



**HAL**  
open science

# Continuous Random Variable Estimation is not Optimal for the Witsenhausen Counterexample

Mael Le Treust, Tobias J Oechtering

► **To cite this version:**

Mael Le Treust, Tobias J Oechtering. Continuous Random Variable Estimation is not Optimal for the Witsenhausen Counterexample. IEEE ISIT 2021, Jul 2021, melbourne, Australia. hal-03223047v2

**HAL Id: hal-03223047**

**<https://hal.science/hal-03223047v2>**

Submitted on 19 May 2021

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Continuous Random Variable Estimation is not Optimal for the Witsenhausen Counterexample

Maël Le Treust  
ETIS UMR 8051,

CY Cergy Paris Université, ENSEA, CNRS  
95014 Cergy-Pontoise CEDEX, France

Tobias J. Oechtering

KTH Royal Institute of Technology EECS  
Div. Inform. Science and Engineering  
10044 Stockholm, Sweden

**Abstract**—Optimal design of distributed decision policies can be a difficult task, illustrated by the famous Witsenhausen counterexample. In this paper we characterize the optimal control designs for the vector-valued setting assuming that it results in an interim state, i.e. the result of the first decision maker action, that can be described by a continuous random variable which has a probability density function. More specifically, we provide a genie-aided outer bound that relies on our previous results for empirical coordination problems. This solution turns out to be not optimal in general, since it consists of a time-sharing strategy between two linear schemes of specific power. It follows that the optimal decision strategy for the original scalar Witsenhausen problem must lead to an interim state that cannot be described by a continuous random variable which has a probability density function.

## I. INTRODUCTION

Distributed decision-making systems arise in many engineering problems where decentralized agents choose actions based on locally available information as to minimize a common cost function. The information at each agent is either locally observed or received from other agents. Since the process of sharing information comes with a cost, agents usually do not have access to the whole information available at all agents. The design of optimal decision strategies for such problems is considered to be notoriously difficult. The Witsenhausen counterexample [1] from 1968 is an outstanding toy example that has significantly helped to understand the fundamental difficulty that actions serve two purposes, a control purpose affecting the system state and a communication purpose providing information to other agents [2]. More generally, this problem is also referred to a team decision problem for which the existence of the optimal solution has been studied in [3].

Although Witsenhausen refuted with his simple two-point counterexample the assertion that a linear policy would be also optimal in such a Gaussian setting, the optimal non-linear policy remains unknown. Many researcher have approached the optimization problem with various methods. In the last decade for instance it has been approached with numerical optimization methods [4], [5], where the latter is based on an iterative source-channel coding approach. An asymptotically

The authors gratefully acknowledge the financial support of SRV ENSEA for visits at KTH in Stockholm in 2017 and 2019, and at ETIS in Cergy in 2018. This research has been conducted as part of the Labex MME-DII (ANR11-LBX-0023-01). Part of the research has been supported by Swedish Research Council (VR) under grant 2020-03884.

optimal approximation approach has been presented in [6]. Analytically, using results from optimal transport theory, it has been shown in [7] that the optimal decision strategy is a strictly increasing unbounded piece-wise real analytic function with a real analytic left inverse. More necessary conditions have been derived in [8] by analyzing an equivalent optimization problem on the space of square-integrable quantile functions. However, it is unclear if the optimal decision policy of the first agent results in an interim state that can be described by a continuous random variable.

In this work, we show that the optimal decision strategy will not lead to an interim state that can be described by a continuous random variable that has a probability density function. The observation points on a subtle point in an outer bound argument, which might be easily overseen. We will further discuss that this observation, and in essence also the Witsenhausen counterexample, can be easily explained by the relation between the MMSE and mutual information considering Gaussian or binary distributed input [9].

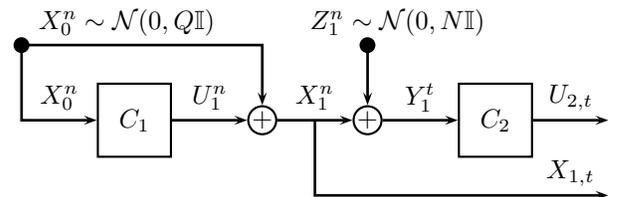


Fig. 1. The state and the channel noise are drawn according to the i.i.d. Gaussian distributions  $X_0^n \sim \mathcal{N}(0, Q\mathbb{I})$  and  $Z_1^n \sim \mathcal{N}(0, N\mathbb{I})$ . The interim state sequence  $X_1^n$  is causally estimated by decision maker  $C_2$ .

Our approach is based on a vector-valued extension of the Witsenhausen counterexample as proposed by Grover and Sahai in [10]. They study a non-causal encoding and decoding strategy that combines a coding scheme with side information and a linear scheme, which has been shown to be optimal by Choudhuri and Mitra in [11]. It has later been observed that such problems can be also approached as an empirical coordination coding problem. In [12], we have provided an overview on the individual findings and completed the missing cases using coding results from [13]. In [14] we have derived an achievability result considering non-causal encoding and causal decoding using a continuous alphabet building on proof

methods from [15]. In this work, we now derive a genie-aided outer bound for this case considering only decision strategies that result in continuous random variables which have a probability density function.

## II. SYSTEM MODEL

In this work, we restrict our study to continuous random variables which have a probability density function (pdf), so that the joint differential entropy is defined according to [16, Chap. 8]. For brevity we only refer to continuous random variables.

We consider the vector-valued Witsenhausen setup in which the sequences of states and channel noises are drawn independently according to the i.i.d. Gaussian distributions  $X_0^n \sim \mathcal{N}(0, Q\mathbb{I})$  and  $Z_1^n \sim \mathcal{N}(0, N\mathbb{I})$  with  $\min(Q, N) > 0$ , where  $\mathbb{I}$  denotes the identity matrix. We denote by  $X_1$  the interim state and  $Y_1$  the output of the noisy channel.

$$X_1 = X_0 + U_1 \quad \text{with } X_0 \sim \mathcal{N}(0, Q), \quad (1)$$

$$Y_1 = X_1 + Z_1 = X_0 + U_1 + Z_1 \quad \text{with } Z_1 \sim \mathcal{N}(0, N). \quad (2)$$

We denote by  $\mathcal{P}_{X_0} = \mathcal{N}(0, Q)$  the Gaussian state distribution and by  $\mathcal{P}_{X_1 Y_1 | X_0 U_1}$  the conditional probability distribution corresponding to equations (1) and (2).

**Definition 1.** For  $n \in \mathbb{N}^* = \mathbb{N} \setminus \{0\}$ , a ‘‘control design’’ with non-causal encoder and causal decoder is a tuple of stochastic functions  $c = (f, \{g_t\}_{t \in \{1, \dots, n\}})$  defined by

$$f : \mathcal{X}_0^n \longrightarrow \mathcal{U}_1^n, \quad g_t : \mathcal{Y}_1^t \longrightarrow \mathcal{U}_2, \quad \forall t \in \{1, \dots, n\}, \quad (3)$$

which induces a distribution over the sequences given by

$$\left( \prod_{t=1}^n \mathcal{P}_{X_{0,t}} \right) f_{U_1^n | X_0^n} \left( \prod_{i=t}^n \mathcal{P}_{X_{1,t} Y_{1,t} | X_{0,t} U_{1,t}} \right) \left( \prod_{t=1}^n g_{U_{2,t} | Y_1^t} \right).$$

We denote by  $\mathcal{C}_d(n)$  the set of control designs with non-causal encoder and causal decoder  $c = (f, \{g_t\}_{t \in \{1, \dots, n\}})$  that induce sequences of continuous random variables.

**Definition 2.** We define the  $n$ -stage costs associated with  $c$  by

$$\gamma_p^n(c) = \begin{cases} \mathbb{E} \left[ \frac{1}{n} \sum_{t=1}^n U_{1,t}^2 \right] & \text{if it exists,} \\ +\infty & \text{otherwise,} \end{cases} \quad (4)$$

$$\gamma_s^n(c) = \begin{cases} \mathbb{E} \left[ \frac{1}{n} \sum_{t=1}^n (X_{1,t} - U_{2,t})^2 \right] & \text{if it exists,} \\ +\infty & \text{otherwise.} \end{cases} \quad (5)$$

The pair of costs  $(P, S) \in \mathbb{R}^2$  is achievable if for all  $\varepsilon > 0$ , there exists  $\bar{n} \in \mathbb{N}^*$ , for all  $n \geq \bar{n}$ , there exists a control design  $c \in \mathcal{C}_d(n)$  such that

$$\left| P - \gamma_p^n(c) \right| + \left| S - \gamma_s^n(c) \right| \leq \varepsilon. \quad (6)$$

**Theorem 1.** The pair of Witsenhausen costs  $(P, S)$  is achievable if and only if there exists continuous random variables with probability distribution that decomposes according to

$$\mathcal{P}_{X_0} \mathcal{Q}_{U_1 W_1 W_2 | X_0} \mathcal{P}_{X_1 Y_1 | X_0 U_1} \mathcal{Q}_{U_2 | W_2 Y_1}, \quad (7)$$

where  $(W_1, W_2)$  are continuous auxiliary random variables such that  $0 \leq I(W_1; Y_1 | W_2) - I(W_1, W_2; X_0)$  and

$$P = \mathbb{E}_{\mathcal{Q}}[U_1^2], \quad S = \mathbb{E}_{\mathcal{Q}}[(X_1 - U_2)^2]. \quad (8)$$

This result is stated in [14, Theorem 1].

**Remark 1.** The probability distribution in (7) satisfies

$$\begin{cases} (X_1, Y_1) \text{---} (X_0, U_1) \text{---} (W_1, W_2), \\ U_2 \text{---} (Y_1, W_2) \text{---} (X_0, X_1, U_1, W_1), \\ X_2 \text{---} (X_1, U_2) \text{---} (X_0, U_1, Y_1, W_1, W_2). \end{cases} \quad (9)$$

The causality condition prevents the controller  $C_2$  to recover  $W_1$  which induces the second Markov chain of (9).

**Definition 3.** The optimal cost considering continuous random variables is characterized by the optimization problem defined as follows

$$S_c(P) = \inf_{\mathcal{Q} \in \mathcal{Q}_c(P)} \mathbb{E}_{\mathcal{Q}}[(X_1 - U_2)^2], \quad (10)$$

$$\mathcal{Q}_c(P) = \left\{ \left( \mathcal{Q}_{U_1 W_1 W_2 | X_0}, \mathcal{Q}_{U_2 | W_2 Y_1} \right) \text{ s.t. } P = \mathbb{E}_{\mathcal{Q}}[U_1^2], \right. \\ \left. I(W_1; Y_1 | W_2) - I(W_1, W_2; X_0) \geq 0 \text{ and } \right. \\ \left. X_0, U_1, W_1, W_2, X_1, Y_1, U_2 \text{ are continuous} \right\}. \quad (11)$$

**Lemma 1** (Lemma 11 in [1]). The best linear scheme is  $U_1 = -\sqrt{\frac{P}{Q}}X_0$  if  $P \leq Q$ , otherwise  $U_1 = -X_0 + \sqrt{P-Q}$ .<sup>1</sup> This induces the estimation cost given by

$$S_\ell(P) = \begin{cases} \frac{(\sqrt{Q}-\sqrt{P})^2 \cdot N}{(\sqrt{Q}-\sqrt{P})^2 + N} & \text{if } P \in [0, Q], \\ 0 & \text{otherwise.} \end{cases} \quad (12)$$

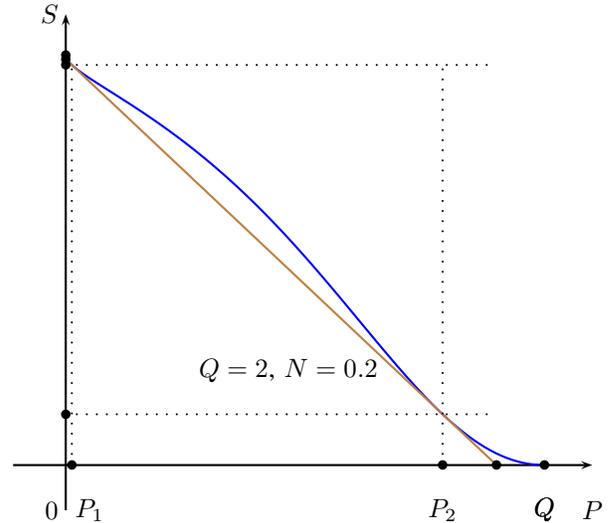


Fig. 2. The curve  $S_\ell(P)$  in (12) and the straight line  $\frac{N \cdot (Q-N-P)}{Q}$ .

<sup>1</sup>If  $P > Q$ , the interim state  $X_1$  can be canceled and the offset  $\sqrt{P-Q}$  is only included to meet the power constraint with equality as in (6).

**Theorem 2.** *The optimal cost with continuous random variables satisfies*

$$S_c(P) = \begin{cases} \frac{N \cdot (Q - N - P)}{Q} & \text{if } Q > 4N \text{ and } P \in [P_1, P_2], \\ S_\ell(P) & \text{otherwise,} \end{cases} \quad (13)$$

$$P_1 = \frac{1}{2} \left( Q - 2N - \sqrt{Q \cdot (Q - 4N)} \right), \quad (14)$$

$$P_2 = \frac{1}{2} \left( Q - 2N + \sqrt{Q \cdot (Q - 4N)} \right). \quad (15)$$

The proof of Theorem 2 is stated in Sec. V-A and V-B. Figure 2 represents the cost of the linear scheme  $S_\ell(P)$  and the line of equation  $P \mapsto \frac{N \cdot (Q - N - P)}{Q}$ . Note that the upper bound in (13) may be obtained by using either a time sharing strategy between the two linear schemes with parameters  $P_1$  and  $P_2$  when  $Q > 4N$  and  $P \in [P_1, P_2]$ , or a linear scheme of power  $P$ . This result shows that memoryless policies are optimal so that these policies are also optimal for the original scalar Witsenhausen counterexample setup restricted to continuous random variables. However, as pointed out by Witsenhausen in [1], such a strategy in the original scalar model is generally not optimal!

### III. WITSENHAUSEN'S TWO-POINTS STRATEGY

The two-point strategy described in [1, pp. 141] outperforms the optimal cost with continuous random variables  $S_c(P)$  of Theorem 2. We consider  $Q = 10$ ,  $N = 1$  and the sender's strategy with parameter  $a \geq 0$  given by

$$U_1 = a \cdot \text{sign}(X_0) - X_0, \quad (16)$$

which induces  $X_1 = a \cdot \text{sign}(X_0)$ . For all  $a \geq 0$ , the pair of costs are given by

$$P_t(a) = Q + a \left( a - 2\sqrt{\frac{2Q}{\pi}} \right), \quad (17)$$

$$S_t(a) = \frac{a^2}{\sqrt{2\pi}} e^{-\frac{a^2}{2}} \int \frac{e^{-\frac{y^2}{2}}}{\cosh ay} dy. \quad (18)$$

Fig. 3 shows that for some  $a \in [0.05, 5]$ , the pair of costs  $(P_t(a), S_t(a))$  Pareto-dominates  $S_c(P)$  of Theorem 2, for some  $P$ .

From the previous, we have the following Theorem.

**Theorem 3.** *There exists values for  $Q$ ,  $N$ ,  $a$  for which we have  $S_t(a) \leq S_c(P_t(a))$ .*

#### Discussion

In the following we also briefly want to explain the Witsenhausen counterexample result in terms of the I-MMSE formula by Guo, Shamai and Verdu [9]. The formula has been used to illustrate in [9, Fig. 1] that binary inputs in additive Gaussian noise channels result in a lower MMSE than Gaussian distributed inputs with the same SNR. In other words, to achieve the same MMSE, binary distributed input requires less channel input power than Gaussian distributed input. Exactly this power gain has been exploited in the Witsenhausen counterexample scheme in which the interim

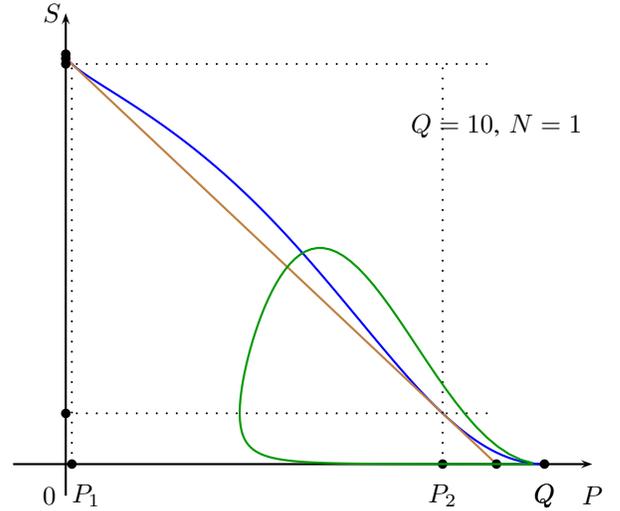


Fig. 3. The lasso-shaped curve corresponds to the Witsenhausen's two-point strategy with parametric equations (17) and (18) for  $a \in [0.05, 5]$ . The upper curve corresponds to the best linear scheme  $S_\ell(P)$  in (12) and the straight line  $\frac{N \cdot (Q - N - P)}{Q}$  provides the value of the time-sharing strategy for  $P \in [P_1, P_2]$ .

state  $X_1$  is binary so that the resulting MMSE outperforms the MMSE of the best linear scheme. Analytically, it is interesting to see that the MMSE formulas (17) and (18) directly relate to [9, Equations (13) and (17)] with the corresponding signal powers and noise power.

### IV. CONCLUSION

Our results show that information theoretic methods, in particular coordination coding results, lead to new insights on the Witsenhausen counterexample. Vice versa, we believe that our observation makes the Witsenhausen counterexample also interesting for other source-channel coding problems. In more detail, we show that a convex combination of linear memoryless policies is optimal for the vector-valued Witsenhausen problem with causal decoder restricted to the space of continuous random variables. Since the policy is memoryless, it follows that the linear policy is also optimal for the original Witsenhausen problem restricted to the space of continuous random variables which have a pdf. However, Witsenhausen's two-points strategy outperforms the best linear strategy, which implies that the hypothesis of a continuous random variable which has a pdf is an active restriction for the Witsenhausen counter-example setup. According to the Lebesgue's decomposition Theorem, every probability distribution on the real line is a mixture of discrete part, singular part, and an absolutely continuous part. Accordingly, we conclude that the optimal decision strategy for the unrestricted Witsenhausen's counterexample must lead to an interim state that cannot be described by a continuous random variable which has a pdf. In future works, we will consider policies that result in interim states described by more general probability distributions.

## V. PROOFS

### A. Lower bound for the Theorem 2

The Markov chain  $Y_1 \text{---} (X_0, U_1) \text{---} (W_1, W_2)$  implies

$$\begin{aligned} & I(W_1; Y_1 | W_2) - I(W_1, W_2; X_0) \\ & \leq I(W_1; Y_1 | W_2, X_0) - I(W_2; X_0) \end{aligned} \quad (19)$$

$$\leq I(U_1; Y_1 | W_2, X_0) - I(W_2; X_0). \quad (20)$$

Therefore

$$S_c(P) \geq \min_{\substack{\mathcal{Q}_{U_1, W_2 | X_0} \in \mathcal{Q}_1(P), \\ \mathcal{Q}_{U_2 | W_2, Y_1}}} \mathbb{E}_{\mathcal{Q}}[(X_1 - U_2)^2] \quad (21)$$

$$\geq \min_{\mathcal{Q}_{U_1, W_2 | X_0} \in \mathcal{Q}_1(P)} \mathbb{E}_{\mathcal{Q}}\left[\left(X_1 - \mathbb{E}[X_1 | W_2, Y_1]\right)^2\right], \quad (22)$$

where

$$\begin{aligned} \mathcal{Q}_1(P) = \left\{ \mathcal{Q}_{U_1, W_2 | X_0} \text{ s.t. } P = E_{\mathcal{Q}}[U_1^2], \right. \\ \left. \begin{aligned} & I(U_1; Y_1 | W_2, X_0) - I(W_2; X_0) \geq 0, \\ & (X_0, U_1, W_2, X_1, Y_1, U_2) \text{ are continuous} \end{aligned} \right\}, \end{aligned} \quad (23)$$

The random variables  $(X_0, W_2, U_1)$  drawn according to  $\mathcal{Q}'_{U_1, W_2 | X_0}$  which is optimal for (22), have covariance matrix

$$K = \begin{pmatrix} Q & \rho_1 \sqrt{QV} & \rho_2 \sqrt{QP} \\ \rho_1 \sqrt{QV} & V & \rho_3 \sqrt{VP} \\ \rho_2 \sqrt{QP} & \rho_3 \sqrt{VP} & P \end{pmatrix}, \quad (24)$$

where the correlation coefficients  $(\rho_1, \rho_2, \rho_3) \in [-1, 1]^3$  are such that  $QVP \cdot (1 - \rho_1^2 - \rho_2^2 - \rho_3^2 + 2\rho_1\rho_2\rho_3) \geq 0$ , i.e.  $K$  is semi-definite positive.

We denote by  $\mathcal{Q}_{U_1, W_2 | X_0}$  the Gaussian conditional distribution such that  $(X_0, W_2, U_1) \sim \mathcal{N}(0, K)$ . According to [17, Maximum Differential Entropy Lemma, pp. 21], we have

$$\mathbb{E}_{\mathcal{Q}'}\left[\left(X_1 - \mathbb{E}[X_1 | Y_1, W_2]\right)^2\right] \geq \frac{1}{2\pi e} \cdot 2^{2h(X_1 | Y_1, W_2)} \quad (25)$$

$$= \mathbb{E}_{\mathcal{Q}}\left[\left(X_1 - \mathbb{E}[X_1 | Y_1, W_2]\right)^2\right]. \quad (26)$$

Moreover, both  $\mathcal{P}_{X_0, \mathcal{Q}'_{U_1, W_2 | X_0}}$  and  $\mathcal{P}_{X_0, \mathcal{Q}_{U_1, W_2 | X_0}}$  satisfy

$$0 \leq I_{\mathcal{Q}'}(U_1; Y_1 | W_2, X_0) - I_{\mathcal{Q}'}(W_2; X_0) \quad (27)$$

$$\begin{aligned} & = h_{\mathcal{Q}'}(Y_1 | W_2, X_0) - h_{\mathcal{Q}}(Y_1 | W_2, X_0, U_1) \\ & \quad - h_{\mathcal{Q}}(X_0) + h_{\mathcal{Q}'}(X_0 | W_2) \end{aligned} \quad (28)$$

$$\begin{aligned} & \leq h_{\mathcal{Q}}(Y_1 | W_2, X_0) - h_{\mathcal{Q}}(Y_1 | W_2, X_0, U_1) \\ & \quad - h_{\mathcal{Q}}(X_0) + h_{\mathcal{Q}}(X_0 | W_2) \end{aligned} \quad (29)$$

$$= I_{\mathcal{Q}}(U_1; Y_1 | W_2, X_0) - I_{\mathcal{Q}}(W_2; X_0), \quad (30)$$

where (28) comes from  $h_{\mathcal{Q}'}(X_0) = h_{\mathcal{Q}}(X_0)$  and  $h_{\mathcal{Q}'}(Y_1 | W_2, X_0, U_1) = h_{\mathcal{Q}}(Z) = h_{\mathcal{Q}}(Y_1 | W_2, X_0, U_1)$ , and (29) comes from [16, (8.92), pp. 256].

**Lemma 2.** Assume that  $(X_0, W_2, U_1) \sim \mathcal{N}(0, K)$ , then

$$\begin{aligned} & I(U_1; Y | X_0, W_2) - I(X_0; W_2) \\ & = \frac{1}{2} \log_2 \left( \frac{P}{N} \cdot (1 - \rho_1^2 - \rho_2^2 - \rho_3^2 + 2\rho_1\rho_2\rho_3) + (1 - \rho_1^2) \right), \end{aligned} \quad (31)$$

$$\begin{aligned} & \mathbb{E}_{\mathcal{Q}}\left[\left(X_1 - \mathbb{E}[X_1 | Y_1, W_2]\right)^2\right] \\ & = \frac{N \cdot \left(Q \cdot (1 - \rho_1^2) + P \cdot (1 - \rho_3^2) + 2\sqrt{QP} \cdot (\rho_2 - \rho_1\rho_3)\right)}{N + \left(Q \cdot (1 - \rho_1^2) + P \cdot (1 - \rho_3^2) + 2\sqrt{QP} \cdot (\rho_2 - \rho_1\rho_3)\right)}. \end{aligned} \quad (32)$$

The proof of Lemma 2 is stated in Sec. V-C. Note that the equations (31) and (32) do not depend on the variance parameter  $V$  of the auxiliary random variable  $W_2$ . Also, when (31) is positive the matrix  $K$  is semi-definite positive.

By using Lemma 2, we reformulate (22) and since the function  $x \rightarrow \frac{N \cdot x}{N+x}$  is strictly increasing for all  $x \geq 0$ , the optimal parameters  $(\rho_1^*, \rho_2^*, \rho_3^*) \in [-1, 1]^3$  minimize

$$Q \cdot (1 - \rho_1^2) + P \cdot (1 - \rho_3^2) + 2\sqrt{QP} \cdot (\rho_2 - \rho_1\rho_3), \quad (33)$$

under the constraint

$$\frac{P}{N} \cdot (1 - \rho_1^2 - \rho_2^2 - \rho_3^2 + 2\rho_1\rho_2\rho_3) - \rho_1^2 \geq 0 \quad (34)$$

$$\iff (1 - \rho_1^2) \cdot (1 - \rho_3^2) - \frac{N}{P} \cdot \rho_1^2 \geq (\rho_2 - \rho_1\rho_3)^2, \quad (35)$$

which yields

$$\rho_2^* = \rho_1\rho_3 - \sqrt{(1 - \rho_1^2) \cdot (1 - \rho_3^2) - \frac{N}{P} \cdot \rho_1^2}. \quad (36)$$

**Lemma 3.** If  $Q > 4N$  and  $P \in [P_1, P_2]$ , then

$$\rho_1^{*2} = \frac{P \cdot Q - (P + N)^2}{(P + N) \cdot Q}, \quad \rho_3^{*2} = 0. \quad (37)$$

If  $Q \leq 4N$  or if  $Q > 4N$  and  $P \in [0, P_1] \cup [P_2, Q]$ , then

$$\rho_1^{*2} = 0, \quad \rho_3^{*2} = 0. \quad (38)$$

If  $P > Q$ , then

$$\rho_1^{*2} = 0, \quad \rho_3^{*2} = \frac{P - Q}{P}. \quad (39)$$

The proof of Lemma 3 is stated in Sec. V-D. We obtain the lower bound by replacing the optimal parameters  $(\rho_1^*, \rho_2^*, \rho_3^*) \in [-1, 1]^3$  in (32), that is if  $P \in [0, Q]$

$$S_c(P) \geq \begin{cases} \frac{N \cdot (Q - N - P)}{Q} & \text{if } Q > 4N \text{ and } P \in [P_1, P_2], \\ \frac{(\sqrt{Q} - \sqrt{P})^2 \cdot N}{(\sqrt{Q} - \sqrt{P})^2 + N} & \text{otherwise.} \end{cases} \quad (40)$$

### B. Upper bound for the Theorem 2

1) *Linear Scheme:* By considering the best linear scheme

$$U_1 = \begin{cases} -\sqrt{\frac{P}{Q}} \cdot X_0, & \text{if } P < Q, \\ -X_0 + \sqrt{P - Q}, & \text{if } P \geq Q, \end{cases} \quad (41)$$

we have that  $S_c(P) \leq S_\ell(P)$ , for all  $P \geq 0$ .

2) *Case where  $Q > 4N$  and  $P \in [P_1, P_2]$ :* The upper bound of Theorem 2 can be obtained by using a time sharing strategy between the two linear schemes with parameters  $P_1$  and  $P_2$ . We obtain the same result by assuming that the random variables  $(U_1, W_1, W_2)$  are drawn according to

$$W_2 = \sqrt{\frac{P+N}{PQ - (P+N)^2}} \cdot (X_0 - Z_0) \sim \mathcal{N}(0, 1),$$

with  $Z_0 \sim \mathcal{N}\left(0, \frac{QN + (P+N)^2}{P+N}\right)$  and  $Z_0 \perp W_2$ , (42)

$$W_1 = \frac{PQ - (P+N)^2}{Q(P+N)} \cdot X_0 + U_0, \text{ with } U_0 \perp (X_0, W_2)$$

and  $U_0 \sim \mathcal{N}\left(0, \frac{N \cdot (PQ - (P+N)^2)}{QN + (P+N)^2}\right)$ , (43)

$$U_1 = -\frac{(P+N)^2}{QN + (P+N)^2} \cdot X_0 + U_0$$

$$+ \sqrt{\frac{PQ - (P+N)^2}{P+N}} \cdot \frac{(P+N)^2}{QN + (P+N)^2} \cdot W_2. \quad (44)$$

Then we have

$$I(W_1; Y_1, W_2) - I(W_1; X_0, W_2) = I(U_1; Y_1 | X_0, W_2) \quad (45)$$

$$= I(X_0; W_2) = \frac{1}{2} \log_2 \left( 1 + \frac{PQ - (P+N)^2}{QN + (P+N)^2} \right), \quad (46)$$

and

$$S_c(P) \leq \frac{N \cdot (Q - P - N)}{Q}. \quad (47)$$

#### C. Proof of Lemma 2

We consider  $(X_0, W_2, U_1) \sim \mathcal{N}(0, K)$  with  $K$  defined in (24), which together with (2), induces the Gaussian random variables  $(X_0, W_2, Y_1)$  whose entropy is

$$h(X_0, W_2, Y) = \frac{1}{2} \log_2 \left( (2\pi e)^3 \cdot QV \quad (48)$$

$$\times \left( P \cdot (1 - \rho_1^2 - \rho_2^2 - \rho_3^2 + 2\rho_1\rho_2\rho_3) + N \cdot (1 - \rho_1^2) \right) \right). \quad (49)$$

Therefore we have

$$I(U_1; Y | X_0, W_2) - I(X_0; W_2)$$

$$= h(X_0, W_2, Y) - h(Y | U_1, X_0, W_2) - h(X_0) - h(W_2) \quad (50)$$

$$= \frac{1}{2} \log_2 \left( \frac{P}{N} \cdot (1 - \rho_1^2 - \rho_2^2 - \rho_3^2 + 2\rho_1\rho_2\rho_3) + (1 - \rho_1^2) \right). \quad (51)$$

According to (1) and (2) the entropy of  $(X_1, W_2, Y_1)$  writes

$$h(X_1, W_2, Y) = \frac{1}{2} \log_2 \left( (2\pi e)^3 \cdot V \cdot N \quad (52)$$

$$\times \left( Q \cdot (1 - \rho_1^2) + P \cdot (1 - \rho_3^2) + 2\sqrt{QP} \cdot (\rho_2 - \rho_1\rho_3) \right) \right), \quad (53)$$

and hence

$$\mathbb{E} \left[ (X_1 - \mathbb{E}(X_1 | Y, W_2))^2 \right]$$

$$= \frac{N \cdot \left( Q \cdot (1 - \rho_1^2) + P \cdot (1 - \rho_3^2) + 2\sqrt{QP} \cdot (\rho_2 - \rho_1\rho_3) \right)}{N + \left( Q \cdot (1 - \rho_1^2) + P \cdot (1 - \rho_3^2) + 2\sqrt{QP} \cdot (\rho_2 - \rho_1\rho_3) \right)}. \quad (54)$$

#### D. Proof of Lemma 3

We replace  $\rho_2^*$  in (33) and we define

$$f(\rho_1^2, \rho_3^2) = Q \cdot (1 - \rho_1^2) + P \cdot (1 - \rho_3^2)$$

$$- 2\sqrt{QP} \cdot \sqrt{(1 - \rho_1^2) \cdot (1 - \rho_3^2) - \frac{N}{P} \cdot \rho_1^2}. \quad (55)$$

Note that  $f$  is well defined if  $\rho_1^2 \leq \frac{P}{P+N}$  and  $\rho_3^2 \leq 1 - \frac{N}{P} \cdot \frac{\rho_1^2}{1 - \rho_1^2}$ .

$$\frac{\partial f(\rho_1^2, \rho_3^2)}{\partial \rho_3^2} = \sqrt{PQ} \cdot \frac{1 - \rho_1^2}{\sqrt{(1 - \rho_1^2) \cdot (1 - \rho_3^2) - \frac{N}{P} \cdot \rho_1^2}} - P, \quad (56)$$

then for all  $\rho_1^2 \leq \frac{P}{P+N}$ , the optimal  $\rho_3^{2*}(\rho_1^2)$  is

$$\rho_3^{2*}(\rho_1^2) = \max \left( 1 - \left( \frac{Q}{P} \cdot (1 - \rho_1^2) + \frac{N}{P} \cdot \frac{\rho_1^2}{1 - \rho_1^2} \right), 0 \right). \quad (57)$$

We introduce the parameters

$$\rho_a = \frac{2Q - (P+N) - \sqrt{(P+N)^2 - 4QN}}{2Q}, \quad (58)$$

$$\rho_b = \frac{2Q - (P+N) + \sqrt{(P+N)^2 - 4QN}}{2Q}, \quad (59)$$

and we define the function

$$F(\rho_1^2) = f(\rho_1^2, \rho_3^{2*}(\rho_1^2))$$

$$= \begin{cases} Q \cdot (1 - \rho_1^2) + P \\ \quad - 2\sqrt{QP} \cdot \sqrt{1 - \rho_1^2 \cdot \frac{P+N}{P}} & \text{if } 0 \leq \rho_1^2 \leq \rho_a, \\ N \cdot \frac{\rho_1^2}{1 - \rho_1^2} & \text{if } \rho_a \leq \rho_1^2 \leq \rho_b, \\ Q \cdot (1 - \rho_1^2) + P \\ \quad - 2\sqrt{QP} \cdot \sqrt{1 - \rho_1^2 \cdot \frac{P+N}{P}} & \text{if } \rho_b \leq \rho_1^2 \leq \frac{P}{P+N}. \end{cases} \quad (60)$$

The function  $F(\rho_1^2)$  is continuous in  $\rho_a$  and  $\rho_b$ . We define

$$\rho^* = \frac{P \cdot Q - (P+N)^2}{(P+N) \cdot Q}. \quad (61)$$

• If  $Q > 4N$  and  $P \in [P_1, P_2]$ , then the function  $F(\rho_1^2)$  is decreasing over the interval  $\rho_1^2 \in [0, \rho^*]$  and increasing over the interval  $\rho_1^2 \in [\rho^*, \frac{P}{P+N}]$ , then the optimal parameters are

$$\rho_1^2 = \rho^*, \quad \rho_3^2 = 0. \quad (62)$$

• If  $Q \leq 4N$  or if  $Q > 4N$  and  $P \in [0, P_1] \cup [P_2, Q]$ , then the optimal parameters are  $\rho_1^2 = \rho_3^2 = 0$ .

• If  $P > Q$ , then

$$\rho_1^2 = 0, \quad \rho_3^2 = \frac{P-Q}{P}. \quad (63)$$

## REFERENCES

- [1] H. Witsenhausen, "A counterexample in stochastic optimum control," *SIAM Journal on Control*, vol. 6, no. 1, pp. 131–147, 1968.
- [2] S. Yüksel and T. Başar, *Stochastic Networked Control Systems: Stabilization and Optimization under Information Constraints*, ser. Systems & Control Foundations & Applications. New York, NY: Springer, 2013.
- [3] A. Gupta, S. Yüksel, T. Başar and C. Langbort, "On the existence of optimal policies for a class of static and sequential dynamic teams," *SIAM Journal on Control and Optimization*, vol. 53, no. 3, pp. 1681–1712, 2015.
- [4] S.-H. Tseng and A. Tang, "A Local Search Algorithm for the Witsenhausens Counterexample," in *Proc. IEEE CDC*, 2017.
- [5] J. Karlsson, A. Gattami, T. J. Oechtering, and M. Skoglund, "Iterative source-channel coding approach to Witsenhausens counterexample," in *Proc. IEEE ACC*, 2011.
- [6] N. Saldı, S. Yüksel and T. Linder, "Finite model approximations and asymptotic optimality of quantized policies in decentralized stochastic control," in *Proc. IEEE CDC*, 2016.
- [7] Y. Wu and S. Verdú, "Witsenhausen's counterexample: A view from optimal transport theory," in *Proc. IEEE CDC*, 2011.
- [8] W. M. McEneaney and S. H. Han, "Optimization formulation and monotonic solution method for the Witsenhausen problem," *Automatica*, 2015.
- [9] D. Guo, S. Shamai and S. Verdú, "Mutual information and minimum mean-square error in Gaussian channels," in *IEEE Trans. Inform. Theory*, vol. 51, no. 4, pp. 1261–1282, 2005.
- [10] P. Grover and A. Sahai, "Witsenhausen's counterexample as assisted interference suppression," in *Int. J. Syst., Control Commun.*, vol. 2, nos. 1-3, pp. 197–237, 2010.
- [11] C. Choudhuri and U. Mitra, "On Witsenhausen's counterexample: the asymptotic vector case," *2012 IEEE ITW*, 2012.
- [12] M. Le Treust and T. J. Oechtering, "Optimal Control Designs for Vector-valued Witsenhausen Counterexample Setups," *IEEE 56th Annual Allerton Conf. on Commun., Control, and Comp.*, 2018.
- [13] M. Le. Treust, "Joint Empirical Coordination of Source and Channel," *IEEE Trans. Inf. Theory*, vol. 63, no. 8, pp. 5087–5114, 2017.
- [14] T. J. Oechtering and M. Le Treust, "Coordination Coding with Causal Decoder for Vector-valued Witsenhausen Counterexample Setups," *IEEE Information Theory Workshop (ITW)*, 2019.
- [15] M. T. Vu, T. J. Oechtering, and M. Skoglund, "Hierarchical Identification With Pre-Processing," *IEEE Trans. Inf. Theory*, 2020.
- [16] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed. Wiley & Sons, 2006.
- [17] A. El Gamal and Y.-H. Kim, *Network Information Theory*. Cambridge University Press, 2011.