2023 62nd IEEE Conference on Decision and Control (CDC)
December 13-15, 2023. Marina Bay Sands, Singapore

Optimal Symmetric Strategies in Multi-Agent Systems with Decentralized Information

Sagar Sudhakara and Ashutosh Nayyar

***Abstract*— We consider a cooperative multi-agent system consisting of a team of agents with decentralized information. Our focus is on the design of *symmetric* (i.e. identical) strategies for the agents in order to optimize a finite horizon team objective. We start with a general information structure and then consider some special cases. The constraint of using symmetric strategies introduces new features and complications in the team problem. For example, we show in a simple example that randomized symmetric strategies may outperform deterministic symmetric strategies. We also discuss why some of the known approaches for reducing agents' private information in teams may not work under the constraint of symmetric strategies. We then adopt the common information approach for our problem and modify it to accommodate the use of symmetric strategies. This results in a common information based dynamic program where each step involves minimization over a single function from the space of an agent's private information to the space of probability distributions over actions. We present specialized models where private information can be reduced using simple dynamic program based arguments.**

I. Introduction

The problem of sequential decision-making by a team of collaborative agents operating in an uncertain environment has received significant attention in the recent control (e.g. [1]–[5]) and artificial intelligence (e.g. [6]–[10]) literature. The goal in such problems is to design decision/control strategies for the multiple agents in order to optimize a performance metric for the team.

In some cooperative multi-agent (or team) problems, the agents are essentially identical and interchangeable. For example, consider a team of autonomous agents operating in an environment. The agents may have identical sensors that they use to observe their local surroundings and they may have identical action spaces. For teams with such identical agents, it may be convenient for the designer to design identical decision/control strategies for the agents. This would be particularly helpful if the number of agents is large – instead of designing n different strategies for n agents in a team, the designer needs to design just one strategy for all agents. Identical strategies may also be necessary for other practical and regulatory reasons. For example, a self-driving car company would be expected to have the same control algorithm on all its cars. Another reason for using identical strategies arises in situations where agents don't have any individualized identities. This can happen in settings where the population of the agents is not fixed and agents are unaware of the total number of agents currently present or their own index in the population. An example of such a situation for a multi-access communication problem is described in [11]. When an agent doesn't know its own index ("Am I agent 1 or agent 2?"), it makes sense to use symmetric (i.e. identical) strategies for all agents irrespective of their index. In this paper, we will focus on the design of identical strategies for a team of cooperative agents. We will refer to such strategies as *symmetric strategies*.

Our focus is on designing symmetric strategies to optimize a finite horizon team objective. We start with a general information structure and then consider some special cases. The constraint of using symmetric strategies introduces new features and complications in the team problem. For example, when agents in a team are free to use individualized strategies, it is well-known that agents can be restricted to deterministic strategies without loss of optimality [12]. However, we show in a simple example that randomized strategies may be helpful when the agents are constrained to use symmetric strategies.

We adopt the common information approach [2] for our problem and modify it to accommodate the use of symmetric strategies. This results in a common information based dynamic program where each step involves minimization over a single function from the space of an agent's private information to the space of probability distributions over actions. The complexity of this dynamic program depends in large part on the size of the private information space. We discuss some known approaches for reducing agents' private information and why they may *not* work under the constraint of symmetric strategies. We present two specialized models where private information can be reduced using simple dynamic program based arguments.

Notation: Random variables are denoted by upper case letters (e.g. X), their realization with lower case letters (e.g. x), and their space of realizations by script letters (e.g. $\mathcal{X}$). Subscripts denote time and superscripts denote agent index; e.g., X^i_t denotes the state of agent i at time t. The short hand notation $X^i_{1:t}$ denotes the collection $(X^i_1, X^i_2, ..., X^i_t)$. $\triangle(\mathcal{X})$ denotes the probability simplex for the space $\mathcal{X}$. $\mathbb{P}(A)$ denotes the probability of an event A. $\mathbb{E}[X]$ denotes the expectation of a random variable X. $\mathbb{1}_A$ denotes the indicator function of event A. For simplicity of notation, we use $\mathbb{P}(x_{1:t}, u_{1:t-1})$ to denote $\mathbb{P}(X_{1:t} = x_{1:t}, U_{1:t-1} = u_{1:t-1})$ and a similar notation for conditional probability. For a strategy pair (g^1, g^2), we use $\mathbb{P}^{(g^1,g^2)}(\cdot)$ (resp. $\mathbb{E}^{(g^1,g^2)}[\cdot]$) to indicate that the probability (resp. expectation) depends on the choice of the strategy pair. We use $-i$ to denote agent/agents other than agent i. $U \sim \lambda$ indicates that U is

S. Sudhakara and A. Nayyar are with the Department of Electrical & Computer Engineering, University of Southern California, Los Angeles, CA 90089. (E-mail: sagarsud@usc.edu; ashutosn@usc.edu). This work was supported by NSF grants ECCS 2025732 and ECCS 1750041.

979-8-3503-0123-6/23/$31.00 ©2023 IEEE

randomly distributed according to the distribution λ.

II. Problem Formulation

Consider a discrete-time system with two agents. The system state consists of three components - a shared state and two local states, one for each agent. $X_t^i \in \mathcal{X}$ denotes the local state of agent i, $i = 1, 2$, at time t and $X_t^0 \in \mathcal{X}^0$ denotes the shared state at time t. X_t denotes the triplet (X_t^0, X_t^1, X_t^2). Let $U_t^i \in \mathcal{U}$ denote the control action of agent i at time t. U_t denotes the pair (U_t^1, U_t^2). The dynamics of the shared and local states are as follows:

$$X_{t+1}^0 = f_t^0(X_t^0, U_t, W_t^0), \tag{1}$$

$$X_{t+1}^i = f_t(X_t^i, X_t^0, U_t, W_t^i), \quad i = 1, 2, \tag{2}$$

where $W_t^0 \in \mathcal{W}^0$ and $W_t^i \in \mathcal{W}$ are random disturbances with W_t^0 having the probability distribution p_W^0 and $W_t^i, i = 1, 2$, having the probability distribution p_W. We use W_t to denote the triplet (W_t^0, W_t^1, W_t^2). Note that the next local state of agent i depends on its own current local state, the shared state and the control actions of both the agents. Also note that the function f_t in (2) is the same for both agents. The initial states X_1^0, X_1^1, X_1^2 are independent random variables with X_1^0 having the probability distribution α^0 and $X_1^i, i = 1, 2$, having the probability distribution α. The initial states X_1^0, X_1^1, X_1^2 and the disturbances $W_t^0, W_t^i, t \geq 1, i = 1, 2$, are independent discrete random variables. These will be referred to as the primitive random variables of the system. In this paper, we assume that all system variables take value in discrete sets.

A. Information structure and strategies

The information available to agent i , $i = 1, 2$, at time t consists of two parts:

1) Common information C_t - This information is available to both agents[1]. C_t takes values in the set $\mathcal{C}_t$.
2) Private information P_t^i - Any information available to agent i at time t that is not included in C_t is included in P_t^i. P_t^i takes values in $\mathcal{P}_t$. (Note that the space of private information is the same for both agents.) We use P_t to denote the pair (P_t^1, P_t^2).

C_t should be viewed as an ordered list (or row vector) of some of the system variables that are known to both agents. Similarly, P_t^i should be viewed as an ordered list (or row vector).

We assume that C_t is non-decreasing with time, i.e., any variable included in C_t is also included in C_{t+1}. Let Z_{t+1} be the increment in common information from time t to $t+1$. We assume the following dynamics for Z_{t+1} and P_{t+1}^i ($i = 1, 2$):

$$Z_{t+1} = \zeta_t(X_t, P_t, U_t, W_t); P_{t+1}^i = \xi_t^i(X_t, P_t, U_t, W_t),$$

Agent i uses its information at time t to select a probability distribution δU_t^i on the action space $\mathcal{U}$. We will refer to δU_t^i as agent i's *behavioral action* at time t. The action U_t^i is then randomly generated according to the chosen distribution, i.e., $U_t^i \sim \delta U_t^i$. Thus, we can write

$$\delta U_t^i = g_t^i(P_t^i, C_t), \tag{3}$$

where g_t^i is a mapping from $\mathcal{P}_t \times \mathcal{C}_t$ to $\Delta(\mathcal{U})$. The function g_t^i is referred to as the control strategy of agent i at time t. The collection of functions $g^i := (g_1^i, \dots, g_T^i)$ is referred to as the control strategy of agent i. Let $\mathcal{G}$ denote the set of all possible strategies for agent i. (Note that the set of all possible strategies is the same for the two agents since the private information space, the common information space and the action space are the same for the two agents.)

We use (g^1, g^2) to denote the pair of strategies being used by agent 1 and agent 2 respectively. We are interested in the finite horizon total expected cost incurred by the system which is defined as:

$$J(g^1, g^2) := \mathbb{E}^{(g^1, g^2)} \left[\sum_{t=1}^{T} k_t(X_t, U_t) \right], \tag{4}$$

where k_t is the cost function at time t. Our focus will be on the case of *symmetric strategies*, i.e., the case where both agents use the same control strategy. When referring to symmetric strategies, we will drop the superscript i in g^i and denote a symmetric strategy pair by (g, g).

*Symmetric strategy optimization problem (**Problem P1**):* Our objective is to find a symmetric strategy pair that achieves the minimum total expected cost among all symmetric strategy pairs. That is, we are looking for a strategy $g \in \mathcal{G}$ such that

$$J(g, g) \leq J(h, h), \quad \forall h \in \mathcal{G}. \tag{5}$$

Remark 1 a) We assume that the randomization at each agent is done independently over time and independently of the other agent [13]. b) We have formulated the problem with two agents for simplicity. The number of agents can in fact be any positive integer n or even a deterministic time-varying sequence n_t. Our results extend to these cases with only notational modifications.

Remark 2 If the private information space, the common information space and the action space are finite, then it can be shown that the strategy space $\mathcal{G}$ is a compact space and that $J(g, g)$ is a continuous function of $g \in \mathcal{G}$. Thus, an optimal g satisfying (5) exists.

Remark 3 Note that we are not claiming that use of symmetric strategies is always optimal – it is not. We are simply focusing on the design of symmetric strategies for reasons mentioned in the introduction.

B. Some specific information structures

We will be particularly interested in the special cases of Problem P1 described below. Each case corresponds to a different information structure. In each case, the shared state history until time t, $X_{1:t}^0$, and the action history until $t-1$, $U_{1:t-1}$ are part of common information C_t.

1. *One-step delayed sharing information structure:* In this case, each agent knows its own local state history until time t and the local state history of the other agent until time $t-1$.

[1] C_t does not have to be the *entirety* of information that is available to both agents; it simply cannot include anything that is not available to both agents.

Thus, the common and private information available to agent i at time t is given by

$$C_t = (X^0_{1:t}, U_{1:t-1}, X^{1,2}_{1:t-1});\ P^i_t = X^i_t. \tag{6}$$

We refer to the instance of Problem P1 with this information structure as **Problem P1a**.

2. *Full local history information structure:* In this case, each agent knows its own local state history until time t but does not observe the local states of the other agent. Thus, the common and private information available to agent i at time t is given by

$$C_t = (X^0_{1:t}, U_{1:t-1});\ P^i_t = X^i_{1:t}. \tag{7}$$

This information structure corresponds to the control sharing information structure of [3]. We refer to the instance of Problem P1 with this information structure as **Problem P1b**.

3. *Reduced local history information structure:* In this case, each agent knows its own *current* local state but does not recall its past local states and does not observe the local states of the other agent. Thus, the common and private information available to agent i at time t is given by

$$C_t = (X^0_{1:t}, U_{1:t-1});\ P^i_t = X^i_t. \tag{8}$$

We refer to the instance of Problem P1 with this information structure as **Problem P1c**.

Another special case of Problem P1 that might be of interest is the following: Consider a situation where the state dynamics are governed not by the vector of agents' actions but only by an aggregate effect of agents' actions. Let $A_t = a(U^1_t, U^2_t)$ denote the aggregate action. We refer to $a(\cdot,\cdot)$ as the aggregation function. Some examples of a could be the sum or the maximum function. The state dynamics are as described in equations (1) and (2) except with U_t replaced by A_t. The common and private information are given as: $C_t = (X^0_{1:t}, A_{1:t-1});\ P^i_t = X^i_t$. While we don't address this case in this paper, it may be of interest for future work.

C. Why are randomized strategies needed?

In team problems, it is well-known that one can restrict agents to deterministic strategies without loss of optimality [12]. However, since the agents are restricted to use symmetric strategies in our setup, randomization can help. This can be illustrated by the following simple example.

Example 1: Let $T = 1$ and let $(X^0_1, X^1_1, X^2_1) = (0, 0, 0)$ with probability 1. The action space is $\mathcal{U} = \{0, 1\}$. The information structure is that of Problem P1c described in II-B. The cost at $t = 1$ is given by, $k_1(X_1, U_1) = \mathbb{1}_{\{U^1_1 = U^2_1\}}$.

Note that the cost function penalizes the agents for taking the same action. In this case, each agent has only two *deterministic* strategies – taking action 0 or taking action 1 at time 1. If both agents use the same deterministic strategy, then, clearly, $U^1_1 = U^2_1$ and hence the expected cost incurred is 1.

Consider now the following randomized strategy for each agent: $U^i_1 = 1$ with probability p and $U^i_1 = 0$ with probability $(1 - p)$. When the two agents use this randomized strategy, the expected cost is $p^2 + (1 - p)^2$. With $p = 0.5$, this cost is 0.5 which is less than the expected cost achieved by any deterministic symmetric strategy pair. Thus, when agents are restricted to use the same strategy, they can benefit from randomization.

III. Common information approach

We adopt the common information approach [2] for Problem P1. This approach formulates a new decision-making problem from the perspective of a coordinator that knows the common information. At each time, the coordinator selects prescriptions that map each agent's private information to its action. The behavioral action of each agent in this problem is simply the prescription evaluated at the current realization of its private information. Since Problem P1 requires symmetric strategies for the two agents, we will require the coordinator to select *identical prescriptions for the two agents*. To make things precise, let $\mathcal{B}_t$ denote the space of all functions from $\mathcal{P}_t$ to $\Delta(\mathcal{U})$. Let $\Gamma_t \in \mathcal{B}_t$ denote the prescription selected by the coordinator at time t. Then, the behavioral action of agent i, $i = 1, 2$, is given by: $\delta U^i_t = \Gamma_t(P^i_t)$.

As in Problem P1, agent i's action U^i_t is generated according to the distribution δU^i_t using independent randomization. The coordinator selects its prescription at time t based on the common information at time t and the history of past prescriptions. Thus, we can write: $\Gamma_t = d_t(C_t, \Gamma_{1:t-1})$, where d_t is a mapping from $\mathcal{C}_t \times \mathcal{B}_1 \ldots \times \mathcal{B}_{t-1}$ to $\mathcal{B}_t$. The collection of mappings $d := (d_1, \ldots, d_T)$ is referred to as the coordination strategy. The coordinator's objective is to choose a coordination strategy that minimizes the finite horizon total expected cost:

$$\mathcal{J}(d) := \mathbb{E}^d \left[\sum_{t=1}^{T} k_t(X_t, U_t) \right]. \tag{9}$$

The following lemma establishes the equivalence of the coordinator problem formulated above and the problem Problem P1. The use of identical prescriptions by the coordinator is needed to connect the coordinator's strategy to symmetric strategies for the agents in Problem P1.

Lemma 1 *Problem P1 and the coordinator's problem are equivalent in the following sense:*
(i) For any symmetric strategy pair (g, g), consider the following coordination strategy: $d_t(C_t) = g_t(\cdot, C_t)$. Then, $J(g, g) = \mathcal{J}(d)$. (ii) Conversely, for any coordination strategy d, consider the symmetric strategy pair defined as follows: $g_t(\cdot, C_t) = d_t(C_t, \Gamma_{1:t-1})$, where $\Gamma_k = d_k(C_k, \Gamma_{1:k-1})$ for $k = 1, \ldots, t - 1$.

Proof The proof is based on Proposition 3 of [2] and the fact that the use of identical prescriptions for the two agents by the coordinator corresponds to the use of symmetric strategies in Problem P1. ■

We now proceed with finding a solution for the coordinator's problem. As shown in [2], the coordinator's belief on (X_t, P_t) can serve as its information state (sufficient statistic) for selecting prescriptions. At time t, the coordinator's belief is given as:

$$\Pi_t(x,p) = \mathbb{P}(X_t = x, P_t = p | C_t, \Gamma_{1:(t-1)}), \qquad (10)$$

for all $x \in \mathcal{X}^0 \times \mathcal{X} \times \mathcal{X}, p \in \mathcal{P}_t \times \mathcal{P}_t$. The belief can be sequentially updated by the coordinator as described in Lemma 2 below. The lemma follows from arguments similar to those in Lemma 2 of [13] (or Theorem 1 of [2]).

Lemma 2 *For any coordination strategy d, the coordinator's belief Π_t evolves almost surely as*

$$\Pi_{t+1} = \eta_t(\Pi_t, \Gamma_t, Z_{t+1}), \qquad (11)$$

where η_t is a fixed transformation that does not depend on the coordination strategy.

Using the results in [2], we can write a dynamic program for the coordinator's problem. Recall that $\mathcal{B}_t$ is the space of all functions from $\mathcal{P}_t$ to $\Delta(\mathcal{U})$. For a $\gamma \in \mathcal{B}_t$ and $p \in \mathcal{P}_t$, $\gamma(p)$ is a probability distribution on $\mathcal{U}$. Let $\gamma(p;u)$ denote the probability assigned to $u \in \mathcal{U}$ under the probability distribution $\gamma(p)$.

Theorem 1 *The value functions for the coordinator's dynamic program are as follows: Define $V_{T+1}(\pi_{T+1}) = 0$ for every π_{T+1}. For $t \leq T$ and for any realization π_t of Π_t, define*

$$\begin{aligned} V_t(\pi_t) = \min_{\gamma_t \in \mathcal{B}_t} \mathbb{E}[k_t(X_t, U_t) + \\ V_{t+1}(\eta_t(\pi_t, \gamma_t, Z_{t+1})) | \Pi_t = \pi_t, \Gamma_t = \gamma_t] \end{aligned} \qquad (12)$$

The coordinator's optimal strategy is to pick the minimizing prescription for each time and each π_t.

Proof The coordinator's problem can be seen as a POMDP [2]. The theorem is the corresponding dynamic program. ■

Remark 4 The expectation in (12) should be interpreted as follows: Z_{t+1} is given by (II-A), $U_t^i, i = 1,2$, is independently randomly generated according to the distribution $\gamma_t(P_t^i)$ and the joint distribution on (X_t, P_t) is π_t.

Remark 5 It can be established by backward induction that the term being minimized in (12) is a continuous function of γ_t. This can be shown using an argument very similar to the one used in the proof of Lemma 3 in [14]. This continuity property along with the fact that $\mathcal{B}_t$ is a compact set ensures that the minimum in (12) is achieved.

For the instances of Problem P1 described in Problems P1a - P1c (see Section II), the private information of an agent includes its current local state. Consequently, for these instances, the coordinator's belief is just on the private information of the agents and the current shared state. This belief can be factorized as shown in the following lemma.

Lemma 3 *In Problems 1a - 1c, for any realization x^0 of the shared state and any realizations p^1, p^2 of the agents' private information,*

$$\Pi_t(x^0, p^1, p^2) = \delta_{X_t^0}(x^0)\Pi_t^1(p^1)\Pi_t^2(p^2), \qquad (13)$$

where Π_t is the coordinator's belief (see (10)*), Π_t^1, Π_t^2 are the marginals of Π_t for each agent's private information and $\delta_{X_t^0}(\cdot)$ is a delta distribution located at X_t^0. (Recall that X_t^0 is part of the common information in Problems P1a-P1c.)*

Further, for any coordination strategy d, $\Pi_t^i, i = 1,2$, evolves almost surely as

$$\Pi_{t+1}^i = \eta_t^i(X_t^0, \Pi_t^i, \Gamma_t, Z_{t+1}), \qquad (14)$$

where η_t^i is a fixed transformation that does not depend on the coordination strategy.

Proof See Appendix I in [15]. ■

Because of the above lemma, we can replace Π_t (and its realizations π_t) by $(\Pi_t^1, \Pi_t^2, X_t^0)$ (and the corresponding realizations $(\pi_t^1, \pi_t^2, x_t^0)$) in the dynamic program of Theorem 1 for Problems P1a -P1c.

IV. Comparison of Problems 1b and 1c

The information structures in Problems P1b and P1c differ only in the private information available to the agents – in P1b, each agent know its entire local state history whereas in P1c each agent knows only its current local state. *If the agents were not restricted to use the same strategies,* it is known that the two information structures are equivalent. That is, if a (possibly asymmetric) strategy pair (g^1, g^2) is optimal for the information structure in Problem P1c, then it is also optimal for the information structure in Problem P1b [3]. This effectively means that agents can ignore their past local states without any loss in performance. However, such an equivalence of the two information structures may not hold when agents are restricted to use symmetric strategies. In other words, an optimal symmetric strategy in Problem P1c may not be optimal for Problem P1b; and the optimal performance in Problem P1c may be strictly worse than the optimal performance in Problem P1b. We explore this point in more detail below.

One approach for establishing that agents can ignore parts of their private information that has been commonly used in prior literature on multi-agent/decentralized systems is the agent-by-agent (or person-by-person approach) [2], [16]. This approach works as follows: We start by fixing strategies of all agents other than agent i to arbitrary choices and then show that agent i can make decisions based on a subset or a function of its private information without compromising performance. If this reduction in agent i's information holds for any arbitrary strategy of other agents, we can conclude that this reduction would hold for globally optimal strategies as well. By repeating this argument for all agents, one can reduce the private information of all agents without losing performance. The problem with this approach is that it cannot accommodate the restriction to symmetric strategies. The reduced-information based strategies obtained using this approach may or may not be symmetric. Thus, we cannot adopt this approach for reducing agents' private information in Problem P1b.

Another approach for reducing private information that has been used in some game-theoretic settings [14] involves the use of conditional probabilities of actions given reduced information. To see how this approach can be used, let's consider an arbitrary (possible asymmetric) strategy pair (g^1, g^2) for the information structure of Problem P1b and define the following conditional probabilities for $i = 1, 2$:

$$\mathbb{P}^{(g^1,g^2)}[U_t^i = u|X_t^i = x, C_t = c_t]. \tag{15}$$

Note that (15) specifies a probability distribution on $\mathcal{U}$ for each x and c_t. Thus, it can be viewed as a valid strategy for agent i *under the information structure of Problem P1c*. This observation lets us define the following reduced-information strategies for the agents: for $i = 1, 2$,

$$\bar{g}_t^i(x, c_t) := \mathbb{P}^{(g^1,g^2)}[U_t^i = \cdot|X_t^i = x, C_t = c_t]. \tag{16}$$

Further, it can be shown that the above construction ensures that the joint distributions of (X_t, U_t, C_t) under strategies (g^1, g^2) and $(\bar{g}^1, \bar{g}^2)$ are the same for all t. This, in turn, implies that $J(\bar{g}^1, \bar{g}^2) = J(g^1, g^2)$. This argument establishes that there is a reduced-information strategy pair with the same performance as an arbitrary full-information strategy pair. Thus, the optimal performance with reduced-information strategies must be the same as the optimal performance with full-information strategies for the information structure of Problem P1b.

We can try to use the above argument for symmetric strategy pairs. We start with an arbitrary symmetric strategy pair (g, g) in Problem P1b and use (16) to define a reduced-information strategy pair that achieves the same performance as (g, g). The problem with this argument is that even though we started with a symmetric strategy pair (g, g), the reduced-information strategy pair constructed by (16) need not be symmetric. Hence, this reduced-information strategy pair may not be a valid solution for Problem P1c. We illustrate this point in the following example.

Example 2: Consider a setting where there is no shared state, the action space is $\mathcal{U} = \{a, b\}$ and the local states are i.i.d. (across time and across agents). Each local state is a Bernoulli (1/2) random variable. Consider the symmetric strategy pair (g, g) for Problem P1b where g_1 (the strategy at $t = 1$) is: $g_1(u_1^i = a|x_1^i) = \alpha(1 - x_1^i) + \beta x_1^i$, where $0 \leq \alpha, \beta \leq 1$. And g_2 (the strategy at $t = 2$) is:

$$g_2(u_2^i = a|x_1^i, x_2^i, u_1^1, u_1^2) = \begin{cases} \alpha, & \text{if } x_1^i = x_2^i \\ \beta, & \text{if } x_1^i \neq x_2^i. \end{cases} \tag{17}$$

We now use (16) to define a reduced-information strategy. Even though we started with a symmetric strategy pair for the two agents, the conditional probability on the right hand side of (16) may be different for the two agents. To see this, consider $t = 2$ and $C_2 = (U_1^1, U_1^2) = (a, b)$ and $X_2^i = 0$. Then, for agent 1:

$$\begin{aligned} &\mathbb{P}^{(g,g)}(U_2^1 = a|X_2^1 = 0, U_1^1 = a, U_1^2 = b) \\ &= \sum_{x=0,1} \mathbb{P}^{(g,g)}(U_2^1 = a, X_1^1 = x|X_2^1 = 0, U_1^1 = a, U_1^2 = b) \\ &= \alpha\left(\frac{\alpha}{\alpha+\beta}\right) + \beta\left(\frac{\beta}{\alpha+\beta}\right) \end{aligned} \tag{18}$$

On the other hand, a similar calculation for agent 2 shows that: $\mathbb{P}^{(g,g)}(U_2^2 = a|X_2^2 = 0, U_1^1 = a, U_1^2 = b)$

$$= \alpha\left(\frac{1-\alpha}{2-\alpha-\beta}\right) + \beta\left(\frac{1-\beta}{2-\alpha-\beta}\right). \tag{19}$$

The expressions in (18) and (19) are clearly different (e.g. with $\alpha = 1/4$ and $\beta = 1/2$). Thus, the reduced-information strategies constructed by (16) are not symmetric and, therefore, invalid for Problem P1c.

A. Special cases

In this section, we present two special cases under which Problems P1b and P1c can be shown to be equivalent, i.e., we can show that an optimal strategy for Problem P1c is also optimal for Problem P1b.

1) Specialized cost: We assume that the cost function at each time t is non-negative, i.e., $k_t(X_t^0, X_t^1, X_t^2, U_t^1, U_t^2) \geq 0$. Further, we assume that for each possible local state x^i of agent i there exists an action $m(x^i)$ such that $k_t(x^0, x^1, x^2, m(x^1), m(x^2)) = 0$ for all $x^0 \in \mathcal{X}^0$. An example of such a cost function is $k_t(X_t^0, X_t^1, X_t^2, U_t^1, U_t^2) = (X_t^1 - U_t^1)^2[(X_t^2 - U_t^2)^2 + 1] + (X_t^2 - U_t^2)^2$, where the states and actions are integer-valued.

Recall that in Problem P1b the prescription space at time t is the space of functions from $\mathcal{X}^t$ to $\Delta(\mathcal{U})$ and in Problem P1c the prescription space is the space of functions from $\mathcal{X}$ to $\Delta(\mathcal{U})$. Using the dynamic programs for Problems P1b and P1c with the specialized cost above, we can show that optimal prescriptions in both problems effectively coincide with the mapping m from $\mathcal{X}$ to $\mathcal{U}$.

Lemma 4 *The value functions for the coordinator's dynamic programs in Problems P1b and P1c can be written as follows: For $t \leq T$ and for any realization π_t^1, π_t^2, x_t^0 of Π_t^1, Π_t^2, X_t^0,*

$$V_t(\pi_t^1, \pi_t^2, x_t^0) := \min_{\gamma_t \in \mathcal{B}_t} Q_t(\pi_t^1, \pi_t^2, x_t^0, \gamma_t), \tag{20}$$

where the function Q satisfies

$$Q_t(\pi_t^1, \pi_t^2, x_t^0, \gamma_t) \geq Q_t(\pi_t^1, \pi_t^2, x_t^0, m) = 0, \tag{21}$$

Consequently, the coordinator's optimal prescription is m at each time.

PROOF See Appendix III in [15]. ■

Since the coordinator's optimal strategy is identical in Problems P1b and P1c, it follows that the optimal symmetric strategy for the agents in the two problems is also the same, namely $U_t^i = m(X_t^i)$.

2) Specialized dynamics: We consider a specialized dynamics where the local states $X_{1:T}^i, i = 1, 2$, are i.i.d. uncontrolled random variables with probability distribution α and there is no shared state. The following lemma shows the equivalence between Problems P1b and P1c.

Lemma 5 *The optimal performance in Problem P1c is the same as the optimal performance in Problem P1b. Further, the optimal symmetric strategy for Problem P1c is optimal for Problem P1b as well.*

PROOF See Appendix IV in [15]. ■

In summary, for the specialized cases described above, agents' private information can be reduced without losing performance, even with the restriction to symmetric strategies.

V. Comparison of Problems 1a and 1c

The information structures in Problems P1a and P1c differ only in the common information available to the agents – in P1a, each agent has an additional part in the common information consisting of the local state history of both agents. Since agents in Problem P1c have less information that their counterparts in Problem P1a, an optimal symmetric strategy in Problem P1c may not be optimal for Problem P1a; and the optimal performance in Problem P1c may be strictly worse then the optimal performance in Problem P1a.

Example 3: Consider a setting where there is no shared state, the action space is $\mathcal{U} = \{0,1\}$ and state space is $\mathcal{X} = \{0,1\}$. Let $T = 2$ and the local states of each agent are stationary across time. The initial states X_1^1, X_1^2 are independent random variables with probability distribution Bernoulli (1/2). The cost at time $t = 1$ is given by $k_1(X_1, U_1) = 10\mathbb{1}_{\{U_1^1 \neq 0, U_1^2 \neq 0\}}$ and cost at time $t = 2$ is given by:

$$k_2(X_2, U_2) = \begin{cases} 0, \text{ if } U_2^1 = X_2^2 \textit{ and } U_2^2 = X_2^1 \\ 1, \text{ otherwise}, \end{cases} \tag{22}$$

Consider the symmetric strategy pair (g, g) for Problem P1a where g_1 (the strategy at $t = 1$) is: $U_1^1 = 0, U_1^2 = 0$. At time $t = 2$, each agent uses the following strategy: $U_2^1 = X_1^1, U_2^2 = X_1^2$ if $X_1^1 = X_1^2$ and $U_2^1 = 1 - X_1^1, U_2^2 = 1 - X_1^2$ if $X_1^1 \neq X_1^2$. This results in optimal expected cost of 0 in Problem P1a. In Problem P1c, it can be shown that the optimal strategy at time $t = 1$ is $U_1^1 = 0, U_1^2 = 0$ and at time $t = 2$, U_2^i follows probability distribution Bernoulli (1/2). The optimal expected cost is 0.75 for the Problem P1c, which is strictly worse than optimal performance in Problem P1a.

Special Case: We present a special dynamics under which Problems P1a and P1c can be shown to be equivalent. The dynamics of the shared and local states in the specialized dynamics problem are as follows:

$$X_{t+1}^0 = f_t^0(X_t^0, U_t, W_t^0), \tag{23}$$

$$X_{t+1}^i = f_t(X_t^0, U_t, W_t^i), \quad i = 1, 2. \tag{24}$$

In this case, we have the following result:

Lemma 6 *For the specialized dynamics described in* (23) - (24)*, an optimal symmetric strategy in Problem P1c is also optimal for Problem P1a and, consequently, the optimal performance in the two problems are the same.*

PROOF See Appendix V in [15]. ■

VI. Conclusion

In this paper, we focused on designing symmetric strategies to optimize a finite horizon team objective. We started with a general information structure and then considered some special cases. We showed in a simple example that randomized symmetric strategies may outperform deterministic symmetric strategies. We also discussed why some of the known approaches for reducing agents' private information in teams may not work under the constraint of symmetric strategies. We modified the common information approach to obtain optimal symmetric strategies for the agents. This resulted in a common information based dynamic program whose complexity depends in large part on the size of the private information space. We presented two specialized models where private information can be reduced using simple dynamic program based arguments.

References

[1] A. Nayyar, A. Mahajan, and D. Teneketzis, "Optimal control strategies in delayed sharing information structures," *IEEE Transactions on Automatic Control*, vol. 56, no. 7, pp. 1606–1620, 2010.

[2] A. Nayyar, A. Mahajan, and D. Teneketzis, "Decentralized stochastic control with partial history sharing: A common information approach," *IEEE Transactions on Automatic Control*, vol. 58, no. 7, pp. 1644–1658, 2013.

[3] A. Mahajan, "Optimal decentralized control of coupled subsystems with control sharing," *IEEE Transactions on Automatic Control*, vol. 58, no. 9, pp. 2377–2382, 2013.

[4] S. Sudhakara, D. Kartik, R. Jain, and A. Nayyar, "Optimal communication and control strategies in a multi-agent MDP problem," *arXiv preprint arXiv:2104.10923*, 2021.

[5] D. Kartik, S. Sudhakara, R. Jain, and A. Nayyar, "Optimal communication and control strategies for a multi-agent system in the presence of an adversary," *arXiv preprint arXiv:2209.03888*, 2022.

[6] S. Seuken and S. Zilberstein, "Formal models and algorithms for decentralized decision making under uncertainty," *Autonomous Agents and Multi-Agent Systems*, vol. 17, no. 2, pp. 190–250, 2008.

[7] A. Kumar, S. Zilberstein, and M. Toussaint, "Probabilistic inference techniques for scalable multiagent decision making," *Journal of Artificial Intelligence Research*, vol. 53, pp. 223–270, 2015.

[8] T. Rashid, M. Samvelyan, C. Schroeder, G. Farquhar, J. Foerster, and S. Whiteson, "Qmix: Monotonic value function factorisation for deep multi-agent reinforcement learning," in *International Conference on Machine Learning*, pp. 4295–4304, PMLR, 2018.

[9] H. Hu and J. N. Foerster, "Simplified action decoder for deep multi-agent reinforcement learning," in *International Conference on Learning Representations*, 2019.

[10] D. Szer, F. Charpillet, and S. Zilberstein, "MAA* a heuristic search algorithm for solving decentralized POMDPs," in *Proceedings of the Twenty-First Conference on Uncertainty in Artificial Intelligence*, pp. 576–583, 2005.

[11] M. J. Neely, "Repeated games, optimal channel capture, and open problems for slotted multiple access," *arXiv preprint arXiv:2110.09638*, 2021.

[12] S. Yüksel and T. Basar, "Stochastic networked control systems: Stabilization and optimization under information constraints.," *Springer Science & Business Media*, 2013.

[13] D. Kartik and A. Nayyar, "Upper and lower values in zero-sum stochastic games with asymmetric information," *Dynamic Games and Applications*, vol. 11, no. 2, pp. 363–388, 2021.

[14] D. Kartik, A. Nayyar, and U. Mitra, "Common information belief based dynamic programs for stochastic zero-sum games with competing teams," *arXiv preprint arXiv:2102.05838*, 2021.

[15] S. Sudhakara and A. Nayyar, "Optimal symmetric strategies in multi-agent systems with decentralized information," *arXiv preprint arXiv:2307.07150*, 2023.

[16] Y.-C. Ho, "Team decision theory and information structures," *Proceedings of the IEEE*, vol. 68, no. 6, pp. 644–654, 1980.