



**HAL**  
open science

# Optimal Control Designs for Vector-valued Witsenhausen Counterexample Setups

Mael Le Treust, Tobias Oechtering

► **To cite this version:**

Mael Le Treust, Tobias Oechtering. Optimal Control Designs for Vector-valued Witsenhausen Counterexample Setups. Allerton Conference 2019, Oct 2018, Monticello, Illinois, United States. hal-02278220

**HAL Id: hal-02278220**

**<https://hal.science/hal-02278220>**

Submitted on 4 Sep 2019

**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.

# Optimal Control Designs for Vector-valued Witsenhausen Counterexample Setups

Maël Le Treust

ETIS UMR 8051, Université Paris Seine,  
Université Cergy-Pontoise, ENSEA, CNRS  
95014 Cergy-Pontoise, France

Tobias J. Oechtering

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

**Abstract**—In this work, necessary and sufficient conditions for empirical coordination of vector-valued Witsenhausen counterexample two terminal setups with non-classical information structure are derived. Vector-valued processing allows to involve coding in the design of the control strategies. Optimal characterizations are obtained for the non-causal encoding and causal decoding case as well as causal encoding and non-causal decoding case. Necessary and sufficient conditions are provided for the case with both non-causal encoding and decoding. The feasible set of target distributions can serve as optimization domain for characterizing the optimal average cost, in particular using Witsenhausen’s cost function.

**Index Terms**—Witsenhausen counterexample, empirical coordination, feasible target distributions, non-causal and causal coding strategies

## I. INTRODUCTION

The design of optimal decision strategies in distributed stochastic networks with non-classical information structures is a long-standing difficult problem. The famous counterexample of Witsenhausen introduced in 1968 in [1] showed that non-linear strategies can outperform the best linear strategy [2]. Until today the setup serves as primary study object to develop a better understanding on the impact of the information structure, on the optimal decision strategy design problem [3]. The fundamental structure of the best decision strategies for the Witsenhausen counterexample problem have been reproduced in several different numerical optimisation attempts, while the currently best result is achieved with an iterative source-channel coding approach in [4].

Another approach, more information-theoretic, is to consider a multi-letter version of the problem. In a series of works, [5], [6], [7] to mention a few, Grover et al. studied the setup where the decision makers have access to a sequence of observations enabling block-coding strategies. This allowed them to transfer advanced coding techniques [8], [9], [10], [11] to the vector-valued Witsenhausen counterexample problem. In one of their last works [7], which also provides a good literature overview, they extended the concept of *dual* (role of) control to *triple* roles by adding an explicit communication

Maël Le Treust thanks the supports of DIM-RFSI under grant EX032965, and of Labex MME-DII (ANR11-LBX-0023-01). Tobias Oechtering thanks the support of the Swedish Research Council under grants 2016-03853 and E0628201. The authors gratefully acknowledge the financial support of SRV ENSEA for visits at KTH Stockholm in 2017 and at ETIS Cergy in 2018.

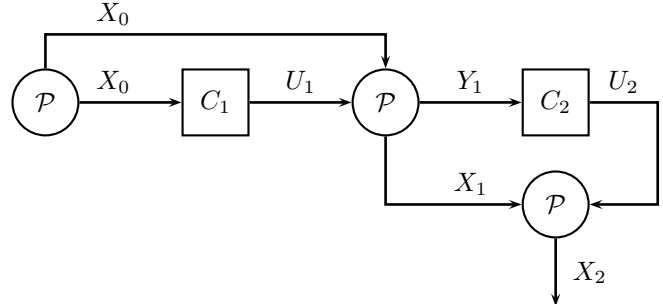


Fig. 1. The information source  $\mathcal{P}(x_0)$ , the channel  $\mathcal{P}(x_1, y_1|x_0, u_1)$  and the final state  $\mathcal{P}(x_2|x_1, u_2)$  are i.i.d. distributed. The encoder  $C_1$  and the decoder  $C_2$  may be causal or non-causal

task to the problem highlighting the fundamental tension among the tasks. Previously, in [6] approximately lattice-based optimal solutions were obtained for the finite-length vector case. Recently, improved asymptotic bounds have been found in [12] using a new vector quantization scheme. Choudhuri and Mitra characterized the optimal power-distortion trade-off for the vector-valued Witsenhausen problem in [13], which prove the optimality of the coding scheme by Grover et al. in [5] combining linear coding and Costa’s dirty-paper-coding [14]. Much less work has been done considering the vector-valued Witsenhausen problem with causal processing although in several coding techniques have been extended to the causal case, e.g. causal state communication in [15] or estimation with a helper in [16].

In [17], Cuff and Zhao considered empirical coordination for a *cascade of controllers* that act on its observed signals where the empirical coordination criterion is a probabilistic statement on the statistics of the joint sequences. In particular, they point out that given a reward function, then the optimal average reward can be found by optimizing over the coordination set. An extension to more abstract alphabets has been done in [18] introducing a new definition of typical sequences and deriving properties using the Glivenko-Cantelli theorem. Originally, the problem of coordination between agents was introduced in [19], Gossner et al. for a two-player team game with asymmetric information. They also discussed the case of noisy observations, which relates to the case of noisy communication channels. In [20] the concept has been generalized

and the notion of coordination capacity (region) has been introduced, which can be used to characterize the joint behavior of distributed nodes, given communication constraints. In particular, results for simple multi-source settings considering empirical and strong coordination have been obtained. In [21], Cuff and Schieler investigated the case where the action of terminal one has also to be coordinated with the state and terminal two's action. Interestingly, the achievability proof relies on a hybrid coding strategy which conceptually can be seen as the multi-letter extension of the best Witsenhausen counterexample decision strategy.

In [22], [23], [24], [25], and [26] empirical coordination capacity results for two terminal settings with side information have been derived considering state-dependent channels as well as causal and non-causal encoding and decoding. In [23], necessary and sufficient conditions for the non-causal encoding and decoding case for a *cascade setting* have been presented, which includes the results of the lossless decoding case with correlated source and state presented in [22]. Optimal results have been obtained for the perfect channel case with two-sided side information and the case with independent source and channel. Further, optimal results have been presented for causal/non-causal encoding and decoding including sketches of the achievability and converse proofs, while full proofs were provided in [24] and [27]. In contrast, the authors of [25] characterize optimal conditions for a setting where both terminals (a.k.a. agents) provide a channel input whose output, that also depends on the system state, is observed by terminal two only. Next, they consider the special case without any channel input from terminal two for which they characterize, in [25, Theorem 3], the optimal solution for non-causal encoder and causal decoder.

An improved understanding of the fundamental distributed decision making problem is of great value due to its wide applications. For instance, Larrousse et al. applied in [28], [29] the coordination approach to a two agents distributed power allocation problem where only one agent is knowledgeable about a state and informs the second agent through its actions. The idea of (state) communication through actions is usually known as *dual control* where control actions have a second purpose. In [30] coordination in a two-agent setting with common average payoff function where each agent can control only one variable is considered. Such payoff function includes the Witsenhausen cost function as special case. The authors assumed standard Borel spaces to justify the transfer of coding results, from finite alphabets to continuous alphabets.

In this work we also consider two terminal settings with finite alphabets motivated by the Witsenhausen counterexample as illustrated in Figure 1. Terminal  $C_1$  has access to the independent and identically distributed (i.i.d.) system state  $X_0$ , which are non-causally or causally encoded into control actions  $U_1$ . A new system state  $X_1$  and observation at terminal  $C_2$  are probabilistically generated based on the current state  $X_0$  and action  $U_1$ . Terminal  $C_2$  causally or non-causally decodes the observations  $Y_1$  and decides on control action  $U_2$ , which probabilistically leads to the new system state  $X_2$ . In the

following we adapt and extend results in the literature on coordination to this Witsenhausen counterexample setup. In more detail, in Section II, necessary and sufficient conditions for optimal control designs with non-causal encoding and causal decoding are derived and in Section III, optimal necessary and sufficient conditions for the control designs for causal encoding and non-causal decoding are presented. Lastly, in Section IV both encoder and decoder are assumed to be non-causal. Achievability results are deduced from previous results in the literature [21], [23], [24], [25], [27], the converses are outlined in the appendices. The set of feasible target distributions can be then used to evaluate feasible average values for a given cost function. Accordingly, in Section V the optimization problem for the average cost function of Witsenhausen counterexample is provided.

## II. NON-CAUSAL ENCODER AND CAUSAL DECODER

The results in this section are closely related to the corresponding empirical coordination results in [23], [24], [25], [27]. In more detail, the result has been stated without proof for an almost similar setting without post-processing in [25, Theorem 3] as well as in [23, eq. (16)]. Closely related proofs have been shown in [24], [27]. We present here the result adapted to the Witsenhausen setting.

**Definition II.1** A “control design” with non-causal encoder and causal decoder is a tuple of functions  $c = (f, \{g_i\}_{i \in \{1, \dots, n\}})$  defined by:

$$f : \mathcal{X}_0^n \longrightarrow \mathcal{U}_1^n, \quad (1)$$

$$g_i : \mathcal{Y}_1^i \longrightarrow \mathcal{U}_2, \quad \forall i \in \{1, \dots, n\}. \quad (2)$$

We denote by  $\mathcal{C}_d(n)$  the set of control designs with non-causal encoder and causal decoder. This code induces a probability distribution over the sequences given by:

$$\begin{aligned} & \prod_{i=1}^n \mathcal{P}(x_{0,i}) \times \mathbb{1}(u_1^n = f(x_0^n)) \times \prod_{i=1}^n \mathcal{P}(x_{1,i}, y_{1,i} | x_{0,i}, u_{1,i}) \\ & \times \prod_{i=1}^n \mathbb{1}(u_{2,i} = g_i(y_1^i)) \times \prod_{i=1}^n \mathcal{P}(x_{2,i} | x_{1,i}, u_{2,i}). \end{aligned} \quad (3)$$

**Definition II.2** Given a sequence  $x_0^n \in \mathcal{X}_0^n$ , the empirical frequency of symbol  $x_0 \in \mathcal{X}_0$  is defined by:

$$Q^n(x_0) = \frac{1}{n} \cdot \sum_{i=1}^n \mathbb{1}(x_0 = x_{0,i}) \in [0, 1], \quad \forall x_0 \in \mathcal{X}_0, \quad (4)$$

where  $x_{0,i}$  denotes the  $i$ -th symbol in the sequence  $x_0^n$ . This defines the empirical distribution  $Q^n \in \Delta(\mathcal{X}_0)$  of the sequence  $x_0^n$ , where  $\Delta(\mathcal{X}_0)$  denotes the set of probability distributions over the finite set  $\mathcal{X}_0$ .

This definition extends to a vector of sequences  $(x_0^n, u_1^n, y_1^n, x_1^n, u_2^n, x_2^n)$  whose empirical distribution is given by:  $Q^n \in \Delta(\mathcal{X}_0 \times \mathcal{U}_1 \times \mathcal{X}_1 \times \mathcal{Y}_1 \times \mathcal{U}_2 \times \mathcal{X}_2)$ . With a slight abuse of notation, we denote by  $Q^n(x_0) \in \Delta(\mathcal{X}_0)$  and  $Q^n(x_0, u_1, x_1, y_1, u_2, x_2) \in \Delta(\mathcal{X}_0 \times \mathcal{U}_1 \times \mathcal{X}_1 \times \mathcal{Y}_1 \times \mathcal{U}_2 \times \mathcal{X}_2)$ ,

the respective empirical distributions of the sequences  $x_0^n$  and  $(x_0^n, u_1^n, y_1^n, x_1^n, u_2^n, x_2^n)$ .

**Definition II.3** A probability distribution  $\mathcal{Q} \in \Delta(\mathcal{X}_0 \times \mathcal{U}_1 \times \mathcal{X}_1 \times \mathcal{Y}_1 \times \mathcal{U}_2 \times \mathcal{X}_2)$  is achievable if for all  $\varepsilon > 0$ , there exists an  $\bar{n} \in \mathbb{N}$  such that for all  $n \geq \bar{n}$ , there exists a control design with non-causal encoder and causal decoder  $c \in \mathcal{C}_d(n)$  such that:

$$\mathcal{P}_e(c) = \mathcal{P}_c \left( \left\| Q^n - \mathcal{Q} \right\|_{\text{tv}} \geq \varepsilon \right) \leq \varepsilon, \quad (5)$$

where  $Q^n \in \Delta(\mathcal{X}_0 \times \mathcal{U}_1 \times \mathcal{X}_1 \times \mathcal{Y}_1 \times \mathcal{U}_2 \times \mathcal{X}_2)$  denotes the random variable of the empirical distribution of the sequences of symbols  $(X_0^n, U_1^n, X_1^n, Y_1^n, U_2^n, X_2^n)$  induced by the control design  $c \in \mathcal{C}_d(n)$  and the probability distributions of the source  $\mathcal{P}(x_0)$ , of the channel  $\mathcal{P}(x_1, y_1 | x_0, u_1)$  and of the final state  $\mathcal{P}(x_2 | x_1, u_2)$ .

The topology of the network induces several restriction regarding the set of achievable probability distributions. A first restriction is the one imposed by the marginals probability distribution of the source and channel.

**Definition II.4** A target probability distribution  $\mathcal{Q}(x_0, u_1, x_1, y_1, u_2, x_2)$  satisfies the marginal conditions of the source  $\mathcal{P}(x_0)$ , of the channel  $\mathcal{P}(x_1, y_1 | x_0, u_1)$  and of the final state  $\mathcal{P}(x_2 | x_1, u_2)$ , and decomposes as follows:

$$\begin{aligned} \mathcal{Q}(x_0, u_1, x_1, y_1, u_2, x_2) \\ = \mathcal{P}(x_0) \times \mathcal{Q}(u_1 | x_0) \times \mathcal{P}(x_1, y_1 | x_0, u_1) \\ \times \mathcal{Q}(u_2 | x_0, u_1, y_1) \times \mathcal{P}(x_2 | x_1, u_2). \end{aligned} \quad (6)$$

**Theorem II.5 (Non-Causal Encoder and Causal Decoder)** A target probability distribution  $\mathcal{P}(x_0) \times \mathcal{Q}(u_1 | x_0) \times \mathcal{P}(x_1, y_1 | x_0, u_1) \times \mathcal{Q}(u_2 | x_0, u_1, y_1) \times \mathcal{P}(x_2 | x_1, u_2)$  is achievable with causal encoder and non-causal decoder if and only if there exists a pair of auxiliary random variables  $(W_1, W_2)$  drawn according to:

$$\begin{aligned} \mathcal{P}(x_0) \times \mathcal{Q}(u_1, w_1, w_2 | x_0) \times \mathcal{P}(x_1, y_1 | x_0, u_1) \\ \times \mathcal{Q}(u_2 | w_2, y_1) \times \mathcal{P}(x_2 | x_1, u_2), \end{aligned} \quad (7)$$

whose marginals equal the target probability distribution and such that:

$$I(X_0; W_2) \leq I(W_1; Y_1, W_2) - I(W_1; X_0, W_2) \quad (8)$$

$$\iff 0 \leq I(W_1; Y_1 | W_2) - I(W_1, W_2; X_0). \quad (9)$$

The supports of the auxiliary random variables  $(W_1, W_2)$  are bounded by  $\max(|\mathcal{W}_1|, |\mathcal{W}_2|) \leq d+4$  with  $d = |\mathcal{X}_0 \times \mathcal{U}_1 \times \mathcal{Y}_1 \times \mathcal{U}_2|$ . We denote by  $\mathbb{Q}_d$  the set of achievable target probability distributions  $\mathcal{Q}(x_0, u_1, x_1, y_1, u_2, x_2)$ .

The achievability proof of Theorem II.5 comes from [24, Theorem V.1] that generalizes [27, Theorem 4]. The converse proof of Theorem II.5 is stated in App. A. This result was also stated in [25, Theorem 3].

**Remark II.6** The probability distribution of equation (7) satisfies:

$$\begin{cases} (X_1, Y_1) \text{---} (X_0, U_1) \text{---} (W_1, W_2), \\ U_2 \text{---} (Y_1, W_2) \text{---} (X_0, X_1, U_1, W_1), \\ X_2 \text{---} (X_1, U_2) \text{---} (X_0, U_1, Y_1, W_1, W_2). \end{cases} \quad (10)$$

The first and third Markov chain in (10) correspond to the channels  $\mathcal{P}(x_1, y_1 | x_0, u_1)$  after encoder  $C_1$  and  $\mathcal{P}(x_2 | x_1, u_2)$  after decoder  $C_2$ . The second Markov chain is due to the causal decoding which prevents the action  $U_2$  to depend on  $W_1$  as well.

### III. CAUSAL ENCODER AND NON-CAUSAL DECODER

The results in this section are closely related to the corresponding empirical coordination results in [21], [23], [24], which are based on the optimal coding scheme developed by Choudhuri and Mitra in [15]. Here we extend and adapt the results to the Witsenhausen setting.

**Definition III.1** A ‘‘control design’’ with causal encoder and non-causal decoder  $c \in \mathcal{C}_e(n)$  is a tuple of functions  $c = (\{f_i\}_{i \in \{1, \dots, n\}}, g)$  defined by:

$$f_i : \mathcal{X}_0^i \longrightarrow \mathcal{U}_1, \quad \forall i \in \{1, \dots, n\}, \quad (11)$$

$$g : \mathcal{Y}_1^n \longrightarrow \mathcal{U}_2^n. \quad (12)$$

This code induces a probability distribution over the sequences given by:

$$\begin{aligned} \prod_{i=1}^n \mathcal{P}(x_{0,i}) \times \prod_{i=1}^n \mathbb{1}(u_{1,i} = f_i(x_0^i)) \times \prod_{i=1}^n \mathcal{P}(x_{1,i}, y_{1,i} | x_{0,i}, u_{1,i}) \\ \times \mathbb{1}(u_2^n = g(y_1^n)) \times \prod_{i=1}^n \mathcal{P}(x_{2,i} | x_{1,i}, u_{2,i}). \end{aligned} \quad (13)$$

The notion of achievable target probability distribution with causal encoder and non-causal decoder is defined similarly as in Definition II.3, by replacing  $c \in \mathcal{C}_e(n)$  instead of  $c \in \mathcal{C}_d(n)$ .

**Theorem III.2 (Causal Encoder and Non-Causal Decoder)**

A target probability distribution  $\mathcal{P}(x_0) \times \mathcal{Q}(u_1 | x_0) \times \mathcal{P}(x_1, y_1 | x_0, u_1) \times \mathcal{Q}(u_2 | x_0, u_1, y_1) \times \mathcal{P}(x_2 | x_1, u_2)$  is achievable with causal encoder and non-causal decoder if and only if there exists a pair of auxiliary random variables  $(W_1, W_2)$  drawn according to:

$$\begin{aligned} \mathcal{P}(x_0) \times \mathcal{Q}(w_1) \times \mathcal{Q}(w_2 | x_0, w_1) \times \mathcal{Q}(u_1 | x_0, w_1) \\ \times \mathcal{P}(x_1, y_1 | x_0, u_1) \times \mathcal{Q}(u_2 | w_1, w_2, y_1) \times \mathcal{P}(x_2 | x_1, u_2), \end{aligned} \quad (14)$$

whose marginals equal the target probability distribution and such that:

$$0 \leq I(W_1, W_2; Y_1) - I(W_2; X_0 | W_1) \quad (15)$$

$$\iff 0 \leq I(W_1, W_2; Y_1) - I(W_1, W_2; X_0). \quad (16)$$

The supports of the auxiliary random variables  $(W_1, W_2)$  are bounded by  $\max(|\mathcal{W}_1|, |\mathcal{W}_2|) \leq d+4$  with  $d = |\mathcal{X}_0 \times \mathcal{U}_1 \times \mathcal{Y}_1 \times \mathcal{U}_2|$ .

$\mathcal{U}_2$ . We denote by  $\mathbb{Q}_e$ , the set of achievable target probability distributions  $\mathcal{Q}(x_0, u_1, x_1, y_1, u_2, x_2)$ .

The achievability proof of Theorem III.2 comes from [23, Theorem V.1], see also [24, Theorem VI.1]. The converse proof of Theorem III.2 is stated in App. B.

**Remark III.3** The probability distribution of equation (14) satisfies:

$$\begin{cases} X_0 \text{ independent of } W_1, \\ U_1 \text{ } \perp\!\!\!\perp (X_0, W_1) \text{ } \perp\!\!\!\perp W_2, \\ (X_1, Y_1) \text{ } \perp\!\!\!\perp (X_0, U_1) \text{ } \perp\!\!\!\perp (W_1, W_2), \\ U_2 \text{ } \perp\!\!\!\perp (W_1, W_2, Y_1) \text{ } \perp\!\!\!\perp (X_0, X_1, U_1), \\ X_2 \text{ } \perp\!\!\!\perp (X_1, U_2) \text{ } \perp\!\!\!\perp (X_0, U_1, Y_1, W_1, W_2). \end{cases} \quad (17)$$

The third and fifth Markov chains in (17) again correspond to the *channels*. The first two conditions are due the causal encoding. The fourth Markov chain shows that the action of decoder  $C_2$  can depend on both auxiliary random variables  $(W_1, W_2)$  and the noisy observation  $Y_1$  due the non-causal decoding.

#### IV. BOTH NON-CAUSAL ENCODER AND DECODER

In this section we adapt the empirical coordination results from [21], [23], [24] to the Witsenhausen counterexample setting.

**Definition IV.1** A “control design” with non-causal encoder and non-causal decoder  $c \in \mathcal{C}_{nc}(n)$  is a pair of functions  $c = (f, g)$  defined by:

$$f : \mathcal{X}_0^n \longrightarrow \mathcal{U}_1^n, \quad (18)$$

$$g_i : \mathcal{Y}_1^n \longrightarrow \mathcal{U}_2^n. \quad (19)$$

This code induces a probability distribution over the sequences given by:

$$\begin{aligned} & \prod_{i=1}^n \mathcal{P}(x_{0,i}) \times \mathbb{1}(u_1^n = f(x_0^n)) \times \prod_{i=1}^n \mathcal{P}(x_{1,i}, y_{1,i} | x_{0,i}, u_{1,i}) \\ & \times \mathbb{1}(u_2^n = f(y_1^n)) \times \prod_{i=1}^n \mathcal{P}(x_{2,i} | x_{1,i}, u_{2,i}). \end{aligned} \quad (20)$$

The notion of achievable target probability distribution with non-causal encoder and non-causal decoder is defined similarly as in Definition II.3, by replacing  $c \in \mathcal{C}_{nc}(n)$  instead of  $c \in \mathcal{C}_d(n)$ .

#### Theorem IV.2 (Both Non-Causal Encoder and Decoder)

• If the target probability distribution  $\mathcal{P}(x_0) \times \mathcal{Q}(u_1|x_0) \times \mathcal{P}(x_1, y_1|x_0, u_1) \times \mathcal{Q}(u_2|x_0, u_1, y_1) \times \mathcal{P}(x_2|x_1, u_2)$  is achievable with non-causal encoder and non-causal decoder then there exists an auxiliary random variable  $W$  drawn according to:

$$\begin{aligned} & \mathcal{P}(x_0) \times \mathcal{Q}(w|x_0) \times \mathcal{Q}(u_1|x_0, w) \times \mathcal{P}(x_1, y_1|x_0, u_1) \\ & \times \mathcal{Q}(u_2|w, y_1) \times \mathcal{P}(x_2|x_1, u_2), \end{aligned} \quad (21)$$

whose marginals equal the target probability distribution and such that:

$$0 \leq I(U_1, X_0; Y_1) - I(W; X_0). \quad (22)$$

• If there exists an auxiliary random variable  $W$  drawn according to (21), whose marginals equal the target probability distribution and such that:

$$0 \leq I(W; Y_1) - I(W; X_0), \quad (23)$$

then the target probability distribution  $\mathcal{P}(x_0) \times \mathcal{Q}(u_1|x_0) \times \mathcal{P}(x_1, y_1|x_0, u_1) \times \mathcal{Q}(u_2|x_0, u_1, y_1) \times \mathcal{P}(x_2|x_1, u_2)$  is achievable with non-causal encoder and non-causal decoder.

• The supports of the auxiliary random variables  $W$  are bounded by  $|\mathcal{W}| \leq d + 4$  with  $d = |\mathcal{X}_0 \times \mathcal{U}_1 \times \mathcal{Y}_1 \times \mathcal{U}_2|$ . We denote by  $\mathbb{Q}_i$  the set of target probability distributions  $\mathcal{Q}(x_0, u_1, x_1, y_1, u_2, x_2)$  for which there exists an auxiliary random variable  $W$  satisfying (21) and (23) and we denote by  $\mathbb{Q}_o$  the set of target probability distributions  $\mathcal{Q}(x_0, u_1, x_1, y_1, u_2, x_2)$  for which there exists an auxiliary random variable  $W$  satisfying (21) and (22).

The achievability proof of Theorem II.5 comes from [23, Theorem III.1] and [24, Theorem I.1]. The converse proof of Theorem II.5 is stated in App. C.

**Remark IV.3** The probability distribution of equation (21) satisfies:

$$\begin{cases} (X_1, Y_1) \text{ } \perp\!\!\!\perp (X_0, U_1) \text{ } \perp\!\!\!\perp W, \\ U_2 \text{ } \perp\!\!\!\perp (Y_1, W) \text{ } \perp\!\!\!\perp (X_0, X_1, U_1), \\ X_2 \text{ } \perp\!\!\!\perp (X_1, U_2) \text{ } \perp\!\!\!\perp (X_0, U_1, Y_1, W). \end{cases} \quad (24)$$

The first and third Markov chains in (24) are again due to the two *channels*. The second Markov chain shows that the channel output  $Y_1$  and the auxiliary random variable  $W$  are a sufficient statistic for deciding on the action  $U_2$ .

#### V. WITSENHAUSEN’S COST FUNCTION

The four sets  $\mathbb{Q}_d$ ,  $\mathbb{Q}_e$ ,  $\mathbb{Q}_i$ ,  $\mathbb{Q}_o$  are convexes (see for example [27, Theorem 3, pp. 5093]) and Witsenhausen’s cost function is linear w.r.t. the probability distribution  $\mathcal{Q}(x_0, u_1, x_1, y_1, u_2, x_2)$ . Hence the four minimum costs  $\phi_d$ ,  $\phi_e$ ,  $\phi_i$ ,  $\phi_o$  are achieved by solving this problem:

$$\phi = \min_{\mathcal{Q} \in \mathbb{Q}} \mathbb{E}_{\mathcal{Q}} \left[ k^2 \cdot U_1^2 + (X_1 - U_2)^2 \right], \quad (25)$$

where  $k \in \mathbb{R}$  is a weight parameter and  $\mathbb{Q}$  is one of the sets  $\mathbb{Q}_d$ ,  $\mathbb{Q}_e$ ,  $\mathbb{Q}_i$ ,  $\mathbb{Q}_o$ .

#### APPENDIX A

##### SKETCH OF THE CONVERSE PROOF OF THEOREM II.5

From the converse proof in [24, Sec.V-B], we identify the auxiliary random variables  $W_{1,i} = X_{0,i+1}^n$  and  $W_{2,i} = Y_1^{i-1}$ , so as to have:

$$0 \leq \sum_{i=1}^n I(W_{1,i}; Y_{1,i} | W_{2,i}) - \sum_{i=1}^n I(W_{1,i}, W_{2,i}; X_{0,i}). \quad (26)$$

The two random variables  $(W_{1,i}, W_{2,i})$  satisfy the following Markov Chains corresponding to the set of probability distributions (10) of Theorem II.5:

$$(X_{1,i}, Y_{1,i}) \ominus (X_{0,i}, U_{1,i}) \ominus (W_{1,i}, W_{2,i}), \quad (27)$$

$$U_{2,i} \ominus (Y_{1,i}, W_{2,i}) \ominus (X_{0,i}, X_{1,i}, U_{1,i}, W_{1,i}), \quad (28)$$

$$X_{2,i} \ominus (X_{1,i}, U_{2,i}) \ominus (X_{0,i}, U_{1,i}, Y_{1,i}, W_{1,i}, W_{2,i}). \quad (29)$$

Eq. (27) comes from the memoryless property of the channel  $\mathcal{P}(x_1, y_1|x_0, u_1)$ .

Eq. (28) comes from the causal decoding: the output of the decoder  $U_{2,i}$  depends on the symbols  $(X_{0,i}, X_{1,i}, U_{1,i}, X_{0,i+1}^n)$  only through the past and current channel outputs  $(Y_{1,i}, Y_{1,i+1}^n)$ .

Eq. (29) comes from the memoryless property of the channel  $\mathcal{P}(x_2|x_1, u_2)$ .

## APPENDIX B

### SKETCH OF THE CONVERSE PROOF OF THEOREM III.2

From the converse proof in [24, Sec.VI-B], we identify the auxiliary random variables  $W_{1,i} = X_0^{i-1}$  and  $W_{2,i} = Y_{1,i+1}^n$ , so as to have:

$$0 \leq \sum_{i=1}^n I(W_{1,i}, W_{2,i}; Y_{1,i}) - \sum_{i=1}^n I(W_{2,i}; X_{0,i}|W_{1,i}). \quad (30)$$

The two random variables  $(W_{1,i}, W_{2,i})$  satisfy the following properties corresponding to the set of probability distributions (17) of Theorem III.2:

$$X_{0,i} \text{ independent of } W_{1,i}, \quad (31)$$

$$U_{1,i} \ominus (X_{0,i}, W_{1,i}) \ominus W_{2,i}, \quad (32)$$

$$(X_{1,i}, Y_{1,i}) \ominus (X_{0,i}, U_{1,i}) \ominus (W_{1,i}, W_{2,i}), \quad (33)$$

$$U_{2,i} \ominus (W_{1,i}, W_{2,i}, Y_{1,i}) \ominus (X_{0,i}, X_{1,i}, U_{1,i}), \quad (34)$$

$$X_{2,i} \ominus (X_{1,i}, U_{2,i}) \ominus (X_{0,i}, U_{1,i}, Y_{1,i}, W_{1,i}, W_{2,i}). \quad (35)$$

Eq. (31) comes from the i.i.d property of the source.

Eq. (32) comes from the causal encoding function that implies  $U_{1,i}$  is a deterministic function of  $X_0^i$  which is equal to:  $(X_{0,i}, W_{1,i})$ .

Eq. (33) comes from the memoryless property of the channel  $\mathcal{P}(x_1, y_1|x_0, u_1)$ .

Eq. (34) comes from the causal encoding and non-causal decoding, as stated in [24, Lemma 3].

Eq. (35) comes from the memoryless property of the channel  $\mathcal{P}(x_2|x_1, u_2)$ .

## APPENDIX C

### SKETCH OF THE CONVERSE PROOF OF THEOREM IV.2

$$0 \leq I(U_1^n, X_0^n; Y_1^n) - I(X_0^n; Y_1^n) \quad (36)$$

$$= \sum_{i=1}^n I(U_{1,i}, X_{0,i}; Y_1^n | U_1^{i-1}, X_0^{i-1}) - \sum_{i=1}^n I(X_{0,i}; Y_1^n, X_0^{i-1}) \quad (37)$$

$$\leq \sum_{i=1}^n I(U_{1,i}, X_{0,i}; Y_1^n, U_1^{i-1}, X_0^{i-1}) - \sum_{i=1}^n I(X_{0,i}; Y_1^{i-1}, Y_{1,i+1}^n) \quad (38)$$

$$= \sum_{i=1}^n I(U_{1,i}, X_{0,i}; Y_{1,i}) - \sum_{i=1}^n I(X_{0,i}; W_i). \quad (39)$$

Eq. (37) comes from the i.i.d. property of  $X_0^n$ .

Eq. (38) comes from the property of the mutual information.

Eq. (39) comes from the memoryless property of the channel  $\mathcal{P}(x_1, y_1|x_0, u_1)$  and the identification of the auxiliary random variable  $W_i = (Y_1^{i-1}, Y_{1,i+1}^n)$  which satisfies the Markov chains of (24):

$$(X_{1,i}, Y_{1,i}) \ominus (X_{0,i}, U_{1,i}) \ominus W_i, \quad (40)$$

$$U_{2,i} \ominus (W_i, Y_{1,i}) \ominus (X_{0,i}, X_{1,i}, U_{1,i}), \quad (41)$$

$$X_{2,i} \ominus (X_{1,i}, U_{2,i}) \ominus (X_{0,i}, U_{1,i}, Y_{1,i}, W_i). \quad (42)$$

Eq. (40) comes from the memoryless property of the channel  $\mathcal{P}(x_1, y_1|x_0, u_1)$ .

Eq. (41) comes from the non-causal decoding.

Eq. (42) comes from the memoryless property of the channel  $\mathcal{P}(x_2|x_1, u_2)$ .

## REFERENCES

- [1] H. Witsenhausen, "A counterexample in stochastic optimum control," *SIAM Journal on Control*, vol. 6, no. 1, pp. 131–147, 1968.
- [2] R. Bansal and T. Basar, "Stochastic teams with nonclassical information revisited: When is an affine law optimal?" *IEEE Transactions on Automatic Control*, vol. 32, no. 6, pp. 554–559, Jun 1987.
- [3] S. Yüksel and T. Basar, *Stochastic Networked Control Systems: Stabilization and Optimization under Information Constraints*, ser. Systems & Control Foundations & Applications. New York, NY: Springer, 2013.
- [4] J. Karlsson, A. Gattami, T. J. Oechtering, and M. Skoglund, "Iterative source-channel coding approach to Witsenhausen's counterexample," in *Proceedings of the 2011 American Control Conference*, June 2011, pp. 5348–5353.
- [5] P. Grover and A. Sahai, "Witsenhausen's counterexample as assisted interference suppression," in *Int. J. Syst., Control Commun.*, vol. 2, no. 1-3, pp. 197–237, 2010.
- [6] P. Grover, S. Y. Park and A. Sahai, "Approximately optimal solutions to the finite-dimensional Witsenhausen counterexample," *IEEE Transactions on Automatic Control*, vol. 58, no. 9, pp. 2189–2204, Sept 2013.
- [7] P. Grover, A. B. Wagner and A. Sahai, "Information embedding and the triple role of control," *IEEE Trans. on Inform. Theory*, vol. 61, no. 4, pp. 1539–1549, April 2015.
- [8] S. I. Gel'fand and M. S. Pinsker, "Coding for channel with random parameters," *Problems Control Inf. Theory*, vol. 9, no. 1, pp. 19–31, 1980.
- [9] Y.-H. Kim, A. Sutivong and T. M. Cover, "State amplification," *IEEE Trans. on Inform. Theory*, vol. 54, no. 5, pp. 1850–1859, May 2008.
- [10] O. Sumszyk and Y. Steinberg, "Information embedding with reversible stegotext," in *Proc. IEEE Int. Symp. Inf. Theory*, Seoul, Korea, Jun./Jul. 2009, pp. 2728–2732.

- [11] A. Sutivong, M. Chiang, T. M. Cover and Y.-H. Kim, "Channel capacity and state estimation for state-dependent Gaussian channels," *IEEE Trans. on Inform. Theory*, vol. 51, no. 4, pp. 1486–1495, April 2005.
- [12] S. Molavipour, G. Bassi and M. Skoglund, "Improved performance bounds for the infinite-dimensional witsenhausen problem," in *2017 American Control Conference (ACC)*, May 2017, pp. 1067–1072.
- [13] C. Choudhuri and U. Mitra, "On Witsenhausen's counterexample: the asymptotic vector case," *2012 IEEE Information Theory Workshop (ITW)*, pp. 162–166, July 2012.
- [14] M. H. M. Costa, "Writing on dirty paper," *IEEE Trans. on Inform. Theory*, vol. 29, no. 3, pp. 439–441, 1983.
- [15] C. Choudhuri, Y.-H. Kim and U. Mitra, "Causal state communication," *IEEE Trans. on Inform. Theory*, vol. 59, no. 6, pp. 3709–3719, June 2013.
- [16] Y. K. Chia, R. Soundararajan and T. Weissman, "Estimation With a Helper Who Knows the Interference," *IEEE Trans. on Inform. Theory*, vol. 59, no. 11, pp. 7097–7117, Nov 2013.
- [17] P. Cuff and L. Zhao, "Coordination using implicit communication," in *2011 IEEE Information Theory Workshop*, Oct 2011, pp. 467–471.
- [18] M. Raginsky, "Empirical processes, typical sequences, and coordinated actions in standard borel spaces," *IEEE Transactions on Information Theory*, vol. 59, no. 3, pp. 1288–1301, March 2013.
- [19] O. Gossner, P. Hernandez, and A. Neyman, "Optimal use of communication resources," *Econometrica*, vol. 74, no. 6, pp. 1603–1636, 2006.
- [20] P. W. Cuff, H. H. Permuter and T. M. Cover, "Coordination capacity," *IEEE Transactions on Information Theory*, vol. 56, no. 9, pp. 4181–4206, Sept 2010.
- [21] P. Cuff and C. Schieler, "Hybrid codes needed for coordination over the point-to-point channel," in *IEEE 49th Annual Allerton Conference on Communication, Control, and Computing*, Sept. 2011, pp. 235–239.
- [22] M. Le Treust, "Correlation between channel state and information source with empirical coordination constraint," in *IEEE Information Theory Workshop (ITW)*, 2014, pp. 272–276.
- [23] M. Le Treust, "Empirical coordination with two-sided state information and correlated source and state," in *IEEE International Symposium on Information Theory*, 2015, pp. 466–470.
- [24] M. Le Treust, "Coding theorems for empirical coordination," *Technical report*, DOI: 10.25327/etis.2018.001, available: <https://cloud.ensea.fr/index.php/s/X9e5x8EzJf17I4Q>, Apr. 2015.
- [25] B. Larrousse, S. Lasaulce and M. Wigger, "Coordination in state-dependent distributed networks: The two-agent case," in *2015 IEEE International Symposium on Information Theory (ISIT)*, June 2015, pp. 979–983.
- [26] B. Larrousse, S. Lasaulce and M. Wigger, "Coordinating partially-informed agents over state-dependent networks," in *2015 IEEE Information Theory Workshop (ITW)*, April 2015, pp. 1–5.
- [27] M. Le Treust, "Joint Empirical Coordination of Source and Channel," *IEEE Trans. Inf. Theory*, vol. 63, no. 8, pp. 5087–5114, Aug. 2017.
- [28] B. Larrousse, A. Agrawal and S. Lasaulce, "Implicit coordination in two-agent team problems; application to distributed power allocation," in *WNC3 2014: Int. Workshop on Wireless Networks: Communication, Cooperation and Competition*, 2014, pp. 579–584.
- [29] B. Larrousse, S. Lasaulce and M. R. Bloch, "Coordination in distributed networks via coded actions with application to power control," *IEEE Trans. on Inform. Theory*, vol. 64, no. 5, pp. 3633–3654, May 2018.
- [30] A. Agrawal, F. Danard, B. Larrousse and S. Lasaulce, "Implicit coordination in two-agent team problems with continuous action sets. application to the witsenhausen cost function," in *2015 European Control Conference (ECC)*, July 2015, pp. 1854–1859.