Information constraints in multiple agent problems with i.i.d. states

Samson Lasaulce, Sophie Tarbouriech

Information constraints in multiple agent problems with i.i.d. states

Samson Lasaulce and Sophie Tarbouriech

Abstract In this chapter we describe several recent results on the problem of coordination among agents when they have partial information about a state which affects their utility, payoff, or reward function. The state is not controlled and rather evolves according to an independent and identically distributed (i.i.d.) random process. This random process might represent various phenomena. In control, it may represent a perturbation or model uncertainty. In the context of smart grids, it may represent a forecasting noise [1]. In wireless communications, it may represent the state of the global communication channel. The approach used is to exploit Shannon theory to characterize the achievable long-term utility region. Two scenarios are described. In the first scenario, the number of agents is arbitrary and the agents have causal knowledge about the state. In the second scenario, there are only two agents and the agents have some knowledge about the future of the state, making its knowledge non-causal.

Chapter overview

This chapter concerns the problem of coordination among agents. Technically, the problem is as follows. We consider a set of $K \geq 2$ agents. Agent k has a utility, payoff, or reward function $u_k(x_0, x_1, \ldots, x_K)$ where x_k, $k \geq 1$, is the action of Agent k while x_0 is the action of an agent called Nature. The Nature's actions correspond to the system state and is assumed to be non-controlled; more precisely, Nature corresponds to an independent and identically distributed (i.i.d.) random process. The problem studied in this chapter is to characterize the long-term utility region

Samson Lasaulce
L2S, 3 rue Joliot Curie, 91191 Gif-sur-Yvette, France. e-mail: lasaulce@l2s.centralesupelec.fr

Sophie Tarbouriech
LAAS-CNRS, Université de Toulouse, CNRS, Toulouse, France. e-mail: sophie.tarbouriech@laas.fr

under various assumptions in terms of observation at the agents. By long-term utility for Agent k we mean the following quantities:

$$U_k(\sigma_1,...,\sigma_K) = \lim_{T\to+\infty} \frac{1}{T}\mathbb{E}\left[\sum_{t=1}^{T} u_i(X_0(t),...,X_K(t))\right] \quad (1)$$

where $\sigma_k = (\sigma_{k,t})_{t\geq 1}$ is a sequence of functions which represent the strategy of Agent k, $x_k(t)$ is the action chosen by Agent k at time or stage $t \geq 1$, t being the time or stage index; concerning notations, as far as random variables are concerned, capital letters will stand for random variables whereas, small letters will stand for realizations. Note that, implicitly, we assume sufficient conditions (such as utility boundedness) under which the above limit exists. The functions $\sigma_{k,t}$, $k \in \{1,...,K\}$, map the available knowledge to the action of the considered agent. The available knowledge depends on the information assumptions made (e.g., the knowledge of the state can be causal or non-causal). We will distinguish between **two scenarios**. In the first scenario, agents are assumed to have some causal knowledge (in the wide sense) about the state whereas, in the second scenario non-causal knowledge (i.e., some knowledge about the future) about the state is assumed. The second scenario is definitely the most difficult one technically, which is why only two agents will be assumed.

Remarkably, the long-term utility region, whenever available, can be characterized in terms of elegant information constraints. For instance, in the scenario of non-causal state information, determining the long-term utility region amounts to solving a convex optimization problem whose non-trivial constraints are the derived information-theoretic constraints.

1 Introduction

An important example, which illustrates well how the results reported in this chapter can be used, is given by the problem of power control in wireless networks (see Fig. 1). Each transmitter has to adapt its transmit power not only to the fluctuations of the quality of the link (or channel gain) between itself and its respective receiver but also to the transmit power levels of the other transmitters that uses the same radio resources (and therefore create interference). This problem is a multi-agent problem where the agents are the transmitters, the actions of the agents are their transmit power level, and the system state is given by the set of channel gains of the various links in presence; channel gains are typically non-controlled variables (they do not depend on the transmit power levels) and evolve in a random manner; in practice, each transmitter has a partial and imperfect knowledge of the system state. Now, if the agents (namely, the transmitters in the considered example) have a certain performance criterion, which will be referred to as a utility function for the general setup considered in the chapter, the important problem of knowing the best achievable utilities appears. For instance, a transmitter might be designed to maximize its communication rate. The best data rate of a given transmitter would be obtained if all the other transmitters would be silent (i.e., when they don't transmit) and when the transmitter perfectly adapts its power to the channel gain fluctuations of the link between itself and its intended receiver. Obviously, in the real life, several transmitters will transmit at the same time, hence the need to coordinate as well as possible, which leads to the problem of characterizing the best performance possible in terms of coordination. This precisely corresponds to the problem of characterizing the *long-term utility region* i.e., the set of possible achievable points $(U_1, U_2, \ldots, U_K)$ for a given definition for the strategies. In Sections. 3 and 4, we will consider two different definitions for the strategies, each of them corresponding to a given observation structure that is, to some given information assumptions.

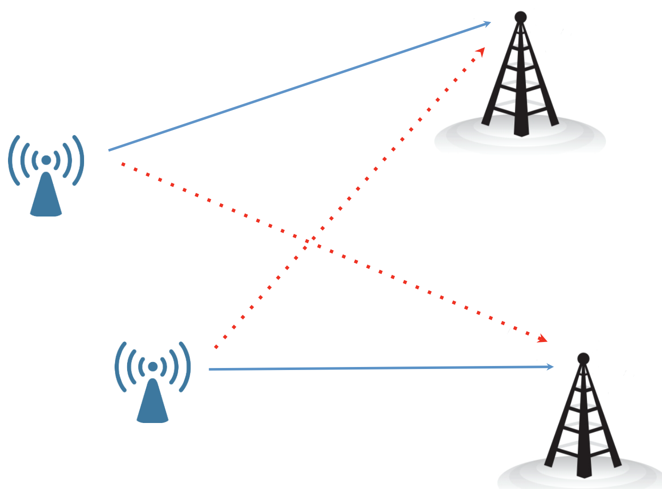

Fig. 1 The problem of power control in wireless networks is a typical application for the results provided in this chapter. The agents are the transmitters, the agents' actions are given by the transmit power level, and the agent utility function may be its communication rate with its intended receiver.

2 General problem formulation

This chapter aims at describing a few special instances of a general problem which has been addressed in several recent works [2], [3], [4], [5], [6], [7], [8].

We consider $K \geq 2$ agents, where Agent $k \in \{1,\ldots,K\}$ produces time-t action $x_k(t) \in \mathscr{X}_k$[1] for $t \in \{1,\ldots,T\}$, $T \geq 1$, the set $\mathscr{X}_k$ representing the set of actions for Agent k. Each agent has access to some observations associated with the chosen actions and the realization of a random process $\{X_{0,t}\}_{t=1}^{T} = \{X_{0,1},...,X_{0,T}\} \in \mathscr{X}_0^T$. In the motivating example described in the introduction, the random process was given by the global wireless channel state i.e., the set of qualities of all the links in presence. In a control problem, the random process may represent a non-controlled perturbation or some uncertainty. All agents' actions and the random process also affect the agents' individual *stage or instantaneous utility functions* $u_1,...,u_K$ where for all $k \in \{1,...,K\}$ the function u_k writes:

$$u_k : \begin{array}{ccc} \mathscr{X}_0 \times \mathscr{X}_1 \times ... \times \mathscr{X}_K & \to & \mathbb{R} \\ (x_0,x_1,...,x_K) & \mapsto & u_k(x_0,x_1,...,x_K) \end{array}. \tag{2}$$

One of the main goals of the chapter is to explain how to determine the set of feasible *expected long-term utilities*:

$$U_k^{(T)} = \mathbb{E}\left[\frac{1}{T}\sum_{t=1}^{T} u_k(X_{0,t},X_{1,t},...,X_{K,t})\right], \tag{3}$$

that are reachable by some strategies for the agents. The set of feasible utilities is fully characterized by the set of feasible averaged joint probability distributions on the $(K+1)-$uple $\{(X_{0,t},X_{1,t},...,X_{K,t})\}_{t=1}^{T}$. Indeed, denoting by $P_{X_{0,t}X_{1,t}\ldots X_{K,t}}$ the joint probability distribution of the time $(K+1)-$uple $(X_{0,t},X_{1,t},...,X_{K,t})$, we have

$$\begin{aligned} U_k^{(T)} &= \frac{1}{T}\sum_{t=1}^{T}\mathbb{E}\left[u_k(X_{0,t},X_{1,t},...,X_{K,t})\right] \\ &= \frac{1}{T}\sum_{t=1}^{T}\sum_{x_0,\ldots,x_K} P_{X_{0,t}X_{1,t}\ldots X_{K,t}}(x_0,x_1,...,x_K)u_k(x_0,x_1,...,x_K) \\ &= \sum_{x_0,\ldots,x_K} u_k(x_0,x_1,...,x_K)\frac{1}{T}\sum_{t=1}^{T} P_{X_{0,t}X_{1,t}\ldots X_{K,t}}(x_0,x_1,...,x_K). \end{aligned}$$

Therefore the problem of characterizing the long-term utility region amounts to determining the set of averaged distributions

$$P^{(T)}(x_0,x_1,...,x_K) = \frac{1}{T}\sum_{t=1}^{T} P_{X_{0,t}X_{1,t}\ldots X_{K,t}}(x_0,x_1,...,x_K) \tag{4}$$

that can be induced by the agents' strategies. For simplicity, and in order to obtain closed-form expressions, we shall focus on the case where $T \to \infty$ [11], [2].

We consider two types of scenarios with two different observation structures. In the first scenario, referred to as the *non-causal state information* scenario, the agents

[1] Throughout the chapter, we assume that all the alphabets such as $\mathscr{X}_k$ are finite.

observe the system states non-causally. That means, at each stage $t \in \{1,\dots,T\}$ they have some knowledge about the entire state sequence $X_0^T = (X_{0,1},\dots,X_{0,T})$. In the second scenario, referred to as the *causal state information* scenario, the agents learn the states only causally and therefore, at any stage t, the agents have some knowledge about the sequence $X_0^t = (X_{0,1},\dots,X_{0,t})$, where throughout the chapter we use the shorthand notations A^m and a^m for the tuples $(A_1,\dots,A_m)$ and $(a_1,\dots,a_m)$, when m is a positive integer.

3 Coordination among agents having causal state information

3.1 Limiting performance characterization

Firstly, we define the information structure under consideration. At every instant or stage t, Agent k is assumed to have an image or a partial observation $S_{k,t} \in \mathscr{S}_k$ of the nature state $X_{0,t}$ with respect to which all agents are coordinating. In the case of the wireless power control example described in the introduction, this might be the knowledge of local channel state information, e.g., a noisy estimate of the direct channel between the transmitter and the associated receiver. The observations $S_{k,t}$ are assumed to be generated by a memoryless channel. By memoryless it is meant that the joint conditional probability on sequences of realizations factorizes the product of individual conditional probabilities. Denoting by $\daleth_k$ the transition probability for the observation structure of Agent k, the memoryless condition can be written as:

$$P(s_K^T|x_0^T) = \prod_{t=1}^{T} \daleth_k(s_k(t)|x_0(t)). \tag{5}$$

The strategy or the sequence of decision functions for Agent k, $\sigma_{k,t}$, is defined by:

$$\sigma_{k,t} \quad : \quad \mathscr{S}_k^t \longrightarrow \mathscr{X}_k \tag{6}$$

$$(s_k(1), s_k(2), ..., s_k(t)) \longmapsto x_k(t) \tag{7}$$

where $\mathscr{S}_k$ is observation alphabet for Agent k.

As mentioned in Section 2, the problem of characterizing the long-term utility region amounts to determining the achievable correlations measured in terms of joint distribution, hence the notion of implementability for a distribution.

Definition 1. (Implementability) The probability distribution $Q(x_0,x_1,...,x_N)$ is implementable if there exist strategies $(\sigma_{1,t})_{t\geq 1},...,\sigma_{K,t})_{t\geq 1}$ such that as $T \to +\infty$, we have for all $x \in \mathscr{X}$,

$$\frac{1}{T}\sum_{t=1}^{T} P_{X_{0,t}\dots X_{K,t}}(x_0,...,x_K) \longrightarrow Q(x_0,...,x_K) \tag{8}$$

where $P_{X_{0,t}\dots X_{K,t}}$ is the joint distribution induced by the strategies at stage t.

The following theorem is precisely based on the notion of implementability and characterizes the achievable long-term utilities that are implementable under the information structure (6); for this, we first define the *weighted utility function* w as a convex combination of the individual utilities u_k:

$$w = \sum_{k=1}^{K} \lambda_k u_k. \tag{9}$$

Theorem 1. *[5] Assume the random process $X_{0,t}$ to be i.i.d. following a probability distribution ρ and the available information to the transmitters $S_{k,t}$ to be the output of a discrete memoryless channel obtained by marginalizing the joint conditional probability $\urcorner$. An expected payoff $\overline{w}$ is achievable in the limit $T \to \infty$ if and only if it can be written as:*

$$\begin{array}{l} \overline{w} = \displaystyle\sum_{\substack{x_0,x_1,\dots x_N,\\ u,s_1,\dots s_N}} \rho(x_0) P_U(u) \urcorner(s_1,\dots,s_K|x_0) \times \\ \left(\prod_{k=1}^{K} P_{X_k|S_k,U}(x_k|s_k,u)\right) w(x_0,x_1,\dots,x_K). \end{array} \tag{10}$$

where U is an auxiliary variable, which can be optimized, and $P_{X_k|S_k,U}(x_k|s_k,u)$ is the probability that Agent k, chooses action x_k after observing s_k,u.

The auxiliary variable U is an external lottery known to the agents beforehand, which can be used to achieve better coordination e.g., in presence of individual constraints or at equilibrium. Theorem 1 allows us find all the achievable utility vectors $(U_1,\dots,U_K)$. Indeed, the long-term utility region being convex (this readily follows from a time-sharing argument), its Pareto boundary can be found by maximizing the weighted utility w. Of course, remains the problem of determining the strategies allowing to operate at a given arbitrary point of the utility region. Since, this problem is non-trivial and there does not exist any methodology for this, we provide an algorithm which allows one to find a suboptimal strategies. Indeed, the associated multilinear optimization problem is too complex to be solved and to overcome this we resort to an iterative technique which is much less complex but is suboptimal.

3.2 An algorithm to determine suboptimal strategies

One of the merits of Theorem 1 is to provide the best performance achievable in terms of long-term utilities when agents have an arbitrary observation structure. However, Theorem 1 does not provide practical strategies which would allow a given utility vector to be reached. Finding ”optimal” strategies consists in finding good sequences of functions as defined per (6), which is an open and promising direction to be explored. More pragmatically, the authors of [6] proposed to restrict to stationary strategies which are merely functions of the form $f_k : \mathscr{S}_k \to \mathscr{X}_k$. This choice is motivated by practical considerations such as computational complexity and it is also coherent with the current state of the literature. The water-filling solution is a special instance of this class of strategies. To find good decision functions, the idea, which is proposed in [6], is to exploit Theorem 1. This is precisely the purpose of this section.

The first observation we make is that the best performance only depends on the vector of conditional probabilities $P_{X_1|S_1,U},\dots,P_{X_K|S_K,U}$ and the auxiliary variable probability distribution P_U, the other quantities being fixed. It is therefore relevant to try to find an optimum vector of lotteries for every action possible and use it to take decisions. Since this task is typically computationally demanding, a possible and generally suboptimal approach consists in applying a distributed algorithm to maximize the expected weighted utility. The procedure proposed in [6] is to use the sequential best response dynamics (see e.g., [10]). The idea is to fix all the variables (that are probability distributions here) expect one and maximize the expected weighted utility with respect to the only possible degree of freedom. This operation is then repeated by considering another variable. The key observation to be made is then to see that when the distributions of the other agents are fixed, the best distribution for Agent k boils down to a function of s_k, giving us a candidate for a decision function which can be used in practice.

To describe the algorithm of [6] (see also Fig. 2), we first rewrite the expected weighted utility in the following manner:

$$\begin{aligned}\overline{W} = & \sum_{x_0,x_1,\dots x_K,u,s_1,\dots s_K} \rho(x_0)P_U(u)\times && (11)\\ & \Gamma(s_1,\dots,s_K|x_0,x_1,\dots,x_K)\times && (12)\\ & \left(\prod_{k=1}^{K} P_{X_k|S_k,U}(x_k|s_k,u)\right) w(x_0,x_1,\dots,x_K) \\ = & \sum_{i_k,j_k,u} \delta_{i_k,j_k,u} P_{X_k|S_k,U}(x_k|s_k,u) && (13)\end{aligned}$$

where i_k,j_k,u are the respective indices of x_k,s_k,u and

$$\delta_{i_k,j_k,u} = \Bigg[\sum_{i_0}\rho(x_{i_0})\Gamma_k(s_k|x_{i_0})\sum_{i_{-k}}u_k(x_{i_0},x_{i_1},\dots x_{i_K})\times \sum_{j_{-k}}\prod_{k'\neq k}\Gamma_{k'}(s_{j_{k'}}|x_0)\prod_{k'\neq k}P_{X_k|S_k,U}(x_k|s_k,u)\Bigg]P_U(u) \quad (14)$$

where i_{-k},j_{-k} are the indices which represent i_k,j_k being constant, while all the other indices are summed over. To make the description of the algorithm clearer, we have also assumed the independence of the observation channels as well as independence of the signal with the strategies chosen by the agents, i.e., $\Gamma(s_1,\dots,s_K|x_0,x_1,\dots,x_K) = \Gamma_1(s_1|x_0)\times\dots\times\Gamma_K(s_K|x_0)$. Written under this form, for every agent, optimizing the expected weighted utility in a distributed manner implies giving a probability 1 for the optimal coefficient $\delta_{i_k,j_k,u}$, and every player does that turn by turn.

Algorithm 1: Proposed decentralized Algorithm for finding an optimal point for (7)

inputs: $\mathcal{X}_k \quad \forall k \in \{0,...,K\}, w(x_0,x_1,...x_K) \quad \forall \underline{x}$, $\rho(x_0), \Gamma_{\underline{S}|\underline{X}_0}(\underline{s}|\underline{x}_0) \quad \forall x_0, f_k^{init} \quad \forall k \in \{1...K\}$, ϵ

output: $f_k^*(s_k) \quad \forall k \in \{1...K\}$

Initialization - $f_k^0 = f_k^{init}, \quad iter = 0$

while $\forall k \ \|f_k^{(iter-1)} - f_k^{iter}\|^2 \leq \epsilon$ *OR iter=0* **do**
- iter = iter+1;
- **forall the** $k \in \{1,\dots,K\}$ **do**
  - **forall the** $s_k \in \mathcal{S}_k$ **do**
    - **forall the** $w \in \mathcal{W}$ **do**
      - **forall the** $x_k \in \mathcal{X}_k$ **do**
        - Find the optimal coefficient $\delta_{i_k,j_k,w}$ using (11);
        - Update the function $f_k^{iter}(s_k) \in \arg\min_{i_k} \quad \delta_{i_k,j_k,w}$;
      - **end**
    - **end**
  - **end**
- **end**

end

Fig. 2 Pseudo-code of the Algorithm proposed in [6] to find suboptimal strategies.

To conclude, note that the above algorithm always converges. This can be proved e.g., by induction or by calling for an exact potential game property (see e.g., [9], [10]).

4 Coordination between two agents having noncausal state information

4.1 Limiting performance characterization

As explained previously, the problem of characterizing the utility region in the case where the state is known non-causally to the agents is much more involved technically. Even in the case of two agents, one may have to face with an open problem, depending on the observation structure assumed for the agents. Here, we consider an important case for which the problem can be solved, as shown in [5]. Therein, the authors consider an asymmetric observation structure. In the case of non-causal state information, agents' strategies are sequences of functions that are defined as follows. For Agent 1 the strategy is defined by:

$$\sigma_{1,t} \quad : \quad \mathscr{S}_1^T \times \mathscr{Y}_1^{t-1} \longrightarrow \mathscr{X}_1 \quad (15)$$
$$(s_1(1),...,s_K(T),y_1(1),...,y_1(t-1)) \longmapsto x_1(t) \quad (16)$$

and for Agent 2 the strategy is defined by:

$$\sigma_{2,t} \quad : \quad \mathscr{S}_2^T \times \mathscr{Y}_2^{t-1} \longrightarrow \mathscr{X}_2 \quad (17)$$
$$(s_2(1),...,s_2(t),y_2(1),...,y_2(t-1)) \longmapsto x_2(t) \quad (18)$$

where $y_k(t) \in \mathscr{Y}_k$ is the observation Agent k has about the triplet $(x_0(t),x_1(t),x_2(t))$ whereas $s_k(t) \in \mathscr{S}_k$ is the observation Agent k has about the state $x_0(t)$. Note that distinguishing between the two observations s_k and y_k is instrumental. Indeed, it does not make any sense physically speaking to assume that an agent might have some future knowledge about the actions of the other agents, which is why the feedback signal is strictly causal. On the other hand, assuming some knowledge about the future of the non-controlled state x_0 perfectly makes sense, as motivated the chapter abstract and the works quoted in the list of references. More precisely the observation is assumed to be the output of a memoryless channel whose transition law is denoted by Γ:

$$\Pr\left[Y_1(t)=y_1(t),Y_2(t)=y_2(t)\middle|X_0^t=s_0^t,X_1^t=x_1^t,X_2^t=x_2^t,Y_1^{t-1}=y_1^{t-1},Y_2^{t-1}=y_2^{t-1}\right]$$
$$=\Gamma(y_1(t),y_2(t)|sx_0(t),x_1(t),x_2(t)). \quad (19)$$

We now provide the characterization of the set of implementable probability distributions both for the considered non-causal strategies.

Theorem 2. *[5] The distribution Q is implementable if and only if it satisfies the following condition*[2]

[2] The notation $I_Q(A;B)$ indicates that the mutual information should be computed with respect to the probability distribution Q.

$$I_Q(S_1;U) \leq I_Q(V;Y_2|U) - I_Q(V;S_1|U). \tag{20}$$

where U and V are auxiliary random variables and Q is any joint distribution that factorizes as

$$\begin{array}{l} Q(x_0,s_1,s_2,u,v,x_1,x_2,y_1,y_2) = \\ \rho(x_0)\urcorner(s_1,s_2|x_0)P_{UVX_1|S_1}(u,v,x_1|s_1)P_{X_2|US_2}(x_2|u,s_2)\Gamma(y_1,y_2|x_0,x_1,x_2) \end{array} \tag{21}$$

In practice, to plot the utility region, one typically has to solve a convex optimization problem. To be illustrative, we consider the special case of [2] namely, $\mathscr{Y} = \mathscr{X}_1$. Denoting by H the entropy function, the problem of finding the Pareto-frontier of the utility region exactly corresponds to solving the following optimization problem:

$$\begin{array}{ll} \text{minimize} & -\sum_{x_0,x_1,x_2} Q(x_0,x_1,x_2)w(x_0,x_1,x_2) \\ \text{subject to} & H_Q(X_0)+H_Q(X_2)-H_Q(X_0,X_1,X_2) \leq 0 \\ & -Q(x_0,x_1,x_2) \leq 0 \\ & -1+\sum_{x_0,x_1,x_2} Q(x_0,x_1,x_2) = 0 \\ & -\rho(x_0)+\sum_{x_1,x_2} Q(x_0,x_1,x_2) = 0 \end{array}.$$

The above problem can be shown to be convex (see [2]). In the next section we exploit this result to assess the performance gain brought by implementing coordination for distributed power control in wireless networks.

4.2 Application to distributed power control

Here we apply the results of the previous section to the wireless power control problem.

A flat-fading interference channel (IC) with two transmitter-receiver pairs is considered. Transmissions are assumed to be time-slotted and synchronized; the time-slot or stage index is denoted by $t \in \mathbb{N}^*$. For $k \in \{1,2\}$ and "$\ell = -k$" ($-k$ stands for the terminal other than k), the signal-to-noise plus interference ratio (SINR) at Receiver k on a given stage writes as $\text{SINR}_k = \frac{g_{kk}x_k}{\sigma^2+g_{\ell k}x_{-k}}$ where $x_k \in \mathscr{X}_i^{\text{IC}} = \{0,P_{\max}\}$ is the power level chosen by Transmitter k, $g_{k\ell}$ represents the channel gain of link $k\ell$, and σ^2 the noise variance. If Transmitter 1 is fully informed of $x_0 = (g_{11},g_{12},g_{21},g_{22})$ for the next stage and Transmitter 2 has no transmit CSI while both transmitters want to maximize the average of a common stage payoff which is $w^{\text{IC}}(x_0,x_1,x_2) = \sum_{k=1}^{2} f(\text{SINR}_k(x_0,x_1,x_2))$, there may be an incentive for Transmitter 1 to inform Transmitter 2 what to do for the next stage; a typical choice for f is $f(a) = \log(1+a)$. Since Transmitter 1 knows the optimal pair of power levels to be chosen on the next stage, say $(x_1^*,x_2^*) \in \arg\max_{(x_1,x_2)} w(x_0,x_1,x_2)$, a simple coded power control (CPC) policy for Transmitter 1 consists in transmitting on stage t at the level Transmitter 2 should transmit on stage $t+1$. Therefore, if Transmitter 2 is able to observe the actions of Transmitter 1, power levels will be optimally tuned half of the time. Such a simple policy, which will be referred to as semi-coordinated PC (SPC), may outperform (in terms of average payoff) pragmatical PC policies such as the one for which the maximum power level is always chosen by both Transmitters ($(x_1, x_2) = (P_{\max}, P_{\max})$ is the Nash equilibrium of the static game whose individual utilities are $u_k = f(\mathrm{SINR}_k)$).

The channel gain of the link between Transmitter k and Receiver ℓ is assumed to be Bernouilli distributed: $g_{k\ell} \in \{g_{\min}, g_{\max}\}$ is i.i.d. and Bernouilli distributed $g_{k\ell} \sim \mathscr{B}(p_{k\ell})$ with $P(g_{k\ell} = g_{\min}) = p_{k\ell}$. The utility function is either $f(a) = \log(1+a)$ or $f(a) = a$. We define $\mathrm{SNR[dB]} = 10\log_{10}\frac{P_{\max}}{\sigma^2}$ and set $g_{\min} = 0.1$, $g_{\max} = 1.9$, $\sigma^2 = 1$. The low and high interference regimes (LIR for low interference regime, HIR and for high interference regime) are respectively defined by $(p_{11}, p_{12}, p_{21}, p_{22}) = (0.5, 0.9, 0.9, 0.5)$ and $(p_{11}, p_{12}, p_{21}, p_{22}) = (0.5, 0.1, 0.1, 0.5)$. At last, $Y \equiv X_1$ and we define two reference PC policies : full power control (FPC) policy $x_k = P_{\max}$ for every stage ; the semi-coordinated PC (SPC) policy $x_2 = P_{\max}$, $x_1^\dagger \in \arg\max_{x_1} w^{\mathrm{IC}}(x_0, x_1, P_{\max})$. Fig. 3 and 4 depict the relative gain in % in terms of average payoff versus SNR[dB] which is obtained by costless optimal coordination and information-constrained coordination. Compared to FPC, gains are very significant whatever the interference regime and provided the SNR has realistic values. Compared to SPC, the gain is of course less impressive since SPC is precisely a coordinated PC scheme but, in the HIR and when the communication cost is negligible, gains as high as 25% can be obtained with $f(a) = \log(1+a)$ and 45% with $f(a) = a$.

5 Conclusion

In this chapter, we have described an information-theoretic framework to characterize the limiting performance of a multiple agent problem. More precisely, the theoretical performance analysis has been conducted in terms of long-term utility region. We have seen that the problem amounts to finding the set of implementable joint distribution over the system state and actions. Both in the scenarios of causal and non-causal state information, auxiliary random variables appear in the characterization of implementable joint distribution. To be able to assess numerically the limiting performance for given utility functions, an optimization problem has to be solved. In the causal state information scenario, the problem is multilinear and the challenge is due to the dimension of the vectors involved. In the non-causal state information scenario, the problem to be solved is a convex problem; more precisely, the information constraint function which translates the agent capabilities in terms of coordination is a convex function of the joint distribution. Note that although the state is not controlled and evolves randomly, the general problem of characterizing the utility region for any number of agents is not trivial. Of course, the problem

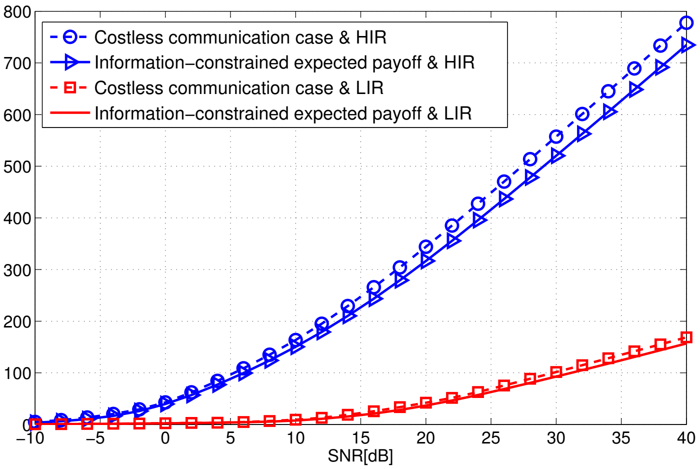

Fig. 3 Relative gain in terms of expected payoff ("CPC/FPC - 1" in [%]) vs SNR[dB] obtained with CPC (with and without communication cost) when the reference power control policy is to transmit at full power (FPC).

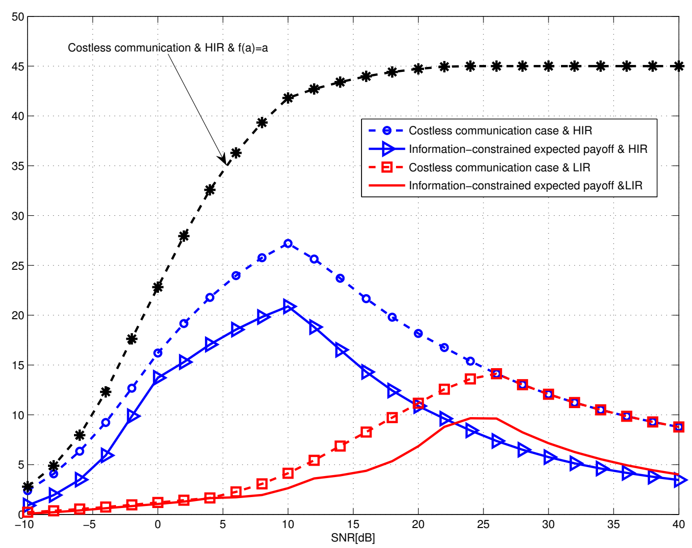

Fig. 4 The difference with Fig. 3 is that the reference power control policy is the semi-coordinated power control policy (SPC), which is already a CPC policy. Additionally, the top curve is obtained with $f(a) = a$.

is even more difficult in the case of controlled states, which therefore constitutes one possible non-trivial extension of the results reported in this chapter. Another interesting research direction would be to consider the case where the state and actions are continuous. A first attempt to this has be made in [12]. Interestingly, the corresponding problem can be shown to be strongly connected to the famous Witsenhausen problem [13], [14], which is a typical decentralized control problem where control and communication intervene in an intricate manner.

References

1. O. Beaude, A. Agrawal, and S. Lasaulce, ”A framework for computing power consumption scheduling functions under uncertainty”, 6th IEEE International Conference on Smart Grid Communications (SmartGridComm 2015), Miami, Florida, USA, Nov. 2015.
2. B. Larrousse and S. Lasaulce, ”Coded Power Control: Performance Analysis”, IEEE Intl. Symposium on Information Theory (ISIT), Istanbul, Turkey, Jul. 2013.
3. B. Larrousse, A. Agrawal, and S. Lasaulce, ”Implicit coordination in two-agent team problems. Application to distributed power allocation”, IEEE 12th Intl. Symposium on Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks (WiOpt), Hammamet, Tunisia, May 2014.
4. B. Larrousse, S. Lasaulce, and M. Wigger, ”Coordination in state-dependent distributed networks: The two-agent case”, IEEE International Symposium on Information Theory (ISIT), Hong Kong, June 2015.
5. B. Larrousse, S. Lasaulce, and M. Wigger, ”Coordination in State-Dependent Distributed Networks”, IEEE Proc. of the Information Theory Workshop (ITW), Jerusalem, Israel, Apr.-May 2015.
6. A. Agrawal, S. Lasaulce, O. Beaude, and R. Visoz, ”A framework for decentralized power control with partial channel state information”, IEEE Proc. of the Fifth International Conference on Communications and Networking (ComNet 2015), Hammamet, Tunisia, November 4-7, 2015.
7. B. Larrousse, S. Lasaulce, and M. Bloch, ”Coordination in distributed networks via coded actions with application to power control”, IEEE Transactions on Information Theory, available on arXiv, https://arxiv.org/pdf/1501.03685.pdf
8. M. Le Treust, ”Joint empirical coordination of source and channel”, IEEE Transactions on Information Theory, available on arXiv, https://arxiv.org/pdf/1406.4077.pdf
9. D. Mondrere, ”Potential games”, Games and Economic Behavior, Vol. 14, pp. 124143.
10. S. Lasaulce and H. Tembine, ”Game Theory and Learning for Wireless Networks: Fundamentals and Applications”, Academic Press, Elsevier, pp. 1-336, Oct. 2011, ISBN 978-0123846983.
11. O. Gossner, P. Hernandez, and A. Neyman, Optimal use of communication resources, Econometrica, vol. 74, no. 6, pp. 16031636, 2006.
12. A. Agrawal, F. Danard, B. Larrousse, and S. Lasaulce, ”Implicit coordination in two-agent team problems”, European Conference on Control (ECC), Linz, Austria, July 2015.
13. H. S. Witsenhausen, ”A counterexample in stochastic optimum control”, SIAM Journal on Control, Vol. 6, pp. 131-147, 1968.
14. A. Sahai, and P. Grover, ”Demystifying the Witsenhausen counterexample”, IEEE Control Systems, Vol. 30, pp. 20-24, 2010.