Is Witsenhausen’s counterexample a relevant toy?

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

Abstract—This paper answers a question raised by Doyle on the relevance of the Witsenhausen counterexample as a toy distributed control problem. The question has two sides, the first of which focuses on the lack of an external channel in the counterexample. Using existing results, we argue that the core difficulty in the counterexample is retained even in the presence of such a channel. The second side questions the LQG formulation of the counterexample. We consider frameworks alternative to the LQG formulation and show that the understanding developed for the LQG case guides the investigation for these other cases as well. Specifically, we consider 1) a variation on the original counterexample with general bounded noise distributions, and 2) an adversarial extension with bounded disturbance and quadratic costs. For each of these formulations, we show that quantization-based nonlinear strategies outperform linear strategies by an arbitrarily large factor. Further, these nonlinear strategies also perform within a constant factor of the optimal, uniformly over all possible parameter choices.

As an aside, the assumption of bounded noise results in a significant simplification of proofs as compared to those for the LQG formulation. Therefore, the results in this paper are also of pedagogical interest.

I. INTRODUCTION

Recently, we provided the first provably approximately optimal solution to the Witsenhausen counterexample and its vector extensions [1], [2]. The solutions are obtained using techniques from information theory that help understand the implicit communication between the controllers — the ability of one controller to ‘talk’ to the other by making changes to the state of the system. The counterexample was discussed quite a bit in the symposium on ‘Paths ahead in the science of information and decision systems’ held at MIT LIDS in honor of Prof. Sanjoy Mitter. The ensuing discussions led Prof. John Doyle to question the relevance of the counterexample as a toy problem in distributed control. The goal of this paper is to convince the reader of the relevance.

It is hard to define what constitutes a useful and relevant toy problem. In order to obtain better understanding of what such a problem could be, it is useful to look at the following toy problem from the neighboring field of information theory: communicating a source across a power-constrained AWGN channel to minimize the average quadratic distortion in reconstructing the source. The problem is a toy for three reasons: real world communication problems today are almost always bigger than just point-to-point links, noise is non-Gaussian in most practical situations, and a quadratic distortion cost is not the perceptually ‘correct’ cost criterion for most sources. Even though assumptions of the problem make it a toy, it is a useful toy: insights gained from the problem provide a foundation upon which understanding was developed for larger communication problems (e.g. see [3]) including those with multiple transmitters and receivers, multiple antennas, and non-Gaussian noise.

Back to Doyle’s question, his first argument rests on the work of Rotkowitz and Lall [4], which shows that with extremely fast, infinite-capacity, and perfectly reliable external channels, the optimal controllers are linear not just for the Witsenhausen counterexample (which is a simple observation), but for more general problems as well. Given that using an external channel is often a valid engineering option in distributed control problems, Doyle argued that Witsenhausen’s counterexample may be artificially hard because it does not allow the controllers to talk over an external channel.

In practice, however, such a channel never has infinite capacity or perfect reliability. In fact, a growing body of the control theory literature (for example [5], [6]) addresses the same general issue: control over noisy and finite-capacity communication channels. In the presence of an imperfect external channel connecting the two controllers in Witsenhausen’s counterexample, Martins [7] shows that while finding optimal solutions continues to be hard, one can design signaling-based nonlinear strategies guided by those developed for the original counterexample. A closer inspection of the problem in [7] reveals that such signaling-based nonlinear strategies can outperform linear ones by an arbitrarily large factor for any fixed SNR on the external channel. In fact, Martins also shows that in some cases, nonlinear strategies outperform the optimal linear even without using an external channel. Provisioning for a very high SNR external channel may therefore be unnecessary as long as nonlinear control techniques are used.

Doyle’s second argument is about the relevance of the LQG framework in Witsenhausen’s counterexample. Linearity is fine, but do we believe that primitive random variables are Gaussian? Or that the designer is wedded to quadratic costs? The answer is no! Primitive random variables are

Martins shows that nonlinear strategies that do not even use the external channel can outperform linear ones that do use the channel where the external channel SNR is high. As is suggested by what David Tse calls the “deterministic perspective” (along the lines of [8]–[10]), linear strategies do not make good use of the external channel because they only communicate the “most significant bits” — which can anyway be estimated reliably at the second controller. So if the uncertainty in the initial state is large, the external channel is only of limited help and there may be substantial advantage in having the controllers talk through the plant. A similar problem is considered by Shoharnejad et al in [11], where noisy side information of the source is available at the receiver. Since this formulation is even more constrained than that in [7], it is clear that nonlinear strategies outperform linear for this problem as well.
almost never Gaussian, and the cost function is chosen more freely by the designer — the quadratic case is one possible formalization of the intuition that the cost increases at an increasing rate. As suggested by Doyle, of interest here is the work of Rotkowitz [12]. Rotkowitz shows that for the adversarial $L_2$-induced norm, instead of the original expected quadratic cost in Witsenhausen’s formulation, linear control laws are optimal and easy to find. At the same time, noise and initial state realizations can be completely arbitrary. Doyle’s implicit argument, based on Rotkowitz’s observation, is that because there is nothing sacred about the choice of a norm, viewed through the lens of a different norm (and with fewer assumptions), Witsenhausen’s problem is no longer hard! From an induced-norm perspective, the problem seems no more intriguing than other team-theoretic problems with two controllers.

The rest of this paper addresses this second argument. Our contention is that Rotkowitz’s induced norm formulation is one with paranoid controllers — they budget against the initial state and noise chosen adversarially without any restrictions. Usually one knows something about the initial state and noise, say their typical values, or a bound on their magnitudes. Assuming a known bound on the magnitude of noise, we consider two formulations (introduced in Section II) 1) a variation on the original counterexample with general noise distribution (Section III), and 2) an adversarial (robust control) extension with quadratic costs (Section IV).

For each of these formulations, using the known bound on noise magnitude, we show that quantization-based nonlinear strategies outperform linear strategies by an arbitrarily large factor. Further, these strategies also attain costs that are within a constant factor of the optimal cost. Using the two formulations, we argue that within the LQ framework, a relaxation of Gaussianity of the noise does not change the core issue that made the problem hard in the first place. A closer examination of our proofs reveals that quadratic nature of the cost function is not of critical importance either — it is merely convenient when finding lower bounds on the cost function. We believe that the results would extend to some other cost functions as well.

Thus, if the noise takes values within a bounded set (as is often the case in practice, and is also frequently assumed in robust control formulations [14, Ch.8, 9]), then there is a significant loss in restricting the control techniques to be linear. Further, in such situations, the intuitions and techniques used to obtain approximately optimal solutions within the LQG framework [1], [2] extend naturally to provide approximately optimal solutions to these new problems. In fact, the proofs for bounded noise formulations considered in this paper are substantially simpler than those for the LQG formulation — a finite-length analysis in the style of [2] is not needed to show approximate optimality.

II. Notation and Problem Statement

Vectors are denoted in bold, with the superscript to denote their length (e.g. $x^m$ is a vector of length $m$). Upper case tends to be used for random variables or vectors, while lower case symbols represent their realizations. Hats (\(\hat{}\)) on the top of random variables denote the estimates of the random variables. The block-diagram for both of the formulations considered in the paper is shown in Fig. 1.

The state evolution is as shown in Fig. 1. A control strategy is denoted by $\gamma = (\gamma_1, \gamma_2)$, where $\gamma_i$ is the function that maps the observation $y^m_i$ at $C_i$ to the control input $u^m_i$. The observations are given by $y^m_1 = x^m_0$ and $y^m_2 = x^m_1 + z^m$, where $z^m$ is the disturbance, or the noise at the input of the second controller. For the first two formulations, the total cost is a quadratic function of the state and the input given by:

$$J(\gamma)(x^m_0, z^m) = \frac{1}{m}k^2\|u^m_1\|^2 + \frac{1}{m}\|u^m_2\|^2,$$

where $u^m_1 = \gamma_1(x^m_0)$, $x^m_2 = x^m_0 + \gamma_1(x^m_0) - u^m_1$ and $u^m_2 = \gamma_2(x^m_0 + \gamma_1(x^m_0) + z^m)$. The cost expression is normalized by the vector-length $m$ to allow for natural comparisons between different vector-lengths.

We now provide the two problem formulations that are addressed in the paper.

A. A stochastic formulation

The initial state $X^m_0$ is Gaussian, distributed $\mathcal{N}(0, \sigma^2_{X^m_0})$, where $\|m|$ is the identity matrix of size $m \times m$. The observation noise $Z^m$ is distributed iid according to distribution $f_Z(z)$ with finite differential entropy $h(Z)$, finite variance $\sigma^2_Z$, and bounded support contained in $(-a, a)$, where $(-a, a)$ is also the smallest interval symmetric about zero that contains the support. Without loss of generality, we assume that $\sigma^2_Z = 1$. For example, for a uniformly distributed $Z$, $\sigma^2_Z = 1$ for $a = \sqrt{3}$.

The control objective is to minimize the expected quadratic cost $J(\gamma)$,

$$J(\gamma) = \mathbb{E}\left[J(\gamma)\right] = \frac{1}{m}k^2\mathbb{E}\left[\|U^m_1\|^2\right] + \frac{1}{m}\mathbb{E}\left[\|X^m_2\|^2\right].$$

Even though a finite-length analysis is needed to obtain tighter bounds on the associated constant factors.
over the choice of $\gamma$. The cost is averaged over the random realizations of $X_m^0$ and $Z^m$. We use the variable $P := \frac{1}{m} \mathbb{E} \left[ \| U_1^m \|^2 \right]$ to denote the power of the input $u_1^m$, and $MMSE = \frac{1}{m} \mathbb{E} \left[ \| X_1^m \|^2 \right] = \frac{1}{m} \mathbb{E} \left[ \| X_1^m - U_2^m \|^2 \right]$ to denote the second stage cost.

**B. An adversarial formulation with quadratic cost**

The block-diagram is the same as that for the stochastic problem. The total cost is still the same function given by (1), however, the cost for a strategy $\gamma$ is given by the maximum cost under the constraint that $|z_i| < \sqrt{3}$ for all $i$. That is,

$$J_{\text{adv, quad}}(\gamma) = \sup_{x_0^m, \| z^m \| < \sqrt{3}} J(\gamma)(x_0^m, z^m).$$

**III. STOCHASTIC MODELS FOR STATE AND NOISE**

**A. Upper bound on costs**

**Theorem 1:** An upper bound on the costs for the stochastic problem of Section II-A is given by

$$J_{\text{opt}} \leq \min \left\{ k^2 a^2, \frac{\sigma^2}{2}, k^2 \sigma^2 \right\}.$$ (4)

**Proof:** We consider the following three strategies 1) an essentially scalar quantization strategy that quantizes the entire real line with bins of sizes $2a$ in each dimension, 2) the zero-input strategy, followed by LLSE estimation at the second controller, and 3) the zero-forcing strategy. For a given $(k, \sigma)$-pair, the strategy with minimum cost is chosen.

For quantization strategy, the input forces the state to the nearest quantization point. The magnitude of the input is therefore bounded by $a$. Since the bins are disjoint, there are never any errors at the second controller (because the noise is smaller than $a$). The cost is therefore upper bounded by $k^2 a^2$.

For zero-input strategy with LLSE estimation, the cost is the same as that in the Gaussian case zero-input strategy of [1] of $\frac{\sigma^2}{2}$ (because MMSE and LLSE are the same in the Gaussian formulation). For zero-forcing, the input is forced to zero, and thus cost is $k^2 \sigma^2$. This completes the proof.

**B. A lower bound on the costs**

**Theorem 2:** A lower bound on the costs for the stochastic problem of Section II-A with observation noise $Z$ of variance 1 and differential entropy $h(Z)$ is given by

$$J_{\text{opt}} \geq \inf_{P \geq 0} k^2 P + \left( \sqrt{\kappa(P)} - \sqrt{P} \right)^2.$$ (5)

where

$$\kappa(P) = \frac{\sigma^2 + 2h(P)}{2\pi e \left( \sigma + \sqrt{P} \right)^2 + 1}. $$ (6)

**Proof:** The proof proceeds along the lines of the proof of Theorem 3 in [1]. That is, for a fixed $P := \frac{1}{m} \mathbb{E} \left[ \| U_1^m \|^2 \right]$, we first obtain a lower bound on the $MMSE$.

To that end, we need the following lemma [1, Lemma 3].

**Lemma 1:** For any three vector random variables $A$, $B$ and $C$,

$$\sqrt{\mathbb{E} \left[ \| B - C \|^2 \right]} \geq \left| \sqrt{\mathbb{E} \left[ \| A - C \|^2 \right]} - \sqrt{\mathbb{E} \left[ \| A - B \|^2 \right]} \right|.$$ (7)

**Proof:** See [1].

Substituting $X_m^0$ for $A$, $X_1^m$ for $B$, and $U_2^m$ for $C$ in Lemma 1, we get

$$\sqrt{\mathbb{E} \left[ \| X_m^0 - U_2^m \|^2 \right]} \geq \sqrt{\mathbb{E} \left[ \| X_0^m - X_1^m \|^2 \right]} - \sqrt{\mathbb{E} \left[ \| X_0^m - X_1^m \|^2 \right]}.$$ (8)

We wish to lower bound $\mathbb{E} \left[ \| X_m^0 - U_2^m \|^2 \right]$. The second term on the RHS is smaller than $\sqrt{m} \mathbb{E}$. Therefore, it suffices to lower bound the first term on the RHS of (8). To that end, we will interpret $U_2^m$ as an estimate for $X_m^0$. The mutual information $I(X_m^0 : Y_2^m)$ is bounded as follows

$$I(X_m^0 : Y_2^m) = h(Y_2^m) - h(Y_2^m | X_m^0) \leq \sum_i h(Y_{2,i}) - h(Y_{2,i} | X_{1,i}) \leq \sum_i I(X_{1,i} ; Y_{2,i}) \leq m I(X_1 ; Y_2).$$

where $X_1 = X_{1,i}$ if $Q = i$ (and $Y_2$ is defined similarly), and $Q$ is distributed uniformly on the discrete set $\{1, 2, \ldots, m\}$. Now,

$$Y_2 = X_1 + Z = X_0 + U_1 + Z.$$

The variance of $Y_2$ is maximized when $X_0$ and $U_1$ are aligned, and it equals $(\sigma + \sqrt{P})^2 + 1$. Thus,

$$I(X_1 ; Y_2) = h(Y_2) - h(Y_2 | X_1) \leq h(Y_2) - h(Z) \leq \frac{1}{2} \log_2 \left( 2 \pi e \left( \sigma + \sqrt{P} \right)^2 + 1 \right) - h(Z) \leq \frac{1}{2} \log_2 \left( 2 \pi e \left( \sigma + \sqrt{P} \right)^2 + 1 \right),$$

where $(a)$ follows from the observation that for given second moment of the random variable, the distribution that maximizes the differential entropy is Gaussian.

Pretending we wish to communicate $X_m^0$ across the $X_1 - Y_2$ channel (instead of $X_m^0$), we can obtain a lower bound on the distortion in reconstructing $X_m^0$ as follows: $X_0^m$ is a Gaussian source that needs to be communicated across a
channel of mutual information (and hence also the capacity) upper bounded by the expression in (9). The distortion in reconstructing $X_0^m$ is therefore lower bounded by $D_\sigma^2(R)$ where $D_\sigma^2(R) := \sigma^2 2^{-2R}$ is the distortion-rate function [15, Ch. 13] of a Gaussian source, and $C_{X_1 - Y_2}$ is the capacity across the $X_1 - Y_2$ channel.

Thus, the mean-squared distortion in reconstructing $X_0^m$ is lower bounded by

$$\frac{1}{m} \mathbb{E} \left[ \|X_0^m - U_r^m\|^2 \right] \geq \frac{D_\sigma^2(C_{X_1 - Y_2})}{\sigma^2 2^{2h(z)}} \geq \frac{2\pi e}{2\pi e \left( (\sigma + \sqrt{P})^2 + 1 \right)}.$$

A lower bound on the MMSE follows from (8) and (10). The theorem follows from the minimizing the sum of $k^2 P$ and MMSE over non-negative values of $P$.

C. Quantization-based strategies are approximately optimal

The following theorem shows that quantization-based strategies, complemented by linear strategies, are approximately optimal for the problem.

**Theorem 3:** For the problem as stated in Section II-A, the mean-squared distortion of the optimal for the problem.

$$\inf_{P \geq 0} k^2 P + \left( \left( \sqrt{\kappa(P)} - \sqrt{P} \right)^2 \right) \leq \bar{J}_{\text{opt}}$$

$$\leq \mu \left( \inf_{P > 0} k^2 P + \left( \left( \sqrt{\kappa(P)} - \sqrt{P} \right)^2 \right) \right),$$

where $\mu \leq \frac{200 a^2}{2 \pi e}$, and the upper bound is achieved by quantization-based strategies, complemented by linear strategies. For example, for $Z \sim \mathcal{U}(-\sqrt{3}, \sqrt{3})$, the uniform distribution of variance 1, $\mu \leq 50$.

**Proof:** The proof is along the lines of proof of Theorem 1 of [1]. We consider two cases:

**Case 1:** $\sigma^2 < \frac{1}{2}$. If $P > \frac{2^{2h(z)}}{200}$, using the zero-forcing upper bound of $k^2 \sigma^2$, the ratio is smaller than $\frac{200}{2^{2h(z)}}$.

If $P \leq \frac{2^{2h(z)}}{200}$, then

$$\kappa(P) = \frac{\sigma^2 2^{2h(z)}}{2\pi e \left( (\sigma + \sqrt{P})^2 + 1 \right)} \geq \frac{2\pi e \left( 2^{2h(z)} \right)^2}{200} \geq \frac{2\pi e \left( 1 + \sqrt{2^{2h(z)}} \right)^2 + 1}{46},$$

where (a) follows from the fact that $h(Z) \leq \frac{1}{2} \log_2 (2\pi e)$, the differential entropy for the $N(0, 1)$ random variable. Thus,

$$\left( \left( \kappa - \sqrt{P} \right)^2 \right) \geq \frac{\sigma^2 2^{2h(z)} (1 + \sqrt{2^{2h(z)}})^2}{200 \pi e} \geq \frac{16}{200} \geq \frac{\sigma^2 2^{2h(z)}}{200}.$$

Using the zero-input upper bound of $\frac{\sigma^2}{\sigma^2 + 1} < 1$, the ratio in this case is bounded by $\frac{200}{2^{2h(z)}}$.

**Case 2:** $\sigma^2 > 1$. If $P > \frac{2^{2h(z)}}{200}$, using the upper bound of $k^2 \sigma^2$, the ratio of upper and lower bounds is smaller than $\frac{\kappa^2 \sigma^2 + 1}{k^2 \sigma^2 + \frac{1}{2}} = \frac{200 a^2}{2^{2h(z)}}$. If $P \leq \frac{2^{2h(z)}}{200}$, (again, because Gaussian distribution maximizes the differential entropy for given variance),

$$\kappa(P) = \frac{\sigma^2 2^{2h(z)}}{2\pi e \left( (\sigma + \sqrt{P})^2 + 1 \right)} \geq \frac{2\pi e \left( 1 + \sqrt{P} \right)^2 + 1}{46} \geq \frac{\sigma^2 2^{2h(z)}}{200}.$$

Thus, the following lower bound holds for the MMSE error

$$\text{MMSE} \geq 2^{2h(z)} \left( \frac{1}{46} - \frac{1}{200} \right)^2 \geq 2^{2h(z)} 0.0058.$$

Using the zero-input upper bound, the ratio is smaller than

$$\frac{2^{2h(z)} \left( \frac{1}{46} - \frac{1}{200} \right)^2}{2^{2h(z)} 0.0058} < \frac{1}{173}$$

Using the fact that $a > 1$, we get the theorem.

Note that the result is not asymptotic — the constant factor is uniform over all vector lengths (we note that the ratio improves as the vector length increases to infinity because both upper and lower bounds improve due to concentration).

**Remark:** The constant factor of $\frac{200 a^2}{2^{2h(z)}}$ is not really uniform over all problem parameters, since it is a function of $h(Z)$ and $a$. However, scaling the distribution by a factor of $\beta$ would increase both the numerator and the denominator by a factor of $\beta^2$, keeping the ratio constant. Thus, fixing the noise distribution and the initial state distribution, scaling either of them is not going to alter the constant factor. We also note that tighter bounds on the constant factor, that depend only on the variance of the noise (and not on $a$), can be derived in the limit of long vector lengths using laws of large numbers (in the style of [1]).

D. Quantization-based strategies outperform linear strategies by an unbounded factor

Consider the scalar case. A linear constraint on the second controller forces it to perform an LLSE estimation on the output $Y_2$ in order to estimate $X_1$. The first controller, also linear, scales the state either up or down by a factor $\alpha$. The resulting state $X_1$ has variance $\sigma^2 = \sigma^2 (1 + \alpha^2)$. The mean-squared estimation error is, therefore, $\frac{\sigma^2}{\alpha^2 + 1}$. Since this is an increasing function of $\sigma^2$, the optimizing $\alpha$ is negative, and $P = \alpha^2 \sigma^2$. Thus the total cost for the optimal linear strategy is

$$J_{\text{lin}} = k^2 P + \frac{\left( (\sigma - \sqrt{P})^2 \right)^2}{\left( \sigma - \sqrt{P} \right)^2 + 1}.$$

(11)
Clearly, this cost remains the same in the vector case as well.

We now consider two cases. If \( P < \frac{\sigma^2}{4} \),
\[
\hat{J}_{\text{lin}} \geq \frac{\left( \sigma - \sqrt{P} \right)^2}{\left( \sigma - \sqrt{P} \right)^2 + 1}
\]
where \((a)\) follows from the fact that \( P < \frac{\sigma^2}{4} \). In the limit of \( \sigma^2 \to \infty \) and \( k \to 0 \), this lower bound increases to 1, whereas the quantization upper bound of \( 3k^2 \) decreases to zero.

Alternatively, if \( P \geq \frac{\sigma^2}{4} \),
\[
\hat{J}_{\text{lin}} \geq k^2 P \geq \frac{k^2 \sigma^2}{4}.
\]
Thus, the ratio of the costs attained by the optimal linear strategy and those attained by the quantization upper bound is larger than \( \frac{k^2 \sigma^2}{4 \pi} = \frac{\sigma^2}{12} \), which diverges to infinity as \( k \to 0 \), \( \sigma^2 \to \infty \).

IV. ADVERSARIAL MODEL FOR NOISE AND STATE

**Theorem 4:** The optimal cost for adversarially modeled initial state and (bounded) noise \( Z \in (-\sqrt{3}, \sqrt{3}) \) with quadratic costs (as defined in Section II-B) is bounded as follows
\[
\inf_{P \geq 0} k^2 P + \left( \frac{6}{\pi e} - \sqrt{P} \right)^2 \leq J_{\text{adv,quad}} \leq 2\pi e \left( \inf_{P \geq 0} k^2 P + \left( \frac{6}{\pi e} - \sqrt{P} \right)^2 \right),
\]
where the upper bound is achieved using quantization-based strategies complemented by linear strategies. Further, in the regime of \( k \to 0 \), the ratio of the costs attained by the best linear strategy to that attained by appropriate quantization-based nonlinear strategies diverges to infinity.

**Proof:** *Upper bound:* If \( k^2 \leq 1 \), we use a uniform quantization strategy with bin size \( 2\sqrt{3} \). Since the noise amplitude is smaller than \( \sqrt{3} \), there are no errors at the second controller. The cost of this strategy is therefore \( 3k^2 \), which is attained in the event when the initial state is exactly at the edge of one of the quantization bins.

If \( k^2 > 1 \), we use the zero-input strategy — the first controller inputs zero, and the second controller chooses \( U_2^m = Y_2^m \) as the estimate of \( X^m \). Since noise amplitude is bounded by \( \sqrt{3} \), the normalized error for this strategy is bounded by 3.

The upper bound is therefore given by \( \min\{3k^2, 3\} \).

*Lower bound:* Even though the noise is chosen adversarially (and deterministically), we first assume that the noise behaves as a random variable with distribution \( U\{-\sqrt{3}, \sqrt{3}\} \), and the initial state behaves as a Gaussian with variance \( \sigma^2 \), for some \( \sigma^2 > 0 \). We assume that the adversary declares this strategy in advance (which can only reduce the costs). If the first controller chooses an average power \( P \), then the \( \text{MMSE} \) at the second controller is lower bounded by
\[
\text{MMSE} \geq \left( \sqrt{\kappa_{\text{quad}}(P) - \sqrt{P}} \right)^2, 
\]
where
\[
\kappa_{\text{quad}}(P) = \frac{\sigma^2 2^{2h(Z)}}{2\pi e (\sigma + \sqrt{P})^2 + 1}. \tag{14}
\]
Since this lower bound holds for all \( \sigma^2 \), we let \( \sigma^2 \to \infty \), and obtain the following bound,
\[
\text{MMSE} \geq \left( \sqrt{\lim_{\sigma^2 \to \infty} \kappa_{\text{quad}}(P) - \sqrt{P}} \right)^2, 
\]
\[
\geq \left( \frac{2^{2h(Z)}}{2\pi e - \sqrt{P}} \right)^2, 
\]
\[
= \left( \frac{6}{\pi e} - \sqrt{P} \right)^2, 
\]
where \((a)\) follows from observing that the optimizing \( P \leq 3 \), since at \( P = 3 \), the upper bound on \( \text{MMSE} \) itself is zero.

A lower bound on the average costs (averaged over the initial state and noise realizations) for this problem is therefore given by
\[
\hat{J}_{\text{opt}} \geq \inf_{P \geq 0} k^2 P + \left( \sqrt{\frac{6}{\pi e} - \sqrt{P}} \right)^2. \tag{15}
\]
We now invoke an argument inspired by the probabilistic method [16]. Since this lower bound holds for any given choice of control strategy averaged over the realizations of the initial state and noise, there exists at least one (deterministic) pair of realizations of initial state and noise such that this lower bound holds for the given control strategy \( \gamma \). Thus this lower bound holds for a deterministic adversary as well.

**Bounded ratios:**

*Case 1:* \( P < \frac{6}{3\pi e} \).

In this case,
\[
\text{MMSE} \geq \left( \sqrt{\frac{6}{\pi e} - \sqrt{P}} \right)^2, 
\]
\[
\geq \left( \sqrt{\frac{6}{\pi e} - \frac{1}{2} \frac{6}{\pi e}} \right)^2, 
\]
\[
= \frac{6}{\pi e} - \frac{3}{2\pi e}, 
\]
which is also a lower bound on the cost. Thus the ratio of the zero-input upper bound (which is 3) and this lower bound is smaller than \( \frac{3\times 2\pi e}{3\times 2\pi e} = \frac{3}{2\pi e} \).

*Case 2:* \( P \geq \frac{3}{\frac{6}{\pi e}} \).

In this case, the cost is no smaller than \( k^2 P = k^2 \frac{6}{3\pi e} \). Thus
the ratio of quantization-based upper bound (which is $3k^2$) and this lower bound is smaller than $\frac{3k^2 \times 2\pi e}{3k^2} = 2\pi e$.

The ratio of the upper and lower bound is therefore always smaller than $2\pi e \approx 17.08$.

**Nonlinear strategies can outperform linear by an arbitrary factor:**

As in the proof of constant factor optimality, assume that the noise behaves as $U(-\sqrt{3}, \sqrt{3})$, and the initial state behaves as $N(0, \sigma^2)$. The average costs attained by any linear strategy are bounded by $\inf P k^2 P + \frac{(\sigma - \sqrt{P})^2}{(\sigma - \sqrt{P})^2 + 1}$. Again, by the argument of the probabilistic method, there exists a realization of noise that attains a lower bound no smaller than the average. Also, the costs attained by quantization-based strategies is bounded by $3k^2$. Thus the ratio of costs attained by the optimal linear strategy to that of quantization-based strategies diverges to infinity as $k \to 0$ and $\sigma^2 \to \infty$ (with the remaining proof the same as that in the stochastic case).

**ACKNOWLEDGMENTS**

We thank Gireeja Ranade for pointing errors in an early version of this paper, and for helpful comments. This research is supported by NSF grants CCF-0729122, CCF-0917212 and CNS-0932410.

**REFERENCES**


