Empirical Coordination with Channel Feedback and Strictly Causal or Causal Encoding

Maël Le Treust

Empirical Coordination with Channel Feedback and Strictly Causal or Causal Encoding

Maël Le Treust

ETIS, UMR 8051 / ENSEA, Université Cergy-Pontoise, CNRS,
6, avenue du Ponceau,
95014 CERGY-PONTOISE CEDEX,
FRANCE
Email: mael.le-treust@ensea.fr

***Abstract*—In multi-terminal networks, feedback increases the capacity region and helps communication devices to coordinate. In this article, we deepen the relationship between coordination and feedback by considering a point-to-point scenario with an information source and a noisy channel. Empirical coordination is achievable if the encoder and the decoder can implement sequences of symbols that are jointly typical for a target probability distribution. We investigate the impact of feedback when the encoder has strictly causal or causal observation of the source symbols. For both cases, we characterize the optimal information constraints and we show that feedback improves coordination possibilities. Surprisingly, feedback also reduces the number of auxiliary random variables and simplifies the information constraints. For empirical coordination with strictly causal encoding and feedback, the information constraint does not involve auxiliary random variable anymore.**

***Index Terms*—Shannon Theory, Feedback, Empirical Coordination, Joint Source-Channel Coding, Empirical Distribution of Symbols, Strictly Causal and Causal Encoding.**

I. Introduction

Feedback does not increase the capacity of a memoryless channel [1]. However, it has a significant impact when considering problems of empirical coordination. In this framework, encoder and decoder are considered as autonomous agents [2], that implement a coding scheme in order to coordinate their sequences of actions, *i.e.* channel inputs and decoder outputs, with a sequence of source symbols. The problem of empirical coordination [3], [4], [5] consists in determining the set of joint probability distributions, that are achievable for empirical frequencies of symbols. Empirical coordination provides a single-letter solution that simplifies the analysis of optimization problems such as minimal source distortion, minimal channel cost or maximal utility function of a decentralized communication network [6]. For example, the optimal distortion level is the minimum of the expected distortion function, taken over the set of achievable joint probability distributions.

In the framework of multi-terminal networks, feedback increases the capacity region of the multiple-access channel [7], [8] and of the broadcast channel [9], [10]. In the literature of game theory, feedback is considered from a strategic point-of-view. In [2], a player observes the past actions of another player through a monitoring structure involving perfect or imperfect feedback. In [11], the authors investigate a four-player coordination game with imperfect feedback and provide a subset of achievable joint probability distributions. Empirical coordination is a first step toward a better understanding of decentralized communication network. The set of achievable joint distributions was characterized for strictly causal and causal decoding in [6], with two-sided state information in [12] and with feedback from the source in [13]. From a practical perspective, coordination with polar codes was considered in [14]. Lossless decoding with correlated information source and channel states is solved in [15]. Empirical coordination for multi-terminal source coding is treated in [16] and in [17].

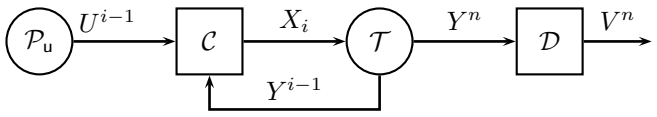

Fig. 1. Strictly causal encoding function with feedback $f_i : \mathcal{U}^{i-1} \times \mathcal{Y}^{i-1} \to \mathcal{X}$, for all $i \in \{1, \dots, n\}$ and non-causal decoding function $g : \mathcal{Y}^n \to \mathcal{V}^n$.

In this article, we consider the point-to-point scenario of [18] with channel feedback, as represented by Fig. 1 and 2. The encoder has perfect feedback from the channel and strictly causal or causal observation of the symbols of source. In both cases, we characterize the set of achievable joint probability distributions over the symbols of source and channel. We show that the information constraints are larger than the ones stated in [18]. Surprisingly, feedback also reduces the number of auxiliary random variables and simplifies the information constraints. For empirical coordination with strictly causal encoding and feedback, the information constraint does not involve auxiliary random variable anymore. There is an analogy with strictly causal decoding [6], [13], since no auxiliary random variable is needed when the decoder has feedback from the source. Feedback allows to remove auxiliary random variables of information constraints, for empirical coordination problems.

System model and definitions are stated in Sec. II and characterizations of achievable joint distributions are stated in Sec. III. Comparison with previous works and an example are stated in Sec. IV and V. Conclusions and sketches of proofs are stated in Sec. VI and in Appendix A, B, C.

II. System model

Figure 1 represents the problem under investigation. Random variable U is denoted by capital letter, lowercase letter $u \in \mathcal{U}$ designates the realization and $\mathcal{U}^n$ corresponds to the n-time cartesian product. U^n, X^n, Y^n, V^n stands for sequences of random variables of source symbols $u^n = (u_1, \ldots, u_n) \in \mathcal{U}^n$, inputs of the channel $x^n \in \mathcal{X}^n$, outputs of the channel $y^n \in \mathcal{Y}^n$ and decoder's output $v^n \in \mathcal{V}^n$. The sets $\mathcal{U}$, $\mathcal{X}$, $\mathcal{Y}$, $\mathcal{V}$ are discrete. The set of probability distributions $\mathcal{P}(X)$ over $\mathcal{X}$ is denoted by $\Delta(\mathcal{X})$. Notation $||\mathcal{Q} - \mathcal{P}||_{\mathsf{tv}} = 1/2 \cdot \sum_{x \in \mathcal{X}} |\mathcal{Q}(x) - \mathcal{P}(x)|$ stands for the total variation distance between probability distributions $\mathcal{Q}$ and $\mathcal{P}$. Notation $Y \multimap X \multimap U$ stands for the Markov chain property corresponding to $\mathcal{P}(y|x,u) = \mathcal{P}(y|x)$ for all (u,x,y). Information source is i.i.d. distributed with $\mathcal{P}_{\mathsf{u}}$ and the channel is memoryless with transition probability $\mathcal{T}_{\mathsf{y|x}}$. Encoder $\mathcal{C}$ and decoder $\mathcal{D}$ know the statistics $\mathcal{P}_{\mathsf{u}}$ and $\mathcal{T}_{\mathsf{y|x}}$ of the source and channel. The coding process is deterministic.

Definition II.1 *A code $c \in \mathcal{C}(n)$ with strictly-causal encoder and feedback is a tuple of functions $c = (\{f_i\}_{i=1}^n, g)$ defined by equations* (1) *and* (2)*:*

$$f_i \quad : \mathcal{U}^{i-1} \times \mathcal{Y}^{i-1} \longrightarrow \mathcal{X}, \qquad i = 1, \ldots, n, \tag{1}$$

$$g \quad : \mathcal{Y}^n \longrightarrow \mathcal{V}^n. \tag{2}$$

The number of occurrence of symbol $u \in \mathcal{U}$ in sequence u^n is denoted by $\mathsf{N}(u|u^n)$. The empirical distribution $Q^n \in \Delta(\mathcal{U} \times \mathcal{X} \times \mathcal{Y} \times \mathcal{V})$ of sequences (u^n, x^n, y^n, v^n) is defined by:

$$\begin{aligned} Q^n(u,x,y,v) &= \frac{\mathsf{N}(u,x,y,v|u^n,x^n,y^n,v^n)}{n}, \\ \forall (u,x,y,v) &\in \mathcal{U} \times \mathcal{X} \times \mathcal{Y} \times \mathcal{V}. \end{aligned} \tag{3}$$

Fix a target probability distribution $\mathcal{Q} \in \Delta(\mathcal{U} \times \mathcal{X} \times \mathcal{Y} \times \mathcal{V})$, the error probability of the code $c \in \mathcal{C}(n)$ is defined by:

$$\mathcal{P}_{\mathsf{e}}(c) = \mathcal{P}_c\left(\left|\left| Q^n - \mathcal{Q} \right|\right|_{\mathsf{tv}} \geq \varepsilon \right), \tag{4}$$

where $Q^n \in \Delta(\mathcal{U} \times \mathcal{X} \times \mathcal{Y} \times \mathcal{V})$ is the random variable of the empirical distribution induced by the probability distributions $\mathcal{P}_{\mathsf{u}}$, $\mathcal{T}_{\mathsf{y|x}}$ and the code $c \in \mathcal{C}(n)$.

Definition II.2 *The probability distribution $\mathcal{Q} \in \Delta(\mathcal{U} \times \mathcal{X} \times \mathcal{Y} \times \mathcal{V})$ is achievable if for all $\varepsilon > 0$, there exists a $\bar{n} \in \mathbb{N}$ s.t. for all $n \geq \bar{n}$, there exists a code $c \in \mathcal{C}(n)$ that satisfies:*

$$\mathcal{P}_{\mathsf{e}}(c) = \mathcal{P}_c\left(\left|\left| Q^n - \mathcal{Q} \right|\right|_{\mathsf{tv}} \geq \varepsilon \right) \leq \varepsilon. \tag{5}$$

The error probability $\mathcal{P}_{\mathsf{e}}(c)$ is small if the total variation distance between the empirical frequency of symbols $Q^n(u,x,y,v)$ and the target probability distribution $\mathcal{Q}(u,x,y,v)$ is small, with large probability. In that case, the sequences of symbols $(U^n, X^n, Y^n, V^n) \in A_\varepsilon^{\star n}(\mathcal{Q})$ are jointly typical, *i.e.* coordinated, for the target probability distribution $\mathcal{Q}$ with large probability.

As mentioned in [6] and [15], the performance of the coordination can be evaluated using an objective function $\Phi : \mathcal{U} \times \mathcal{X} \times \mathcal{Y} \times \mathcal{V} \mapsto \mathbb{R}$. We denote by $\mathcal{A}^\star$, the set of joint probability distributions $\mathcal{Q} \in \mathcal{A}^\star$ that are achievable. Based on the expectation $\mathbb{E}_{\mathcal{Q} \in \mathcal{A}^\star}\Big[\Phi(U,X,Y,V)\Big]$, it is possible to derive the minimal channel cost $\Phi(u,x,y,v) = c(x)$, the minimal distortion level $\Phi(u,x,y,v) = d(u,v)$ or the maximal utility of a decentralized network [2], using a single-letter characterization.

III. Characterization of achievable distributions

This section presents the two main results of this article. Theorem III.1 characterizes of the set of achievable joint probability distributions for strictly causal encoding with feedback, represented in Fig. 1.

Theorem III.1 (Strictly causal encoding with feedback) 1) *If the joint probability distribution $\mathcal{Q}(u,x,y,v)$ is achievable, then it decomposes as follows:*

$$\begin{cases} \mathcal{Q}(u) = \mathcal{P}_{\mathsf{u}}(u), & \mathcal{Q}(y|x) = \mathcal{T}(y|x), \\ U \text{ independent of } X, & Y \multimap X \multimap U. \end{cases} \tag{6}$$

2) Joint probability distribution $\mathcal{P}_{\mathsf{u}}(u) \otimes \mathcal{Q}(x) \otimes \mathcal{T}(y|x) \otimes \mathcal{Q}(v|u,x,y)$ is achievable if:

$$I(X;Y) - I(U;V|X,Y) > 0, \tag{7}$$

3) Joint probability distribution $\mathcal{P}_{\mathsf{u}}(u) \otimes \mathcal{Q}(x) \otimes \mathcal{T}(y|x) \otimes \mathcal{Q}(v|u,x,y)$ is not achievable if:

$$I(X;Y) - I(U;V|X,Y) < 0, \tag{8}$$

Sketch of proof of Theorem III.1 is stated in Appendix A. Equation (7) comes from Theorem 3 in [18] by replacing the auxiliary random variable by decoder's output V and the observation of the encoder by the pair of information source and channel feedback (U,Y).

A causal encoding function is defined by $f_i : \mathcal{U}^i \times \mathcal{Y}^{i-1} \to \mathcal{X}$, $\forall i \in \{1, \ldots, n\}$. Theorem III.2 characterizes of the set of achievable joint probability distributions for causal encoding with feedback, represented in Fig. 2.

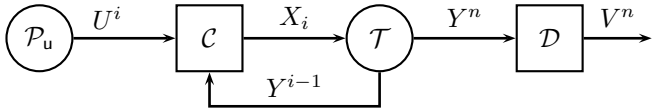

Fig. 2. Causal encoding function with feedback $f_i : \mathcal{U}^i \times \mathcal{Y}^{i-1} \to \mathcal{X}$, for all $i \in \{1, \ldots, n\}$ and non-causal decoding function $g : \mathcal{Y}^n \to \mathcal{V}^n$.

Theorem III.2 (Causal Encoding with Feedback) 1) *If the joint probability distribution $\mathcal{Q}(u,x,y,v)$ is achievable, then it decomposes as follows:*

$$\mathcal{Q}(u) = \mathcal{P}_{\mathsf{u}}(u), \;\; \mathcal{Q}(y|x) = \mathcal{T}(y|x), \;\; Y \multimap X \multimap U, \tag{9}$$

2) Joint probability distribution $\mathcal{P}_{\mathsf{u}}(u) \otimes \mathcal{Q}(x|u) \otimes \mathcal{T}(y|x) \otimes \mathcal{Q}(v|u,x,y)$ is achievable if:

$$\max_{\mathcal{Q} \in \mathbb{Q}} \Big(I(W;Y) - I(U;V|W,Y) \Big) > 0, \tag{10}$$

3) Joint probability distribution $\mathcal{P}_{\mathsf{u}}(u) \otimes \mathcal{Q}(x|u) \otimes \mathcal{T}(y|x) \otimes \mathcal{Q}(v|u,x,y)$ is not achievable if:

$$\max_{\mathcal{Q}\in\mathbb{Q}} \Big(I(W;Y) - I(U;V|W,Y) \Big) < 0, \tag{11}$$

where $\mathbb{Q}$ is the set of probability distributions $\mathcal{Q} \in \Delta(\mathcal{U}\times\mathcal{W}\times\mathcal{X}\times\mathcal{Y}\times\mathcal{V})$ with auxiliary random variable W that satisfies:

$$\begin{cases} \sum_{w\in\mathcal{W}} \mathcal{Q}(u,w,x,y,v) \\ \quad = \mathcal{P}_{\mathsf{u}}(u) \otimes \mathcal{Q}(x|u) \otimes \mathcal{T}(y|x) \otimes \mathcal{Q}(v|u,x,y), \\ U \text{ independent of } W, \\ Y \multimap X \multimap (U,W), \\ V \multimap (U,Y,W) \multimap X. \end{cases}$$

The probability distribution $\mathcal{Q} \in \mathbb{Q}$ decomposes as follows:

$$\mathcal{P}_{\mathsf{u}}(u) \otimes \mathcal{Q}(w) \otimes \mathcal{Q}(x|u,w) \otimes \mathcal{T}(y|x) \otimes \mathcal{Q}(v|u,y,w).$$

The support of W is bounded by $|\mathcal{W}| \leq |\mathcal{U}\times\mathcal{X}\times\mathcal{Y}\times\mathcal{V}| + 2$.

Sketch of proofs of Theorem III.2 are stated in Appendix B and C. The random variable V is directly correlated with the pair (U,Y) of source and channel output. Feedback implies that V is extracted from the Markov chain $Y \multimap X \multimap (U,W)$ of the memoryless channel.

IV. Feedback improves empirical coordination

In this section, we investigate the impact of the feedback on the set of achievable joint distributions stated in Theorems III.1 and III.2. Considering strictly causal encoding, we evaluate the difference between information constraint stated in equation (7) and the one stated in Theorem 3 in [18] without feedback.

$$\begin{aligned} & I(X;Y) - I(U;V|X,Y) && (12)\\ - & \max_{\mathcal{Q}\in\mathbb{Q}_{\mathsf{se}}} \Big(I(X;Y) - I(U;W_2|X) \Big) && (13)\\ = & \min_{\mathcal{Q}\in\mathbb{Q}_{\mathsf{se}}} I(U;W_2|X) - I(U;V|X,Y) && (14)\\ = & H(U|V,X,Y) - \max_{\mathcal{Q}\in\mathbb{Q}_{\mathsf{se}}} H(U|X,W_2) \geq 0. && (15) \end{aligned}$$

$\mathbb{Q}_{\mathsf{se}}$ is the set of probability distributions $\mathcal{Q} \in \Delta(\mathcal{U}\times\mathcal{W}_2\times\mathcal{X}\times\mathcal{Y}\times\mathcal{V})$ with auxiliary random variable W_2 that satisfies:

$$\mathcal{P}_{\mathsf{u}}(u) \otimes \mathcal{Q}(x) \otimes \mathcal{Q}(w_2|u,x) \otimes \mathcal{T}(y|x) \otimes \mathcal{Q}(v|y,x,w_2).$$

• Equation (15) is equal to zero if (U,V) is independent of (X,Y), this corresponds to the lossy transmission without coordination in which the feedback does not increase the channel capacity [1].

• Equation (15) is equal to zero when the decoder output V is empirically coordinated with (U,X) and not with the channel output Y, because in that case $W_2 = V$. Since the auxiliary random variable W_2 should satisfy $\mathcal{Q}(v|y,x,u) = \sum_{w_2\in\mathcal{W}_2} \mathcal{Q}(w_2|u,x) \cdot \mathcal{Q}(v|y,x,w_2)$, equation (12) provides an upper bound to equation (13) that is easier to evaluate

There is a strong analogy between strictly causal *encoding* with channel feedback and strictly causal *decoding* with source feedback. Equation (16) corresponds to strictly causal decoding without feedback from the source, stated in [6].

$$\max_{\mathcal{Q}\in\mathbb{Q}_{\mathsf{sd}}} \Big(I(W_1;Y|V) - I(U;V,W_1) \Big) > 0. \tag{16}$$

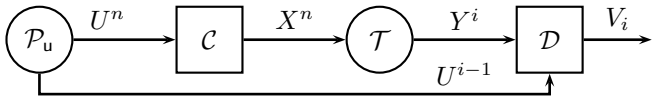

Fig. 3. Non-causal encoding $f : \mathcal{U}^n \to \mathcal{X}^n$ and causal decoding $g_i : \mathcal{Y}^i \times \mathcal{U}^{i-1} \to \mathcal{V}$ for all $i \in \{1,\dots,n\}$ with feedback from the source.

$\mathbb{Q}_{\mathsf{sd}}$ is the set of probability distributions $\mathcal{Q} \in \Delta(\mathcal{U}\times\mathcal{W}_1\times\mathcal{X}\times\mathcal{Y}\times\mathcal{V})$ with auxiliary random variable W_1, that satisfy:

$$\mathcal{P}_u(u) \otimes \mathcal{Q}(x,v|u) \otimes \mathcal{Q}(w_1|u,x,v) \otimes \mathcal{T}(y|x).$$

Equation (17) corresponds to strictly causal decoding with feedback from the source, characterized in [13].

$$I(X;Y|U,V) - I(U;V) > 0. \tag{17}$$

Equation (17) can be deduced from equation (16), by replacing the auxiliary random variable W_1 by X and the observation of the decoder Y by the pair (U,Y).

This analysis extends to causal decoding with feedback from the source, represented by Fig. 3 and characterized by (18).

$$\max_{\mathcal{Q}\in\mathbb{Q}_{\mathsf{df}}} \Big(I(X;Y|U,W_3) - I(U;W_3) \Big) > 0. \tag{18}$$

$\mathbb{Q}_{\mathsf{df}}$ is the set of probability distributions $\mathcal{Q} \in \Delta(\mathcal{U}\times\mathcal{W}_3\times\mathcal{X}\times\mathcal{Y}\times\mathcal{V})$ with auxiliary random variable W_3, that satisfy:

$$\mathcal{P}_u(u) \otimes \mathcal{Q}(x,w_3|u) \otimes \mathcal{T}(y|x) \otimes \mathcal{Q}(v|y,w_3).$$

The proof is in [19]. Theorems III.1 and III.2 also extend to two-sided state information by replacing (U,S) by (U,S,Y) in the results of [12], for strictly causal and causal encoding.

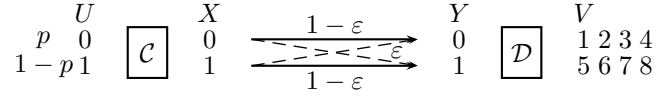

Fig. 4. Binary information source and binary symmetric channel with parameters $p = 1/2$ and $\varepsilon \in [0, 0.5]$

V. Example: binary source and channel

We consider a binary information source and a binary symmetric channel represented by Fig. 4. The set of symbols are

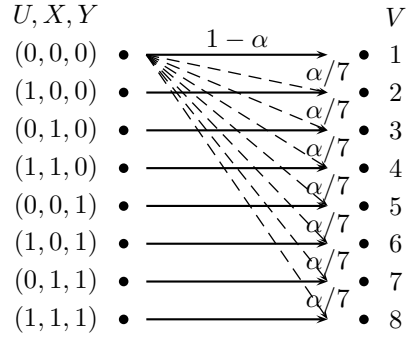

Fig. 5. Conditional probability distribution $\mathcal{Q}_{\mathsf{v}|\mathsf{uxy}}$ depending on parameter $\alpha \in [0, 7/8]$ where $\mathcal{Q}\big(V=1\big|(U,X,Y)=(0,0,0)\big) = 1-\alpha$ and $\mathcal{Q}\big(V=2\big|(U,X,Y)=(0,0,0)\big) = \alpha/7$. For $\alpha = 7/8$, the probability distribution is uniform over the set $\mathcal{V} = \{1,\dots,8\}$ and independent of the triple (U,X,Y). For $\alpha = 0$, the output V corresponds exactly to the triple (U,X,Y).

given by $\mathcal{U} = \mathcal{X} = \mathcal{Y} = \{0,1\}$ and $\mathcal{V} = \{1,2,3,4,5,6,7,8\}$. We assume the parameter $p \in [0,1]$ of the information source

is equal to 1/2. The probability distribution of channel input is uniform $\mathcal{Q}(X=0)=\mathcal{Q}(X=1)=1/2$. The transition probability of the channel depends on a noise parameter $\varepsilon \in [0, 0.5]$. Since the input distribution is uniform and the channel is symmetric, the output probability distribution is also uniform $\mathcal{Q}(Y=0)=\mathcal{Q}(Y=1)=1/2$. We investigate a class of achievable conditional probability distributions $\mathcal{Q}_{\mathsf{v}|\mathsf{uxy}}$ described by Fig. 5.

We consider strictly causal encoding with feedback. The information constraint (7) of Theorem III.1 writes:

$$\begin{aligned} & I(X;Y) - I(U;V|X,Y) \\ = \;& H(Y) - H(Y|X) - H(V|X,Y) + H(V|U,X,Y) \\ = \;& 1 - H_b(\varepsilon) - H_b\left(\frac{6\alpha}{7}\right) - 1 - \frac{6\alpha}{7}\cdot\log_2 3 + H_b(\alpha) + \alpha\cdot\log_2 7 \\ = \;& H_b(\alpha) - H_b(\varepsilon) - H_b\left(\frac{6\alpha}{7}\right) + \alpha\cdot\left(\log_2 7 - \frac{6}{7}\cdot\log_2 3\right). \end{aligned}$$

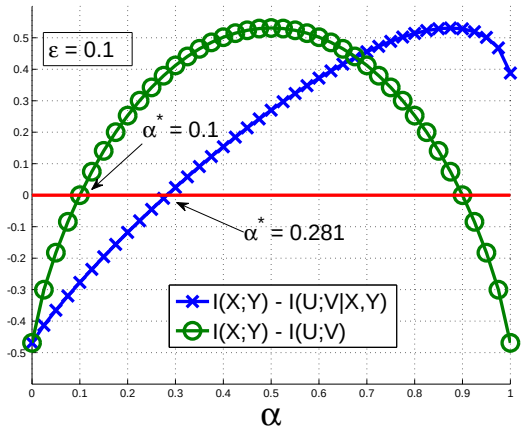

Fig. 6. Comparison between the information constraint for empirical coordination with feedback $I(X;Y)-I(U;V|X,Y)$ and the information constraint $I(X;Y)-I(U;V)$ for lossy transmission.

In Fig. 6, we compare the information constraint for empirical coordination with feedback (7) and information constraint for lossy transmission without coordination (19), where α is the distortion parameter of conditional distribution $\mathcal{Q}_{\mathsf{v}|\mathsf{u}}$:

$$\begin{aligned} I(X;Y) - I(U;V) &= 1 - H_b(\varepsilon) - 1 + H_b(\alpha) && (19) \\ &= H_b(\alpha) - H_b(\varepsilon). && (20) \end{aligned}$$

The minimal coordination parameter $\alpha^\star \simeq 0.281 > 0.1$ is much larger for empirical coordination than for lossy compression. This restriction comes from the additional correlation requirement between the decoder output V and the random variables (X,Y) of the channel. Fig. 7 provides the minimal value of parameter $\alpha^\star \in [0, 0.875]$ for empirical coordination, depending on the level of noise of the channel $\varepsilon \in [0, 0.5]$.

VI. Conclusion

We investigate the relationship between coordination and feedback by considering a point-to-point scenario with strictly causal and causal encoder. For both cases, we characterize the optimal solutions and we show that feedback simplifies the information constraints by reducing the number of auxiliary random variables. For empirical coordination with strictly causal encoding and feedback, the information constraint does not involve auxiliary random variable anymore.

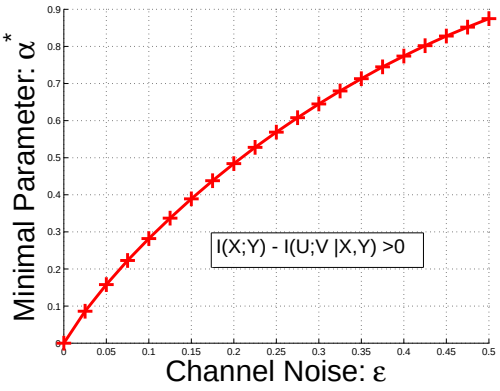

Fig. 7. Minimal value of parameter $\alpha^\star \in [0, 0.875]$ for the information constraint $I(X;Y) - I(U;V|X,Y) > 0$ to be positive, depending on the noise of the channel $\varepsilon \in [0, 0.5]$. It corresponds to the higher level of coordination between the random variable V and the triple (U,X,Y).

Appendix

The full versions of the proofs are stated in [19].

A. Sketch of proof of Theorem III.1

Achievability proof can be obtained from the proof of Theorem III.2 stated in Appendix B, by replacing the auxiliary random variable W by X.

For the converse proof, we consider code $c(n) \in \mathcal{C}$ with small error probability $\mathcal{P}_{\mathsf{e}}(c)$.

$$\begin{aligned} 0 &= I(U^n;Y^n) - I(U^n;Y^n,V^n) && (21) \\ &= I(U^n;Y^n) - \sum_{i=1}^n I(U_i;Y^n,V^n,U^{i-1}) && (22) \\ &\le \sum_{i=1}^n \Big(I(Y_i;U^n,X_i|Y^{i-1}) - I(U_i;Y^n,V^n,U^{i-1},X_i) \Big) && (23) \\ &\le \sum_{i=1}^n \Big(H(Y_i) - H(Y_i|X_i) - I(U_i;Y_i,V_i,X_i) \Big) && (24) \\ &\le \sum_{i=1}^n \Big(H(Y_i) + H(U_i|Y_i,V_i,X_i) \Big) - n\Big(H(Y|X) + H(U) \Big) && (25) \\ &\le n\Big(I(X;Y) - I(U;V|X,Y) \Big). && (26) \end{aligned}$$

Equation (21) comes from the non-causal decoding that induces the Markov chain: $U^n \multimap Y^n \multimap V^n$.

Equation (22) comes from the i.i.d. properties of the information source U that implies: $I(U_i;U^{i-1})=0$.

Equation (23) comes from the channel feedback and the strictly causal encoding function: $X_i = f_i(U^{i-1},Y^{i-1})$.

Equations (24) and (25) are due to the properties of i.i.d. information source and of memoryless channel.

Equation (26) comes from the concavity of the entropy function and from the hypothesis of small error probability $\mathcal{P}_{\mathsf{e}}(c)$.

B. Sketch of achievability proof of Theorem III.2

Consider $\mathcal{Q} \in \mathbb{Q}$ that achieves the maximum in equation (10). There exists a $\delta > 0$ and a rate $\mathsf{R} > 0$ such that:

$$\begin{aligned} \mathsf{R} &\ge I(U,Y;V|W) + \delta, && (27) \\ \mathsf{R} &\le I(W;Y) + I(V;Y|W) - \delta = I(W,V;Y) - \delta. && (28) \end{aligned}$$

We define a block-Markov random code $c \in \mathcal{C}(n)$ over $B \in \mathbb{N}$ blocks of length $n \in \mathbb{N}$.

- *Random codebook.* We generate $|\mathcal{M}| = 2^{n\mathsf{R}}$ sequences $W^n(m)$ drawn from $\mathcal{Q}_{\mathsf{w}}^{\otimes n}$ with index $m \in \mathcal{M}$. For each index $m \in \mathcal{M}$, we generate the same number $|\mathcal{M}| = 2^{n\mathsf{R}}$ of sequences $V^n(m, \hat{m})$ with index $\hat{m} \in \mathcal{M}$, drawn from $\mathcal{Q}_{\mathsf{v}|\mathsf{w}}^{\otimes n}$ depending on $W^n(m)$.
- *Encoding function.* It recalls m_{b-1} and finds $m_b \in \mathcal{M}$ s.t. sequences $(U_{b-1}^n, Y_{b-1}^n, W^n(m_{b-1}), V^n(m_{b-1}, m_b)) \in A_\varepsilon^{\star n}(\mathcal{Q})$ are jointly typical in block $b-1$. It deduces $W^n(m_b)$ for block b and sends X_b^n drawn from $\mathcal{Q}_{\mathsf{x}|\mathsf{uw}}^{\otimes n}$ depending on $(U_b^n, W^n(m_b))$.
- *Decoding function.* It recalls m_{b-1} and finds $m_b \in \mathcal{M}$ s.t. sequences $(Y_b^n, W^n(m_b)) \in A_\varepsilon^{\star n}(\mathcal{Q})$ and $(Y_{b-1}^n, W^n(m_{b-1}), V^n(m_{b-1}, m_b)) \in A_\varepsilon^{\star n}(\mathcal{Q})$ are jointly typical. It returns $V^n(m_{b-1}, m_b)$ over block $b-1$.
- *First block at the encoder.* An arbitrary index $m_1 \in \mathcal{M}$ of $W^n(m_1) \in \mathcal{W}^n$ is given to encoder and decoder. Encoder sends $X_{b_1}^n$ drawn from $\mathcal{Q}_{\mathsf{x}|\mathsf{uw}}^{\otimes n}$ depending on $(U_{b_1}^n, W^n(m_1))$. At the beginning of the second block b_2, encoder finds index m_2 such that $(U_{b_1}^n, Y_{b_1}^n, W^n(m_1), V^n(m_1, m_2)) \in A_\varepsilon^{\star n}(\mathcal{Q})$. It sends $X_{b_2}^n$ drawn from $\mathcal{Q}_{\mathsf{x}|\mathsf{wu}}^{\otimes n}$ depending on $(U_{b_2}^n, W^n(m_2))$.
- *First block at the decoder.* At the end of second block b_2, the decoder finds the index m_2 such that $(Y_{b_2}^n, W^n(m_2)) \in A_\varepsilon^{\star n}(\mathcal{Q})$ and $(Y_{b_1}^n, W^n(m_1), V^n(m_1, m_2)) \in A_\varepsilon^{\star n}(\mathcal{Q})$. Over the first bloc, decoder $\mathcal{D}$ returns $V^n(m_1, m_2) \in \mathcal{V}^n$. Sequences $(U_{b_1}^n, W^n(m_1), X_{b_1}^n, Y_{b_1}^n, V^n(m_1, m_2)) \in A_\varepsilon^{\star n}(\mathcal{Q})$ are jointly typical over the first block b_1.
- *Last bloc.* Sequences are not jointly typical.

Equations (27), (28) imply for all $n \geq \bar{n}$, for a large number of blocks $B \in \mathbb{N}$, the sequences are jointly typical with large probability.

$$\mathbb{E}_c\Big[\mathcal{P}\Big(U^n \notin A_\varepsilon^{\star n}(\mathcal{Q})\Big)\Big] \leq \varepsilon,$$
$$\mathbb{E}_c\Big[\mathcal{P}\Big(\forall m \in \mathcal{M},\ (U_{b-1}^n, Y_{b-1}^n, W^n(m_{b-1}), V^n(m_{b-1}, m)) \notin A_\varepsilon^{\star n}(\mathcal{Q})\Big)\Big] \leq \varepsilon,$$
$$\mathbb{E}_c\Big[\mathcal{P}\Big(\exists m' \neq m,\ \text{s.t.}\ \big\{(Y_b^n, W^n(m')) \in A_\varepsilon^{\star n}(\mathcal{Q})\big\} \cap \big\{(Y_{b-1}^n, W^n(m_{b-1}), V^n(m_{b-1}, m')) \in A_\varepsilon^{\star n}(\mathcal{Q})\big\}\Big)\Big] \leq \varepsilon.$$

C. Sketch of Converse Proof of Theorem III.2

Consider code $c(n) \in \mathcal{C}$ with small error probability $\mathcal{P}_{\mathsf{e}}(c)$.

$$\begin{aligned}
0 &\leq \sum_{i=1}^n I(U^{i-1}, Y^{i-1}, Y_{i+1}^n; Y_i) - \sum_{i=1}^n I(Y_{i+1}^n; U_i, Y_i | U^{i-1}, Y^{i-1}) && (29)\\
&= \sum_{i=1}^n I(U^{i-1}, Y^{i-1}; Y_i) - \sum_{i=1}^n I(Y_{i+1}^n; U_i | U^{i-1}, Y^{i-1}, Y_i) && (30)\\
&= \sum_{i=1}^n I(U^{i-1}, Y^{i-1}; Y_i) - \sum_{i=1}^n I(Y_{i+1}^n, V_i; U_i | U^{i-1}, Y^{i-1}, Y_i) && (31)\\
&\leq \sum_{i=1}^n I(U^{i-1}, Y^{i-1}; Y_i) - \sum_{i=1}^n I(V_i; U_i | U^{i-1}, Y^{i-1}, Y_i) && (32)\\
&= \sum_{i=1}^n I(W_i; Y_i) - \sum_{i=1}^n I(V_i; U_i | W_i, Y_i). && (33)\\
&\leq n \cdot \max_{\mathcal{Q} \in \mathbb{Q}} \Big(I(W; Y) - I(V; U | W, Y) \Big). && (34)
\end{aligned}$$

Eq. (29), (30) are due to Csiszár Sum Identity, prop. of MI.
Eq. (31) is due to the non-causal decoding function $V^n = g(Y^n)$, that implies: $I(V_i; U_i | U^{i-1}, Y^{i-1}, Y_i, Y_{i+1}^n) = 0$.
Eq. (32) is due to the properties of the mutual information.
Eq. (33) is due to the introduction of auxiliary random variables $W_i = (U^{i-1}, Y^{i-1})$ satisfying properties of set $\mathbb{Q}$.
Eq. (34) comes from taking the maximum over the set $\mathbb{Q}$.

$$U_i \text{ is independent of } W_i, \quad (35)$$
$$Y_i \multimap X_i \multimap (U_i, W_i), \quad (36)$$
$$V_i \multimap (U_i, Y_i, W_i) \multimap X_i. \quad (37)$$

• Eq. (35) is due to the i.i.d. property of the source that implies U_i is independent of U^{i-1}. The causal encoding with feedback $X_i = f_i(U^i, Y^{i-1})$ and the memoryless property of the channel implies that Y^{i-1} is independent of U_i.
• Eq. (36) comes from the memoryless property of the channel and the fact that Y_i is not included in W_i.
• Eq. (37) comes from the causal encoding with feedback function that implies that X_i is a deterministic function of (U_i, U^{i-1}, Y^{i-1}) which is included in (U_i, Y_i, W_i).

REFERENCES

[1] C. E. Shannon, "The zero error capacity of a noisy channel," *IRE Trans. Inf. Theory*, vol. 2, no. 3, pp. 8–19, 1956.
[2] O. Gossner, P. Hernandez, and A. Neyman, "Optimal use of communication resources," *Econometrica*, vol. 74, no. 6, pp. 1603–1636, 2006.
[3] G. Kramer and S. Savari, "Communicating probability distributions," *IEEE Trans. on Information Theory*, vol. 53, no. 2, pp. 518–525, 2007.
[4] P. Cuff, H. Permuter, and T. Cover, "Coordination capacity," *IEEE Trans. on Information Theory*, vol. 56, no. 9, pp. 4181–4206, 2010.
[5] P. Cuff and L. Zhao, "Coordination using implicit communication," *IEEE Information Theory Workshop (ITW)*, pp. 467–471, 2011.
[6] M. Le Treust, "Empirical coordination for the joint source-channel coding problem," *submitted to IEEE Transactions on Information Theory, http://arxiv.org/abs/1406.4077*, 2014.
[7] N. Gaarder and J. Wolf, "The capacity region of a multiple-access discrete memoryless channel can increase with feedback," *IEEE Transactions on Information Theory*, vol. 21, no. 1, pp. 100–102, Jan 1975.
[8] L. Ozarow, "The capacity of the white gaussian multiple access channel with feedback," *IEEE Transactions on Information Theory*, vol. 30, no. 4, pp. 623–629, Jul 1984.
[9] G. Dueck, "The capacity region of the two-way channel can exceed the inner bound," *Information and Control*, vol. 40, pp. 258–266, 1979.
[10] L. Ozarow and S. Leung-Yan-Cheong, "An achievable region and outer bound for the gaussian broadcast channel with feedback (corresp.)," *IEEE Trans. on Info. Theory*, vol. 30, no. 4, pp. 667–671, Jul 1984.
[11] M. Le Treust, A. Zaidi, and S. Lasaulce, "An achievable rate region for the broadcast wiretap channel with asymmetric side information," *49th Annual Allerton Conference on Communication, Control, and Computing*, pp. 68–75, 2011.
[12] M. Le Treust, "Empirical coordination with two-sided state information and correlated source and states," in *IEEE Internat. Sym. Info. Th.*, 2015.
[13] B. Larrousse, S. Lasaulce, and M. Bloch, "Coordination in distributed networks via coded actions with application to power control," *Submitted to IEEE Trans. on Info. Theory, http://arxiv.org/abs/1501.03685*, 2014.
[14] M. Bloch, L. Luzzi, and J. Kliewer, "Strong coordination with polar codes," *50th Annual Allerton Conference on Communication, Control, and Computing*, pp. 565–571, 2012.
[15] M. Le Treust, "Correlation between channel state and information source with empirical coordination constraint," *IEEE Information Theory Workshop (ITW)*, pp. 272–276, 2014.
[16] A. Bereyhi, M. Bahrami, M. Mirmohseni, and M. Aref, "Empirical coordination in a triangular multiterminal network," *IEEE International Symposium on Information Theory (ISIT)*, pp. 2149–2153, 2013.
[17] Z. Goldfeld, H. H. Permuter, and G. Kramer, "The ahlswede-korner coordination problem with one-sided encoder cooperation," in *IEEE Internat. Symp. on Info. Th. (ISIT)*, pp. 1341–1345, July 2014.
[18] P. Cuff and C. Schieler, "Hybrid codes needed for coordination over the point-to-point channel," *49th Annual Allerton Conference on Communication, Control, and Computing*, pp. 235–239, 2011.
[19] M. Le Treust, "Empirical coordination with feedback," internal technical report, to be submitted, 2015.