) \\
& \stackrel{(q)}{=} \left(p_u - p_u \frac{12}{20} - \frac{2}{16} \right) \left(\frac{4\sqrt{P}}{2} \right)^2 \\
& \quad \frac{1}{\sqrt{2\pi}} \left(\frac{1}{20\sqrt{P}} - \frac{4}{(20\sqrt{P})^3} \right) \exp(-50P) \\
& \stackrel{(r)}{\geq} \left(p_u - p_u \frac{12}{20} - \frac{2}{16} \right) \left(\frac{4\sqrt{P}}{2} \right)^2 \\
& \quad \frac{1}{\sqrt{2\pi}} \left(\frac{1}{20\sqrt{P}} - \frac{1}{60\sqrt{P}} \right) \exp(-50P) \\
& \stackrel{(s)}{\geq} \frac{1}{274} \sqrt{P} \exp(-50P). \tag{15}
\end{aligned}$$

(p): By (13); (q): Since $P \leq \frac{\sigma_0^2}{100}$ and using Lemma 2

(r): Since $\frac{3}{100} \leq P$

(s): $p_u = \frac{2}{\sqrt{2\pi}} \exp(-\frac{1}{2})$, $p_u - p_u \frac{12}{20} - \frac{2}{16} \approx \frac{1}{14.5}$, and

$$\left(\frac{4}{2}\right)^2 \frac{1}{\sqrt{2\pi}} \left(\frac{1}{20} - \frac{1}{60}\right) \approx \frac{1}{18.8}$$

When $P \leq \frac{3}{100}$, by [9, Theorem 3], $\mathbb{E}[C_2(\mathbf{X}_2; \gamma_1, \gamma_2^{(g)})]$

$$\begin{aligned}
& \geq \left(\left(\sqrt{\frac{\sigma_0^2}{(\sigma_0 + \sqrt{P})^2 + 1}} - \sqrt{P} \right)^+ \right)^2 \\
& \geq \left(\frac{1}{\left(1 + \frac{\sqrt{3}}{10}\right)^2 + 1} - \frac{\sqrt{3}}{10} \right)^2 \geq 0.06
\end{aligned}$$

where the second inequality uses $\sigma_0^2 \geq 1$. Using the above, (14), and (15), for $\sigma_0^2 \geq 1$

$$\begin{aligned}
\bar{C}^*(k, \sigma_0) & \geq \min \left\{ \frac{k^2 \sigma^2}{100}, 0.06, \right. \\
& \quad \left. \min_{\frac{3}{100} \leq P \leq \frac{\sigma_0^2}{100}} k^2 P + \frac{1}{274} \sqrt{P} \exp(-50P) \right\} \tag{16} \\
& \geq \min \left\{ \frac{k^2 \sigma^2}{100}, 0.06, m_k \left(1, \frac{1}{274}, 50 \right) \right\}.
\end{aligned}$$

Note that if the minimization of the third term in (16) is infeasible, it is ∞ by default. ■

C. Proof of Theorem 1

To prove that the two bounds track each other, we need some technical lemmas about $m_k(a, b, c)$.

Lemma 3: If $a_1 \leq a_2$, $b_1 \leq b_2$, and $c_1 \geq c_2$, then $m_k(a_1, b_1, c_1) \leq m_k(a_2, b_2, c_2)$.

Proof: Let $B^* \geq \sqrt{\frac{3}{2c_2}}$ achieve the minimum in the expression for $m_k(a_2, b_2, c_2)$.

$$\begin{aligned}
m_k(a_2, b_2, c_2) & = a_2 k^2 B^{*2} + b_2 B^* \exp(-B^{*2} c_2) \\
& \geq a_1 k^2 B^{*2} + b_1 B^* \exp(-B^{*2} c_2) \\
& = a_1 k^2 B^{*2} + b_1 B^* \exp\left(-\frac{c_2}{c_1} B^{*2} c_1\right) \\
& \geq a_1 k^2 B^{*2} + b_1 B^* \exp(-B^{*2} c_1) \\
& \stackrel{(a)}{\geq} m_k(a_1, b_1, c_1).
\end{aligned}$$

(a): $B^* \geq \sqrt{\frac{3}{2c_2}} \geq \sqrt{\frac{3}{2c_1}}$ ■

The following lemma shows that $m_k(a, b, c)$ and $m_k(1, 1, 1)$ are within a constant factor of each other regardless of k .

Lemma 4: For positive real numbers a, b , and c ,

$$t_{lower} m_k(1, 1, 1) \leq m_k(a, b, c) \leq t_{upper} m_k(1, 1, 1),$$

where

$$\begin{aligned}
t_{lower} & = \min \left\{ \frac{a}{c}, \frac{b}{\sqrt{c}} \right\}, t_{upper} = \max \{a, b\} \text{ when } c \geq 1 \\
t_{lower} & = \min \{a, b\}, t_{upper} = \max \left\{ \frac{a}{c}, \frac{b}{\sqrt{c}} \right\} \text{ when } c < 1.
\end{aligned}$$

Proof: Assume $c \geq 1$. By Lemma 3, we know

$$m_k(a, b, c) \leq m_k(\max\{a, b\}, \max\{a, b\}, c) \\ = \max\{a, b\} m_k(1, 1, c) \leq \max\{a, b\} m_k(1, 1, 1). \quad (17)$$

Let the minimum in the expression for $m_k(a, b, c)$ be achieved by B^* and denote $B' := \sqrt{c}B^*$.

$$m_k(a, b, c) = \min_{B \geq \sqrt{\frac{3}{2c}}} ak^2B^2 + bB \exp(-B^2c) \\ = ak^2B^{*2} + bB^* \exp(-B^{*2}c) \\ = \frac{a}{c}k^2B'^2 + \frac{b}{\sqrt{c}}B' \exp(-B'^2) \\ \geq \min\left\{\frac{a}{c}, \frac{b}{\sqrt{c}}\right\} (k^2B'^2 + B' \exp(-B'^2)) \\ \stackrel{(a)}{\geq} \min\left\{\frac{a}{c}, \frac{b}{\sqrt{c}}\right\} \\ \min_{B \geq \sqrt{\frac{3}{2}}} (k^2B^2 + B \exp(-B^2)) \\ = \min\left\{\frac{a}{c}, \frac{b}{\sqrt{c}}\right\} m_k(1, 1, 1). \quad (18)$$

$$(a): B' = \sqrt{c}B^* \geq \sqrt{c}\sqrt{\frac{3}{2c}} = \sqrt{\frac{3}{2}}.$$

By (17) and (18),

$$\min\left\{\frac{a}{c}, \frac{b}{\sqrt{c}}\right\} m_k(1, 1, 1) \leq m_k(a, b, c) \\ \leq \max\{a, b\} m_k(1, 1, 1).$$

By essentially symmetric reasoning, the case $c < 1$ can be obtained. ■

Proof: [Of Theorem 1] The upper bound follows from Theorem 1. For the lower bound, first consider $\sigma_0^2 > 1$.

$$\bar{C}^*(k, \sigma_0) \stackrel{(a)}{\geq} \min\left\{\frac{k^2\sigma^2}{100}, 0.06, m_k\left(1, \frac{1}{274}, 50\right)\right\} \\ \stackrel{(b)}{\geq} \min\left\{\frac{k^2\sigma^2}{100}, 0.06, \min\left\{\frac{1}{50}, \frac{1}{274\sqrt{50}}\right\} m_k(1, 1, 1)\right\} \\ \stackrel{(c)}{\geq} \min\left\{\frac{k^2\sigma^2}{100}, 0.06, \frac{\min\left\{\frac{1}{50}, \frac{1}{274\sqrt{50}}\right\}}{\max\{2, 1.6\sqrt{8}\}} m_k\left(\frac{1}{4}, 1.6, \frac{1}{8}\right)\right\} \\ \geq \frac{1}{8769} \min\left\{k^2\sigma_0^2, \frac{\sigma_0^2}{\sigma_0^2 + 1}, m_k\left(\frac{1}{4}, 1.6, \frac{1}{8}\right)\right\}$$

(a): By Theorem 3

(b): By the lower bound of Lemma 4

(c): By the upper bound of Lemma 4

The case $\sigma_0^2 \leq 1$ can be found in [9, Appendix V]. ■

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] Y.-C. Ho, "Review of the Witsenhausen problem," in *47th IEEE Conference on Decision and Control*, Dec. 9–11, 2008, pp. 1611–1613.
- [3] T. Basar, "Variations on the theme of the Witsenhausen counterexample," in *47th IEEE Conference on Decision and Control*, Dec. 9–11, 2008, pp. 1614–1619.
- [4] A. Nayyar and D. Teneketzis, "On globally optimal real-time encoding and decoding strategies in multi-terminal communication systems," in *47th IEEE Conference on Decision and Control*, Dec. 9–11, 2008, pp. 1620–1627.
- [5] G. Lipsa and N. C. Martins, "Finite horizon optimal memoryless control of a delay in Gaussian noise: A simple counterexample," in *47th IEEE Conference on Decision and Control*, Dec. 9–11, 2008, pp. 1628–1635.
- [6] P. Grover and A. Sahai, "A vector version of Witsenhausen's counterexample: A convergence of control, communication and computation," in *47th IEEE Conference on Decision and Control*, Cancun, Dec. 9–11, 2008, pp. 1636–1641.
- [7] M. Rotkowitz, "On information structures, convexity, and linear optimality," in *47th IEEE Conference on Decision and Control*, Dec. 9–11, 2008, pp. 1642–1647.
- [8] 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.
- [9] P. Grover and A. Sahai, "The vector Witsenhausen problem as assisted interference cancellation," *To appear in the International Journal on Systems, Control and Communications (IJSCC)*, 2009. [Online]. Available: <http://www.eecs.berkeley.edu/~sahai/>
- [10] P. Grover, S. Y. Park, and A. Sahai, "Approximately optimal solutions for all finite-dimensional Witsenhausen counterexamples," *IEEE Trans. Automat. Contr.*, To Be Submitted Sep 2009.
- [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, June 1987.
- [12] M. Baglietto, T. Parisini, and R. Zoppoli, "Nonlinear approximations for the solution of team optimal control problems," *IEEE Conference on Decision and Control*, pp. 4592–4594, 1997.
- [13] 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.
- [14] 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.
- [15] M. Rotkowitz, "Linear controllers are uniformly optimal for the Witsenhausen counterexample," *45th IEEE Conference on Decision and Control*, pp. 553–558, Dec. 2006.
- [16] M. Rotkowitz and S. Lall, "A characterization of convex problems in decentralized control," *IEEE Trans. Automat. Contr.*, vol. 51, no. 2, Feb. 2006.
- [17] N. C. Martins, "Witsenhausen's counter example holds in the presence of side information," *45th IEEE Conference on Decision and Control*, pp. 1111–1116, 2006.
- [18] 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.
- [19] G. Ausiello, P. Crescenzi, G. Gambosi, V. Kann, A. Marchetti-Spaccamela, and M. Protasi, *Complexity and Approximation: Combinatorial optimization problems and their approximability properties*. Springer Verlag, 1999.
- [20] R. Cogill, S. Lall, and J. P. Hespanha, "A constant factor approximation algorithm for event-based sampling," in *American Control Conference, 2007. ACC '07*, July 2007, pp. 305–311.
- [21] R. Etkin, D. Tse, and H. Wang, "Gaussian interference channel capacity to within one bit," *Submitted to IEEE Transactions on Information Theory*, Feb. 2007.
- [22] P. Grover, S. Y. Park, and A. Sahai, "The finite-dimensional Witsenhausen counterexample," in *ConCom 2009*, Seoul, Korea, Mar. 2009.
- [23] A. Sahai and P. Grover, "The price of certainty : "waterslide curves" and the gap to capacity," *Submitted to IEEE Transactions on Information Theory*, Dec. 2007. [Online]. Available: <http://arXiv.org/abs/0801.0352v1>
- [24] W. Feller, *An Introduction to Probability Theory and its Applications*. New York, NY: Wiley, 1950, vol. 2.
- [25] A. Sahai, "Any-time information theory," Ph.D. dissertation, Massachusetts Institute of Technology, Cambridge, MA, 2001.
- [26] A. Sahai and S. K. Mitter, "Source coding and channel requirements for unstable processes," *IEEE Trans. Inform. Theory*, Submitted, 2006. [Online]. Available: <http://www.eecs.berkeley.edu/~sahai/Papers/anytime.pdf>