The common-information approach to decentralized stochastic control

Ashutosh Nayyar, Aditya Mahajan, and Demosthenis Teneketzis

1 Introduction

Many modern technological systems, such as cyber-physical systems, communication, transportation and social networks, smart grids, sensing and surveillance systems are informationally decentralized. A key feature of informationally decentralized systems is that decisions are made by multiple decision makers that have access to different information. This feature violates the fundamental assumption upon which centralized stochastic control theory is based, namely, that all decisions are made by a centralized decision maker who has access to all the information and perfectly recalls all past observations and decisions/actions. Consequently, techniques from centralized stochastic control cannot be directly applied to decentralized stochastic control problem primarily for the following reason. In centralized stochastic control, the controller's belief on the current state of the system is a sufficient statistic for decision making. A similar sufficient statistic does not work for decentralized stochastic control because controllers have different information and hence their beliefs on the state of the system are not consistent.

Nevertheless, two general approaches that use ideas from centralized stochastic control theory have been used for the solution of decentralized control problems: (i) *the person-by-person approach*; and (ii) *the designer's approach*. A detailed discussion of the features and merits of these approaches, as well as their application

---

Ashutosh Nayyar
Department of Electrical and Computer Engineering, University of California, Berkeley, CA, USA, e-mail: anayyar@berkeley.edu

Aditya Mahajan
Department of Electrical and Computer Engineering, McGill University, Montreal, QC, Canada, e-mail: aditya.mahajan@mcgill.ca

Demosthenis Teneketzis
Department of Electrical and Computer Engineering, University of Michigan, Ann Arbor, MI, USA, e-mail: teneket@umich.edu

to various classes of problems appears in [17]. Here we briefly present the key characteristics of each approach.

The person-by-person approach investigates the decentralized control problem from the viewpoint of one decision-maker, say the i-th decision-maker and proceeds as follows: (i) arbitrarily fix the strategy of all other decision-makers; and (ii) use centralized stochastic control to derive structural properties for the optimal best-response strategy of the i-th decision-maker. The person-by-person approach can, in several problem instances [9, 14, 18–20, 26–30, 33–37, 41, 42], identify qualitative properties of globally optimal control strategies; furthermore it provides an iterative method to obtain person-by-person optimal strategies [7] which, in general, are not globally optimal.

The designer's approach looks at the decentralized control problem from the point of view of a system designer who knows the system model and the statistics of the primitive random variables, and chooses control/decision strategies for all decision makers. This approach leads to a centralized planning problem whose solution results in globally optimal control strategies. Such strategies are determined by a dynamic program where each step is a functional optimization problem (in contrast to the usual centralized dynamic program where each step is a parameter optimization problem). Thus, the determination of globally optimal strategies via the designer's approach is a computationally formidable problem [40].

In several instances of decentralized control problems [14, 36, 37] the person-by-person approach is used first to identify qualitative properties of globally optimal strategies; then, the designer's approach is employed to determine globally optimal strategies with the identified qualitative property.

In addition to the above mentioned approaches, other methods that exploit system's information structure have been developed for the solution of decentralized control problems. Specifically, solution approaches for systems with partially nested information structure have appeared in [4, 8, 10, 11, 23]; a generalization of partial nestedness called stochastic nestedness was defined and studied in [44]. In [8] it was shown that for linear quadratic Gaussian (LQG) control problems with partially nested information structure, there is an affine control strategy that is globally optimal. In general, the problem of determining optimal control strategies within the class of affine control policies may not be a convex optimization problem; conditions under which it is convex were identified in [2, 24].

Decentralized stochastic control problems with specific models of information sharing among controllers, such as delayed information sharing [1,16,31,43], periodic information sharing [21], broadcast information structure [42], control sharing [3,12], and systems with common and private observations [13] have also been investigated in the literature.

In [17] a new general model of decentralized stochastic control, called *partial history sharing* information structure, was presented. In this model it is assumed that: (i) controllers sequentially share part of their past data (observations and control actions) with one another by means of a shared memory; and (ii) all controllers have perfect recall of the commonly available data (also called the common information). This model subsumes a large class of decentralized control models where information is shared among the controllers. A solution methodology for this model was presented in [17]. This solution methodology is based on the *common information approach* developed in [15] which is applicable to all sequential decision making problems. The common information approach provides a unified framework for several decentralized control problems that had previously been addressed using problem specific solution techniques. The key idea behind this approach is the reformulation of the original decentralized control problem into an equivalent centralized problem from the perspective of a coordinator. The coordinator knows the common information and selects prescriptions that map each controller's local information to its control actions. The optimal control problem at the coordinator is a partially observable Markov decision process (POMDP) that can solved using techniques from Markov decision theory. This approach provides: (i) structural results (qualitative properties) for optimal strategies; and (ii) a dynamic program for obtaining *globally optimal* strategies for all controllers in the original decentralized problem. Notably, the structural results of optimal control strategies obtained by the common information approach cannot be obtained by the person-by-person approach (see [17, Sec III-A]; and the dynamic program obtained by the common information approach is simpler than that obtained by the designer's approach (see [17, Sec III-A].

In this chapter we present the common information approach to decentralized stochastic control. Our objective is to demonstrate that this approach is conceptually powerful as it overcomes some of the fundamental difficulties in decentralized decision making, it has broad applicability, it can resolve a long-standing open problem in decentralized stochastic control, and it can simplify the search for globally optimal strategies.

This chapter is organized as follows. In Section 2 we first present two examples, one for a one stage decentralized control problem (static team) and the other for a two stage decentralized control problem (dynamic team), that illustrate how the common information approach simplifies the search of globally optimal strategies; then we describe the key steps of the approach. In Section 3 we present a brief recap of partially observed Markov decision processes (POMDPs) which play a key role in the common information approach. In Section 4 we illustrate how the common information approach can be used to solve problems that arise in control, communication, and queueing systems. In Section 5, we demonstrate how our approach can resolve a long-standing open problem [39] in decentralized stochastic control. We conclude in Section 6 by discussing how the common information circumvents the conceptual difficulties associated with decentralized stochastic control.

Terminology

Decentralized stochastic control problems are also referred to as *team problems* and further classified as *static* and *dynamic* teams. In dynamic teams, the information observed by a decision maker depends on the control actions of other decision makers, while in static teams it does not. See [7,8] for details. Decentralized stochastic control problems are typically dynamic team problems.

Notation

Random variables are denoted by upper case letters; their realization by the corresponding lower case letter. For integers $a \leq b$ and $c \leq d$, $X_{a:b}$ is a short hand for the vector $(X_a, X_{a+1}, \dots, X_b)$. When $a > b$, $X_{a:b}$ equals the empty set. In general, subscripts are used as time index while superscripts are used to index controllers. $\mathbb{P}(\cdot)$ is the probability of an event, $\mathbb{E}[\cdot]$ is the expectation of a random variable. For a collection of functions g, we use $\mathbb{P}^g(\cdot)$ and $\mathbb{E}^g[\cdot]$ to denote that the probability measure/expectation depends on the choice of functions in g.

2 The common information approach to decentralized stochastic control

The main idea of the common information approach to decentralized stochastic control is to formulate and analyze an alternative but equivalent centralized stochastic control problem. To illustrate this idea, we start with two of the simplest examples of decentralized stochastic control: (i) a two controller static team problem; and (ii) a two controller two-stage dynamic team problem. For both these examples, we show how the common information approach works and simplifies the search of globally optimal strategies. After presenting these examples, we present a high-level description of the main steps of the common information approach.

2.1 Illustrative Example 1: A two controller static team

The following example, which is adapted from [17], illustrates how the common information approach decomposes a static team problem into several smaller subproblems that are easier to solve.

Consider a two controller static team. Nature selects a random variable W. Controller i, $i = 1,2$, observes a common observation C and a local observation M^i. The observations (C, M^1, M^2) are a function of W.

The controllers select their control actions U^1 and U^2 using control laws g^1 and g^2 of the form

$$U^1 = g^1(C, M^1), \quad U^2 = g^2(C, M^2).$$

The system incurs a loss $\ell(W, U^1, U^2)$.

Suppose all system variables are finite valued and W, C, M^i, U^i take values in finite sets $\mathscr{W}$, $\mathscr{C}$, $\mathscr{M}^i$, and $\mathscr{U}^i$, $i = 1,2$, respectively. The objective is to choose control laws

$$g^1 : \mathscr{C} \times \mathscr{M}^1 \mapsto \mathscr{U}^1, \quad g^2 : \mathscr{C} \times \mathscr{M}^2 \mapsto \mathscr{U}^2$$

to minimize

$$J(g^1, g^2) = \mathbb{E}^{(g^1,g^2)}[\ell(W, U^1, U^2)].$$

Since all system variables are finite valued, one solution approach is to find the globally optimal control laws (g^1, g^2) by a brute force search over all possible $\prod_{i=1}^2 |\mathscr{U}^i|^{|\mathscr{C}||\mathscr{M}^i|}$ control laws. For example, if all system variables are binary valued, we need to search over $2^4 \times 2^4 = 256$ possibilities.

The common information approach reduces the number of possibilities that need to be searched. The main idea of this approach is that instead of specifying the control laws (g^1, g^2) directly, we specify them indirectly as follows. Consider an alternative *coordinated system* in which a *coordinator* observes the common information C and chooses prescriptions (Γ^1, Γ^2), where Γ^i is a mapping from *local information* M^i to control action U^i, according to a *coordination law* ψ that is of the form

$$(\Gamma^1, \Gamma^2) = \psi(C).$$

The coordinator communicates these prescriptions (Γ^1, Γ^2) to the controllers who use them to generate control actions as follows:

$$U^1 = \Gamma^1(M^1), \quad U^2 = \Gamma^2(M^2).$$

The objective of the coordinated system is to find a coordination law ψ to minimize

$$\tilde{J}(\psi) = \mathbb{E}^{\psi}[\ell(W, U^1, U^2)].$$

It is easy to verify that there is an one-to-one correspondence between the control laws (g^1, g^2) of the original system and the coordination law ψ of the coordinated system. The optimization problem at the coordinator is a centralized stochastic optimization problem in which the coordinator is the only decision-maker. To solve this centralized stochastic optimization problem, consider any coordination law ψ and for any $c \in \mathscr{C}$, let $(\gamma_c^1, \gamma_c^2) = \psi(c)$. Write the expected loss $\tilde{J}(\psi)$ as

$$\sum_{c \in \mathscr{C}} \mathbb{P}(C = c) \mathbb{E}[\ell(W, \gamma_c^1(M^1), \gamma_c^2(M^2)) \mid C = c]$$

Minimizing $\tilde{J}(\psi)$ is equivalent to separately minimizing, for each value of $c \in \mathscr{C}$, the expected conditional loss $\mathbb{E}[\ell(W, \gamma_c^1(M^1), \gamma_c^2(M^2)) \mid C = c]$ over the choice of (γ_c^1, γ_c^2). One solution approach to solve each of these latter minimizations is by a brute force search over all possible $\prod_{i=1}^2 |\mathscr{U}^i|^{|\mathscr{M}^i|}$ possibilities. Thus, this approach requires searching over $|\mathscr{C}| \prod_{i=1}^2 |\mathscr{U}^i|^{|\mathscr{M}^i|}$ possibilities. For example, if all system variables are binary valued, we need to search over $2 \times 2^2 \times 2^2 = 32$ possibilities. Contrast this by the 256 possibilities that need to be evaluated for a brute force search in the original setup. In general, for this example, the common information approach provides an exponential simplification by reducing the search complexity from

$$\left(\prod_{i=1}^{2}|\mathscr{U}^i|^{|\mathscr{M}^i|}\right)^{|\mathscr{C}|} \quad \text{to} \quad |\mathscr{C}|\prod_{i=1}^{2}|\mathscr{U}^i|^{|\mathscr{M}^i|}.$$

2.2 Illustrative Example 2: A two-stage two-controller dynamic team

The following example illustrates how the common information approach provides a dynamic programming decomposition in a multi-stage dynamic team problem. Consider a two-stage two-controller dynamic team that evolves as follows.

- At $t = 1$, nature selects a random variable W_1. Controller i, $i = 1,2$, observes a common observation C_1 and a local observation M_1^i. The observations (C_1, M_1^1, M_1^2) are a function of W_1.
  The controllers select their control actions U_1^1 and U_1^2 using control laws g_1^1 and g_1^2 of the form
  $$U_1^1 = g_1^1(C_1, M_1^1), \quad U_1^2 = g_1^2(C_1, M_1^2).$$
- At $t = 2$, nature selects a random variable W_2 that may be correlated with W_1. As in stage 1, controller i, $i = 1,2$, observes a common observation C_2 and a local observation M_2^i. The difference from stage 1 is that the observations (C_2, M_2^1, M_2^2) are a function of (W_2, U_1^1, U_1^2).
  The controllers select their control actions U_2^1 and U_2^2 using control laws g_2^1 and g_2^2 of the form
  $$U_2^1 = g_2^1(C_1, C_2, M_1^1, M_2^1), \quad U_2^2 = g_2^2(C_1, C_2, M_1^2, M_2^2).$$
- At the end of the two stages, the system incurs a loss $\ell(W_1, W_2, U_1^1, U_1^2, U_2^1, U_2^2)$.

Suppose all system variables are finite valued and W_t, C_t, M_t^i, U_t^i take values in finite sets $\mathscr{W}_t$, $\mathscr{C}_t$, $\mathscr{M}_t^i$, and $\mathscr{U}_t^i$, $i = 1,2$, $t = 1,2$. The objective is to choose control laws

$$g_1^i : \mathscr{C}_1 \times \mathscr{M}_1^i \mapsto \mathscr{U}_1^i, \quad g_2^i : \mathscr{C}_1 \times \mathscr{C}_2 \times \mathscr{M}_1^i \times \mathscr{M}_2^i \mapsto \mathscr{U}_1^i, \quad i = 1,2$$

to minimize

$$J(g_1^1, g_2^1, g_1^2, g_2^2) = \mathbb{E}^{(g_1^1, g_2^1, g_1^2, g_2^2)}[\ell(W_1, W_2, U_1^1, U_1^2, U_1^2, U_2^2)].$$

Since all system variables are finite valued, one solution approach is to find globally optimal control strategies $(g_1^1, g_2^1, g_1^2, g_2^2)$ by a brute force search over all possible

$$\prod_{i=1}^{2} |\mathscr{U}_1^i|^{|\mathscr{C}_1||\mathscr{M}_1^i|} |\mathscr{U}_2^i|^{|\mathscr{C}_1||\mathscr{C}_2||\mathscr{M}_1^i||\mathscr{M}_2^i|}$$

control strategies. For example, if all system variables are binary valued, we need to search over $(2^4 \times 2^{16})^2 = 2^{40}$ possibilities.

The common information approach enables us to decompose the above multi-stage optimization problem using a dynamic program. As in the static case, the main idea of the common information approach is that instead of specifying the control strategies $(g_1^1, g_2^1, g_1^2, g_2^2)$ directly, we specify them indirectly as follows. Consider an alternative two-stage *coordinated system* in which a coordinator *with perfect recall* observes the common information C_t at time t and chooses prescriptions (Γ_t^1, Γ_t^2) where Γ_1^i is a mapping from *local information* M_1^i to control action U_1^i while Γ_2^i is a mapping from *local information* (M_1^i, M_2^i) to control action U_2^i. These prescriptions are chosen according to a *coordination strategy* (ψ_1, ψ_2) that is of the form

$$(\Gamma_1^1, \Gamma_1^2) = \psi_1(C_1), \quad (\Gamma_2^1, \Gamma_2^2) = \psi_2(C_1, C_2).$$

At time t, the coordinator communicates prescriptions (Γ_t^1, Γ_t^2) to the controllers who use them to generate control actions as follows:

$$U_1^i = \Gamma_1^i(M_1^i), \quad U_2^i = \Gamma_2^i(M_1^i, M_2^i), \quad i = 1, 2.$$

The objective of the coordinated system is to find coordination strategy (ψ_1, ψ_2) to minimize

$$\tilde{J}(\psi_1, \psi_2) = \mathbb{E}^{(\psi_1, \psi_2)}[\ell(W_1, W_2, U_1^1, U_1^2, U_1^2, U_2^2)].$$

It is easy to verify that there is a one-to-one correspondence between the control strategies $(g_1^1, g_2^1, g_1^2, g_2^2)$ of the original system and the coordination strategy (ψ_1, ψ_2) of the coordinated system. The multi-stage optimization problem at the coordinator is a centralized stochastic control problem in which the coordinator is the only decision maker and has perfect recall. To solve this centralized stochastic control problem, proceed as follows. Consider any coordination strategy (ψ_1, ψ_2) and any realization $(c_1, c_2) \in \mathscr{C}_1 \times \mathscr{C}_2$ of the common information. Suppose the prescriptions $(\gamma_1^1, \gamma_1^2) = \psi_1(c_1)$ are fixed. Given this information, what is the best choice of the prescriptions $(\gamma_2^1, \gamma_2^2) = \psi_2(c_1, c_2)$ at time $t = 2$? For any choice $(\tilde{\gamma}_2^1, \tilde{\gamma}_2^2)$ of the prescriptions at time $t = 2$, the expected conditional loss is given by

$$\mathbb{E}[\ell(W_1, W_2, U_1^1, U_1^2, U_1^2, U_2^2) \mid c_1, c_2, \gamma_1^1, \gamma_1^2, \tilde{\gamma}_2^1, \tilde{\gamma}_2^2].$$

Since all the prescriptions are specified, the control actions $(U_1^1, U_1^2, U_1^2, U_2^2)$ are well-defined random variables, and the above conditional expectation is well-defined. To obtain the best choice of the prescriptions at time $t = 2$, minimize the above conditional expectation over all possible choices of $(\tilde{\gamma}_2^1, \tilde{\gamma}_2^2)$ and define the minimum value as

$$V(c_1, c_2, \gamma_1^1, \gamma_1^2) = \min_{\tilde{\gamma}_2^1, \tilde{\gamma}_2^2} \mathbb{E}[\ell(W_1, W_2, U_1^1, U_1^2, U_1^2, U_2^2) \mid c_1, c_2, \gamma_1^1, \gamma_1^2, \tilde{\gamma}_2^1, \tilde{\gamma}_2^2]. \tag{1}$$

One solution approach to solve the above minimization is by a brute force search over all possible $\prod_{i=1}^{2} |\mathscr{U}_2^i|^{|\mathscr{M}_1^i||\mathscr{M}_2^i|}$ prescription pairs. To find the optimal coordination law ψ_2, we need to solve the above minimization problem *for all possible realizations of the common information* (c_1, c_2) *and choices of past prescription* (γ_1^1, γ_1^2). Thus, we need to solve $|\mathscr{C}_1||\mathscr{C}_2| \prod_{i=1}^{2} |\mathscr{U}_1^i|^{|\mathscr{M}_1^i|}$ minimization problems, each requiring the evaluation of $\prod_{i=1}^{2} |\mathscr{U}_2^i|^{|\mathscr{M}_1^i||\mathscr{M}_2^i|}$ conditional expectations.

Now that we know how the coordinator selects optimal prescriptions at time $t = 2$, what is the best choice of prescriptions (γ_1^1, γ_1^2) at time $t = 1$? For any realization $c_1 \in \mathscr{C}_1$ and any choice of coordination law $\tilde{\psi}_2$, the expected conditional loss at the coordinator when the prescriptions at time $t = 1$ are $(\tilde{\gamma}_1^1, \tilde{\gamma}_1^2)$ is given as

$$\begin{aligned} &\mathbb{E}[\ell(W_1, W_2, U_1^1, U_1^2, U_1^2, U_2^2) \mid c_1, \tilde{\gamma}_1^1, \tilde{\gamma}_1^2] \\ &\qquad = \mathbb{E}\Big[\mathbb{E}[\ell(W_1, W_2, U_1^1, U_1^2, U_1^2, U_2^2) \mid c_1, C_2, \tilde{\gamma}_1^1, \tilde{\gamma}_1^2, \tilde{\psi}_2] \,\Big|\, c_1, \tilde{\gamma}_1^1, \tilde{\gamma}_1^2\Big]. \end{aligned} \tag{2}$$

Use the optimal prescription at time $t = 2$, which is given by (1), to lower bound the conditional expected cost in (2) as follows:

$$\begin{aligned} &\mathbb{E}\Big[\mathbb{E}[\ell(W_1, W_2, U_1^1, U_1^2, U_1^2, U_2^2) \mid c_1, C_2, \tilde{\gamma}_1^1, \tilde{\gamma}_1^2, \tilde{\psi}_2] \,\Big|\, c_1, \tilde{\gamma}_1^1, \tilde{\gamma}_1^2\Big] \\ &\qquad\qquad \geq \mathbb{E}\Big[V(c_1, C_2, \tilde{\gamma}_1^1, \tilde{\gamma}_1^2) \,\Big|\, c_1, \tilde{\gamma}_1^1, \tilde{\gamma}_1^2\Big] \end{aligned} \tag{3}$$

with equality if the coordinator uses the optimal prescriptions at time $t = 2$, which are given by (1).

One solution approach to select the best prescriptions at time $t = 1$ is to evaluate the conditional expectation in (3) for all $\prod_{i=1}^{2} |\mathscr{U}_1^i|^{|\mathscr{M}_1^i|}$ choices of $(\tilde{\gamma}_1^1, \tilde{\gamma}_1^2)$. To find the optimal coordination law ψ_1, we need to solve the above minimization problem *for all possible realizations of* c_1. Thus, we need to solve $|\mathscr{C}_1|$ minimization problems, each requiring the evaluation of $\prod_{i=1}^{2} |\mathscr{U}_1^i|^{|\mathscr{M}_1^i|}$ conditional expectations.

The above *dynamic program* based on the common information approach requires

$$|\mathscr{C}_1| \prod_{i=1}^{2} |\mathscr{U}_1^i|^{|\mathscr{M}_1^i|} + |\mathscr{C}_1||\mathscr{C}_2| \prod_{i=1}^{2} |\mathscr{U}_2^i|^{|\mathscr{M}_1^i||\mathscr{M}_2^i|}$$

evaluations.[1] For example, when all variables are binary valued, we need to evaluate 2^{14} conditional expectations. Contrast this with 2^{40} possibilities that need to be evaluated for a brute force search in the original setup. In general, for this example, the common information approach provides an exponential simplification by reducing the search complexity from

[1] We assume that evaluating expected loss or expected conditional loss requires same computational effort irrespective of the cost function and the probability measure. This analysis is meant to provide a general idea of reduction in complexity, and is not a strict evaluation of the computational benefits of the common information approach.

$$\left(\prod_{i=1}^{2}|\mathscr{U}_1^i|^{|\mathscr{M}_1^i|}\left(|\mathscr{U}_2^i|^{|\mathscr{M}_1^i||\mathscr{M}_2^i|}\right)^{|\mathscr{C}_2|}\right)^{|\mathscr{C}_1|} \quad \text{to} \quad |\mathscr{C}_1|\left(\prod_{i=1}^{2}|\mathscr{U}_1^i|^{|\mathscr{M}_1^i|}+|\mathscr{C}_2|\prod_{i=1}^{2}|\mathscr{U}_2^i|^{|\mathscr{M}_1^i||\mathscr{M}_2^i|}\right)$$

In general, it is possible to improve the computational advantage of the common information approach by:

1. *Identifying irrelevant information at the controllers:* One way of reducing the complexity of coordinator's problem is to show that part of local information is irrelevant for controllers. If this can be established (often by using the person-by-person approach described in Section 1), then the coordinator's prescription are mappings from the reduced local information to control actions. This reduces the number of possible prescription choices to be considered by the coordinator.
2. *Identifying an information state for the coordinator:* An information state serves as a sufficient statistic for the data available to the coordinator. Instead of finding best prescriptions for all possible realizations of coordinator's data, we only need to find best prescriptions for each each realization of coordinator's information state. If the coordinator's decision problem can be shown to be equivalent to some known models of centralized stochastic control (such as Markov decision problems or partially observed Markov decision problem), then we can use stochastic control techniques to find an information state for the coordinator.

2.3 The common information approach

The previous two examples illustrate how the common information approach works for simple static and dynamic teams. We generalize this approach to to a broad class of decentralized stochastic control systems by proceeding as follows:

1. ***Construct a coordinated system***
   The first step of the approach is to identify the common information at the controllers. The common information at time t must be known to all controllers at t. Define the local information at a controller to be the information left after subtracting the common information from all the data available at that controller. If the common information is non-empty, construct a coordinated system in which at each time a coordinator has access to the common information at that time and selects a set of prescriptions that map each controllers' local information to its control action. The loss function of the coordinated system is the same as the loss function of the original system. The objective of the coordinator is to choose a coordination strategy (i.e., a sequence of coordination laws) to minimize the expected total loss.
2. ***Formulate the coordinated system as a POMDP***
   If the system model is such that the data available at the coordinator—the common information— is increasing with time, then the decision problem at the coordinator is centralized stochastic control problem. The second step of the approach is to formulate this centralized stochastic control problem as a partially observable

Markov decision process (POMDP). To do so, we need to identify the (unobserved) state for input-output mapping for the coordinated system. In general, the vector consisting of the state of the original system and the local information of all controllers (or an appropriate subset of this vector) is a state for input-output mapping for the coordinated system.

3. ***Solve the resultant POMDP***
   The third step of the approach is to use Markov decision theory to identify the structure of optimal coordination strategies in the coordinated system and to identify a dynamic program to obtain an optimal coordination strategy with such structure.
4. ***Show equivalence between the original system and the coordinated system***
   The fourth step of the approach is to show that the two models are equivalent. In particular, for any coordination strategy in the coordinated system, there exists a control strategy in the original system that yields the same expected loss, and vice-versa.
5. ***Translate the solution of the coordinated system to the original system***
   The fifth step of the approach is to use the equivalence of the fourth step to translate the structural results and the dynamic program obtained in the third step for the coordinated system to structural results and dynamic program for the original system.

In Sections 4 and 5, we illustrate how the above methodology applies to problems is communication, control and queueing systems. Before we present these applications, we briefly review the POMDP model and results.

3 A brief recap of partially observable Markov decision processes (POMDPs)

A partially observable Markov decision process (POMDP) is a model of centralized (single decision-maker) stochastic control. It consists of a state process $\{S_t\}_{t=1}^{T}$, an observation process $\{O_t\}_{t=1}^{T}$, and an action process $\{A_t\}_{t=1}^{T}$. For simplicity, assume that all system variables are finite valued and S_t, O_t, A_t takes value in time-homogeneous finite sets $\mathscr{S}$, $\mathscr{O}$, and $\mathscr{A}$. A POMDP has the following features:

1. The decision maker perfectly recalls its past observations and actions and chooses the action as a function of its observation and action history, that is,
   $$A_t = d_t(O_{1:t}, A_{1:t-1}),$$
   where d_t is the decision rule at time t.
2. The state, observation, and action processes satisfy the following controlled Markov property

$$\mathbb{P}(S_{t+1}, O_{t+1} \mid S_{1:t}, O_{1:t}, A_{1:t}) = \mathbb{P}(S_{t+1}, O_{t+1} \mid S_t, A_t).$$

3. At each time, the system incurs an instantaneous cost $\ell(S_t, A_t)$.
4. The objective of the decision-maker is to choose a decision strategy $\mathbf{d} := (d_1, \dots, d_T)$ to minimize a total cost which is given by

$$J(\mathbf{d}) = \mathbb{E}\Big[\sum_{t=1}^{T} \ell(S_t, A_t)\Big].$$

The following standard result from Markov decision theory identifies the structure of globally optimal decision strategies and a dynamic program to find optimal strategies with that structure. See [38] for details.

Theorem 1 (POMDP Result). *Let Θ_t be the conditional probability distribution of the state S_t at time t given the observations $O_{1:t}$ and actions $A_{1:t-1}$,*

$$\Theta_t(s) := \mathbb{P}(S_t = s \mid O_{1:t}, A_{1:t-1}), \quad s \in \mathscr{S}.$$

Then,

1. *$\Theta_{t+1} = \eta_t(\Theta_t, A_t, O_{t+1})$, where η_t is the standard non-linear filter described as follows: If θ_t, a_t, o_{t+1} are the realizations of Θ_t, A_t and O_{t+1}, then the realization of s^{th} element of the vector Θ_{t+1} is*

$$\begin{aligned}\theta_{t+1}(s) &= \frac{\sum_{s'} \theta_t(s')\mathbb{P}(S_{t+1} = s, O_{t+1} = o_{t+1} \mid S_t = s', A_t = a_t)}{\sum_{s'', \tilde{s}} \theta_t(s'')\mathbb{P}(S_{t+1} = \tilde{s}, O_{t+1} = o_{t+1} | S_t = s'', A_t = a_t)} \\ &=: \eta_t^s(\theta_t, a_t, o_{t+1}).\end{aligned}$$

The function $\eta_t(\theta_t, a_t, o_{t+1})$ is the vector of functions $(\eta_t^s(\theta_t, a_t, o_{t+1}))_{s \in \mathscr{S}}$.

2. *There exists an optimal decision strategy of the form*

$$A_t = \hat{d}_t(\Theta_t).$$

Furthermore, the following dynamic program determines such an optimal strategy: Define

$$V_T(\theta) := \min_a \mathbb{E}[\ell(S_T, a) \mid \Theta_T = \theta],$$

and for $t = T-1, T-2, \dots, 1$, recursively define

$$V_t(\theta) := \min_a \mathbb{E}\big[\ell(S_t, a) + V_{t+1}(\eta_t(\theta, a, O_{t+1})) \bigm| \Theta_t = \theta, A_t = a\big].$$

Then, for each time t and each realization of θ of Θ_t, the optimal action $\hat{d}_t(\theta)$ is the minimizer in the definition of $V_t(\theta)$.

4 Applications of the common information approach to communication, networked control, and queueing systems

In this section we illustrate how the common information approach provides a unified framework for solving problems that arise in various disciplines such as communication, networked control, and queueing systems. These problems have been previous investigated using problem specific solution techniques.

4.1 Point-to-point real-time communication with feedback

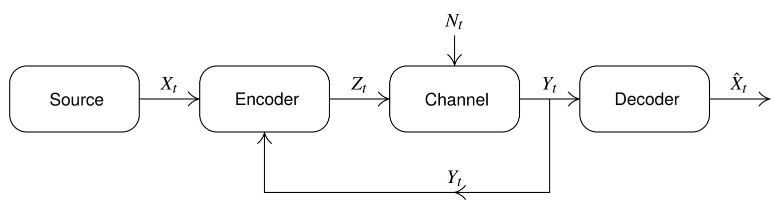

Fig. 1 A real-time communication system with noiseless feedback

Communication problems can be thought of as team problems with the encoders and the decoders as the decision-makers in the team. Point-to-point feedback communication, in particular, is a dynamic team problem because: (a) the encoder (and in some cases the decoder as well) has to make decisions over time based on information that is changing with time, and (b) the decoder's information is directly affected by the decisions (i.e., the transmitted symbols) selected at the encoder. We will consider the point-to-point feedback communication with the real-time constraint, that is, we will require the decoder to produce estimates of the current state of the source in real-time. We describe the model and the common information approach below.

Problem description

Consider the model of real-time communication with noiseless feedback, shown in Figure 1, that was investigated in [37]. The source $X_t \in \mathscr{X}$, $t = 1,2,\dots,T$ is a discrete-time, finite state Markov chain with a fixed transition probability matrix, $P^S(\cdot|\cdot)$, and a fixed distribution on the initial state. At each time instant, the encoder can send a symbol $Z_t \in \mathscr{Z}$ to the decoder over a memoryless noisy channel that is characterized by the transition probability matrix $P^C(\cdot|\cdot)$. The received symbol $Y_t \in \mathscr{Y}$ at the decoder is fed back noiselessly to the encoder. At the end of each time instant t, the decoder produces an estimate $\hat{X}_t \in \mathscr{X}$ of the current state of the Markov source. A distortion metric $\rho(X_t, \hat{X}_t)$ measures the accuracy of the decoder's estimate. The order of events at time t is the following (see Figure 2): (i) the state X_t of the Markov source is generated, (ii) the encoder transmits Z_t over the channel, (iii) the channel outputs Y_t to the receiver, (iv) Y_t is fed back to the encoder and (v) the decoder produces the estimate $\hat{X}_t$.

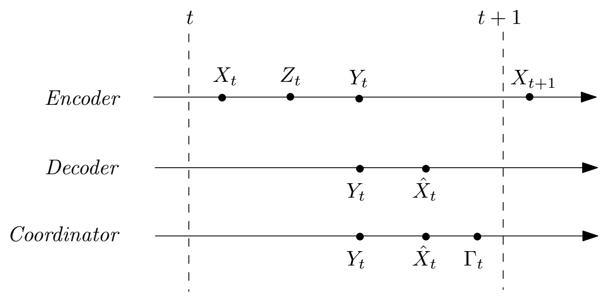

Fig. 2 Timing diagram for the real-time communication system

The encoder and the decoder are the two decision makers in this system. The encoder selects the symbol Z_t to be transmitted according to

$$Z_t = f_t(X_{1:t}, Y_{1:t-1}, Z_{1:t-1}),$$

where f_t is the encoder's decision rule at time t and $\mathbf{f} := (f_1, f_2, \dots, f_T)$ is the encoder's strategy. The decoder selects its estimate according to

$$\hat{X}_t = g_t(Y_{1:t}),$$

where g_t is the decoder's decision rule at time t and $\mathbf{g} := (g_1, g_2, \dots, g_T)$ is the decoder's strategy. The objective is to select $\mathbf{f}, \mathbf{g}$ so as to minimize

$$J(\mathbf{f}, \mathbf{g}) := \mathbb{E}\Big[\sum_{t=1}^{T} \rho(X_t, \hat{X}_t)\Big] \tag{4}$$

Preliminary result: Ignoring irrelevant information

As explained in Section 2.2, one way to extend the scope of the common information approach is to combine it with the person-by-person approach so as to identify and ignore irrelevant information at the decision makers. For the above example, a person-by-person approach was used in [37] to show that irrespective of the decoder's strategy, there is no loss of performance in restricting attention to encoding strategies of the form

$$Z_t = f_t(X_t, Y_{1:t-1}). \tag{5}$$

This result is a consequence of the Markovian nature of the source and the real-time nature of the distortion function. After restricting attention to encoding strategies of the form in (5), we proceed with the common information approach.

Applying the common information approach

We follow the five-step outline in Section 2.3.

1. ***Construct a coordinated system***
   After the time of reception of Y_t at the decoder, the information at the decoder is $I_t^d := \{Y_{1:t}\}$. Just before the time of transmission of Z_{t+1} by the encoder, the information at the encoder is $I_t^e := \{X_{t+1}, Y_{1:t}\}$. Between the time of reception of Y_t and the time of transmission of Z_{t+1}, the common information is then defined as
   $$C_t = I_t^d \cap I_t^e = \{Y_{1:t}\}.$$
   The local information at the encoder is $I_t^e \setminus C_t = X_{t+1}$ and the local information at the decoder is $I_t^d \setminus C_t = \emptyset$.
   The first step of the approach is to construct a coordinated system in which a coordinator observes the common information and selects the prescriptions for the encoder and decoder that map their respective local information to their decisions. Since the decoder has no local information, the coordinator's prescription is simply a prescribed decision $\hat{X}_t$ for the decoder. The prescription for encoder, Γ_t, is a mapping from $\mathscr{X}$ to $\mathscr{Z}$. For each possible value of encoder's local information x_{t+1}, the prescription Γ_t prescribes a decision $z_{t+1} = \Gamma_t(x_{t+1})$. The coordinator selects its prescriptions according to a coordination strategy $(\psi_1^e, \psi_1^d), \dots, (\psi_T^e, \psi_T^d)$ so that
   $$\Gamma_t = \psi_t^e(Y_{1:t}), \quad \hat{X}_t = \psi_t^d(Y_{1:t}). \tag{6}$$
   For this coordinated system, the source dynamics, the distortion metric, and the problem objective are the same as in the original system.
2. ***Formulate the coordinated system as a POMDP***
   The second step of the approach is to formulate the decision problem for the coordinator as a POMDP. In order to do so, we need to identify a state for input-output mapping for the coordinated system. As suggested in step 2 of the common information approach, the state for input-output mapping is a subset of the state of the original dynamic system (in this case, the source) and the local information at each decision maker. In this example, the state of the source X_t is sufficient for input-output mapping. In particular, define the state, action, and observation processes for the coordinator as:
   $$S_t := X_t, \quad A_t = (\Gamma_t, \hat{X}_t), \quad O_t := Y_t.$$
   It is easy to verify that
   $$\mathbb{P}(S_{t+1}, O_{t+1} \mid S_{1:t}, O_{1:t}, A_{1:t}) = \mathbb{P}(S_{t+1}, O_{t+1} \mid S_t, A_t) \tag{7}$$

Furthermore, for specific realization of the random variables involved, the right hand side of (7) can be written as

$$\mathbb{P}(x_{t+1}, y_{t+1} \mid x_t, \gamma_t, \hat{x}_t) = P^C(y_{t+1} \mid \gamma_t(x_{t+1}))P^S(x_{t+1}|x_t)$$

and the distortion cost can be written as

$$\rho(X_t, \hat{X}_t) = \tilde{\rho}(S_t, A_t),$$

with a suitably defined $\tilde{\rho}$. Thus, the coordinator's decision problem can be viewed as an instance of the POMDP model of Section 3.

3. ***Solve the resultant POMDP***
The third step of the common information approach is to solve the resultant POMDP at the coordinated system. Using Theorem 1 for the coordinated system, we get the following structural result and dynamic programming decomposition.

Theorem 2. *Let Θ_t be the conditional probability distribution of the state X_t at time t given the coordinator's observations $Y_{1:t}$ and actions $\Gamma_{1:t-1}, \hat{X}_{1:t-1}$, i.e.,*

$$\Theta_t(x) = \mathbb{P}(X_t = x | Y_{1:t}, \Gamma_{1:t-1}, \hat{X}_{1:t-1}), \quad x \in \mathscr{X}.$$

Then,

a. *If $\theta_t, \gamma_t, \hat{x}_t, y_{t+1}$ are the realizations of $\Theta_t, \Gamma_t, \hat{X}_t$ and Y_{t+1}, the realization of x^{th} element of the vector Θ_{t+1} is*

$$\begin{aligned}\theta_{t+1}(x) &= \frac{\sum_{x'} \theta_t(x')\mathbb{P}(X_{t+1} = x, Y_{t+1} = y_{t+1} | X_t = x', \Gamma_t = \gamma_t, \hat{X}_t = \hat{x}_t)}{\sum_{x'', \tilde{x}} \theta_t(x'')\mathbb{P}(X_{t+1} = \tilde{x}, Y_{t+1} = y_{t+1} | X_t = x'', \Gamma_t = \gamma_t, \hat{X}_t = \hat{x}_t)} \\ &= \frac{\sum_{x'} \theta_t(x')P^C(Y_{t+1} = y_{t+1} | Z_{t+1} = \gamma_t(x))P^S(X_{t+1} = x | X_t = x')}{\sum_{x'', \tilde{x}} \theta_t(x'')P^C(Y_{t+1} = y_{t+1} | Z_{t+1} = \gamma_t(\tilde{x}))P^S(X_{t+1} = \tilde{x} | X_t = x'')} \\ &=: \eta_t^x(\theta_t, \gamma_t, y_{t+1}) \end{aligned} \tag{8}$$

Therefore, we have that $\theta_{t+1} = \eta_t(\theta_t, \gamma_t, y_{t+1})$ where $\eta_t(\theta_t, \gamma_t, y_{t+1})$ is the vector of functions $(\eta_t^x(\theta_t, \gamma_t, y_{t+1}))_{x \in \mathscr{X}}$.

b. *There exists an optimal coordinator strategy of the form*

$$\Gamma_t = \psi_t^e(\Theta_t), \quad \hat{X}_t = \psi_t^d(\Theta_t).$$

Furthermore, the following dynamic program determines such an optimal strategy. Define:

$$V_T(\theta) := \min_{\hat{x}} \mathbb{E}[\rho(X_T, \hat{x}) \mid \Theta_T = \theta],$$

and for $t = T-1, T-2, \dots, 1$, recursively define

$$V_t(\theta) := \min_{\hat{x}, \gamma} \mathbb{E}\big[\rho(X_t, \hat{x}) + V_{t+1}(\eta_t(\theta, \gamma, Y_{t+1})) \bigm| \Theta_t = \theta, \Gamma_t = \gamma\big].$$

Then, for each time t and each realization of θ of Θ_t, the optimal prescriptions $\psi_t^e(\theta), \psi_t^d(\theta)$ are the minimizers in the definition of $V_t(\theta)$.

4. ***Show equivalence between the original system and the coordinated system***
The fourth step of the common information approach is to show the equivalence between the original system and the coordinated system. To show this equivalence, we show that any strategy for the coordinator can be implemented in the original system and vice versa.

Let ψ_t^e, ψ_t^d, $t = 1, 2, \dots, T$, be the coordinator's strategy of the form (6). Define the strategies for the encoder and decoder in the original system as follows:

$$f_{t+1}(\cdot, Y_{1:t}) := \psi_t^e(Y_{1:t}), \quad g_t(Y_{1:t}) := \psi_t^d(Y_{1:t}). \tag{9}$$

For each realization of the common information $y_{1:t}$ and each realization of the source state x_{t+1}, the encoder and decoder strategies as defined by (9) result in the same symbol z_{t+1} being transmitted and same estimate $\hat{x}_t$ being produced as in the coordinated system. Thus, the strategies for the encoder and the decoder defined by (9) will achieve the same expected cost as the coordinator's strategies ψ_t^e, ψ_t^d, $t = 1, 2, \dots, T$.

Conversely, given any strategies $\mathbf{f} = (f_1, \dots, f_T)$, $\mathbf{g} = (g_1, \dots, g_T)$, for the encoder and the decoder in the original system, we can construct strategies for the coordinator that achieve the same expected cost. Simply reverse (9) and define the coordinator's strategy as:

$$\psi_t^e(Y_{1:t}) := f_{t+1}(\cdot, Y_{1:t}), \quad \psi_t^d(Y_{1:t}) := g_t(Y_{1:t}). \tag{10}$$

Then, for each realization of the common information $y_{1:t}$ and each realization of the source state x_{t+1}, the coordinator strategies as defined by (10) will result in the same symbol z_{t+1} being transmitted and same estimate $\hat{x}_t$ being produced as in the original system. Thus, the coordinator's strategies defined by (10) will achieve the same expected cost as the strategies $(\mathbf{f}, \mathbf{g})$.

Consequently, the original system is equivalent to the coordinated system. The equivalence between the two systems implies that translating a globally optimal strategy for the coordinator to the original system (using (9)) will give globally optimal strategies for the original system.

5. ***Translate the solution of the coordinated system to the original system***
The last step of the approach is to translate the result of Theorem 2 to the original system, which gives the following:

Theorem 3. *For the real-time communication problem formulated above, there exist globally optimal encoding and decoding strategies of the form*

$$Z_{t+1} = f_{t+1}^*(X_{t+1}, \Theta_t), \quad \hat{X}_t = g_t^*(\Theta_t),$$

where $\Theta_t = \mathbb{P}^{f_{1:t}}(X_t | Y_{1:t})$ and Θ_t evolves according to equation (8). *Furthermore, if $(\psi_t^{e*}, \psi_t^{d*})$ is an optimal coordination strategy (i.e., the solution of dynamic* *program of Theorem 2), then the optimal encoding and decoding strategies are given by*

$$f^*_{t+1}(\cdot,\Theta_t) = \psi^{e*}_t(\Theta_t), \quad g^*_t(\Theta_t) = \psi^{d*}_t(\Theta_t).$$

The result of Theorem 3 is equivalent to the result of [37, Theorem 2 and (4.4)].

4.2 Networked control systems

In networked control systems, the controller relies on a communication network to gather information from the sensors at the plant and/or to send control actions to actuators at the plant. Communication related imperfections such as rate limited channels, delays and noise can affect the performance of the control system. A key question in such systems is whether the communication system and the control system can be jointly designed for improved performance. We consider a basic model of such a system where a sensor needs to communicate with the controller over a rate limited channel.

Problem description

The structure of the problem above bears considerable similarity to the real-time communication problem formulated in Section 4.1.

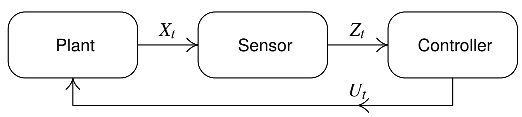

Fig. 3 A networked control system with communication over a rate-limited channel

Consider the model of networked control system with communication over a rate-limited channel shown in Figure 3. A related problem was first considered in [36]. The state of the plant $X_t \in \mathscr{X}$, $t = 1,2,\ldots,T$ is a discrete-time, finite state controlled Markov chain that evolves according to the equation

$$X_{t+1} = h_t(X_t,U_t,W_t),$$

where U_t is the control action applied by the controller and W_t is the random noise. The sensor observes the state of the plant and sends a symbol $Z_t \in \mathscr{Z}$ to the controller. We assume that $\mathscr{Z}$ is finite; thus, the communication link between the sensor and the controller is a rate limited communication link. At the end of each time instant t, the controller selects a control action U_t that is applied to the system. The order of events at time instant t is the following: (i) the state X_t is generated, (ii) the sensor transmits Z_t over the channel, (iii) the controller generates U_t.

The sensor selects the symbol to be transmitted Z_t according to

$$Z_t = f_t(X_{1:t}, Z_{1:t-1}),$$

the controller selects its action according to

$$U_t = g_t(Z_{1:t}).$$

At each time an instantaneous cost $\ell(X_t, U_t)$ is incurred. The objective is to select $\mathbf{f} = (f_1, \dots, f_T), \mathbf{g} = (g_1, \dots, g_T)$ so as to minimize

$$J(\mathbf{f}, \mathbf{g}) := \mathbb{E}\Big[\sum_{t=1}^{T} \ell(X_t, U_t)\Big] \tag{11}$$

Preliminary result: Ignoring irrelevant information

The structure of the problem above bears considerable similarity to the real-time communication problem formulated in Section 4.1. As in that example, using a person-by-person approach, we can show that irrespective of the controller's strategy, there is no loss of performance in restricting attention to sensor strategies of the form

$$Z_t = f_t(X_t, Z_{1:t-1}).$$

This result is analogous to structural result of encoder's strategies in (5) and is derived using similar arguments. See [36] for a proof of a similar result for a slightly different channel model.

Applying the common information approach

We follow the five-step outline in Section 2.3.

1. ***Construct a coordinated system***
   Between the time of reception of Z_t at the controller and the time of transmission of Z_{t+1} by the sensor, the information at the controller is $I_t^c := \{Z_{1:t}\}$, and the information at the sensor is $I_t^s := \{X_{t+1}, Z_{1:t}\}$. The common information is then defined as
   $$C_t = I_t^c \cap I_t^s = \{Z_{1:t}\}.$$
   The local information at the sensor is $I_t^s \setminus C_t = X_{t+1}$ and the local information at the controller is $I_t^c \setminus C_t = \emptyset$.
   The first step of the approach is to construct a coordinated system in which a coordinator observes the common information and selects the prescriptions for the sensor and the controller that map their respective local information to their decisions. Since the controller has no local information, the coordinator's prescription is simply a prescribed action U_t for the controller. The prescription

for the sensor, Γ_t, is a mapping from $\mathscr{X}$ to $\mathscr{Z}$. For each possible value of sensor's local information x_{t+1}, the prescription Γ_t prescribes a decision $z_{t+1} = \Gamma_t(x_{t+1})$. The coordinator selects its prescriptions according to a coordination strategy $(\psi_1^s, \psi_1^c) \dots, (\psi_T^s, \psi_T^c)$ so that

$$\Gamma_t = \psi_t^s(Z_{1:t}), \quad U_t = \psi_t^c(Z_{1:t}). \tag{12}$$

For this coordinated system, the plant dynamics, the loss function and the problem objective are the same as in the original system.

2. ***Formulate the coordinated system as a POMDP***
The second step of the approach is to formulate the decision problem for the coordinator as a POMDP. As in Section 4.1, define the state, action, and observation processes for the POMDP as

$$S_t := X_t, \quad A_t := (\Gamma_t, U_t), \quad O_t := Z_t.$$

It is easy to verify that

$$\mathbb{P}(S_{t+1}, O_{t+1} \mid S_{1:t}, O_{1:t}, A_{1:t}) = \mathbb{P}(S_{t+1}, O_{t+1} \mid S_t, A_t)$$

and the instantaneous cost can be written as

$$\ell(X_t, U_t) = \tilde{\ell}(S_t, A_t),$$

with a suitably defined $\tilde{\ell}$. Thus, the coordinator's decision problem can be viewed as an instance of the POMDP model of Section 3.

3. ***Solve the resultant POMDP***
The third step of the common information approach is to solve the resultant POMDP at the coordinated system. Using Theorem 1 for the coordinated system, we get the following structural result and dynamic programming decomposition.

Theorem 4. *Let Θ_t be the conditional probability distribution of the state X_t at time t given the coordinator's observations $Z_{1:t}$ and actions $\Gamma_{1:t-1}, \hat{X}_{1:t-1}$, i.e.,*

$$\Theta_t(x) = \mathbb{P}(X_t = x \mid Z_{1:t}, \Gamma_{1:t-1}, U_{1:t-1}), \quad x \in \mathscr{X}.$$

Then,

a. The realization θ_t of Θ_t updates according to a non-linear filtering equation,

$$\theta_{t+1} = \eta_t(\theta_t, \gamma_t, u_t, z_{t+1}).$$

b. There exists an optimal decision strategy of the form

$$\Gamma_t = \psi_t^s(\Theta_t), \quad \hat{X}_t = \psi_t^c(\Theta_t).$$

Furthermore, the following dynamic program determines such an optimal strategy: Define

$$V_T(\theta) := \min_u \mathbb{E}[\ell(X_T, u) \mid \Theta_T = \theta],$$

and for $t = T-1, T-2, \dots, 1$, recursively define

$$V_t(\theta) := \min_{u,\gamma} \mathbb{E}\big[\ell(X_t, u) + V_{t+1}(\eta_t(\theta, \gamma, Z_{t+1})) \bigm| \Theta_t = \theta, \Gamma_t = \gamma\big].$$

Then, for each time t and each realization of θ of Θ_t, the optimal prescriptions $(\psi_t^s(\theta), \psi_t^c(\theta))$ are the minimizers in the definition of $V_t(\theta)$.

4. ***Show equivalence between the original system and the coordinated system***
   The fourth step of the common information approach is to show the equivalence between the original system and the coordinated system. This equivalence follows from the same argument used in Section 4.1. In particular, the optimal strategy for the coordinator can be translated to optimal strategies for the sensor and the controller in the original system.
5. ***Translate the solution of the coordinated system to the original system***
   The last step of the approach is to translate the result of Step 3 to the original system, which gives the following result.

   Theorem 5. *For the networked control problem formulated above, there exist globally optimal strategies for the sensor and the controller of the form*

   $$Z_{t+1} = f^*_{t+1}(X_{t+1}, \Theta_t), \quad U_t = g^*_t(\Theta_t)$$

   where $\Theta_t = \mathbb{P}^{f_{1:t}}(X_t | Z_{1:t})$. Furthermore, if ψ_t^{s*}, ψ_t^{c*} is an optimal coordination strategy (i.e., a solution of the dynamic program of Theorem 4), then the optimal sensor and controller strategies are given by

   $$f^*_{t+1}(\cdot, \Theta_t) = \psi_t^{s*}(\Theta_t), \quad g^*_t(\Theta_t) = \psi_t^{c*}(\Theta_t).$$

The result of Theorem 5 is equivalent to the result of [36, Theorem 3.2] when specialized to the above model.

4.3 Paging and registration in cellular networks

In cellular networks, the network needs to keep track of the location of a mobile station. This tracking may be done in two ways: the network may either page the mobile station, or the mobile station may register its location with the network. Both operations have an associated cost. The problem of finding optimal paging and registration strategies can be viewed as a team problem with the mobile station and the network operator as the decision-makers.

Problem description

Consider a cellular network consisting of one mobile station (MS) and one network operator (N). The mobile station's motion is described by a discrete-time, finite state Markov chain $X_t \in \mathscr{X}$, $t = 1,2,\ldots$, with known transition probability matrix. Each state represent a cell of the cellular network. At each time instant t, the MS may or may not register with the network. The cost of registration is r. If the MS registers with the network at time t, the network learns its location X_t. At each time t, the network may receive an exogenous paging request to seek MS's location. The exogenous paging request is an i.i.d. binary process which is independent of the motion of MS. The probability of a paging request at any time t is p. If a paging request arrives, the network operator must decide an order in which the cells are to be searched in order to locate the MS. We assume that if the MS is present in the cell being searched, the network successfully finds it. Further, we assume that the time it takes to search one cell is negligible compared to the time step of MS's motion, so that the paging request is completed within one time step. The cost of paging depends on the number of cells that are searched before MS is located. This model was investigated in [5].

The order of events at time instant t is the following: (i) The MS moves to location X_t according to a probability distribution that depends on its previous location; (ii) a paging request arrives with probability p; (iii) if a paging request arrives, the network operator must decide an order in which the cells are to be searched; (iv) if no paging request is made, the MS decides whether or not to register its location with the network.

Define a random variable Y_t as

$$Y_t = \begin{cases} X_{t-1} & \text{if the network learns MS location either by a paging} \\ & \text{request or by MS registration at time } t-1 \\ \varepsilon & \text{otherwise,} \end{cases}$$

Let $\sigma(\mathscr{X})$ denote the set of all permutations of the locations in $\mathscr{X}$. At the beginning of time t, if the network received a paging request, it selects $U_t^N \in \sigma(\mathscr{X})$ according to

$$U_t^N = g_t(Y_{1:t}).$$

If a paging request does not arrive, the MS makes a decision $U_t^{MS} \in \{0,1\}$ according to

$$U_t^{MS} = f_t(X_{1:t}, Y_{1:t}),$$

where $U_t^{MS} = 1$ represents a decision to register and $U_t^{MS} = 0$ represents a decision to not register with the network. The collection of functions $\mathbf{f} := (f_1, f_2, \ldots, f_T)$ and $\mathbf{g} := (g_1, g_2, \ldots, g_T)$ are the strategies of the MS and the network respectively. The objective is to select $\mathbf{f}, \mathbf{g}$ so as to minimize

$$J(\mathbf{f},\mathbf{g}) := \mathbb{E}\Big[\sum_{t=1}^{T} (1-p) r U_t^{MS} + pk\tau(X_t, U_t^N)\Big], \tag{13}$$

where p is the probability of paging request arrival, r is the cost of registration by MS, k is the cost of searching one cell, $\tau(x,u^N)$ is the position of x in the permutation specified by u^N and, therefore, $\tau(X_t,U_t^N)$ is the number of cells searched by the network before MS is located at time t.

Preliminary result: Ignoring irrelevant information

For the above example, we may use an argument similar to the argument based on the person-by-person approach used in Section 4.1 to show that irrespective of the strategy of the network, there is no loss of performance in restricting attention to the strategies of the MS of the form

$$U_t^{MS} = f_t(X_t, Y_{1:t}). \tag{14}$$

This result is a consequence of the Markovian nature of the MS motion and the fact that a paging request is completed within one time step. After restricting attention to MS strategies of the form in (14), we proceed with the common information approach.

Applying the common information approach

We follow the five-step outline in Section 2.3.

1. ***Construct a coordinated system***
   At the beginning of time t, the information at the network is $I_t^N := \{Y_{1:t}\}$, and the information at the MS is $I_t^{MS} := \{X_t, Y_{1:t}\}$. The common information at time t is

   $$C_t = I_t^N \cap I_t^{MS} = \{Y_{1:t}\}.$$

   The local information at the network is $I_t^N \setminus C_t = \emptyset$ and the local information at the MS is $I_t^{MS} \setminus C_t = X_t$.

   The first step of the approach is to construct a coordinated system in which a coordinator observes the common information and selects the prescriptions for the network and the MS. Since the network has no local information, the coordinator's prescription is simply a prescribed order U_t^N in which to search the cells if a paging request arrives. The prescription Γ_t for the MS is a mapping from $\mathscr{X}$ to $\{0,1\}$. If a paging request does not arrive, the prescription Γ_t prescribes a registration decision $u_t^{MS} = \Gamma_t(x_t)$ for each possible value of MS location x_t. The coordinator selects its prescriptions according to a coordination strategy $(\psi_1^{MS}, \psi_1^N), \dots, (\psi_T^{MS}, \psi_T^N)$ so that

   $$\Gamma_t = \psi_t^{MS}(Y_{1:t}) \quad U_t^N = \psi_t^N(Y_{1:t}).$$

For this coordinated system, the MS motion dynamics, the cost function and the problem objective are the same as in the original system.

2. ***Formulate the coordinated system as a POMDP***
The second step of the approach is to formulate the decision problem for the coordinator as a POMDP. In order to do so, we define the state, action and observation processes of the POMDP as:

$$S_t := X_t, \quad A_t = (\Gamma_t, U_t^N), \quad O_t := Y_t.$$

It is easy to verify that

$$\mathbb{P}(S_{t+1}, O_{t+1} \mid S_{1:t}, O_{1:t}, A_{1:t}) = \mathbb{P}(S_{t+1}, O_{t+1} \mid S_t, A_t).$$

and that the instantaneous cost $(1-p)rU_t^{MS} + pk\tau(X_t, U_t^N)$ can be written as a function of S_t and A_t. Thus, the coordinator's decision problem can be viewed as an instance of the POMDP model of Section 3.

3. ***Solve the resultant POMDP***
The third step of the common information approach is to solve the resultant POMDP at the coordinated system. Using Theorem 1 for the coordinated system, we get the following structural result and dynamic programming decomposition.

Theorem 6. *Let Θ_t be the conditional probability distribution of the state X_t at time t given the coordinator's observations $Y_{1:t}$ and actions $\Gamma_{1:t-1}$, i.e.,*

$$\Theta_t(x) := \mathbb{P}(X_t = x \mid Y_{1:t}, \Gamma_{1:t-1}), \quad x \in \mathscr{X}.$$

Then,

a. The realization θ_t of Θ_t updates according to a non-linear filtering equation,

$$\theta_{t+1} = \eta_t(\theta_t, \gamma_t, y_{t+1}).$$

b. There exists an optimal decision strategy of the form

$$\Gamma_t = \psi_t^{MS}(\Theta_t), \quad U_t^N = \psi_t^N(\Theta_t).$$

Furthermore, the following dynamic program determines such an optimal strategy: Define

$$V_T(\theta) := \min_{\gamma, u^N} \mathbb{E}[(1-p)r\Gamma_T(X_T) + pk\tau(X_T, U_T^N) \mid \Theta_T = \theta, \Gamma_T = \gamma, U_T^n = u^N],$$

and for $t = T-1, T-2, \dots, 1$, recursively define

$$\begin{aligned} V_t(\theta) := \min_{\gamma, u^N} \mathbb{E}\big[&(1-p)r\Gamma_t(X_t) + pk\tau(X_t, U_t^N) \\ &+ V_{t+1}(\eta_t(\theta, \gamma, Y_{t+1})) \,\big|\, \Theta_t = \theta, \Gamma_t = \gamma, U_t^N = u^N\big]. \end{aligned}$$

Then, for each time t and each realization of θ of Θ_t, the optimal prescriptions $(\psi_t^{MS}(\theta), \psi_t^{N}(\theta))$ are the minimizers in the definition of $V_t(\theta)$.

4. ***Show equivalence between the original system and the coordinated system***
   The fourth step of the common information approach is to show the equivalence between the original system and the coordinated system. This equivalence follows from the same argument used in Section 4.1. In particular, the optimal strategy for the coordinator can be translated to optimal strategies for the MS and the network in the original system.

5. ***Translate the solution of the coordinated system to the original system***
   The last step of the approach is to translate the result of Step 3 to the original system, which gives the following result.

   Theorem 7. *For the paging and registration problem formulated above, there exist globally optimal strategies of the form*

   $$U_t^N = g_t^*(\Theta_t), \quad U_t^{MS} = f_t^*(X_t, \Theta_t),$$

   where $\Theta_t = \mathbb{P}^{f_{1:t}}(X_t|Y_{1:t})$. Furthermore, if $(\psi_t^{MS*}, \psi_t^{N*})$ is an optimal coordination strategy (i.e., the solution of dynamic program of Theorem 6), then the optimal paging and registration strategies $(\mathbf{f}^*, \mathbf{g}^*)$ are given by

   $$f_t^*(\cdot, \Theta_t) = \psi_t^{MS*}(\Theta_t), \quad g_t^*(\Theta_t) = \psi_t^{N*}(\Theta_t).$$

The result of Theorem 7 is equivalent to the result of [5, Section III-C]. The dynamic program was using in [5] to identify further structural properties of the optimal paging and registration strategies when the motion of the MS follows a symmetric random walk.

4.4 Multiaccess broadcast systems

In a multiaccess broadcast system, multiple users communicate to a common receiver over a broadcast medium. If more than one user transmits at a time, the transmissions "collide" and the receiver cannot decode the packets due to interference. Such systems can be viewed as team problems in which all users must cooperate to maximize system throughput. In this section we consider a specific variation of a two-user multiaccess broadcast system.

Problem description

Consider a two-user multiaccess broadcast system. At time t, $W_t^i \in \{0,1\}$ packets arrive at each user according to independent Bernoulli processes with $\mathbb{P}(W_t^i = 1) = p^i$, $i = 1,2$. Each user may store only $X_t^i \in \{0,1\}$ packets in a buffer. If a packet arrives when the user-buffer is full, the packet is dropped.

Both users may transmit $U_t^i \in \{0,1\}$ packets over a shared broadcast medium. If only one user transmits at a time, the transmission is successful and the transmitted packet is removed from the queue. If both users transmit simultaneously, packets "collide" and remain in the queue. Thus, the state update for user 1 is given by

$$X_{t+1}^1 = \max(X_t^1 - U_t^1 \cdot (1 - U_t^2) + W_t^1, 1).$$

The state update rule for user 2 is symmetric dual of the above.

Due to the broadcast nature of the communication medium, each user knows the control action of the other user after one-step delay. Thus, each user chooses a transmission decision as

$$U_t = g_t^i(X_{1:t}^i, \mathbf{U}_{1:t-1})$$

where $\mathbf{U}_t = (U_t^1, U_t^2)$. A user can transmit only if it has a packet, thus only actions $U_t^i \le X_t^i$ are feasible.

Instead of costs, it is more natural to work with rewards in this example. The objective is to maximize throughput, or the number of successful packet transmissions. Thus, the per unit reward is $r(\mathbf{X}, \mathbf{U}) = U^1 \oplus U^2$, where $\oplus$ means binary XOR. The objective is to maximize

$$J(\mathbf{g}) = \mathbb{E}\Big[\sum_{t=1}^{T} U_t^1 \oplus U_t^2\Big]$$

which corresponds to the total throughput.

When the arrival rates at both users are the same ($p^1 = p^2$), the above model corresponds to the two-user multiaccess broadcast system considered in [6,12,13,22]. Slight variations of the above model were considered in [25, 32].

Preliminary result: Ignoring irrelevant information

As suggested in Section 2.2, we may use the person-by-person approach to identify and ignore irrelevant information at the decision makers before applying the common information approach. For the above example, a person-by-person approach was used in [12] to show that there is no loss of performance in restricting attention to the transmission strategies of the form

$$U_t^i = g_t^i(X_t^i, \mathbf{U}_{1:t-1}). \tag{15}$$

This result is a consequence of the fact that irrespective of the transmission strategies, the processes $\{X_t^1\}$ and $\{X_t^2\}$ are conditionally independent given $\mathbf{U}_{1:t-1}$. After restricting attention to transmission strategies of the form (15), we proceed with the common information approach.

Applying the common information approach

We follow the five-step outline in Section 2.3.

1. ***Construct a coordinated system***
   At the beginning of time t, the information at user i is $I_t^i = \{X_t^i, \mathbf{U}_{1:t-1}\}$. Thus, the common information at time t is
   $$C_t = I_t^1 \cap I_t^2 = \{\mathbf{U}_{1:t-1}\}.$$
   The local information at user i is $I_t^i \setminus C_t = X_t^i$.
   The first step of the common information approach is to construct a coordinated system in which a coordinator observes the common information and selects the prescriptions that map each user's local information to its actions. The prescription Γ_t^i for user i is a mapping from $\mathscr{X}^i$ to $\mathscr{U}^i$. For each realization x_t^i of the local information, the prescription Γ_t^i prescribes a decision $u_t^i = \Gamma_t^i(x_t^i)$. Since, $\Gamma_t^i(0) = 0$, the prescription Γ_t^i is completely specified by $\Gamma_t^i(1)$, which we denote by $Y_t \in \{0,1\}$. Then, the control action is $U_t^i = X_t^i Y_t^i$. The coordinator selects its prescriptions according to a coordination strategy $(\psi_1, \dots \psi_T)$, so that
   $$(Y_t^1, Y_t^2) = \psi_t(\mathbf{U}_{1:t-1}).$$
   For this coordinated system, the queue dynamics, the reward function, and the problem objective are the same as the original system.
2. ***Formulate the coordinated system as a POMDP***
   The second step of the approach is to formulate the decision problem for the coordinator as a POMDP. Define the state, action, and observation processes for the POMDP as
   $$S_t := (X_t^1, X_t^2), \quad A_t := (Y_t^1, Y_t^2), \quad O_t := (U_{t-1}^1, U_{t-1}^2).$$
   It is easy to verify that
   $$\mathbb{P}(S_{t+1}, O_{t+1} \mid S_{1:t}, O_{1:t}, A_{1:t}) = \mathbb{P}(S_{t+1}, O_{t+1} \mid S_t, A_t)$$
   and the instantaneous reward function can be written as
   $$r(X_t, U_t) = U_t^1 \oplus U_t^2 = X^1 Y^1 \oplus X^2 Y^2 =: \tilde{r}(S_t, A_t).$$
   Thus, the coordinator's decision problem can be viewed as an instance of the POMDP model of Section 3.
3. ***Solve the resultant POMDP***
   The third step of the common information approach is to solve the resultant POMDP at the coordinated system. Using Theorem 1 for the coordinated system, we get the following structural result and dynamic programming decomposition.

Theorem 8. *Let Θ_t be the conditional probability distribution of the state $\mathbf{X}_t = (X_t^1, X_t^2)$ given the coordinator's observations $\mathbf{U}_{1:t-1}$ and actions $\mathbf{Y}_{1:t-1}$, i.e.,*

$$\Theta_t(\mathbf{x}) = \mathbb{P}(\mathbf{X}_t = \mathbf{x} \mid \mathbf{U}_{1:t-1}, \mathbf{Y}_{1:t-1}).$$

Then,

a. The realization θ_t of Θ_t updates according to a non-linear filtering equation,

$$\theta_{t+1} = \eta_t(\theta_t, y_t^1, y_t^2, u_t^1, u_t^2).$$

b. There exists an optimal decision strategy of the form

$$(Y_t^1, Y_t^2) = \psi_t(\Theta_t).$$

Furthermore, the following dynamic program determines such an optimal strategy: Define

$$V_T(\theta) := \min_{(y^1, y^2)} \mathbb{E}[X_T^1 y^1 \oplus X_T^2 y^2 \mid \Theta_T = \theta],$$

and for $t = T-1, T-2, \dots, 1$, recursively define

$$V_t(\theta) := \min_{(y^1, y^2)} \mathbb{E}[X_t^1 y^1 \oplus X_t^2 y^2 + V_{t+1}(\psi_t(\theta, y^1, y^2, X_t^1 y^1, X_t^2 y^2) \mid \Theta_t = \theta]$$

Then, for each time t and each realization θ of Θ_t, the optimal prescription $\psi_t(\theta)$ is the minimizer in the definition of $V_t(\theta)$.

4. ***Show equivalence between the original system and the coordinated system***
The fourth step of the common information approach is to show the equivalence between the original system and the coordinated system. This equivalence follows from the same argument used in Section 4.1. In particular, the optimal strategy for the coordinator can be translated to optimal transmission strategies in the original system.

5. ***Translate the solution of the coordinated system to the original system***
The last step of the approach is to translate the result of Step 3 to the original system, which gives the following result.

Theorem 9. *For the two-user multiaccess broadcast system formulated above, there exists optimal transmission strategies of the form*

$$U_t^i = X_t^i \cdot \psi_t^i(\Theta_t)$$

where $\Theta_t = \mathbb{P}^{\psi_{1:t}}(\mathbf{X}_t | \mathbf{U}_{1:t-1})$. Furthermore, an optimal $\psi_t^* = (\psi^{*,1}, \psi^{*,2})$ is given by the solution of the dynamic program in Theorem 8.

The result of Theorem 9 is equivalent to the result of [12, Proposition 14]. The dynamic program (extended to infinite horizon average reward setup) was used in [12] to explicitly characterize the optimal transmission strategies when $p_1 = p_2$.

5 Application to delayed sharing information structures

In this section, we present the result of [16], in which we use the common information approach to solve a long standing open problem associated with delayed sharing information structures.

In a decentralized control system with delayed sharing information structure, the controllers sharing their observations and control actions with each other after a fixed delay. The delayed sharing information structure is a link between classical information structure, which may be viewed as a degenerate decentralized control system in which controllers instantaneous sharing their observations and control actions, and a completely decentralized information structure, where there is no "lateral" sharing of information.

This information structure was proposed by Witsenhausen in a seminal paper [39] where he conjectured the structure of the globally optimal control strategies. Later Varaiya and Walrand [32] showed that Witsenhausen's assertion is true when the delay in the sharing of information is one (called one-step delayed sharing), but false for larger sharing delay. See [16] for a more detailed history of the problem.

Problem description

The delayed-sharing information structure consists of n controllers. Let X_t denote the state of the system, Y_t^i denote the observations of controller i, and U_t^i denote the control action of controller i. The system evolves according to

$$X_{t+1} = f_t^i(X_t, \mathbf{U}_t, W_t^0)$$

where $\mathbf{U}_t = (U_t^1, \dots, U_t^n)$ and $\{W_t^0\}_{t=1}^T$ is an i.i.d. noise process that is independent of the initial state X_1. The observations of the controllers are given by

$$Y_t^i = h_t^i(X_t, W_t^i), \quad i = 1, \dots, n$$

where $\{W_t^i\}_{t=1}^T$, $i = 1, \dots, n$, are i.i.d. noise process that are independent of each other and also independent of $\{W_t^0\}_{t=1}^T$ and X_1.

The controllers share their observations and control actions with each other after a k-step delay. Thus, the control actions are selected as follows:

$$U_t^i = g_t^i(\mathbf{Y}_{1:t-k}, \mathbf{U}_{1:t-k}, Y_{t-k+1:t}^i, U_{t-k+1:t-1}^i)$$

where $\mathbf{Y}_t = (Y_t^1, \dots, Y_t^n)$.

The instantaneous loss function is given by $\ell(X_t, \mathbf{U}_t)$.

For simplicity, assume that all system variables are finite valued and X_t, Y_t^i, U_t^i, W_t^i take values in time-homogeneous finite sets $\mathscr{X}$, $\mathscr{Y}^i$, $\mathscr{U}^i$, and $\mathscr{W}^i$, respectively.

The objective is to choose control strategies $\mathbf{g}^{1:n}$ where $\mathbf{g}^i = (g_1^i, \dots, g_T^i)$, to minimize the expected total loss

$$J(\mathbf{g}^{1:n}) = \mathbb{E}^{(\mathbf{g}^{1:n})}\Big[\sum_{t=1}^{T} \ell(X_t, \mathbf{U}_t)\Big]$$

Applying the common information approach

1. ***Construct a coordinated system***
   The first step of the approach is to construct the coordinated system. At the beginning of time t, the information at controller i is $I_t^i = (\mathbf{Y}_{1:t-k}, \mathbf{U}_{1:t-k}, Y^i_{t-k+1:t}, U^i_{t-k+1:t-1})$. Thus, the common information at all controllers is $C_t = \bigcap_{i=1}^n I_t^i = (\mathbf{Y}_{1:t-k}, \mathbf{U}_{1:t-k})$ and the local information at controller i is $L_t^i = (Y^i_{t-k+1:t}, U^i_{t-k+1:t-1})$.
   Consider a coordinated system where the coordinator observes the common information and selects prescriptions $(\Gamma_t^1, \dots, \Gamma_t^n)$ for the controllers where Γ_t^i maps the local information L_t^i to control action U_t^i, i.e., for each possible value l_t^i of the local information L_t^i, the prescription Γ_t^i prescribes a control action $u_t^i = \Gamma_t^i(l_t^i)$. For convenience, define $Z_t = (\mathbf{Y}_{t-k}, \mathbf{U}_{t-k})$ so that $C_t = Z_{1:t}$. The coordinator selects its prescriptions according to a coordination law ψ_t so that

   $$(\Gamma_t^1, \dots, \Gamma_t^n) = \psi_t(C_t) = \psi_t(Z_{1:t}).$$

   For this coordinated system, the source dynamics, the loss function, and the problem objective are the same as the original problem.
2. ***Formulate the coordinated system as a POMDP***
   The second step of the approach is to formulate the coordinated system as a POMDP. In order to do so, define the state, observation, and action processes of the POMDP as

   $$S_t = (X_t, \mathbf{L}_t), \quad O_t = Z_t, \quad A_t = (\Gamma_t^1, \dots, \Gamma_t^n).$$

   It is easy to verify that

   $$\mathbb{P}(S_{t+1}, O_{t+1} \mid S_{1:t}, A_{1:t}) = \mathbb{P}(S_{t+1}, O_{t+1} \mid S_t, A_t)$$

   and that the instantaneous loss $\ell(X_t, \mathbf{U}_t) = \tilde{\ell}(S_t, A_t)$ for an appropriately defined $\tilde{\ell}$. Hence, the decision problem at the coordinator is a POMDP.
3. ***Solve the resultant POMDP***
   The third step of the common information approach is to solve the resultant POMDP at the coordinated system. Using Theorem 1 for coordinated system defined above, we get the following structural result and dynamic programming decomposition.

   Theorem 10. *Let Θ_t be the conditional probability distribution of the state S_t given the coordinator's history of observations C_t and actions $(\Gamma^1_{1:t-1}, \dots, \Gamma^n_{1:t-1})$, i.e., for any realization s of S_t*

$$\Theta_t(s) = \mathbb{P}(S_t = s \mid C_t, \Gamma^1_{1:t-1}, \dots, \Gamma^n_{1:t-1}).$$

Then

a. *The realization θ_t of Θ_t updates according to a non-linear filtering equation,*

$$\theta_{t+1} = \eta_t(\theta_t, z_{t+1}, \gamma^1_t, \dots, \gamma^n_t)$$

where $z_{t+1} = (\mathbf{y}_{t-k+1}, \mathbf{u}_{t-k+1})$.

b. *There exists an optimal coordination strategy of the form*

$$(\Gamma^1_t, \dots, \Gamma^n_t) = \psi_t(\Theta_t).$$

Furthermore, the following dynamic program determines such an optimal strategy (recall that $U^i_t = \Gamma^i_t(L^i_t)$): Define

$$V_T(\theta) = \min_{(\gamma^1_T, \dots, \gamma^n_T)} \mathbb{E}\big[\ell(X_T, \mathbf{U}_T) \bigm| \Theta_T = \theta, \Gamma^1_T = \gamma^1_T, \dots, \Gamma^n_T = \gamma^n_T\big]$$

and for $t = T-1, T-2, \dots, 1$, recursively define

$$V_t(\theta) = \min_{(\gamma^1_t, \dots, \gamma^n_t)} \mathbb{E}\big[\ell(X_t, \mathbf{U}_t) + V_{t+1}(\eta_t(\theta, Z_{t+1}, \gamma^1_t, \dots, \gamma^n_t)) \bigm| \Theta_t = \theta, \Gamma^1_t = \gamma^1_t, \dots, \Gamma^n_t = \gamma^n_t\big]. \quad (16)$$

Then, for each time t and each realization θ to Θ_t, the optimal prescription $(\gamma^1_t, \dots, \gamma^n_t)$ is the minimizer in the definition of $V_t(\theta)$.

4. ***Show equivalence between the original system and the coordinated system***
The fourth step of the common information approach is to show the equivalence between the original system and the coordinated system. This equivalence follows from the same argument used in Section 4. As a consequence, we can translate an optimal coordination strategy for the coordinated system to an optimal control strategy for the original system.

5. ***Translate the solution of the coordinated system to the original system***
The last step of the approach is to translate the results of step 3 to the original system, which gives the following result.

Theorem 11. *For the delayed sharing information structure, there exists optimal control strategies of the form*

$$U^i_t = g^i_t(L^i_t, \Theta_t)$$

where $\Theta_t = \mathbb{P}(X_t, \mathbf{L}_t \mid C_t)$. Furthermore, if ψ^* is the optimal coordination strategy (i.e., the solution to the dynamic program of Theorem 10), and $\psi^{*,i}$ denote its i-th component, then the optimal control strategy $\mathbf{g}^*_{1:T}$ is given by

$$g^{*,i}_t(\cdot, \theta) = \psi^{*,i}_t(\theta).$$

6 Conclusion

In centralized stochastic control, the controller's belief on the current state of the system plays a fundamental role for predicting future costs. If the control strategy for the future is fixed as a function of future beliefs, then the current belief is a sufficient statistic for future costs under any choice of current action. Hence, the optimal action at any time t is only a function of the controller's belief on the system state at time t. In decentralized problems, where there are many controllers with different information interacting with each other, the controllers' belief on the system state and their predictions of future costs are not expected to be consistent. Furthermore, since the costs depend both on system state as well as other controllers' actions, any controller's prediction of future costs must involve a belief on system state along with a prediction of other controllers' actions. The above discussion describes the difficulties that arise if one attempts to use a controller's belief on the system state for decision-making in decentralized systems.

The common information approach attempts to address the above difficulties based on two key observations: (1) Beliefs based on common information are consistent among all controllers and can serve as a consistent sufficient statistic. (2) Even though controllers cannot accurately predict each other's control actions, for any realization of common information they can know the exact mapping used by each controller to map its local information to its control actions. These observations motivate the creation of a coordinated system with a fictitious coordinator which observes only the common information, forms its beliefs based on the common information, selects prescriptions (described in Sections IV and V) and has the same objective as the original decentralized stochastic control problem. If the system model is such that the data available at the coordinator —the common information— is increasing with time, then the decision problem at the coordinator is centralized stochastic control problem. This centralized problem is equivalent to the original decentralized stochastic control problem. This equivalence allows the use of results obtained from centralized stochastic control theory to obtain: (i) qualitative properties of optimal strategies for the controllers in the original decentralized stochastic control problem, and (ii) a dynamic program for determining optimal strategies for all controllers. The fictitious coordinator is invented purely for conceptual clarity. It is important to realize that the coordinator's problem can be solved by each controller in the original system. Thus, the presence of the coordinator is not necessary. Nevertheless, its presence allows one to look at the original optimization problem from the view point of a "higher level authority" and simultaneously determine how each controller maps its local information to its action for the given realization of common information.

A key assumption in the common information approach is that common information is increasing with time. This assumption ensures that the coordinator has perfect recall and connects the coordinator's problem with centralized stochastic control and POMDPs. The connection between the coordinator's problem and POMDPs can be used for computational purposes as well. The dynamic program obtained for the coordinator is essentially similar to that for POMDPs. In particular, just as in POMDPs, the value-functions can be shown to be piecewise linear and concave function of the coordinator's belief. This characterization of value functions is utilized to find computationally efficient algorithms for POMDPs. Such algorithmic solutions to general POMDPs are well-studied and can be employed here. We refer the reader to [45] and references therein for a review of algorithms to solve POMDPs.

This chapter illustrates how common information approach can be used to solve decentralized stochastic control/decision-making problems that arise in control, communication and queueing systems and to resolve a long-standing theoretical problem on the structure of optimal control strategies in delayed sharing information structures.

As is the case for the designer's approach discussed in Section 1, the common information approach may be combined with the person-by-person approach as follows. First, use the person-by-person approach to identify qualitative properties of globally optimal strategies (e.g., identifying irrelevant information at controllers). Then, use the common information approach to further refine the qualitative properties and determine globally optimal strategies with those properties. In fact, all the examples of Section 4 used such a combined approach.

In this chapter, and in [17], it is assumed that the system has a partial history sharing information structure in which: (i) part of the past data (observations and control actions) of each controller is commonly available to all controllers; and (ii) all controllers have perfect recall of this commonly available data. Although this particular information structure makes it easier to describe the common information approach, it is not necessary for the approach to work. In particular, the common information approach applies to all sequential decision making problems (see [15] for a complete exposition).

Acknowledgement

A. Mahajan's work was partially supported by Natural Sciences and Engineering Research Council of Canada (NSERC) Discovery Grant RGPIN 402753-11 and D. Teneketzis's work was partially supported by National Science Foundation (NSF) Grant CCF-1111061.

References

1. Aicardi, M., Davoli, F., Minciardi, R.: Decentralized optimal control of Markov chains with a common past information set. IEEE Trans. Autom. Control **32**(11), 1028–1031 (1987)
2. Bamieh, B., Voulgaris, P.: A convex characterization of distributed control problems in spatially invariant systems with communication constraints. Systems and Control Letters **54**(6), 575–583 (2005)
3. Bismut, J.M.: An example of interaction between information and control: The transparency of a game. IEEE Trans. Autom. Control **18**(5), 518–522 (1972)
4. Gattami, A.: Control and estimation problems under partially nested information pattern. In: Proceedings of 48th IEEE Conference on Decision and Control, pp. 5415–5419 (2009)

5. Hajek, B., Mitzel, K., Yang, S.: Paging and registration in cellular networks: Jointly optimal policies and an iterative algorithm. IEEE Trans. Inf. Theory **64**, 608–622 (2008).
6. Hluchyj, M.G., Gallager, R.G.: Multiacces of a slotted channel by finitely many users. In: Proceedings of National Telecommunication Conference, pp. D.4.2.1–D.4.2.7 (1981)
7. Ho, Y.C.: Team decision theory and information structures. Proc. IEEE **68**(6), 644–654 (1980)
8. Ho, Y.C., Chu, K.C.: Team decision theory and information structures in optimal control problems–Part I. IEEE Trans. Autom. Control **17**(1), 15–22 (1972)
9. Kaspi, Y., Merhav, N.: Structure theorem for real-time variable-rate lossy source encoders and memory-limited decoders with side information. In: proceedings of the IEEE Symposium on Information Theory. Austin, TX (2010)
10. Kim, J., Lall, S.: A unifying condition for separable two player optimal control problems. In: Proceedings of 50th IEEE Conference on Decision and Control (2011)
11. Lessard, L., Lall, S.: A state-space solution to the two-player decentralized optimal control problem. In: Proceedings of 49th Annual Allerton Conference on Communication, Control and Computing (2011)
12. Mahajan, A.: Optimal decentralized control of coupled subsystems with control sharing. In: Proc. 50th IEEE Conf. Decision and Control and European Control Conf. (CDC-ECC), pp. 5726–5731. Orlando, FL (2011).
13. Mahajan, A., Nayyar, A., Teneketzis, D.: Identifying tractable decentralized control problems on the basis of information structure. In: Proc. 46th Annual Allerton Conf. Communication, Control, and Computing, pp. 1440–1449. Monticello, IL (2008).
14. Mahajan, A., Teneketzis, D.: On the design of globally optimal communication strategies for real-time noisy communication systems with noisy feedback. IEEE J. Sel. Areas Commun. **26**(4), 580–595 (2008).
15. Nayyar, A.: Sequential decision making in decentralized systems. Ph.D. thesis, University of Michigan, Ann Arbor, MI (2011)
16. Nayyar, A., Mahajan, A., Teneketzis, D.: Optimal control strategies in delayed sharing information structures. IEEE Trans. Autom. Control **56**(7), 1606–1620 (2011).
17. Nayyar, A., Mahajan, A., Teneketzis, D.: Decentralized stochastic control with partial history sharing information structures: A common information approach. IEEE Trans. Autom. Control (in print) (2013)
18. Nayyar, A., Teneketzis, D.: On jointly optimal real-time encoding and decoding strategies in multiterminal communication systems. In: proceedings of 47th IEEE Conference of Decision and Control (2008)
19. Nayyar, A., Teneketzis, D.: Decentralized detection with signaling. In: Proceeding of the Workshop on the Mathematical Theory of Networks and Systems (MTNS) (2010)
20. Nayyar, A., Teneketzis, D.: Sequential problems in decentralized detection with communication. IEEE Trans. Info. Theory **57**(8), 5410–5435 (2011)
21. Ooi, J.M., Verbout, S.M., Ludwig, J.T., Wornell, G.W.: A separation theorem for periodic sharing information patterns in decentralized control. IEEE Trans. Autom. Control **42**(11), 1546–1550 (1997).
22. Ooi, J.M., Wornell, G.W.: Decentralized control of a multiple access broadcast channel: performance bounds. In: Proceedings of the 35th IEEE Conference on Decision and Control, pp. 293–298. Kobe, Japan (1996)
23. Rantzer, A.: Linear quadratic team theory revisited. In: Proceedings of the American Control Confernece, pp. 1637–1641 (2006)
24. Rotkowitz, M., Lall, S.: A characterization of convex problems in decentralized control. IEEE Trans. on Automatic Control **51**(2), 274–286 (2006)
25. Schoute, F.C.: Decentralized control in packet switched satellite communication. IEEE Trans. Autom. Control **AC-23**(2), 362–271 (1976)
26. Teneketzis, D.: On the structure of optimal real-time encoders and decoders in noisy communication. IEEE Trans. Inf. Theory pp. 4017–4035 (2006)
27. Teneketzis, D., Ho, Y.C.: The Decentralized Wald problem. Information and Computation, 73 pp. 23–44 (1987)

28. Teneketzis, D., Varaiya, P.: The decentralized quickest detection problem. IEEE Trans. on Automatic Control **AC-29**(7), 641–644 (1984)
29. Tenney, R.R., N. R. Sandell Jr.: Detection with distributed sensors. IEEE Trans. Aerospace Electron. Systems **AES-17**(4), 501–510 (1981)
30. Tsitsiklis, J.N.: Decentralized detection. In: Advances in Statistical Signal Processing, pp. 297–344. JAI Press (1993)
31. Varaiya, P., Walrand, J.: On delayed sharing patterns. IEEE Trans. Autom. Control **23**(3), 443–445 (1978)
32. Varaiya, P., Walrand, J.: Decentralized control in packet switched satellite communication. IEEE Trans. Autom. Control **AC-24**(5), 794–796 (1979)
33. Veeravalli, V.V.: Decentralized quickest change detection. IEEE Trans. Inform. Theory **47**(4), 1657–1665 (2001)
34. Veeravalli, V.V., Basar, T., Poor, H.: Decentralized sequential detection with a fusion center performing the sequential test. IEEE Trans. Inform. Theory **39**, 433–442 (1993)
35. Veeravalli, V.V., Basar, T., Poor, H.: Decentralized sequential detection with sensors performing sequential tests. Mathematics of Control, Signals and Systems **7**(4), 292–305 (1994)
36. Walrand, J.C., Varaiya, P.: Causal coding and control of Markov chains. System and Control Letters **3**, 189–192 (1983)
37. Walrand, J.C., Varaiya, P.: Optimal causal coding-decoding problems. IEEE Trans. Inf. Theory **29**(6), 814–820 (1983)
38. Whittle, P.: Optimization Over Time, *Wiley Series in Probability and Mathematical Statistics*, vol. 2. John Wiley and Sons (1983)
39. Witsenhausen, H.S.: Separation of estimation and control for discrete time systems. Proc. IEEE **59**(11), 1557–1566 (1971)
40. Witsenhausen, H.S.: A standard form for sequential stochastic control. Mathematical Systems Theory **7**(1), 5–11 (1973)
41. Witsenhausen, H.S.: On the structure of real-time source coders. Bell System Technical Journal **58**(6), 1437–1451 (1979)
42. Wu, J., Lall, S.: A dynamic programming algorithm for decentralized markov decision processes with a broadcast structure. In: Proceedings of the 49th IEEE Conference on Decision and Control, pp. 6143–6148 (2010)
43. Yoshikawa, T.: Dynamic programming approach to decentralized stochastic control problems. IEEE Trans. Autom. Control **20**(6), 796 – 797 (1975).
44. Yüksel, S.: Stochastic nestedness and the belief sharing information pattern. IEEE Trans. Autom. Control pp. 2773–2786 (2009)
45. Zhang, H.: Partially observable markov decision processes: A geometric technique and analysis. Operations Research (2009)