Principled Learning-to-Communicate with Quasi-Classical Information Structures

Xiangyu Liu[†] Haoyi You[†] Kaiqing Zhang[†]

Abstract

Learning-to-communicate (LTC) in partially observable environments has received increasing attention in deep multi-agent reinforcement learning, where the control and communication strategies are jointly learned. On the other hand, the impact of communication on decision-making has been extensively studied in control theory. In this paper, we seek to formalize and better understand LTC by bridging these two lines of work, through the lens of *information structures* (ISs). To this end, we formalize LTC in decentralized partially observable Markov decision processes (Dec-POMDPs) under the common-information-based framework from decentralized stochastic control, and classify LTC problems based on the ISs before (additional) information sharing. We first show that non-classical LTCs are computationally intractable in general, and thus focus on quasi-classical (QC) LTCs. We then propose a series of conditions for QC LTCs, violating which can cause computational hardness in general. Further, we develop provable planning and learning algorithms for QC LTCs, and show that some examples of QC LTCs satisfying the above conditions can be solved with quasi-polynomial time and samples. Along the way, we also establish some relationship between (strictly) QC IS and the condition of having strategy-independent common-information-based beliefs (SI-CIBs), as well as a new result of solving Dec-POMDPs without computationally intractable oracles but beyond those with SI-CIBs, which may be of independent interest.

I. Introduction

The learning-to-communicate (LTC) problem has emerged and gained traction in the area of (deep) multi-agent reinforcement learning (MARL) [1, 2, 3]. Unlike classical MARL, which aims to learn *control* strategies that minimize the expected accumulated costs, LTC seeks to *jointly* minimize over both the *control* and the *communication* strategies of all the agents, as a way to mitigate the challenges due to the agents' partial observability of the environment. Despite the promising empirical successes, theoretical understandings of LTC remain largely underexplored.

On the other hand, in control theory, a rich literature has investigated the role of *communication* in decentralized/networked control [4, 5, 6, 7], inspiring us to rigorously examine LTC from such a principled perspective. Most of these studies, however, focused on linear systems, and did not explore the computational or sample complexity guarantees when the system model is not (fully) known. More recently, a few studies [8, 9] explored the settings with general discrete (non-linear) spaces, under special communication protocols and state transition dynamics.

More broadly, the design of communication strategies dictates the *information structure* (IS) of the control system, which characterizes *who knows what and when* [10]. IS and its impact on the *optimization tractability*, especially for linear systems, have been extensively studied in (decentralized)

[†]The authors are ordered alphabetically, and are affiliated with the University of Maryland, College Park, MD, USA, 20742. Emails: {xyliu999, yuriiyou, kaiqing}@umd.edu.

stochastic control, see [11, 12] for comprehensive overviews. In this work, we seek a more principled understanding of LTC through the lens of information structures, with a focus on the computational and sample complexities of the problem.

Specifically, we formalize LTC in the general framework of decentralized partially observable Markov decision processes (Dec-POMDPs) [13], as in the empirical studies [1, 2, 3]. To achieve finite-time and sample complexity guarantees, we resort to the recent development in [14] on partially observable MARL, based on the common-information-based (CIB) framework [15, 16] from decentralized stochastic control, to model the communication and information sharing protocols among agents. We detail our contributions as follows.

Contributions. (i) We formalize learning-to-communicate in Dec-POMDPs under the common-information-based framework [15, 16, 14], allowing the sharing of *historical* information, and the modeling of communication costs. (ii) We classify LTCs through the lens of *information structures*, according to the ISs before (additional) information sharing, i.e., the *baseline* sharing. We then show that LTCs with *non-classical* [11] IS of the baseline sharing can be computationally intractable. (iii) Given the hardness, we thus focus on *quasi-classical* (QC) LTCs, and propose a series of conditions under which LTCs preserve the QC IS after additional sharing, while violating which can cause computational hardness in general. (iv) We propose both planning and learning algorithms for QC LTCs, by reformulating them as Dec-POMDPs with *strategy-independent common-information-based beliefs* (SI-CIBs) [16], a condition shown to be critical for tractable computation and learning [14]. (v) Quasi-polynomial time and sample complexities of the algorithms are established for QC LTC examples that satisfy the conditions in (iii). Along the way, we also establish some relationship between *(strictly) quasi-classical* ((s)QC) ISs and the SI-CIB condition in the framework of [16] under certain assumptions, as well as a new result of solving general Dec-POMDPs without computationally intractable oracles but beyond those with SI-CIBs, which thus advances the results in [14]. These results may be of independent interest to studying general Dec-POMDPs. We conclude with some experimental results to validate the implementability and effectiveness of our algorithms.

I-A Related Work

Communication-control joint optimization. The joint design of control and communication strategies has been studied in the controls literature [6, 17, 18, 7, 8, 9]. However, even with model knowledge, the computational complexity (and associated necessary conditions) of solving these models remains elusive, let alone the sample complexity when it comes to learning. Moreover, these models mostly have special structures, e.g., with linear systems and sometimes specific, fixed communication strategies (e.g., event-triggered ones) [6, 17, 18, 7], or sharing only instantaneous observations [8, 9].

Information sharing and information structures. Information structure has been extensively studied to characterize *who knows what and when* in (decentralized) stochastic control [11, 12]. Our paper aims to formally understand LTC through the lens of information structures. The common-information-based approaches to formalize information sharing in [15, 16] serve as the basis for our work. In comparison, these results did not *optimize/learn* to share the information, and focused on the *structural results*, without concrete computational (and sample) complexity analysis.

Partially observable MARL theory. Planning and learning in partially observable MARL are known to be hard [19, 20, 21, 13]. Recently, [22, 23] developed polynomial-sample complexity algo-

rithms for partially observable stochastic games, but with computationally intractable oracles; [14] developed quasi-polynomial-time and sample algorithms for such models, leveraging information sharing among the agents. In contrast, our paper focuses on *optimizing/learning to share*, together with control strategy optimization/learning.

II. Preliminaries

Notation. We use $\mathbb{N}, \mathbb{Q}, \mathbb{R}$ to denote the sets of all the natural, rational, and real numbers, respectively. For an integer $m > 0$, we denote $[m] := \{1, 2, \cdots, m\}$. For a finite set $\mathcal{X}$, we use $|\mathcal{X}|$ to denote the cardinality of $\mathcal{X}$, and use $\Delta(\mathcal{X})$ to denote the probability simplex over $\mathcal{X}$. For a random variable X, we use $\sigma(X)$ to denote the sigma-algebra generated by X. For σ-algebras $\mathscr{F}_1$ on the space $\mathcal{X}_1$ and $\mathscr{F}_2$ on the space $\mathcal{X}_2$, we denote by $\mathscr{F}_1 \otimes \mathscr{F}_2$ the product σ-algebra on the space $\mathcal{X}_1 \times \mathcal{X}_2$. We use $\mathbb{1}[]$ to denote the indicator function. Unless otherwise noted, the set $\{\}$ considered is ordered, such that the elements in the set are indexed.

II-A Learning-to-Communicate Formulation

For $n > 1$ agents, a (cooperative) *learning-to-communicate* problem can be described by a tuple $\mathcal{L} = \langle H, \mathcal{S}, \{\mathcal{A}_{i,h}\}_{i\in[n],h\in[H]}, \{\mathcal{O}_{i,h}\}_{i\in[n],h\in[H]}, \{\mathcal{M}_{i,h}\}_{i\in[n],h\in[H]}, \mathbb{T}, \mathbb{O}, \mu_1, \{\mathcal{R}_h\}_{h\in[H]}, \{\mathcal{K}_h\}_{h\in[H]}\rangle$ where H denotes the length of each episode, and other components are introduced as follows.

II-A-1 Decision-making components

We use $\mathcal{S}$ to denote the state space, and $\mathcal{A}_{i,h}$ to denote the control action space of agent i at timestep $h \in [H]$. We denote by $s_h \in \mathcal{S}$ the state and by $a_{i,h}$ the control action of agent i at timestep h. We use $a_h := (a_{1,h}, \cdots, a_{n,h}) \in \mathcal{A}_h := \prod_{i\in[n]} \mathcal{A}_{i,h}$ to denote the joint control action of all the n agents at timestep h. We denote by $\mathbb{T} = \{\mathbb{T}_h\}_{h\in[H]}$ the collection of state transition kernels, where $s_{h+1} \sim \mathbb{T}_h(\cdot|s_h, a_h) \in \Delta(\mathcal{S})$ at timestep h. We use $\mu_1 \in \Delta(\mathcal{S})$ to denote the initial state distribution. We denote by $\mathcal{O}_{i,h}$ the observation space and by $o_{i,h} \in \mathcal{O}_{i,h}$ the observation of agent i at timestep h. We use $o_h := (o_{1,h}, o_{2,h}, \cdots, o_{n,h}) \in \mathcal{O}_h := \prod_{i\in[n]} \mathcal{O}_{i,h}$ to denote the joint observation of all the n agents at timestep h. We use $\mathbb{O} = \{\mathbb{O}_h\}_{h\in[H]}$ to denote the collection of emission functions, where $o_h \sim \mathbb{O}_h(\cdot|s_h) \in \Delta(\mathcal{O}_h)$ at timestep h and state $s_h \in \mathcal{S}$. Also, we denote by $\mathbb{O}_{i,h}(\cdot|s_h)$ the emission for agent i, the marginal distribution of $o_{i,h}$ given $\mathbb{O}_h(\cdot|s_h)$ for all $s_h \in \mathcal{S}$. At each timestep h, agents will receive a common reward $r_h = \mathcal{R}_h(s_h, a_h)$, where $\mathcal{R}_h : \mathcal{S} \times \mathcal{A}_h \to [0, 1]$ denotes the reward function shared by the agents.

II-A-2 Communication components

In addition to reward-driven decision-making, agents also need to decide and learn *(what) to communicate with others*. At timestep h, agents share part of their information $z_h \in \mathcal{Z}_h$ with other agents, where $\mathcal{Z}_h$ denotes the collection of all possible shared information at timestep h. Here we consider a general setting where the shared information z_h may contain two parts, the *baseline-sharing* part z_h^b that comes from some existing sharing protocol among agents, and the *additional-sharing* part $z_{i,h}^a$ for each agent i that comes from explicit communication *to be decided/learned*, with the joint additional-sharing information $z_h^a := \cup_{i=1}^n z_{i,h}^a$. This general setting covers those considered in most empirical studies on LTC [1, 2, 3], with a void baseline sharing part. We kept the baseline sharing since our focus is on the *finite-time* and *sample* tractability of LTC, for which a certain amount of information sharing is known to be necessary [14]. Note that $z_h = z_h^b \cup z_h^a$ and $z_h^b \cap z_h^a = \emptyset$. The shared information

is part of the historical observations and (both *control* and *communication*) actions. We denote by $\mathcal{Z}^b_h, \mathcal{Z}^a_h$, and $\mathcal{Z}^a_{i,h}$ the collections of all possible z^b_h, z^a_h, and $z^a_{i,h}$ at timestep h, respectively.

At timestep h, the *common information* among all the agents is thus defined as the union of all the *shared information* so far: $c_{h^-} = \cup_{t=1}^{h-1} z_t \cup z^b_h$, and $c_{h^+} = \cup_{t=1}^{h} z_t$, where c_{h^-} and c_{h^+} denote the (accumulated) common information *before* and *after* additional sharing, respectively. The *private information* of agent i at timestep h *before* and *after* additional sharing are denoted by p_{i,h^-} and p_{i,h^+}, respectively, where $p_{i,h^-} \subseteq \{o_{i,1}, a_{i,1}, \cdots, a_{i,h-1}, o_{i,h}\} \backslash c_{h^-}, p_{i,h^+} \subseteq \{o_{i,1}, a_{i,1} \cdots, a_{i,h-1}, o_{i,h}\} \backslash c_{h^+}$. We denote by $p_{h^-} := (p_{1,h^-}, \cdots, p_{n,h^-})$ the joint private information *before* additional sharing, by $p_{h^+} := (p_{1,h^+}, \cdots, p_{n,h^+})$ the joint private information *after* additional sharing, at timestep h. We then denote by $\tau_{i,h^-} := p_{i,h^-} \cup c_{h^-}, \tau_{i,h^+} := p_{i,h^+} \cup c_{h^+}$ the *information available* to agent i at timestep h, before and after additional sharing, respectively, with $\tau_{h^-} := p_{h^-} \cup c_{h^-}, \tau_{h^+} := p_{h^+} \cup c_{h^+}$ denoting the associated joint information. We use $\mathcal{C}_{h^-}, \mathcal{C}_{h^+}, \mathcal{P}_{i,h^-}, \mathcal{P}_{i,h^+}, \mathcal{P}_{h^-}, \mathcal{P}_{h^+}, \mathcal{T}_{i,h^-}, \mathcal{T}_{i,h^+}, \mathcal{T}_{h^-}, \mathcal{T}_{h^+}$ to denote, respectively, the corresponding collections of all possible $c_{h^-}, c_{h^+}, p_{i,h^-}, p_{i,h^+}, p_{h^-}, p_{h^+}, \tau_{i,h^-}, \tau_{i,h^+}, \tau_{h^-}, \tau_{h^+}$.

We use $m_{i,h}$ to denote the *communication action* of agent i at timestep h, and it will determine what information $z^a_{i,h}$ she will share, through the way specified later. We denote by $\mathcal{M}_{i,h}$ the space of $m_{i,h}$, and by $m_h := (m_{1,h}, \cdots, m_{n,h}) \in \mathcal{M}_h := \prod_{i=1}^n \mathcal{M}_{i,h}$ the joint communication action of all the agents. $\mathcal{K}_h : \mathcal{Z}^a_h \to [0,1]$ denotes the *communication cost* function, and $\kappa_h = \mathcal{K}_h(z^a_h)$ denotes the incurred communication cost at timestep h, due to additional sharing.

II-A-3 System evolution

The system evolves by alternating between the communication and the control steps as follows.

Communication step: At each timestep h, each agent i observes $o_{i,h}$ and may share part of her private information via baseline sharing, receives the baseline sharing of information from others, and forms p_{i,h^-} and c_{h^-}. Then, each agent i chooses her communication action, which determines the additional sharing of information, receives the additional-sharing of information from others, forms p_{i,h^+} and c_{h^+}, and incurs some communication cost κ_h. Formally, the evolution of information is depicted as follows, which, unless otherwise noted, will be assumed throughout the paper. We follow the convention that any quantity at $h = 0$ is empty/null.

Assumption II.1 (*Information evolution*)**.** For each $h \in [H]$,

(a) (Baseline sharing). $z^b_h = \chi_h(p_{(h-1)^+}, a_{h-1}, o_h)$ for some fixed transformation χ_h;

(b) (Additional sharing). For each agent $i \in [n], z^a_{i,h} = \phi_{i,h}(p_{i,h^-}, m_{i,h})$ for some function $\phi_{i,h}$, given communication action $m_{i,h}$, and $m_{i,h} \in z^a_{i,h}$; and the joint sharing $z^a_h := \cup_{i\in[n]} z^a_{i,h}$ is thus generated by $z^a_h = \phi_h(p_{h^-}, m_h)$, for some function ϕ_h;

(c) (Private information before sharing). For each agent $i \in [n]$, $p_{i,h^-} = \xi_{i,h}(p_{i,(h-1)^+}, a_{i,h-1}, o_{i,h})$ for some fixed transformation $\xi_{i,h}$, and the joint private information thus evolves as $p_{h^-} = \xi_h(p_{(h-1)^+}, a_{h-1}, o_h)$ for some fixed transformation ξ_h;

(d) (Private information after sharing). For each agent $i \in [n]$, $p_{i,h^+} = p_{i,h^-} \backslash z^a_{i,h}$;

(e) ($(\tau_{i,h^-}, \tau_{i,h^+})$-inclusion). For each agent $i \in [n]$, $\tau_{i,h^-} \subseteq \tau_{i,h^+} \subseteq \tau_{i,(h+1)^-}$, and $o_{i,h} \in \tau_{i,h^-}$.

Note that as *fixed transformations* (e.g., χ_h and $\xi_{i,h}$ above), they are not affected by the *realized values* of the random variables, but dictate some *pre-defined* transformation of the input random variables. See [15, 16], and [14] for common examples of baseline sharing that admit such fixed transformations when there is no additional sharing, and examples in §A on how they can be extended to

the LTC setting. It should not be confused with some general *function* (e.g., $\phi_{i,h}$ above), which may depend on the *realized values* of the input random variables. (a) and (c) on baseline sharing follow from those in [16, 14]; (b) and (d) on additional sharing dictate how the communication action affects the sharing based on private information. For example, a common choice of $(\mathcal{M}_{i,h}, \phi_{i,h})$ is that $\mathcal{M}_{i,h} = \{0,1\}^{|p_{i,h^-}|}$, for any $p_{i,h^-} \in \mathcal{P}_{i,h^-}$ and $m_{i,h} \in \mathcal{M}_{i,h}$, $\phi_{i,h}(p_{i,h^-}, m_{i,h})$ consists of the k-th element (with $k \in [|p_{i,h^-}|]$) of p_{i,h^-} if and only if the k-th element of $m_{i,h}$ is 1, while other elements are 0. As $m_{i,h}$ (dictating what to share) will be known given $z^a_{i,h}$ (what has been shared), $m_{i,h}$ is thus also modeled as being shared, i.e., $m_{i,h} \in z^a_{i,h}$. This is also consistent with the models in [8, 9] on control/communication joint optimization. (e) means that the agent has full memory of the information she had in the past and at present. We emphasize that this is closely related, but different from the common notion of *perfect recall* [24], where the agent has to recall all her own *past actions*. Condition (e), in contrast, relaxes the memorization of the actions, but includes the instantaneous observation $o_{i,h}$. This condition is satisfied by all the models and examples in [11, 15, 16, 14]. See also §A for more examples that satisfy this assumption. Note that $o_{i,h} \in \tau_{i,h^-}$ was noted necessary in order to have *closed-loop* ISs in [12], which are the focus of the present paper.

Meanwhile, for both the baseline and additional sharing protocols, we follow the model in the series of studies on partial history/information sharing [15, 16, 14, 8, 9] that, if an agent shares, she will share the information with *all other* agentsas *common information*. Additionally, we follow the convention from the literature on information structures [11, 12], by incorporating the σ-algebra of the random variables. These conventions lead to the following regularity assumption on information sharing.

Assumption II.2. $\forall i_1, i_2 \in [n], h_1, h_2 \in [H], i_1 \neq i_2, h_1 < h_2$, if $\sigma(o_{i_1,h_1}) \subseteq \sigma(\tau_{i_2,h_2^-})$, then $\sigma(o_{i_1,h_1}) \subseteq \sigma(c_{h_2^-})$, and if $\sigma(a_{i_1,h_1}) \subseteq \sigma(\tau_{i_2,h_2^-})$, then $\sigma(a_{i_1,h_1}) \subseteq \sigma(c_{h_2^-})$; if $\sigma(o_{i_1,h_1}) \subseteq \sigma(\tau_{i_2,h_2^+})$, then $\sigma(o_{i_1,h_1}) \subseteq \sigma(c_{h_2^+})$, and if $\sigma(a_{i_1,h_1}) \subseteq \sigma(\tau_{i_2,h_2^+})$, then $\sigma(a_{i_1,h_1}) \subseteq \sigma(c_{h_2^+})$.

Assumptions II.1-II.2 will be made throughout the paper.

Decision-making step: After the communication, each agent i chooses her control action $a_{i,h}$, receives a reward r_h, and the joint action a_h drives the state to $s_{h+1} \sim \mathbb{T}_h(\cdot|s_h, a_h)$.

II-A-4 Strategies and solution concept

At timestep h, each agent i has two strategies, a *control* strategy and a *communication* strategy. We define a control strategy as $g^a_{i,h} : \mathcal{T}_{i,h^+} \to \mathcal{A}_{i,h}$ and a communication strategy as $g^m_{i,h} : \mathcal{T}_{i,h^-} \to \mathcal{M}_{i,h}$. See Lemma G.1 for a formal argument on the use of such *deterministic* strategies without loss of optimality. We denote by $g^a_h = (g^a_{1,h}, \cdots, g^a_{n,h})$ the joint control strategy and by $g^m_h = (g^m_{1,h}, \cdots, g^m_{n,h})$ the joint communication strategy. We denote by $\mathcal{G}^a_{i,h}, \mathcal{G}^m_{i,h}, \mathcal{G}^a_h, \mathcal{G}^m_h$ the corresponding spaces of $g^a_{i,h}, g^m_{i,h}, g^a_h, g^m_h$, respectively.

The objective of the agents in LTC is to maximize the expected accumulated sum of the reward and the negative communication cost from timestep $h = 1$ to H:

$$J_{\mathcal{L}}(g^a_{1:H}, g^m_{1:H}) := \mathbb{E}_{\mathcal{L}}\left[\sum_{h=1}^{H}(r_h - \kappa_h)\middle| g^a_{1:H}, g^m_{1:H}\right],$$

where the expectation $\mathbb{E}_{\mathcal{L}}$ is taken over all the randomness in the system evolution, given the strategies $(g^a_{1:H}, g^m_{1:H})$. With this objective, for any $\epsilon \geq 0$, we can define the solution concept of an *ϵ-team optimum* for $\mathcal{L}$ as follows.

Definition II.3 (ϵ-team optimum)**.** We call a joint strategy $(g^a_{1:H}, g^m_{1:H})$ an ϵ-team optimal strategy of the LTC $\mathcal{L}$ if

$$\max_{\widetilde{g}^a_{1:H} \in \mathcal{G}^a_{1:H}, \widetilde{g}^m_{1:H} \in \mathcal{G}^m_{1:H}} J_{\mathcal{L}}(\widetilde{g}^a_{1:H}, \widetilde{g}^m_{1:H}) - J_{\mathcal{L}}(g^a_{1:H}, g^m_{1:H}) \leq \epsilon.$$

II-B Information Structures of LTC

In decentralized stochastic control, the notion of information structure [25, 11] captures *who knows what and when* as the system evolves. In LTC, as the additional sharing via communication will also affect the IS and is *not* determined *beforehand*, when we discuss the *IS of an LTC problem*, we will refer to that of the problem *with only baseline sharing*. In particular, an LTC $\mathcal{L}$ without additional sharing is essentially a Dec-POMDP (with potential baseline information sharing), as defined in §F for completeness. We formally define such a Dec-POMDP *induced* by $\mathcal{L}$ as follows.

Definition II.4 (Dec-POMDP (with information sharing) induced by LTC)**.** For an LTC $\mathcal{L} = \langle H, \mathcal{S}, \{\mathcal{A}_{i,h}\}_{i\in[n],h\in[H]}, \{\mathcal{O}_{i,h}\}_{i\in[n],h\in[H]}, \{\mathcal{M}_{i,h}\}_{i\in[n],h\in[H]}, \mathbb{T}, \mathbb{O}, \mu_1, \{\mathcal{R}_h\}_{h\in[H]}, \{\mathcal{K}_h\}_{h\in[H]}\rangle$, we call a Dec-POMDP (with information sharing) $\overline{\mathcal{D}}_{\mathcal{L}}$ *the Dec-POMDP (with information sharing) induced by* $\mathcal{L}$ if the agents share information only following the baseline sharing protocol of $\mathcal{L}$, i.e., without additional sharing, which can be characterized by the tuple $\overline{\mathcal{D}}_{\mathcal{L}} := \langle H, \mathcal{S}, \{\mathcal{A}_{i,h}\}_{i\in[n],h\in[H]}, \{\mathcal{O}_{i,h}\}_{i\in[n],h\in[H]}, \mathbb{T}, \mathbb{O}, \mu_1, \{\mathcal{R}_h\}_{h\in[H]}\rangle$. We may refer to it as the *Dec-POMDP induced by LTC* or the *induced Dec-POMDP* for short.

In §II-A, we introduced LTC in the *state-space model*. In contrast, information structure is oftentimes more conveniently discussed under the equivalent framework of *intrinsic models* [25] (see the instantiation for Dec-POMDPs in §F for completeness). In an intrinsic model, each agent only *acts once* throughout the system evolution, and the same agent in the state-space model at different timesteps is now treated as *different agents*. There are thus $n \times H$ agents in total. Formally, for completeness, we extend the intrinsic-model-based reformulation to LTCs in §F.

(Strictly) quasi-classical ISs are important subclasses of ISs, which were first introduced for decentralized stochastic control [25, 26, 12] (see the instantiation for Dec-POMDPs in §F). An IS that is not QC is *non-classical* [11, 12]. We extend such a categorization to LTC problems with different ISs as follows.

Definition II.5 ((Strictly) quasi-classical LTC)**.** We call an LTC $\mathcal{L}$ *(strictly) quasi-classical* if the Dec-POMDP induced by $\mathcal{L}$ (see Definition II.4) is *(strictly) quasi-classical*. Namely, each agent in the intrinsic model of $\overline{\mathcal{D}}_{\mathcal{L}}$ knows the information (and the actions) of the agents who influence her, either directly or indirectly.

Similarly, an LTC $\mathcal{L}$ that is not QC is called *non-classical*. See §A for examples of QC and sQC LTCs. Note that the categorization above is defined based on the ISs *before* additional sharing, as an inherent property of the LTC problem, since additional sharing is the solution *to be* decided/learned. We focus on finding such a solution next.

III. Hardness and Structural Assumptions

It is known that computing an (approximate) team-optimum in Dec-POMDPs, which are LTCs *without* information-sharing, is `NEXP-hard` [13]. The hardness cannot be fully circumvented even when agents are allowed to share information: even if agents share all the information, the LTC problem

becomes a Partially Observable Markov Decision Process (POMDP), which is known to be PSPACE-hard [19, 20]. Hence, additional assumptions are necessary to make LTCs computationally more tractable. We introduce several such assumptions and their justifications below, whose proofs can be found in §B.

Recently, [27] showed that *observable* POMDPs [28], a class of POMDPs with relatively *informative* observations, admit *quasi-polynomial time* algorithms to solve. Such a condition was then extended to Dec-POMDPs with information sharing in [14], which also developed quasi-polynomial time and sample complexity algorithms.

As solving LTCs is at least as hard as solving the Dec-POMDPs considered in [14], we first also make such an observability assumption on the *joint* emission function as in [14], to potentially avoid computationally intractable oracles.

Assumption III.1 (γ-observability [28, 27, 14])**.** There exists a $\gamma > 0$ such that $\forall h \in [H]$, the emission $\mathbb{O}_h$ satisfies that $\forall b_1, b_2 \in \Delta(\mathcal{S})$, $\left\|\mathbb{O}_h^\top b_1 - \mathbb{O}_h^\top b_2\right\|_1 \geq \gamma \left\|b_1 - b_2\right\|_1$.

However, we show next that, Assumption III.1 is not enough when it comes to LTC, if the baseline sharing IS is not favorable, and in particular, *non-classical* [11]. The hardness persists even under a few additional assumptions to be introduced later that will make LTCs more tractable.

Lemma III.2 (Non-classical LTCs are hard)**.** For non-classical LTCs under Assumptions III.1, III.4, III.5, and III.7, finding an $\frac{\epsilon}{H}$-team optimum is PSPACE-hard.

Note that the hardness comes from the intuition that, when communication costs are high, the additional sharing from LTC will be limited, preventing the upgrade of the IS from a non-classical one to a (quasi-)classical one, which is hard with only the *joint* observability of the emission (see Assumption III.1), even along with several other assumptions.

By Lemma III.2, we will hence focus on the *quasi-classical* LTCs hereafter. Indeed, QC is also known to be critical for efficiently solving *continuous-space* and *linear* decentralized control [29, 30]. However, quasi-classicality may not be sufficient for LTCs, since the additional sharing may *break* the QC IS, and introduce computational hardness, as argued below.

Firstly, the breaking of QC IS may result from the *communication strategies*. Specifically, the communication strategy space in §II-A-4 allows the dependence on agents' *private information*, which introduces incentives for *signaling* [11] and can also cause computational hardness, as shown next.

Lemma III.3 (QC LTCs with full-history-dependent communication strategies are hard)**.** For QC LTCs under Assumption III.1, together with Assumptions III.5, and III.7, computing a team-optimum in the general space of $(\mathcal{G}^a_{1:H}, \mathcal{G}^m_{1:H})$ with $\mathcal{G}^m_{i,h} := \{g^m_{i,h} : \mathcal{T}_{i,h^-} \to \mathcal{M}_{i,h}\}$ is NP-hard.

The hardness in Lemma III.3 originates from the fact that when depending on the private/local information, determining the communication action can be made as a *Team Decision problem* (TDP) [31], which is known to be hard. This will be the case even when the instantaneous observations are relatively observable (see Assumptions III.1-III.7).

To avoid this hardness, we thus focus on communication strategies that only condition on the *common information*. Intuitively, this assumption is not unreasonable, as it means that *which historical information to share* is determined by *what has been shared* (in the common information). Note that, this does not lose the generality in the sense that the private information p_{i,h^-} *can still be shared*. It only means that the communication action is not determined based on p_{i,h^-}, and the additional sharing is still dictated by $z^a_{i,h} = \phi_{i,h}(p_{i,h^-}, m_{i,h})$ (see Assumption II.1), depending on p_{i,h^-}.

Assumption III.4 (Common-information-based communication strategy)**.** The communication strategies take *common information* as input, with the following form:

$$\forall i \in [n], h \in [H], \quad g^m_{i,h} : \mathcal{C}_{h^-} \to \mathcal{M}_{i,h}. \tag{III.1}$$

Secondly, the breaking of QC may result from the *control strategies*: if some agent did *not* influence others in the baseline sharing (and thus these other agents did *not* have to access the agent's available information, while still satisfying QC), while she starts to influence others by *sharing* her (*useless*) *control* actions, this will make her *control strategies* relevant. We make the following two assumptions to avoid the related pessimistic cases, each followed by a computational hardness result when (only) the condition is missing.

Specifically, sometimes the action of some agents may not influence the *state transition*. Such actions are thus *useless* in terms of decision-making, when there is *no* information sharing. However, if they were deemed *non-influential*, but shared via additional sharing, then QC may break for the LTC problem. We thus make the following assumption.

Assumption III.5 (Control-useless action is not used)**.** $\forall i \in [n], h \in [H]$, if agent i's action $a_{i,h}$ does not influence the state s_{h+1}, namely, $\forall s_h \in \mathcal{S}, a_h \in \mathcal{A}_h, a'_{i,h} \in \mathcal{A}_{i,h}, a'_{i,h} \neq a_{i,h}, \mathbb{T}_h(\cdot \mid s_h, a_h) = \mathbb{T}_h(\cdot \mid s_h, (a'_{i,h}, a_{-i,h}))$. Then, $\forall h' > h$, the random variable $a_{i,h} \notin \tau_{h'^-}$ and $a_{i,h} \notin \tau_{h'^+}$.

Lemma III.6 (QC LTCs without Assumption III.5 are hard)**.** For QC LTCs under Assumptions III.1, III.4, and III.7, finding a team-optimum is still `NP-hard`.

Note that other than the justification above based on computational hardness, Assumption III.5 has been *implicitly* made in the IS examples in the literature when there are *uncontrolled* state dynamics, see e.g., [16, 14]. Moreover, we emphasize that for common cases where actions *do* affect the state transition, this assumption becomes unnecessary.

Other than not influencing state transition, an action may also be *non-influential* if the emission functions of other agents are *degenerate*: they cannot *sense* the influence from previous agents' actions. We thus make the following assumption on the emissions, followed by a justification result.

Assumption III.7 (Other agents' emissions are non-degenerate)**.** $\forall h \in [H], i \in [n]$, $\mathbb{O}_{-i,h}$ satisfies $\forall b_1, b_2 \in \Delta(\mathcal{S})$ such that $b_1 \neq b_2$, $\mathbb{O}^\top_{-i,h} b_1 \neq \mathbb{O}^\top_{-i,h} b_2$.

Lemma III.8 (QC LTCs without Assumption III.7 are hard)**.** For QC LTCs under Assumptions III.1, III.4, and III.5, finding an ϵ/H-team optimum is still `PSPACE-hard`.

We have justified the above assumptions by showing that missing one of them may cause computational intractability of LTCs in general. More importantly, as we will show later, as another justification, LTCs under Assumptions III.4, III.5, and III.7 can indeed *preserve* the QC/sQC information structure *after* additional sharing, making it possible for the overall LTC problem to be computationally more tractable. More examples that satisfy these assumptions can also be found in §A.

IV. Solving LTC Problems Provably

We now study how to solve LTC provably, via either *planning* (with model knowledge) or *learning* (without model knowledge). The pipeline of our solution is shown in Figure 1, and proofs of the results can be found in §C.

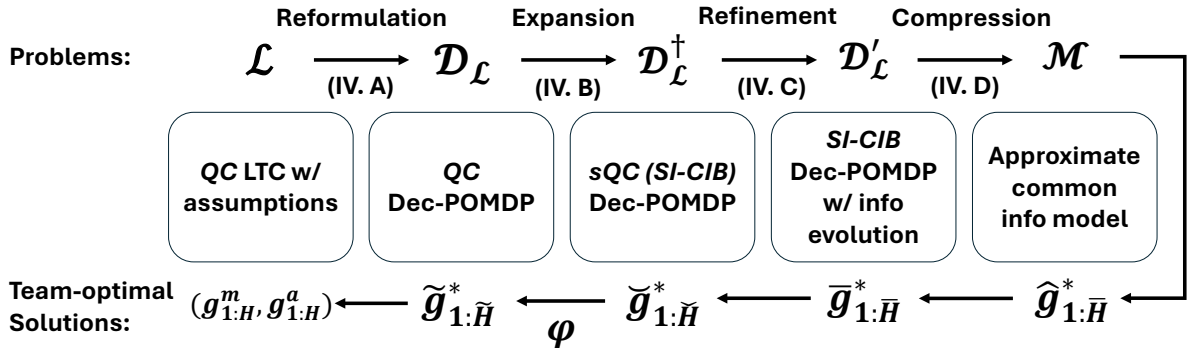


Figure 1: Illustrating the subroutines 1 for solving the LTC problems.

IV-A An Equivalent Dec-POMDP

Given any LTC $\mathcal{L}$, we can define a Dec-POMDP $\mathcal{D}_{\mathcal{L}}$ characterized by $\langle \widetilde{H}, \widetilde{\mathcal{S}}, \{\widetilde{\mathcal{A}}_{i,h}\}_{i\in[n],h\in[\widetilde{H}]}, \{\widetilde{\mathcal{O}}_{i,h}\}_{i\in[n],h\in[\widetilde{H}]}, \{\widetilde{\mathbb{T}}_h\}_{h\in[\widetilde{H}]}, \{\widetilde{\mathbb{O}}_h\}_{h\in[\widetilde{H}]}, \widetilde{\mu}_1, \{\widetilde{\mathcal{R}}_h\}_{h\in[\widetilde{H}]}\rangle$, such that these two are equivalent (under the assumptions in §III): $\forall h \in [H]$,

$$\begin{aligned}
&\widetilde{H} = 2H,\ \ \widetilde{\mathcal{S}} = \mathcal{S},\ \ \widetilde{s}_{2h-1} = \widetilde{s}_{2h} = s_h,\ \ \widetilde{\mathcal{A}}_{i,2h-1} = \mathcal{M}_{i,h},\ \ \widetilde{\mathcal{A}}_{i,2h} = \mathcal{A}_{i,h},\ \ \widetilde{\mathcal{O}}_{i,2h-1} = \mathcal{O}_{i,h},\ \ \widetilde{\mathcal{O}}_{i,2h} = \{\emptyset\},\ \ \widetilde{\mu}_1 = \mu_1,\\
&\widetilde{\mathbb{O}}_{2h-1} = \mathbb{O}_h,\ \ \widetilde{\mathbb{T}}_{2h-1}(\widetilde{s}_{2h}|\widetilde{s}_{2h-1}, \widetilde{a}_{2h-1}) = \mathbb{1}[\widetilde{s}_{2h} = \widetilde{s}_{2h-1}],\ \ \widetilde{\mathbb{T}}_{2h}(\widetilde{s}_{2h+1}|\widetilde{s}_{2h}, \widetilde{a}_{2h}) = \mathbb{T}_h(\widetilde{s}_{2h+1}|\widetilde{s}_{2h}, \widetilde{a}_{2h}),\\
&\widetilde{\mathcal{R}}_{2h-1} = -\mathcal{K}_h,\ \ \widetilde{\mathcal{R}}_{2h} = \mathcal{R}_h,\ \ \widetilde{p}_{i,2h-1} = \emptyset,\ \ \widetilde{p}_{i,2h} = p_{i,h^+},\ \ \widetilde{c}_{2h-1} = c_{h^-},\ \ \widetilde{c}_{2h} = c_{h^+},\ \ \widetilde{z}_{2h-1} = z^b_h,\ \ \widetilde{z}_{2h} = z^a_h, \quad \text{(IV.1)}
\end{aligned}$$

for all $(i,h) \in [n]\times[H]$, $s_h \in \mathcal{S}, a_{i,h} \in \mathcal{A}_{i,h}, o_{i,h} \in \mathcal{O}_{i,h}, m_{i,h} \in \mathcal{M}_{i,h}, p_{i,h^-} \in \mathcal{P}_{i,h^-}, p_{i,h^+} \in \mathcal{P}_{i,h^+}, c_{h^-} \in \mathcal{C}_{h^-}, c_{h^+} \in \mathcal{C}_{h^+}, \tau_{i,h^-} \in \mathcal{T}_{i,h^-}, \tau_{i,h^+} \in \mathcal{T}_{i,h^+}$. Note that we follow the convention of $\widetilde{\tau}_{i,h} := \widetilde{p}_{i,h} \cup \widetilde{c}_h$ for any $h \in [\widetilde{H}]$, and at the odd timestep $2t-1$ for any $t \in [H]$, we have $\widetilde{p}_{i,2t-1} = \emptyset$ under Assumption III.4, i.e., in $\mathcal{D}_{\mathcal{L}}$, each agent only uses the common information so far for decision-making at timestep $2h-1$. Correspondingly, for any $h \in [\widetilde{H}], i \in [n]$, we denote by $\widetilde{g}_{i,h}, \widetilde{g}_h$ the agent i's strategy and the joint strategy, respectively, and denote by $\widetilde{\mathcal{G}}_{i,h}, \widetilde{\mathcal{G}}_h$ their associated spaces. Moreover, to unify the presentation, we define that $\forall h \in [H]$, $\widetilde{r}_{2h-1} = \widetilde{\mathcal{R}}_{2h-1}(\widetilde{s}_{2h-1}, \widetilde{a}_{2h-1}, p_{2h-1}) := -\mathcal{K}_h(\phi_h(p_{2h-1}, \widetilde{a}_{2h-1}))$, with slight abuse of notation, where $p_{2h-1} := p_{h^-}$ can be viewed as part of the underlying state. Similarly, we define $\widetilde{r}_{2h} = \widetilde{\mathcal{R}}_{2h}(\widetilde{s}_{2h}, \widetilde{a}_{2h}, p_{2h}) := \mathcal{R}_h(\widetilde{s}_{2h}, \widetilde{a}_{2h})$, where $p_{2h} := p_{h^+}$. Hence, the objective of $\mathcal{D}_{\mathcal{L}}$ is defined as $J_{\mathcal{D}_{\mathcal{L}}}(\widetilde{g}_{1:\widetilde{H}}) = \mathbb{E}_{\mathcal{D}_{\mathcal{L}}}[\sum_{h=1}^{\widetilde{H}} \widetilde{r}_h | \widetilde{g}_{1:\widetilde{H}}]$.

Essentially, this reformulation splits the H-step control and communication decision-making procedure into a $2H$-step one. A similar splitting of the timesteps was also used in [8, 9]. In comparison, we consider a more general setting, where the state is not decoupled, and agents are allowed to share the observations and actions at the *previous* timesteps, due to the generality of our LTC formulation. The equivalence between $\mathcal{L}$ and $\mathcal{D}_{\mathcal{L}}$ is more formally stated as follows. [kz:where is the proof of this result?]

Proposition IV.1 (Equivalence between $\mathcal{L}$ and $\mathcal{D}_{\mathcal{L}}$)**.** Let $\mathcal{D}_{\mathcal{L}}$ be the reformulated Dec-POMDP from $\mathcal{L}$ satisfying Assumption III.4, then the solutions of the two problems are equivalent, in the sense that $\forall g^m_{1:H} \in \mathcal{G}^m_{1:H}, g^a_{1:H} \in \mathcal{G}^a_{1:H}, i \in [n]$, let $\widetilde{g}_{1:\widetilde{H}} = (g^m_1, g^a_1, \cdots, g^m_H, g^a_H)$, then $J_{\mathcal{D}_{\mathcal{L}}}(\widetilde{g}_{1:\widetilde{H}}) = J_{\mathcal{L}}(g^m_{1:H}, g^a_{1:H})$. Also, $\forall \widetilde{g}_{1:\widetilde{H}} \in \widetilde{\mathcal{G}}_{1:\widetilde{H}}, i \in [n]$, let $g^m_{1:H} = (\widetilde{g}_1, \widetilde{g}_3, \cdots, \widetilde{g}_{\widetilde{H}-1}), g^a_{1:H} = (\widetilde{g}_2, \widetilde{g}_4, \cdots, \widetilde{g}_{\widetilde{H}})$, then $J_{\mathcal{L}}(g^m_{1:H}, g^a_{1:H}) = J_{\mathcal{D}_{\mathcal{L}}}(\widetilde{g}_{1:\widetilde{H}})$.

Also, the Dec-POMDP $\mathcal{D}_{\mathcal{L}}$ preserves the QC information structure of $\mathcal{L}$.

Theorem IV.2 (Preserving (s)QC)**.** If $\mathcal{L}$ is (s)QC and satisfies Assumptions III.4, III.5, and III.7, then the reformulated Dec-POMDP $\mathcal{D}_{\mathcal{L}}$ is also (s)QC.

By Proposition IV.1, it suffices to solve the reformulated $\mathcal{D}_{\mathcal{L}}$ that are QC/sQC, which will be our focus next.

IV-B Strict Expansion of $\mathcal{D}_{\mathcal{L}}$

However, being QC does not necessarily imply $\mathcal{D}_{\mathcal{L}}$ can be solved *without* computationally intractable oracles. Note that this is different from the continuous-space, linear quadratic case, where QC problems can be reformulated and solved efficiently [29, 30]. With nonlinear, discrete spaces, the recent results in [14] established a concrete *quai-polynomial*-time complexity for planning, under the *strategy independence* assumption [16] on the common-information-based beliefs [15, 16]. This SI-CIB assumption was shown critical for *computational tractability* [14]: it eliminates the need to *enumerate* the past strategies in dynamic programming, which would otherwise be prohibitively large. Thus, we need to connect QC IS to the SI-CIB condition for better computational tractability.

Interestingly, under certain conditions, one can connect these two conditions for the reformulated Dec-POMDP $\mathcal{D}_{\mathcal{L}}$. As the first step, we will *expand* the QC $\mathcal{D}_{\mathcal{L}}$ by adding the *actions* of the agents who influence the later agents in the intrinsic model of $\mathcal{D}_{\mathcal{L}}$ to the shared information. We denote the strictly expanded Dec-POMDP as $\mathcal{D}^{\dagger}_{\mathcal{L}}$. We replace the $\widetilde{\ }$ notation in $\mathcal{D}_{\mathcal{L}}$ by the $\check{\ }$ notation in $\mathcal{D}^{\dagger}_{\mathcal{L}}$. All the elements remain the same, except the set of common information $\check{c}_h$: for any $h \in [\widetilde{H}]$

$$\check{c}_h = \widetilde{c}_h \cup \{\widetilde{a}_{j,t} \,|\, \forall j \in [n], t < h, \sigma(\widetilde{\tau}_{j,t}) \subseteq \sigma(\widetilde{c}_h), \widetilde{a}_{j,t} \text{ influences } \widetilde{s}_{t+1}\} \tag{IV.2}$$

and we follow the convention to define $\check{\tau}_{i,h} := \check{p}_{i,h} \cup \check{c}_h$ and $\check{z}_h = \check{c}_h \backslash \check{c}_{h-1}$. It is not hard to verify the following.

Lemma IV.3. If $\mathcal{D}_{\mathcal{L}}$ is QC, then $\mathcal{D}^{\dagger}_{\mathcal{L}}$ is sQC.

In contrast to the reformulation in §IV-A, the expansion here cannot guarantee the equivalence between $\mathcal{D}_{\mathcal{L}}$ and $\mathcal{D}^{\dagger}_{\mathcal{L}}$: the strategy spaces of $\mathcal{D}^{\dagger}_{\mathcal{L}}$ are larger than those of $\mathcal{D}_{\mathcal{L}}$, as each agent can now access more information, i.e., $\widetilde{\tau}_{i,h} \subseteq \check{\tau}_{i,h}$. Fortunately, the team-optimal value and strategy of both Dec-POMDPs are related, as shown in the following theorem.

Theorem IV.4. Let $\mathcal{D}_{\mathcal{L}}$ be the QC Dec-POMDP reformulated from a QC LTC $\mathcal{L}$, and $\mathcal{D}^{\dagger}_{\mathcal{L}}$ be the sQC expansion of $\mathcal{D}_{\mathcal{L}}$. Then, for any ϵ-team-optimal strategy $\check{g}^*_{1:\check{H}}$ of $\mathcal{D}^{\dagger}_{\mathcal{L}}$, there exists a function φ such that $\widetilde{g}^*_{1:\widetilde{H}} = \varphi(\check{g}^*_{1:\check{H}}, \mathcal{D}_{\mathcal{L}})$ is an ϵ-team-optimal strategy of $\mathcal{D}_{\mathcal{L}}$, with $J_{\mathcal{D}_{\mathcal{L}}}(\widetilde{g}^*_{1:\widetilde{H}}) = J_{\mathcal{D}^{\dagger}_{\mathcal{L}}}(\check{g}^*_{1:\check{H}})$.

Theorem IV.4 shows that one can solve the QC $\mathcal{D}_{\mathcal{L}}$ by first solving the sQC expansion $\mathcal{D}^{\dagger}_{\mathcal{L}}$, and then using an oracle φ to translate the solution back as a solution in the strategy spaces of $\mathcal{D}_{\mathcal{L}}$, without loss of optimality. Importantly, we show in Algorithm 4 how to implement such a φ function efficiently.

As shown below, a benefit of obtaining an *sQC* $\mathcal{D}^{\dagger}_{\mathcal{L}}$ is that, it also has *SI-CIBs*, making it possible to be solved without computationally intractable oracles as in [14].

Theorem IV.5. Let $\mathcal{D}^{\dagger}_{\mathcal{L}}$ be an sQC Dec-POMDP generated from $\mathcal{L}$ that satisfies Assumptions III.4, III.5 and III.7 after reformulation and strict expansion, then $\mathcal{D}^{\dagger}_{\mathcal{L}}$ has *strategy-independent common-information-based beliefs* [16, 14]. More formally, for any $h \in [\check{H}]$, any two different joint strategies $\check{g}_{1:h-1}$ and $\check{g}'_{1:h-1}$, and any common information $\check{c}_h$ that can be reached under both $\check{g}_{1:h-1}$ and $\check{g}'_{1:h-1}$, for any joint private information $\check{p}_h \in \check{\mathcal{P}}_h$ and state $\check{s}_h \in \check{\mathcal{S}}$,

$$\mathbb{P}^{\mathcal{D}^{\dagger}_{\mathcal{L}}}_h(\check{s}_h, \check{p}_h \,|\, \check{c}_h, \check{g}_{1:h-1}) = \mathbb{P}^{\mathcal{D}^{\dagger}_{\mathcal{L}}}_h(\check{s}_h, \check{p}_h \,|\, \check{c}_h, \check{g}'_{1:h-1}). \tag{IV.3}$$

IV-C Refinement of $\mathcal{D}^\dagger_\mathcal{L}$

Despite having SI-CIBs, $\mathcal{D}^\dagger_\mathcal{L}$ is still not eligible for applying the results in [14]: the information evolution rules of $\mathcal{D}^\dagger_\mathcal{L}$ break those in [16, 14]. Specifically, due to Assumption III.4, we set $\widetilde{\tau}_{i,2t-1} = \widetilde{c}_{2t-1}, \widetilde{p}_{i,2t-1} = \emptyset, \forall t \in [H], i \in [n]$ in $\mathcal{D}_\mathcal{L}$, which violates Assumption 1 in [16, 14]. To address this issue, we propose to further *refine* $\mathcal{D}^\dagger_\mathcal{L}$ to obtain a Dec-POMDP $\mathcal{D}'_\mathcal{L}$, which satisfies the information evolution rules. We replace the $\breve{}$ notation in $\mathcal{D}^\dagger_\mathcal{L}$ by the $\bar{}$ notation in $\mathcal{D}'_\mathcal{L}$. The elements in $\mathcal{D}'_\mathcal{L}$ remain the same as those in $\mathcal{D}^\dagger_\mathcal{L}$, except that the private information at odd steps is now refined as: for any $t \in [H], i \in [n], \overline{p}_{i,2t-1} := p_{i,t^-}$, and we define $\overline{\tau}_{i,2t-1} := \overline{p}_{i,2t-1} \cup \overline{c}_{2t-1}$ for any $t \in [H]$. Moreover, we define the reward functions as $\overline{r}_{2t-1} = \overline{\mathcal{R}}_{2t-1}(\overline{s}_{2t-1}, \overline{a}_{2t-1}, \overline{p}_{2t-1}) := -\mathcal{K}_t(\phi_t(\overline{p}_{2t-1}, \overline{a}_{2t-1}))$, and $\overline{r}_{2t} = \overline{\mathcal{R}}_{2t}(\overline{s}_{2t}, \overline{a}_{2t}, \overline{p}_{2t}) := \mathcal{R}_t(\overline{s}_{2t}, \overline{a}_{2t})$, for any $t \in [H]$. The new Dec-POMDP $\mathcal{D}'_\mathcal{L}$ is not equivalent to $\mathcal{D}^\dagger_\mathcal{L}$ in general, since it enlarges the strategy space at odd timesteps. However, if we define new strategy spaces in $\mathcal{D}'_\mathcal{L}$ as $\overline{\mathcal{G}}_{i,2t-1} : \overline{\mathcal{C}}_{2t-1} \to \overline{\mathcal{A}}_{i,2t-1}, \overline{\mathcal{G}}_{i,2t} : \overline{\mathcal{T}}_{i,2t} \to \overline{\mathcal{A}}_{i,2t}$ for each $t \in [H], i \in [n]$, and thus define $\overline{\mathcal{G}}_h$ to be the associated joint strategy space, then solving $\mathcal{D}^\dagger_\mathcal{L}$ is equivalent to finding a *best-in-class* team-optimal strategy of $\mathcal{D}'_\mathcal{L}$ within space $\overline{\mathcal{G}}_{1:\overline{H}}$, as shown below.

Theorem IV.6. Let $\mathcal{D}^\dagger_\mathcal{L}$ be an sQC Dec-POMDP generated from $\mathcal{L}$ after reformulation and strict expansion, and $\mathcal{D}'_\mathcal{L}$ be the refinement of $\mathcal{D}^\dagger_\mathcal{L}$ as introduced above. Then, finding the optimal strategy in $\mathcal{D}^\dagger_\mathcal{L}$ is equivalent to finding the optimal strategy of $\mathcal{D}'_\mathcal{L}$ in the space $\overline{\mathcal{G}}_{1:\overline{H}}$, and $\mathcal{D}'_\mathcal{L}$ satisfies the following information evolution rules: for each $h \in [\overline{H}]$:

$$\overline{c}_h = \overline{c}_{h-1} \cup \overline{z}_h, \quad \overline{z}_h = \overline{\chi}_h(\overline{p}_{h-1}, \overline{a}_{h-1}, \overline{o}_h) \qquad \text{for each } i \in [n], \quad \overline{p}_{i,h} = \overline{\xi}_{i,h}(\overline{p}_{i,h-1}, \overline{a}_{i,h-1}, \overline{o}_{i,h}),$$

with some functions $\{\overline{\chi}_h\}_{h\in[\overline{H}]}, \{\overline{\xi}_{i,h}\}_{i\in[n],h\in[\overline{H}]}$. Furthermore, if Assumptions III.4, III.5 and III.7 hold, then $\mathcal{D}'_\mathcal{L}$ has SI-CIBs with respect to the strategy space $\overline{\mathcal{G}}_{1:\overline{H}}$, i.e., for any $h \in [\overline{H}], \overline{s}_h \in \overline{\mathcal{S}}, \overline{p}_h \in \overline{\mathcal{P}}_h, \overline{c}_h \in \overline{\mathcal{C}}_h, \overline{g}_{1:h-1}, \overline{g}'_{1:h-1} \in \overline{\mathcal{G}}_{1:h-1}$ such that $\overline{c}_h$ is reachable under both $\overline{g}_{1:h-1}$ and $\overline{g}'_{1:h-1}$, it holds that

$$\mathbb{P}_h^{\mathcal{D}'_\mathcal{L}}(\overline{s}_h, \overline{p}_h | \overline{c}_h, \overline{g}_{1:h-1}) = \mathbb{P}_h^{\mathcal{D}'_\mathcal{L}}(\overline{s}_h, \overline{p}_h | \overline{c}_h, \overline{g}'_{1:h-1}). \tag{IV.4}$$

IV-D Planning in QC LTC with Finite-Time Complexity

Now we focus on how to solve the Dec-POMDP $\mathcal{D}'_\mathcal{L}$ that has SI-CIBs without computationally intractable oracles, building upon our results in [14]. Given a Dec-POMDP $\mathcal{D}'_\mathcal{L}$ with SI-CIBs, [14] proposed to construct an (ϵ_r, ϵ_z)-*expected*-approximate common information model $\mathcal{M}$ through *finite memory* (as defined in §C), when $\mathcal{D}'_\mathcal{L}$ is γ-observable. ϵ_r and ϵ_z here denote the approximation errors for rewards and incremental common information[kz:I dont think it is "transition"?], respectively, for which we defer a detailed introduction to §C. However, the Dec-POMDP $\mathcal{D}'_\mathcal{L}$ obtained from LTC has two key differences from the general ones considered in [14]. First, $\mathcal{D}'_\mathcal{L}$ does not satisfy the γ-observability assumption *throughout* the whole $\overline{H} = 2H$ timesteps. Fortunately, since the emissions at odd steps are still γ-observable, while those at even steps are unimportant as the states remain *unchanged* from the previous step, similar results of *belief contraction* and near-optimality of finite-memory truncation as in [14] can still be obtained[kz:pointer to where we have this result?]. Second, the rewards at odd steps can now depend on the *private information* $\overline{p}_h$, instead of the state $\overline{s}_h$. Thanks to the existence of some *consistent* approximate common-information-based beliefs $\{\mathbb{P}_h^{\mathcal{M},c}(\overline{s}_h, \overline{p}_h | \widehat{c}_h)\}_{h\in[\overline{H}]}$ (see Definition C.8), which provides the *joint* probability of $\overline{s}_h$ and $\overline{p}_h$ given the approximate common information $\widehat{c}_h$ compressed from $\overline{c}_h$, we can still properly evaluate the rewards

at the odd steps in the algorithms of [14]. Hence, we can leverage the approaches in [14] to develop a planning algorithm for QC LTC, which approximates the optimal strategy $\overline{g}^*_{1:\overline{H}}$ by finding an optimal prescription $\gamma^*_{1:\overline{H}}$ under each possible $\widehat{c}_{1:\overline{H}}$, with backward induction over the timesteps $h = \overline{H}, \cdots, 1$. See Algorithm 1 for a detailed introduction to the planning algorithm.

Note that in each step of the backward induction (Line 6 of Algorithm 6), a *Team Decision problem* [31] needs to be solved for each $\widehat{c}_h$, which is known to be NP-hard in general [31]:

$$\left(\widehat{g}^*_{1,h}(\cdot|\widehat{c}_h,\cdot),\cdots,\widehat{g}^*_{n,h}(\cdot|\widehat{c}_h,\cdot)\right) \leftarrow \operatorname*{argmax}_{\gamma_h} Q^{*,\mathcal{M}}_h(\widehat{c}_h,\gamma_h), \tag{IV.5}$$

where the Q-value function and the prescriptions γ_h are defined in §C[kz:but those are NOT for $Q^{*,\mathcal{M}}_h$, right? they are for the Q and V in the original problem, not the OPTIMAL ONE IN $\mathcal{M}$?]. Hence, to obtain overall computational tractability, we make the following *one-step* tractability assumption, as in [14].

Assumption IV.7 (One-step tractability of $\mathcal{M}$)**.** The one-step Team Decision problems induced by $\mathcal{M}$ (i.e., Line 6 of Algorithm 6) can be solved in polynomial time[1] for all $h = 2t, t \in [H]$.

Several remarks are in order regarding the assumption. First, it can be viewed as a *minimal* assumption when it comes to computational tractability: even with $H = 1$ and no LTC, one-step TDP requires additional structures in order to be solved efficiently. Second, since the Dec-POMDP here is reformulated from an LTC under Assumption III.4, it suffices to only assume one-step tractability for the *control* (i.e., even) steps, since at odd steps, the strategies do not use private information, and the one-step TDP can thus be solved in polynomial time by searching the maximizer given each $\widehat{c}_h$. [kz:pls learn and check.] Third, even without Assumption IV.7, the SI-CIB property of $\mathcal{D}'_{\mathcal{L}}$ and thus the derivation of *fixed*, *tractable size* dynamic programs to solve $\mathcal{L}$ efficiently still hold. Without such efforts, intractably many TDPs may need to be solved, leaving it less hopeful for computational tractability (even under Assumption IV.7). Finally, such an assumption is satisfied for several classes of Dec-POMDPs with information sharing, see Appendix G-B for more examples. With this assumption, we can obtain a concrete quasi-polynomial time complexity guarantee for LTC as follows. Proof of the theorem can be found in Appendix C-H.

Theorem IV.8. Given any QC LTC problem $\mathcal{L}$ satisfying Assumptions III.1, III.4, III.5, and III.7, we can construct a Dec-POMDP problem $\mathcal{D}'_{\mathcal{L}}$ with SI-CIBs such that for any $\epsilon > 0$, finding an ϵ-team optimal strategy in $\mathcal{D}'_{\mathcal{L}}$ can give us an ϵ-team optimal strategy of $\mathcal{L}$, and the following holds. Fix $\epsilon_r, \epsilon_z > 0$, and given any (ϵ_r, ϵ_z)-expected-approximate common information model $\mathcal{M}$ (see Definition C.6) for $\mathcal{D}'_{\mathcal{L}}$ that satisfies Assumption IV.7 and is consistent with some given approximate beliefs $\{\mathbb{P}^{\mathcal{M},c}_h(\overline{s}_h,\overline{p}_h|\widehat{c}_h)\}_{h\in[\overline{H}]}$ (see Definition C.8), there exists an algorithm that can compute a $(2\overline{H}\epsilon_r + \overline{H}^2\epsilon_z)$-team optimal strategy for the original LTC problem $\mathcal{L}$ with time complexity $\max_{h\in[\overline{H}]}|\widehat{\mathcal{C}}_h| \cdot \mathrm{poly}(|\overline{\mathcal{S}}|,|\overline{\mathcal{A}}_h|,|\overline{\mathcal{P}}_h|,\overline{H})$. In particular, fixed $\epsilon > 0$, if $\mathcal{L}$ has a baseline sharing protocol as one of the examples in §A, then Algorithm 1 can find an ϵ-team optimal strategy for $\mathcal{L}$ in quasi-polynomial time.

As sufficient conditions to ensure the construction of such an $\mathcal{M}$ that satisfies Assumption IV.7, as part of the definition, some examples in §A may need additional structural assumptions on the transition dynamics, emission, and reward/cost functions, while some do not. See §A and Remark C.16 for detailed discussions.

[1] By *polynomial time*, we here mean that the time-complexity depends polynomially on the LTC parameters $|\mathcal{S}|, |\mathcal{O}_h|, |\mathcal{A}_h|, |\mathcal{M}_h|, H$.

IV-E LTC with Finite-Time and Sample Complexities

Based on the planning result, we are now ready to solve the *learning* problem with both time and sample complexity guarantees. In particular, we can treat the samples from $\mathcal{L}$ as the samples from $\mathcal{D}'_{\mathcal{L}}$: the *reformulation* step (§IV-A) does not change the system dynamics, but only maps the information to different random variables; the *expansion* step (§IV-B) only requires agents to share more actions with each other, without changing the input and output of the environment; the *refinement* step (§IV-C) only recovers the private information the agents had in the original $\mathcal{L}$. This way, we can utilize similar algorithmic ideas in [14] to develop a learning algorithm for LTC problems. See §C for more details of the provable LTC algorithms adapted from [14]. The algorithm has the following finite-time and sample complexity guarantees. Proof of the theorem can be found in Appendix C-I.

Theorem IV.9. Given any QC LTC problem $\mathcal{L}$ satisfying Assumptions III.1, III.4, III.5, and III.7, we can construct a Dec-POMDP problem $\mathcal{D}'_{\mathcal{L}}$ with SI-CIBs. Moreover, given any compression functions of common information, there exists an LTC algorithm (Algorithm 2) learning in $\mathcal{D}'_{\mathcal{L}}$, such that if the learned expected-approximate-common-information models in $\widehat{\mathcal{M}}$ in the algorithm satisfy Assumption IV.7, then an approximate team-optimal strategy for $\mathcal{L}$ can be learned with high probability, with time and sample complexities polynomial in the parameters of the models in $\widehat{\mathcal{M}}$. Specifically, if $\mathcal{L}$ has a baseline sharing protocol as one of the examples in §A, then an ϵ-team optimal strategy for $\mathcal{L}$ can be learned with high probability, with both quasi-polynomial time and sample complexities.

Again, as sufficient conditions to ensure the learned models in $\widehat{\mathcal{M}}$ above to satisfy Assumption IV.7, some examples in §A were defined with additional structural assumptions on the transition emission, and reward/cost functions, while some were not. See §A and Remark C.19 for detailed discussions.

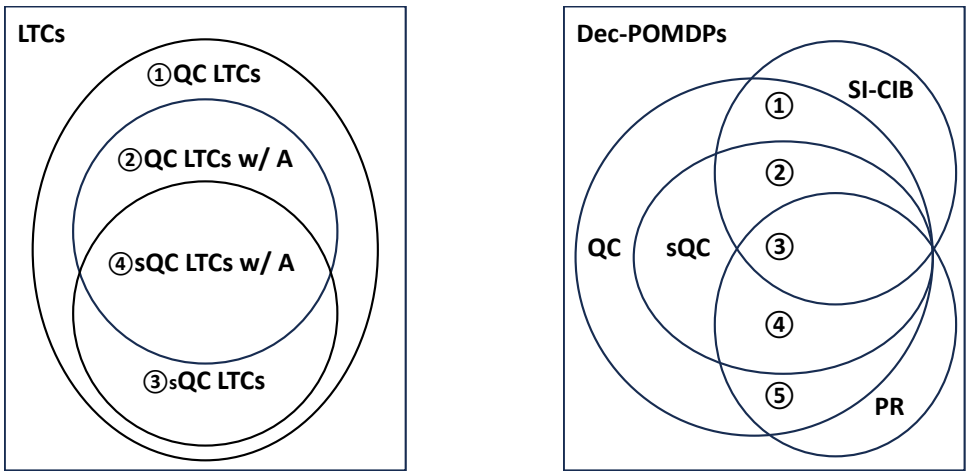
V. Solving General QC Dec-POMDPs

In §IV, we developed a pipeline for solving a special class of QC Dec-POMDPs generated by LTCs, by transforming them into those *with SI-CIBs*. In fact, the pipeline can also be extended to solving general QC Dec-POMDPs, which thus advances the results in [14] that can only address Dec-POMDPs with SI-CIBs, a result of independent interest. Without much confusion given the context, we will adapt the notation for LTCs to studying general Dec-POMDPs: we set $h^+ = h^- = h$ and void the additional sharing protocol. We extend the results in §IV to general QC Dec-POMDPs as follows.

Theorem V.1. Consider a Dec-POMDP $\mathcal{D}$ under Assumptions II.1 (e). If $\mathcal{D}$ is sQC and satisfies Assumption II.2, III.5, and III.7, then it has SI-CIBs. Meanwhile, if $\mathcal{D}$ has SI-CIBs and perfect recall, then it is sQC (up to null sets).

Perfect recall [24] here means that the agents will never forget their own past information and actions (as formally defined in §D). Note that Assumption II.1 (e) is similar to, but different from, perfect recall: it is implied by the latter with $o_{i,h} \in \tau_{i,h}$. Also, Assumptions III.5, III.7, and II.2 were originally made for LTCs, and here we meant to impose them for Dec-POMDPs with $h^+ = h^- = h$. Finally, by *sQC up to null sets*, we meant that if agent (i_1, h_1) influences agent (i_2, h_2) in the intrinsic model of the Dec-POMDP $\mathcal{D}$, then under any strategy $\overline{g}_{1:\overline{H}}$, $\sigma(\overline{\tau}_{i_1,h_1}) \subseteq \sigma(\overline{\tau}_{i_2,h_2})$ except the null sets generated by $\overline{g}_{1:\overline{H}}$, where we add $\overline{\ }$ for all the notation in the Dec-POMDP $\mathcal{D}$. Given Theorem V.1 and the results in §IV, we illustrate the relationship between LTCs and Dec-POMDPs with different assumptions and ISs in Figure 2, which may be of independent interest.

By Theorem V.1, one may start with a QC Dec-POMDP (with information sharing) $\mathcal{D}$ that does not necessarily have SI-CIBs, and then expand it (as §IV-B) to obtain an sQC Dec-POMDP $\mathcal{D}^\dagger$. If $\mathcal{D}$



(a) (b)

Figure 2: (a) Venn diagram of LTCs with different ISs: ① QC LTCs. ② QC LTCs satisfying Assumptions III.4, III.5, and III.7. ③ sQC LTCs. ④ sQC LTCs satisfying Assumptions III.4, III.5, and III.7, whose reformulated Dec-POMDPs have SI-CIBs; (b) Venn diagram of general Dec-POMDPs with different ISs. *PR* denotes *perfect recall*. We construct examples for each area in §H.

satisfies Assumptions II.2, III.5, and III.7, then $\mathcal{D}^\dagger$ has SI-CIBs and can then be solved with finite-time and sample complexity guarantees as in [14].[kz:can we point out to some "non-SI" examples that can be solved in this way (to show the value of this approach)? add some comments/explanation for the examples]

VI. Experimental Results

For the experiments, we validate both the implementability and performance of our LTC algorithms, and conduct ablation studies for LTCs with different communication costs and horizons.

Environment setup. We conduct our experiments on two popular and modest-scale partially observable benchmarks, Dectiger [32] and Grid3x3 [33]. We train the agents in each LTC problem in the two environments with 20 different random seeds and different communication cost functions, and execute them in problems with horizons $[4, 6, 8, 10]$. To fit the setting of LTC in our paper. We regularize the reward between [0,1] and set the base information structure as one-step-delay. As for the communication cost function, we set $\mathcal{K}_h(Z_h^a) = \alpha|Z_h^a|$, and set $\alpha \in [0.01, 0.05, 0.1]$ for the purpose of ablation study. Also, we study 2 baselines under the same environment with information structure of one-step delay and fully-sharing, respectively. The one-step-delay baseline can be regarded as an LTC problem with extremely high communication cost, thus no additional sharing. On the other hand, the fully-sharing baseline is the LTC problem with no communication cost.

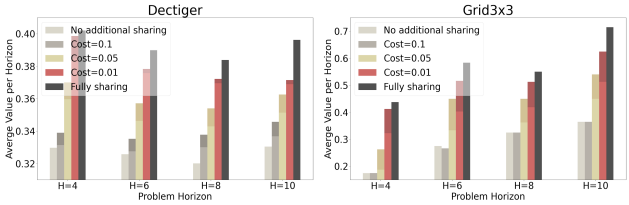
Results and analysis. The results of different horizons and communications costs over 20 random seeds are shown in Table 1 and Table 2. Additionally, the attained average-values are presented in Figure 3, and the learning curves are shown in Figure 4. The results show that communication is beneficial for agents to obtain higher values with better sample efficiency. Also, cheaper communication costs can encourage agents to share more information, and jointly achieve a better strategy.

Horizon/Cost	No Sharing	Cost=0.1	Cost=0.05	Cost=0.01	Fully Sharing
H=4 w/ cost	1.32±0.025	1.33±0.044	1.44±0.034	1.54±0.013	1.57±0.004
H=4 w/o cost	-	1.36±0.032	1.48±0.034	1.59±0.002	-
H=6 w/ cost	1.95±0.009	1.97±0.07	2.08±0.068	2.26±0.012	2.29±0.002
H=6 w/o cost	-	2.01±0.047	2.14±0.072	2.27±0.011	-
H=8 w/ cost	2.56±0.041	2.64±0.078	2.74±0.118	2.96±0.021	3.0±0.002
H=8 w/o cost	-	2.7±0.044	2.83±0.117	2.98±0.02	-
H=10 w/ cost	3.31±0.024	3.37±0.135	3.51±0.153	3.69±0.029	3.87±0.007
H=10 w/o cost	-	3.46±0.069	3.63±0.152	3.71±0.026	-

Table 1: Experimental results for Dectiger.

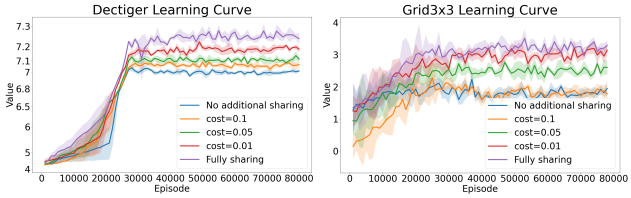
Horizon/Cost	No Sharing	Cost=0.1	Cost=0.05	Cost=0.01	Fully Sharing
H=4 w/ cost	0.14±0.003	0.14±0.019	0.15±0.002	0.26±0.028	-0.48±0.023
H=4 w/o cost	-	0.14±0.019	0.21±0.007	0.33±0.023	-
H=6 w/ cost	0.33±0.02	0.32±0.025	0.4±0.009	0.48±0.059	-0.38±0.075
H=6 w/o cost	-	0.32±0.025	0.54±0.02	0.62±0.075	-
H=8 w/ cost	0.52±0.084	0.52±0.051	0.58±0.072	0.67±0.031	-0.4±0.022
H=8 w/o cost	-	0.52±0.051	0.72±0.035	0.82±0.074	-
H=10 w/ cost	0.73±0.02	0.73±0.037	0.9±0.169	1.03±0.019	-0.15±0.188
H=10 w/o cost	-	0.73±0.037	1.08±0.14	1.25±0.062	-

Table 2: Experimental results for Grid3x3.



(a) (b)

Figure 3: The average-values achieved under different communication costs and horizons. Each full bar, the dark part, and the light part denote the values associated with the reward, the communication cost, and the overall objective (reward minus cost) of the agents, respectively.



(c) (d)

Figure 4: Performance under different communication costs and horizons. Each full bar, the dark part, and the light part denote the reward, communication cost, and overall objective, respectively.

VII. Concluding Remarks

We formalized the learning-to-communicate problem under the Dec-POMDP framework, and proposed a few structural assumptions for LTCs with quasi-classical information structures, violating which can cause computational hardness in general. We then developed provable planning and learning algorithms for QC LTCs. Along the way, we also established some relationship between the strictly quasi-classical information structure and the condition of having strategy-independent common-information-based beliefs, as well as solving general Dec-POMDPs without computationally intractable oracles beyond those with the SI-CIB condition. Our work has opened up many future directions, including the formulation, together with the development of provable planning/learning algorithms, of LTC in non-cooperative (game-theoretic) settings, and the relaxation of (some of) the structural assumptions when it comes to equilibrium computation.

Acknowledgement

The authors acknowledge the valuable feedback from the anonymous reviewers of IEEE CDC 2025, and the support from the Army Research Office (ARO) grant W911NF-24-1-0085 and the NSF CAREER Award 2443704. K.Z. also acknowledges the support from the AFOSR YIP Award FA9550-25-1-0258 and an AI Safety Research Award from Coefficient Giving.

References

[1] J. Foerster, I. A. Assael, N. De Freitas, and S. Whiteson, "Learning to communicate with deep multi-agent reinforcement learning," in *NeurIPS*, 2016.

[2] S. Sukhbaatar, R. Fergus, *et al.*, "Learning multiagent communication with backpropagation," in *NeurIPS*, 2016.

[3] J. Jiang and Z. Lu, "Learning attentional communication for multi-agent cooperation," in *NeurIPS*, 2018.

[4] S. Tatikonda and S. Mitter, "Control under communication constraints," *IEEE Trans. Autom. Control*, vol. 49, pp. 1056–1068, 2004.

[5] G. N. Nair, F. Fagnani, S. Zampieri, and R. J. Evans, "Feedback control under data rate constraints: An overview," *Proceed. of the IEEE*, vol. 95, pp. 108–137, 2007.

[6] L. Xiao, M. Johansson, H. Hindi, S. Boyd, and A. Goldsmith, "Joint optimization of wireless communication and networked control systems," *Switching and Learning Feedback Sys.*, pp. 248–272, 2005.

[7] S. Yüksel, "Jointly optimal LQG quantization and control policies for multi-dimensional systems," *IEEE Trans. Autom. Control*, vol. 59, pp. 1612–1617, 2013.

[8] 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.

[9] 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," in *IEEE Conf. on Dec. and Control*, 2022.

[10] H. S. Witsenhausen, "Separation of estimation and control for discrete time systems," *Proceed. of the IEEE*, vol. 59, pp. 1557–1566, 1971.

[11] A. Mahajan, N. C. Martins, M. C. Rotkowitz, and S. Yüksel, "Information structures in optimal decentralized control," in *IEEE Conf. on Dec. and Control*, 2012.

[12] S. Yüksel and T. Başar, *Stochastic Teams, Games, and Control under Information Constraints*. Springer Nature, 2023.

[13] D. S. Bernstein, R. Givan, N. Immerman, and S. Zilberstein, "The complexity of decentralized control of markov decision processes," *Math. Oper. Res.*, vol. 27, pp. 819–840, 2002.

[14] X. Liu and K. Zhang, "Partially observable multi-agent reinforcement learning with information sharing," *arXiv preprint arXiv:2308.08705 (accepted to SIAM Journal on Control and Optimization (SICON) 2025)*, 2023.

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

[16] A. Nayyar, A. Gupta, C. Langbort, and T. Başar, "Common information based Markov perfect equilibria for stochastic games with asymmetric information: Finite games," *IEEE Trans. Autom. Control*, vol. 59, pp. 555–570, 2013.

[17] L. Zhang and D. Hristu-Varsakelis, "Communication and control co-design for networked control systems," *Automatica*, vol. 42, no. 6, pp. 953–958, 2006.

[18] C. Peng and T. C. Yang, "Event-triggered communication and h_∞ control co-design for networked control systems," *Automatica*, vol. 49, no. 5, pp. 1326–1332, 2013.

[19] C. H. Papadimitriou and J. N. Tsitsiklis, "The complexity of Markov decision processes," *Math. Oper. Res.*, vol. 12, pp. 441–450, 1987.

[20] C. Lusena, J. Goldsmith, and M. Mundhenk, "Nonapproximability results for partially observable Markov decision processes," *J. Artif. Intell. Res.*, pp. 83–103, 2001.

[21] C. Jin, S. Kakade, A. Krishnamurthy, and Q. Liu, "Sample-efficient reinforcement learning of undercomplete pomdps," in *NeurIPS*, 2020.

[22] Q. Liu, C. Szepesvári, and C. Jin, "Sample-efficient reinforcement learning of partially observable Markov games," in *NeurIPS*, 2022.

[23] A. Altabaa and Z. Yang, "On the role of information structure in reinforcement learning for partially-observable sequential teams and games," in *NeurIPS*, 2024.

[24] H. W. Kuhn, "Extensive games and the problem of information," in *Contrib. Theory Games, Vol. II*. Princeton Univ. Press, 1953.

[25] H. S. Witsenhausen, "The intrinsic model for discrete stochastic control: Some open problems," in *Control Theory, Numer. Methods Comput. Syst. Model., Int. Symp., Rocquencourt*, 1975, pp. 322–335.

[26] A. Mahajan and S. Yüksel, "Measure and cost dependent properties of information structures," in *Amer. Control Conf.*, 2010, pp. 6397–6402.

[27] N. Golowich, A. Moitra, and D. Rohatgi, "Planning and learning in partially observable systems via filter stability," in *Proc. 55th Annu. ACM Symp. Theory Comput.*, 2023.

[28] E. Even-Dar, S. M. Kakade, and Y. Mansour, "The value of observation for monitoring dynamic systems," in *IJCAI*, 2007.

[29] Y.-C. Ho *et al.*, "Team decision theory and information structures in optimal control problems – part i," *IEEE Trans. Autom. Control*, vol. 17, pp. 15–22, 1972.

[30] A. Lamperski and L. Lessard, "Optimal decentralized state-feedback control with sparsity and delays," *Automatica*, pp. 143–151, 2015.

[31] J. Tsitsiklis and M. Athans, "On the complexity of decentralized decision making and detection problems," *IEEE Trans. Autom. Control*, vol. 30, pp. 440–446, 1985.

[32] R. Nair, M. Tambe, M. Yokoo, D. Pynadath, and S. Marsella, "Taming decentralized POMDPs: Towards efficient policy computation for multiagent settings," in *IJCAI*, 2003.

[33] C. Amato, J. Dibangoye, and S. Zilberstein, "Incremental policy generation for finite-horizon Dec-POMDPs," in *Proc. Int. Conf. Autom. Plan. Sched. (ICAPS)*, vol. 19, 2009, pp. 2–9.

[34] N. Golowich, A. Moitra, and D. Rohatgi, "Learning in observable pomdps, without computationally intractable oracles," in *NeurIPS*, 2022, pp. 1458–1473.

[35] J. Filar and K. Vrieze, *Competitive Markov decision processes*. Springer, 2012.

[36] Y. Bai and C. Jin, "Provable self-play algorithms for competitive reinforcement learning," in *ICML*, 2020.

[37] J. Peralez, A. Delage, O. Buffet, and J. S. Dibangoye, "Solving hierarchical information-sharing Dec-POMDPs: an extensive-form game approach," *arXiv preprint arXiv:2402.02954*, 2024.

[38] C. Boutilier, "Multiagent systems: Challenges and opportunities for decision-theoretic planning," *AI magazine*, vol. 20, pp. 35–35, 1999.

Appendices

A. Examples of QC LTC Problems

In this section, we introduce 8 examples of QC LTC problems, and 4 of them are extended from the information structures of the baseline sharing protocol considered in the literature [16, 14]. It can be shown that LTC with any of these 8 examples as baseline sharing is QC.

- **Example 1: One-step delayed information sharing:** At timestep $h \in [H]$, agents will share all the action-observation history in the private information until timestep $h-1$. Namely, for any $h \in [H], i \in [n], c_{h^-} = c_{(h-1)^+} \cup \{o_{h-1}, a_{h-1}\}$ and $p_{i,h^-} = \{o_{i,h}\}$.

- **Example 2: State controlled by one controller with asymmetric delayed information sharing:** The state dynamics are controlled by only one agent (without loss of generality, agent 1), i.e., $\forall h \in [H], \mathbb{T}_h(\cdot|s_h, a_{1,h}, a_{-1,h}) = \mathbb{T}_h(\cdot|s_h, a_{1,h}, a'_{-1,h})$ for all $s_h \in \mathcal{S}, a_{1,h} \in \mathcal{A}_{1,h}, a_{-1,h}, a'_{-1,h} \in \mathcal{A}_{-1,h}$, and the reward functions have additive structures, i.e., $\forall h \in [H], \mathcal{R}_h(s_h, a_h) = \sum_{i\in[n]} \mathcal{R}_{i,h}(s_h, a_{i,h})$ for all $s_h \in \mathcal{S}, a_h \in \mathcal{A}_h$. Agent 1 will share all of her information immediately, while others will share their information with a delay of $d \geq 1$ timesteps[2] in the baseline sharing. Namely, for any $h \in [H], i \neq 1$, $c_{h^-} = c_{(h-1)^+} \cup \{a_{1,h-1}, o_{1,h}, o_{-1,h-d}\}, p_{1,h^-} = \emptyset, p_{i,h^-} = (p_{i,(h-1)^+} \cup \{o_{i,h}\})\backslash\{o_{i,h-d}\}$.

- **Example 3: Information sharing with one-directional-one-step-delay:** For convenience, we assume there are 2 agents, and this example can be readily generalized to the multi-agent case. In this case, agent 1 will share the information immediately, while agent 2 will share information with one-step delay. Namely, $c_{1^-} = \{o_{1,1}\}, p_{1,1^-} = \emptyset, p_{2,1^-} = \{o_{2,1}\}$; for any $h \geq 2, i \in [n], c_{h^-} = c_{(h-1)^+} \cup \{o_{1,h}, o_{2,h-1}, a_h\}, p_{1,h^-} = \emptyset, p_{2,h^-} = \{o_{2,h}\}$.

- **Example 4: Uncontrolled state process:** The state transition do not depend on the action of agents, i.e., $\forall h \in [H], \mathbb{T}_h(\cdot|s_h, a_h) = \mathbb{T}_h(\cdot|s_h, a'_h)$ for any $s_h \in \mathcal{S}, a_h, a'_h \in \mathcal{A}_h$, and the reward functions have additive structures, i.e., $\forall h \in [H], \mathcal{R}_h(s_h, a_h) = \sum_{i\in[n]} \mathcal{R}_{i,h}(s_h, a_{i,h})$ for all $s_h \in \mathcal{S}, a_h \in \mathcal{A}_h$. All agents will share their information with a delay of $d \geq 1$. For any $h \in [H], i \in [n]$, $c_{h^-} = c_{(h-1)^+} \cup \{o_{h-d}\}, p_{i,h^-} = (p_{i,(h-1)^+} \cup \{o_{i,h}\})\backslash\{o_{i,h-d}\}$.

- **Example 5: One-step delayed observation sharing:** At timestep $h \in [H]$, each agent has access to observations of all agents until timestep $h-1$ and her present observation. Namely, for any $h \in [H], i \in [n], c_{h^-} = c_{(h-1)^+} \cup \{o_{h-1}\}$ and $p_{i,h^-} = \{o_{i,h}\}$.

- **Example 6: One-step delayed observation and two-step delayed control sharing:** At timestep $h \in [H]$, each agent will share the observation history until timestep $h-1$ and action history until timestep $h-2$ from the private information. Namely, for any $h \in [H], i \in [n], c_{h^-} = c_{(h-1)^+} \cup \{o_{h-1}, a_{h-2}\}, p_{i,h^-} = \{o_{i,h}, a_{i,h-1}\}$.

- **Example 7: State controlled by one controller with asymmetric delayed observation sharing:** The state dynamics and reward are controlled by only one agent (i.e., system dynamics are the same as **Example 2**). Agent 1 will share all of her observations immediately, while others will share their observations with a delay of $d \geq 1$ timesteps in baseline sharing. Namely, for any $h \in [H], i \neq 1, c_{h^-} = c_{(h-1)^+} \cup \{o_{1,h}, o_{-1,h-d}\}, p_{1,h^-} = \emptyset, p_{i,h^-} = (p_{i,(h-1)^+} \cup \{o_{i,h}\})\backslash\{o_{i,h-d}\}$.

[2] Throughout this paper, we view the delay d as a *constant*, although our final bounds in §IV-D and §IV-E also apply for $d = \text{poly}\log H$. See the proofs in §C for more discussions.

- **Example 8: State controlled by one controller with asymmetric delayed observation and two-step delayed action sharing:** [kz:how many agents here? 2?] The state dynamics and reward are controlled by only one agent (i.e., system dynamics are the same as **Example 2**). At timestep $h \in [H]$, agent 1 will share all of her observations immediately and her action history until timestep $h-2$, while others will share their observations with a delay of $d \geq 1$. Namely, for any $h \in [H], i \neq 1, c_{h^-} = c_{(h-1)^+} \cup \{o_{1,h}, a_{1,h-2}, o_{-1,h-d}\}, p_{1,h^-} = \{a_{1,h-1}\}, p_{i,h^-} = (p_{i,(h-1)^+} \cup \{o_{i,h}\}) \backslash \{o_{i,h-d}\}$.

For **Examples 1, 5, 6**, for computational considerations (i.e., when it comes to §IV), we will additionally assume, as part of the example definition, that for any $h \in [H]$, the state s_h can be partitioned into n local states as $s_h = (s_{1,h}, s_{2,h}, \cdots, s_{n,h})$, and the transition kernel and observation emission have the factorized forms of $\mathbb{T}_h(s_{h+1} | s_h, a_h) = \prod_{i=1}^n \mathbb{T}_{i,h}(s_{i,h+1} | s_{i,h}, a_{i,h}), \mathbb{O}_h(o_h | s_h) = \prod_{i=1}^n \mathbb{O}_{i,h}(o_{i,h} | s_{i,h})$. Furthermore, the communication cost and reward functions are assumed to be decoupled as $\mathcal{K}_h(p_h, m_h) = \sum_{i=1}^n \mathcal{K}_{i,h}(p_{i,h}, m_{i,h}), \mathcal{R}_h(s_h, a_h) = \sum_{i=1}^n \mathcal{R}_{i,h}(s_{i,h}, a_{i,h})$.

Remark A.1. These additional conditions on $\mathbb{T}_h, \mathbb{O}_h, \mathcal{K}_h, \mathcal{R}_h$ are used to ensure the one-step tractability of the backward induction when solving the LTC problem (see Assumption IV.7). Note that these **Examples 1, 5, 6** are QC even without these additional structural conditions. Moreover, note that **Examples 2, 3, 4, 7, 8** do not require additional conditions for such computational considerations (see the proofs in §C for more details).

In fact, the first 4 examples are all sQC LTC problems, while the other 4 examples are QC but not sQC problems, as shown in the following lemma.

Lemma A.2. Given an LTC problem $\mathcal{L}$. If the baseline sharing of $\mathcal{L}$ is one of the first 4 examples above, then $\mathcal{L}$ is sQC. If the baseline sharing of $\mathcal{L}$ is one of the last 4 examples above, then $\mathcal{L}$ is QC but not sQC.

Proof. Let $\overline{\mathcal{D}}_{\mathcal{L}}$ denote the Dec-POMDP induced by $\mathcal{L}$ (see Definition II.4). We prove this lemma case by case. For convenience, we use $\dot{\cdot}$ for the notation of the elements in $\overline{\mathcal{D}}_{\mathcal{L}}$.

- **Example 1:** The information in $\overline{\mathcal{D}}_{\mathcal{L}}$ evolves as $\forall h \in [H], i \in [n], \dot{c}_h = \{\dot{o}_{1:h-1}, \dot{a}_{1:h-1}\}$ and $\dot{p}_{i,h} = \{\dot{o}_{i,h}\}$. Then, for any $i_1, i_2 \in [n], h_1, h_2 \in [H], h_1 < h_2$, $\dot{\tau}_{i_1,h_1} = \{\dot{o}_{1:h_1-1}, \dot{a}_{1:h_1-1}, \dot{o}_{i_1,h_1}\} \subseteq \dot{c}_{h_1+1} \subseteq \dot{c}_{h_2} \subseteq \dot{\tau}_{i_2,h_2}$, and $\dot{a}_{i_1,h_1} \subseteq \dot{c}_{h_1+1} \subseteq \dot{c}_{h_2} \subseteq \dot{\tau}_{i_2,h_2}$. Therefore, we have $\sigma(\dot{\tau}_{i_1,h_1}) \subseteq \sigma(\dot{\tau}_{i_2,h_2})$, and thus $\mathcal{L}$ is sQC.

- **Example 2:** The information in $\overline{\mathcal{D}}_{\mathcal{L}}$ evolves as $\forall h \in [H], i \neq 1, \dot{c}_h = \{\dot{a}_{1,1:h-1}, \dot{o}_{1,1:h-1}, \dot{o}_{-1,1:h-d}\}, \dot{p}_{1,h} = \emptyset, \dot{p}_{i,h} = \{o_{i,h-d+1:h}\}$. Then, for any $i_1, i_2 \in [n], h_1, h_2 \in [H], h_1 < h_2$. If $i_1 \neq 1$, then agent (i_1, h_1) will not influence agent (i_2, h_2). If $i_1 = 1$, then $\dot{\tau}_{i_1,h_1} = \{\dot{o}_{1,1:h_1}, \dot{a}_{1,1:h_1-1}, \dot{o}_{-1,1:h_1-d}\} \subseteq \dot{c}_{h_1+1} \subseteq \dot{c}_{h_2} \subseteq \dot{\tau}_{i_2,h_2}$, and $\dot{a}_{i_1,h_1} \subseteq \dot{c}_{h_1+1} \subseteq \dot{c}_{h_2} \subseteq \dot{\tau}_{i_2,h_2}$. Therefore, we have $\sigma(\dot{\tau}_{i_1,h_1}) \subseteq \sigma(\dot{\tau}_{i_2,h_2})$ if agent (i_1, h_1) influences agent (i_2, h_2), and thus $\mathcal{L}$ is sQC.

- **Example 3:** The information in $\overline{\mathcal{D}}_{\mathcal{L}}$ evolves as $\forall h \in [H], \dot{c}_h = \{\dot{o}_{1:h-1}, \dot{a}_{1:h-1}, \dot{o}_{1,h}\}$ and $\dot{p}_{1,h} = \emptyset, \dot{p}_{2,h} = \{\dot{o}_{i,h}\}$. Then, for any $i_1, i_2 \in [n], h_1, h_2 \in [H], h_1 < h_2$, $\dot{a}_{i_1,h_1} \subseteq \dot{c}_{h_1+1} \subseteq \dot{c}_{h_2} \subseteq \dot{\tau}_{i_2,h_2}$. If $i_1 = 1$, then $\dot{\tau}_{i_1,h_1} = \{\dot{o}_{1:h_1-1}, \dot{a}_{1:h_1-1}, \dot{o}_{1,h_1}\} \subseteq \dot{c}_{h_1+1} \subseteq \dot{c}_{h_2} \subseteq \dot{\tau}_{i_2,h_2}$. If $i_1 = 2$, then $\dot{\tau}_{i_1,h_1} = \{\dot{o}_{1:h_1}, \dot{a}_{1:h_1-1}\} \subseteq \dot{c}_{h_1+1} \subseteq \dot{c}_{h_2} \subseteq \dot{\tau}_{i_2,h_2}$. Therefore, we have $\sigma(\dot{\tau}_{i_1,h_1}) \subseteq \sigma(\dot{\tau}_{i_2,h_2})$, and thus $\mathcal{L}$ is sQC.

- **Example 4:** Since in $\overline{\mathcal{D}}_{\mathcal{L}}$, for any $i_1, i_2 \in [n], h_1, h_2 \in [H]$, agent (i_1, h_1) does not influence agent (i_2, h_2), then $\mathcal{L}$ is sQC.

- **Example 5:** The information in $\overline{\mathcal{D}}_{\mathcal{L}}$ evolves as $\forall h \in [H], i \in [n], \dot{c}_h = \{\dot{o}_{1:h-1}\}$ and $\dot{p}_{i,h} = \{\dot{o}_{i,h}\}$. Then, for any $i_1, i_2 \in [n], h_1, h_2 \in [H], h_1 < h_2$, $\dot{\tau}_{i_1,h_1} = \{\dot{o}_{1:h_1-1}, \dot{o}_{i_1,h_1}\} \subseteq \dot{c}_{h_1+1} \subseteq \dot{c}_{h_2} \subseteq \dot{\tau}_{i_2,h_2}$. However, agent $(1,1)$ may influence agent $(1,2)$ but $\sigma(\dot{a}_{1,1}) \nsubseteq \sigma(\dot{\tau}_{1,2})$. Hence, $\mathcal{L}$ is QC but not sQC.

- **Example 6:** The information in $\overline{\mathcal{D}}_{\mathcal{L}}$ evolves as $\forall h \in [H], i \in [n], \dot{c}_h = \{\dot{o}_{1:h-1}, \dot{a}_{1:h-2}\}$ and $\dot{p}_{i,h} = \{\dot{o}_{i,h}, \dot{a}_{i,h-1}\}$. Then, for any $i_1, i_2 \in [n], h_1, h_2 \in [H], h_1 < h_2$, $\dot{\tau}_{i_1,h_1} = \{\dot{o}_{1:h_1-1}, \dot{a}_{1:h_1-2}, \dot{o}_{i_1,h_1}, \dot{a}_{i_1,h_1-1}\} \subseteq \dot{c}_{h_1+1} \subseteq \dot{c}_{h_2} \subseteq \dot{\tau}_{i_2,h_2}$, and $\dot{a}_{i_1,h_1} \subseteq \dot{c}_{h_1+1} \subseteq \dot{c}_{h_2} \subseteq \dot{\tau}_{i_2,h_2}$. However, agent $(1,1)$ [kz:should this be (i_1,h_1)?] may influence agent $(2,2)$ [kz:should this be (i_2,h_2)?] but $\sigma(\dot{a}_{1,1}) \nsubseteq \sigma(\dot{\tau}_{2,2})$[kz:same here. pls check all the examples.]. Hence, $\mathcal{L}$ is QC but not sQC.

- **Example 7:** The information in $\overline{\mathcal{D}}_{\mathcal{L}}$ evolves as $\forall h \in [H], i \neq 1, \dot{c}_h = \{\dot{o}_{1,1:h-1}, \dot{o}_{-1,1:h-d}\}, \dot{p}_{1,h} = \emptyset, \dot{p}_{i,h} = \{o_{i,h-d+1:h}\}$. Then, for any $i_1, i_2 \in [n], h_1, h_2 \in [H], h_1 < h_2$. If $i_1 \neq 1$, then agent (i_1,h_1) will not influence agent (i_2,h_2). If $i_1 = 1$, then $\dot{\tau}_{i_1,h_1} = \{\dot{o}_{1,1:h_1}, \dot{o}_{-1,1:h_1-d}\} \subseteq \dot{c}_{h_1+1} \subseteq \dot{c}_{h_2} \subseteq \dot{\tau}_{i_2,h_2}$. Therefore, we have $\sigma(\dot{\tau}_{i_1,h_1}) \subseteq \sigma(\dot{\tau}_{i_2,h_2})$ if agent (i_1,h_1) influences agent (i_2,h_2). However, agent $(1,1)$ may influence agent $(1,2)$ but $\sigma(\dot{a}_{1,1}) \nsubseteq \sigma(\dot{\tau}_{1,2})$[kz:same here.][kz:we never said there are only 2 agents in this setting, right? why only (1,1) (2,2). pls check all..]. [hy:It has n agents , but as long as (1,1) influences (2,2), it is enough to show breaking QC. We do not need to care other agents such as (1,n). Same for other examples.] Hence, $\mathcal{L}$ is QC but not sQC.

- **Example 8:** The information in $\overline{\mathcal{D}}_{\mathcal{L}}$ evolves as $\forall h \in [H], i \neq 1, \dot{c}_h = \{\dot{o}_{1,1:h-1}, \dot{a}_{1,1:h-2}, \dot{o}_{-1,1:h-d}\}, \dot{p}_{1,h} = \{\dot{a}_{1,h-1}\}, \dot{p}_{i,h} = \{o_{i,h-d+1:h}\}$. Then, for any $i_1, i_2 \in [n], h_1, h_2 \in [H], h_1 < h_2$. If $i_1 \neq 1$, then agent (i_1,h_1) will not influence agent (i_2,h_2). If $i_1 = 1$, then $\dot{\tau}_{i_1,h_1} = \{\dot{o}_{1,1:h_1}, \dot{a}_{1,h_1-1}, \dot{o}_{-1,1:h_1-d}\} \subseteq \dot{c}_{h_1+1} \subseteq \dot{c}_{h_2} \subseteq \dot{\tau}_{i_2,h_2}$. Therefore, we have $\sigma(\dot{\tau}_{i_1,h_1}) \subseteq \sigma(\dot{\tau}_{i_2,h_2})$ if agent (i_1,h_1) influences agent (i_2,h_2). However, agent $(1,1)$ may influence agent $(2,2)$ but $\sigma(\dot{a}_{1,1}) \nsubseteq \sigma(\dot{\tau}_{2,2})$[kz:same..][kz:we never said there are only 2 agents in this setting, right? why only $(1,1)$ and $(2,2)$]. Hence, $\mathcal{L}$ is QC but not sQC.

This completes the proof. □

B. Deferred Details of §III

B-A Supporting Lemmas

We start by proving several supporting lemmas that will be used in the later proofs in this section.

Lemma B.1. Given any QC LTC $\mathcal{L}$, its induced Dec-POMDP $\overline{\mathcal{D}}_{\mathcal{L}}$, and any $i_1, i_2 \in [n], h_1, h_2 \in [H]$. If agent (i_1,h_1) influences agent (i_2,h_2) in the intrinsic model of $\overline{\mathcal{D}}_{\mathcal{L}}$, then for the random variables $\tau_{i_1,h_1^-}, \tau_{i_2,h_2^-}$ in $\mathcal{L}$, we have $\sigma(\tau_{i_1,h_1^-}) \subseteq \sigma(\tau_{i_2,h_2^-})$. Moreover, if $\mathcal{L}$ is sQC, then for random variables $a_{i_1,h_1}, \tau_{i_2,h_2^-}$ in $\mathcal{L}$, we have $\sigma(a_{i_1,h_1}) \subseteq \sigma(\tau_{i_2,h_2^-})$.

Proof. We denote by $\check{\tau}_{i_1,h_1}, \check{\tau}_{i_2,h_2}$ the information of agent $(i_1,h_1),(i_2,h_2)$ in the problem $\overline{\mathcal{D}}_{\mathcal{L}}$. From the definition of $\overline{\mathcal{D}}_{\mathcal{L}}$ being QC, if agent (i_1,h_1) influences agent (i_2,h_2), then $\sigma(\check{\tau}_{i_1,h_1}) \subseteq \sigma(\check{\tau}_{i_2,h_2})$. Since for any $h \in [H], i \in [n]$, $\check{\tau}_{i,h}$ is the information of agent (i,h) without additional sharing, then we know that $\tau_{i,h^-} \backslash \check{\tau}_{i,h} \subseteq \cup_{t=1}^{h-1} z_t^a, \tau_{i,h^+} \backslash \check{\tau}_{i,h} \subseteq \cup_{t=1}^{h} z_t^a$. Therefore, we know that $\sigma(\tau_{i_1,h_1^-} \backslash \check{\tau}_{i_1,h_1}) \subseteq \sigma(\cup_{t=1}^{h-1} z_t^a) \subseteq \sigma(c_{h_1^-}) \subseteq \sigma(c_{h_2^-}) \subseteq \sigma(\tau_{i_2,h_2^-})$. Also, we know $\sigma(\check{\tau}_{i_1,h_1}) \subseteq \sigma(\check{\tau}_{i_2,h_2}) \subseteq \sigma(\tau_{i_2,h_2^-})$. Thus, we can conclude that $\sigma(\tau_{i_1,h_1^-}) \subseteq \sigma(\tau_{i_2,h_2^-})$. Moreover, if $\mathcal{L}$ is sQC, then from the the definition of $\overline{\mathcal{D}}_{\mathcal{L}}$ being sQC and agent (i_1,h_1) influences agent (i_2,h_2) in $\overline{\mathcal{D}}_{\mathcal{L}}$, it holds that $\sigma(a_{i_1,h_1}) \subseteq \sigma(\check{\tau}_{i_2,h_2}) \subseteq \sigma(\tau_{i_2,h_2^-})$. □

Lemma B.2. Let $\mathcal{L}$ be an QC LTC problem satisfying Assumptions III.5 and III.7, and $\mathcal{D}_{\mathcal{L}}$ be the reformulated Dec-POMDP. Then for any $i_1, i_2 \in [n], t_1, t_2 \in [H]$, if agent $(i_1, 2t_1)$ influences agent $(i_2, 2t_2)$ in $\mathcal{D}_{\mathcal{L}}$, then $\sigma(\tau_{i_1,t_1^-}) \subseteq \sigma(\tau_{i_2,t_2^-})$ in $\mathcal{L}$. Moreover, if $\mathcal{L}$ is sQC, then $\sigma(a_{i_1,t_1}) \subseteq \sigma(\tau_{i_2,t_2^-})$.

Proof. We prove this case by case as follows:
1) If a_{i_1,t_1} influences the underlying state s_{t_1+1}, then from Assumption III.7, agent (i_1,t_1) influences

o_{-i_1,t_1+1}, so there must exist $i_3 \neq i_1$, such that agent (i_1,t_1) influences o_{i_3,t_1+1}. From part (e) of Assumption II.1 and $t_1 < t_2$ (since otherwise agent $(i_1,2t_1)$ cannot influence agent $(i_2,2t_2)$ in $\mathcal{D}_{\mathcal{L}}$), we know $o_{i_3,t_1+1} \in \tau_{i_3,(t_1+1)^-} \subseteq \tau_{i_3,t_2^-}$ even under no additional sharing, and then we get agent (i_1,t_1) influences agent (i_3,t_2) in $\overline{\mathcal{D}}_{\mathcal{L}}$ (the Dec-POMDP induced by $\mathcal{L}$). From Lemma B.1, it holds that $\sigma(\tau_{i_1,t_1^-}) \subseteq \sigma(\tau_{i_3,t_2^-})$. From Assumption II.2 and $i_3 \neq i_1$, we know $\sigma(\tau_{i_1,t_1^-}) \subseteq \sigma(c_{t_2^-}) \subseteq \sigma(\tau_{i_2,t_2^-})$. If $\mathcal{L}$ is sQC, by Lemma B.1, we have $\sigma(a_{i_1,t_1}) \subseteq \sigma(\tau_{i_3,t_2^-})$, and then $\sigma(a_{i_1,t_1}) \subseteq \sigma(c_{t_2^-}) \subseteq \sigma(\tau_{i_2,t_2^-})$ from Assumption II.2.
2) If a_{i_1,t_1} does not influence s_{t_1+1}, then from Assumption III.5, for any $t > t_1, a_{i_1,t_1} \notin \tau_{t^-}$ and $a_{i_1,t_1} \notin \tau_{t^+}$, and then agent $(i_1,2t_1)$ does not influence $\widetilde{s}_{2t_1+1}$ and $\widetilde{o}_{2t_1+1}$ in $\mathcal{D}_{\mathcal{L}}$. Thus $\widetilde{a}_{i_1,2t_1} = a_{i_1,t_1}$ does not influence $\widetilde{\tau}_{i,2t_1+1}, \forall i \in [n]$, and then it does not influence $\widetilde{a}_{i,2t_1+1}, \forall i \in [n]$. And hence, it does not influence $\widetilde{\tau}_{i,2t_1+2}$ and $\widetilde{a}_{i,2t_1+2}, \forall i \in [n]$, either. By recursion, we know agent $(i_1,2t_1)$ does not influence agent $(i_2,2t_2)$, which leads to a contradiction to the premise of the lemma. This completes the proof. □

B-B Proof of Lemma III.2

Proof. We first have the following proposition on the hardness of solving POMDPs.

Proposition B.3. For any $\epsilon \in [0,\frac{1}{4})$, such that computing an ϵ-additive optimal strategy in POMDPs with rewards bounded in $[0,1/2]$ is PSPACE-hard.

Proof of Proposition B.3. In the proof of [20, Theorem 4.11], given any $\epsilon \in [0,1)$, it constructed POMDPs from the problem of Stochastic Satisfiability (SSAT) of the Quantified Boolean Formulae. The constructed instances in [20] satisfy that the reward values lie in $\{0,2\}$, and it was then proved that finding an ϵ-relative approximately optimal solution in such POMDPs is PSPACE-hard. Also, one can verify that finding an ϵ-additive approximately optimal solution in such POMDPs is PSPACE-hard.

Then, for any $\epsilon \in [0,\frac{1}{4})$, let $\epsilon_1 = 4\epsilon \in [0,1)$, and leverage the construction in [20, Theorem 4.11] with ϵ_1, but scaling the the reward values by $\frac{1}{4}$ such that rewards are bounded in $[0,\frac{1}{2}]$. Then, finding an ϵ-additive approximately optimal solution in such POMDPs (after scaling) is PSPACE-hard. □

Now we proceed with the proof of Lemma III.2 based on Proposition B.3. Given any POMDP $\mathcal{P} = (\mathcal{S}^{\mathcal{P}},\mathcal{A}^{\mathcal{P}},\mathcal{O}^{\mathcal{P}},\{\mathbb{O}_h^{\mathcal{P}}\}_{h\in[H^{\mathcal{P}}]},\{\mathbb{T}_h^{\mathcal{P}}\}_{h\in[H^{\mathcal{P}}]},\{\mathcal{R}_h^{\mathcal{P}}\}_{h\in[H^{\mathcal{P}}]},\mu_1^{\mathcal{P}})$ with rewards bounded in $[0,1/2]$, we can construct an LTC $\mathcal{L}$ with any $\alpha \in (0,1)$ as follows:

- Number of agents: $n = 3$; length of episode: $H = H^{\mathcal{P}}$.
- Underlying state space: $\mathcal{S} = \mathcal{S}^{\mathcal{P}} \times [2]$. For any $s \in \mathcal{S}$, we can split $s = (s^1,s^2)$, where $s^1 \in \mathcal{S}^{\mathcal{P}}, s^2 \in [2]$. Initial state distribution: $\forall s \in \mathcal{S}, \mu_1(s) = \mu_1^{\mathcal{P}}(s^1)/2$.
- Control action space: For any $h \in [H], \mathcal{A}_{1,h} = \mathcal{A}^{\mathcal{P}}, \mathcal{A}_{2,h} = [2], \mathcal{A}_{3,h} = \{\emptyset\}$.
- Transition: For any $h \in [H], s_h, s_{h+1} \in \mathcal{S}, a_h \in \mathcal{A}_h, \mathbb{T}_h(s_{h+1}\,|\,s_h,a_h) = \mathbb{T}_h^{\mathcal{P}}(s_{h+1}^1\,|\,s_h^1,a_{1,h})\mathbb{1}[s_{h+1}^2 = a_{2,h}]$.
- Observation space: For any $h \in [H], \mathcal{O}_{1,h} = \mathcal{O}_h^{\mathcal{P}}, \mathcal{O}_{2,h} = \mathcal{O}_{3,h} = \mathcal{S}$.
- Emission: For any $h \in [H], o_h \in \mathcal{O}_h, s_h \in \mathcal{S}, \mathbb{O}_h(o_h\,|\,s_h) = \mathbb{O}_h^{\mathcal{P}}(o_{1,h}\,|\,s_h^1)\mathbb{O}_h'(o_{2,h}\,|\,s_h)\mathbb{O}_h'(o_{3,h}\,|\,s_h)$, where $\mathbb{O}_h'$ is defined as
$$\forall o \in \mathcal{S}, s_h \in \mathcal{S}, \qquad \mathbb{O}_h'(o\,|\,s_h) = \begin{cases} \alpha + \dfrac{1-\alpha}{|\mathcal{S}|} & \text{if } o = s_h \\ \dfrac{1-\alpha}{|\mathcal{S}|} & \text{o.w.} \end{cases}.$$
- Baseline sharing: null.

- Communication action space: For any $h \in [H], \mathcal{M}_{1,h} = \mathcal{M}_{2,h} = \{0,1\}^{2h-1}, \mathcal{M}_{3,h} = \{0,1\}^h$. For any $i \in [2], p_{i,h^-} \in \mathcal{P}_{i,h^-}, \phi_{i,h}(p_{i,h^-}, m_{i,h}) = \{o_{i,k} | \forall k \le h$, the $(2k-1)$-th digit of $m_{i,h}$ is 1 and $o_{i,k} \in p_{i,h^-}\} \cup \{a_{i,k} | \forall k \le h-1$, the $2k$-th digit of $m_{i,h}$ is 1 and $a_{i,k} \in p_{i,h^-}\} \cup \{m_{i,h}\}$. For agent 3, $p_{3,h^-} \in \mathcal{P}_{3,h^-}$, $\phi_{3,h}(p_{3,h^-}, m_{3,h}) = \{o_{3,k} | \forall k \le h$, the k-th digit of $m_{3,h}$ is 1 and $o_{3,k} \in p_{3,h^-}\} \cup \{m_{3,h}\}$.
- Reward function: For any $h \in [H]$, $\forall s_h \in \mathcal{S}, a_h \in \mathcal{A}_h, \mathcal{R}_h(s_h, a_h) = \mathcal{R}_h^{\mathcal{P}}(s_h^1, a_{1,h})/H$.
- Communication cost function: For any $h \in [H]$, $\forall z_h^a \in \mathcal{Z}_h^a, \mathcal{K}_h(z_h^a) = \mathbb{1}[z_h^a \neq \{m_h\}]$. It means that the communication cost is 1 unless there is no additional sharing.
- We restrict the communication strategy to only use c_h as input. And for any $t \in [H-1]$, we remove $a_{3,t}$ in τ_{h^-}, τ_{h^+} for all $h > t$.

We first verify that such a construction satisfies Assumptions III.1, III.4, III.5, and III.7, but is not QC.

- $\mathcal{L}$ is not QC, since for agent $(2,1)$ influences state s_2^2 and then influences the information of agent $(3,2)$, but agent $(3,2)$ does not know the information of agent $(2,1)$.
- $\mathcal{L}$ satisfies Assumptions III.1, III.7 because both agent 2 and agent 3 have individual γ-observability with $\gamma = \alpha$. That is, for any $b_1, b_2 \in \Delta(\mathcal{S}), i = 2,3$, we have

$$\begin{aligned}
\|\mathbb{O}_{i,h}^\top (b_1 - b_2)\|_1 &= \sum_{o_{i,h} \in \mathcal{O}_h} |\sum_{s_h \in \mathcal{S}} (b_1(s_h) - b_2(s_h)) \mathbb{O}_{i,h}(o_{i,h} | s_h)| \\
&= \sum_{o_{i,h} \in \mathcal{O}_h} |\sum_{s_h \in \mathcal{S}} (b_1(s_h) - b_2(s_h)) (\frac{1-\alpha}{|\mathcal{S}|} + \alpha \mathbb{1}[o_{i,h} = s_h])| \\
&= \sum_{o_{i,h} \in \mathcal{O}_h} \alpha |b_1(o_{i,h}) - b_2(o_{i,h})| = \alpha \|b_1 - b_2\|_1.
\end{aligned}$$

- $\mathcal{L}$ satisfies Assumption III.4 because we restrict that the communication strategy can only use c_h as input.
- $\mathcal{L}$ satisfies Assumption III.5 since the control actions $a_{3,t}$ (for all $t \in [H-1]$) do not influence the underlying state, and we remove $a_{3,t}$ from τ_h for any $h > t$.
- $\mathcal{L}$ is not QC, since for any $h \in [H-1]$, agent $(2,h)$ influences the state s_{h+1} and then influences $o_{3,h+1}$ and $\tau_{3.h+1}$. However, agent $(3, h+1)$ does not know the information of agent $(2,h)$, i.e., $\sigma(\tau_{2,h}) \nsubseteq \sigma(\tau_{3,h+1})$.

In this LTC problem $\mathcal{L}$, let $(g_{1:H}^{a,*}, g_{1:H}^{m,*})$ be any ϵ-team optimal strategy, with $\epsilon \in [0, 1/4)$. If any agent shares any information through additional sharing, i.e., $\exists h \in [H], \mathbb{P}(z_h^a \neq \{m_h\} | g_{1:H}^{a,*}, g_{1:H}^{m,*}) > 0$, we then choose the h to be the minimal one, i.e., $h = \min\{h' \in [H] | \mathbb{P}(z_{h'}^a \neq \{m_{h'}\} | g_{1:H}^{a,*}, g_{1:H}^{m,*}) > 0\}$ being the first time the additional sharing occurs. This means that at this timestep h, there is no observation or action in c_{h^-} (almost surely), since baseline sharing is null. Then, there exists agent $i \in [2]$ such that $\mathbb{P}(z_{i,h}^a \neq \{m_{i,h}\} | g_{1:H}^{a,*}, g_{1:H}^{m,*}) > 0$.

From the construction of $\mathcal{L}$, we know that agent i chooses $m_{i,h}$ based on c_{h^-}. It means it will always share, i.e., $\mathbb{P}(z_{i,h}^a \neq \{m_{i,h}\} | g_{1:H}^{a,*}, g_{1:H}^{m,*}) = 1$. Therefore, $\mathbb{P}(\kappa_h = 1 | g_{1:H}^{a,*}, g_{1:H}^{m,*}) = 1$, and $\mathcal{J}_{\mathcal{L}}(g_{1:H}^{a,*}, g_{1:H}^{m,*}) = \mathbb{E}[\sum_{t=1}^H r_t - \kappa_t | g_{1:H}^{a,*}, g_{1:H}^{m,*}] \le \mathbb{E}[\sum_{t=1}^H r_t | g_{1:H}^{a,*}, g_{1:H}^{m,*}] - \mathbb{E}[\kappa_h | g_{1:H}^{a,*}, g_{1:H}^{m,*}] \le H \cdot \frac{1}{2H} - 1 \le -\frac{1}{2}$. Note that the rewards in $\mathcal{P}$ is bounded by $[0, \frac{1}{2}]$, and the rewards in $\mathcal{L}$ is bounded by $[0, \frac{1}{2H}]$. Hence, $(g_{1:H}^{a,*}, g_{1:H}^{m,*})$ is not an ϵ-team optimal for any $\epsilon \in [0, 1/4)$.

Therefore, any ϵ-team optimal strategy yields no additional sharing. Then, any $(g^{a,*}_{1:H}, g^{m,*}_{1:H})$ being an $\frac{\epsilon}{H}$-team optimal strategy of $\mathcal{L}$ will directly give an ϵ-optimal strategy of $\mathcal{P}$ as $\{g^{a,*}_{1,h}\}_{h\in[H]}$, since when there is no sharing, the decision process is only controlled by agent 1. From Proposition B.3, we complete the proof. □

B-C Proof of Lemma III.3

Proof. We prove this result by showing a reduction from the Team Decision problem [31].

Definition B.4 (Team decision problem (TDP))**.** Given finite sets Y_1, Y_2, U_1, U_2, a rational probability mass function $p : Y_1 \times Y_2 \to \mathbb{Q}$, and an integer cost function $c : Y_1 \times Y_2 \times U_1 \times U_2 \to \mathbb{N}$, find decision rules $\gamma_i : Y_i \to U_i, i = 1, 2$ that minimize the expected cost

$$J(\gamma_1, \gamma_2) = \sum_{y_1 \in Y_1, y_2 \in Y_2} c(y_1, y_2, \gamma_1(y_1), \gamma_2(y_2)) p(y_1, y_2). \tag{B.1}$$

We show the NP-hardness of solving LTC from the problem of TDP, even with $|U_1| = |U_2| = 2$ [31]. Given any TDP $\mathcal{TD} = (\widetilde{Y}_1, \widetilde{Y}_2, \widetilde{U}_1, \widetilde{U}_2, \widetilde{c}, \widetilde{p}, \widetilde{J})$ with $|\widetilde{U}_1| = |\widetilde{U}_2| = 2$, let $\widetilde{U}_1 = \{1, 2\}, \widetilde{U}_2 = \{1, 2\}$, then we can construct an $H = 3$ and 2-agent LTC $\mathcal{L}$ with two parameters $\alpha_1 \in \mathbb{R}, \alpha_2 \in (0, 1)$ (to be specified later) such that:

- Number of agents: $n = 2$.
- Underlying state: $\mathcal{S} = [2]^4$. For each $s_1 \in \mathcal{S}$, we can split s_1 into 4 parts as $s_1 = (s_1^1, s_1^2, s_1^3)$, where $s_1^1, s_1^2, s_1^3, s_1^4 \in [2]$. Similarly, $s_2, s_3 \in \mathcal{S}$ can be split in the same way.
- Initial state distribution: $\forall s_1 \in \mathcal{S}, \mu_1(s_1) = \frac{1}{16}$.
- Control action space: For the first 2 timesteps, $\forall i = 1, 2, \mathcal{A}_{i,1} = \mathcal{A}_{i,2} = \{\emptyset\}$; for $h = 3, \mathcal{A}_{1,3} = [2], \mathcal{A}_{2,3} = \{\emptyset\}$.
- Transition: $\forall s \in \mathcal{S}, a_1 \in \mathcal{A}_1, a_2 \in \mathcal{A}_2, a_3 \in \mathcal{A}_3, \mathbb{T}_1(s|s, a_1) = \mathbb{T}_2(s|s, a_2) = \mathbb{T}_3(s|s, a_3) = 1$. Note that under the transition dynamics above, $s_1 = s_2 = s_3$ always holds, for any $s_1 \in \mathcal{S}$.
- Observation space: $\mathcal{O}_{1,1} = \mathcal{O}_{2,1} = \mathcal{O}_{1,2} = \mathcal{O}_{2,2} = [2] \times \mathcal{S}, \mathcal{O}_{1,3} = \widetilde{Y}_1 \times \mathcal{S}, \mathcal{O}_{2,3} = \widetilde{Y}_2 \times \mathcal{S}$. Hence, for each $i \in [2], h \in [2], o_{i,h} \in \mathcal{O}_{i,h}$, we can split $o_{i,h}$ into 2 parts as $o_{i,h} = (o^1_{i,h}, o^2_{i,h})$, where $o^1_{i,h} \in [2], o^2_{i,h} \in \mathcal{S}$. For each $i \in [n], o_{i,3} \in \mathcal{O}_{i,3}$, we can similarly split $o_{i,3}$ into 2 parts as $o_{i,3} = (o^1_{i,3}, o^2_{i,3})$, where $o^1_{i,3} \in \widetilde{Y}_i, o^2_{i,3} \in \mathcal{S}$.
- Baseline sharing: null.
- Communication action space: For $i \in [2], h \in \{1, 2\}, \mathcal{M}_{i,h} = \{0, 1\}^h, \mathcal{M}_{i,3} = \{1, 2\}$; $\phi_{i,h}$ is defined as: $\forall h \in \{1, 2\}, \phi_{i,h}(p_{i,h^-}, m_{i,h}) = \{o_{i,k} | \forall k \le h,$ the k-th digit of $m_{i,h}$ is 1 and $o_{i,k} \in p_{i,h^-}\}$. For $h = 3$, $\mathcal{M}_{i,h} = \{1, 2\}$, and if $m_{i,3} = 1$, then $\phi_{i,3}(p_{i,3^-}, m_{i,3}) = \{o_{i,1}, o_{i,3}, m_{i,3}\}$; if $m_{i,3} = 2$, then $\phi_{i,3}(p_{i,3^-}, m_{i,3}) = \{o_{i,2}, o_{i,3}, m_{i,3}\}$.
- Emission: For any $i \in [2], h \in [2], s_h \in \mathcal{S}, o_{i,h} \in \mathcal{O}_{i,h}, \mathbb{O}_h(o_h|s_h) = \Pi_{i=1}^2 \mathbb{O}_{i,h}(o_{i,h}|s_h)$ and $\mathbb{O}_{i,h}(o_{i,h}|s_h)$ is defined as:

$$\mathbb{O}_{i,h}(o_{i,h}|s_h) = \begin{cases} \frac{1-\alpha_2}{16} & o^1_{i,h} = s_h^{i+2h-2}, o^2_{i,h} \neq s_h \\ \frac{1-\alpha_2}{16} + \alpha_2 & o^1_{i,h} = s_h^{i+2h-2}, o^2_{i,h} = s_h \\ 0 & \text{o.w.} \end{cases}.$$

For $i \in [2], h = 3$, $s_3 \in \mathcal{S}, o_3 \in \mathcal{O}_3, \mathbb{O}_3(o_3 \mid s_3) = \widetilde{p}(o^1_{1,3}, o^1_{2,3}) \Pi^2_{i=1} \mathbb{O}^2_{i,3}(o^2_{i,3} \mid s_3)$, where

$$\mathbb{O}^2_{i,3}(o^2_{i,3} \mid s_3) = \begin{cases} \frac{1-\alpha_2}{16} & o^2_{i,3} \neq s_3 \\ \frac{1-\alpha_2}{16} + \alpha_2 & o^2_{i,3} = s_3 \end{cases}.$$

The reward functions are defined as:

$$\mathcal{R}_1(s_1, a_1) = \mathcal{R}_2(s_2, a_2) = 0, \quad \forall s_1, s_2 \in \mathcal{S}, a_1 \in \mathcal{A}_1, a_2 \in \mathcal{A}_2;$$
$$\mathcal{R}_3(s_3, a_3) = \begin{cases} 1 & \text{if } (a_{1,3} = s^2_3 \text{ or } a_{1,3} = s^4_3) \text{ and } (a_{2,3} = s^1_3 \text{ or } a_{2,3} = s^3_3) \\ 0 & \text{o.w.} \end{cases};$$

The communication cost functions are defined as:

$$\forall h \in \{1, 2, 4\}, z^a_h \in \mathcal{Z}^a_h, \qquad \mathcal{K}_h(z^a_h) = 1 \text{ if } z^a_h \neq \{m_{1,h}, m_{2,h}\}, \text{ else } 0;$$
$$\mathcal{K}_3(z^a_3) = \begin{cases} \widetilde{c}(o^1_{1,3}, o^1_{2,3}, 1, 1)/\alpha_1 & \text{if } \{o_{1,1}, o_{2,1}\} \subseteq z^a_3 \text{ and } \{o_{1,2}, o_{2,2}\} \cap z^a_3 = \emptyset \\ \widetilde{c}(o^1_{1,3}, o^1_{2,3}, 2, 1)/\alpha_1 & \text{if } \{o_{1,2}, o_{2,1}\} \subseteq z^a_3 \text{ and } \{o_{1,1}, o_{2,2}\} \cap z^a_3 = \emptyset \\ \widetilde{c}(o^1_{1,3}, o^1_{2,3}, 1, 2)/\alpha_1 & \text{if } \{o_{1,1}, o_{2,2}\} \subseteq z^a_3 \text{ and } \{o_{1,2}, o_{2,1}\} \cap z^a_3 = \emptyset \\ \widetilde{c}(o^1_{1,3}, o^1_{2,3}, 2, 2)/\alpha_1 & \text{if } \{o_{1,2}, o_{2,2}\} \subseteq z^a_3 \text{ and } \{o_{1,1}, o_{2,1}\} \cap z^a_3 = \emptyset \\ 0 & \text{o.w.} \end{cases},$$

where we let $\alpha_0 = \max_{y_1, y_2, u_1, u_2} \widetilde{c}(y_1, y_2, u_1, u_2)$, and set $\alpha_1 = 2\alpha_0$. Note that without loss of optimality, we suppose $\alpha_1 > 0$, since if $\alpha_1 = 0$, then $\widetilde{c}(y_1, y_2, u_1, u_2) = 0, \forall y_1 \in \widetilde{Y}_1, y_2 \in \widetilde{Y}_2, u_1 \in \widetilde{U}_1, u_2 \in \widetilde{U}_2$, which is a trivial instance that cannot be the one that leads to the hardness in [31]. Hence, $0 \leq \kappa_3 = \mathcal{K}_3(z^a_3) \leq \frac{1}{2}$ for any $z^a_3 \in \mathcal{Z}^a_3$ always holds. Also, we remove $a_{i,t}$ in τ_{h^-} and τ_{h^+} for any $t \in [2], i \in [2], h > t$. Under such a construction, $\mathcal{L}$ satisfies the following conditions:

- Problem $\mathcal{L}$ is QC[kz:we need to have such discussions in other lower-bound lemmas – the discussion of the "baseline info structure".]: For all $i_1, i_2 \in [2], h_1, h_2 \in [4]$ with $i_1 \neq i_2$, agent (i_1, h_1) does not influence (i_2, h_2) because agent (i_1, h_1) cannot influence the observations of agent (i_2, h_2), and the baseline sharing is null. For the same agent with $i_1 = i_2$, the information-inclusion assumption in Assumption II.1 (e) ensures that $\mathcal{L}$ is QC.
- Problem $\mathcal{L}$ satisfies Assumptions III.1 and III.7: We prove this by showing that each agent $i \in [2]$

satisfies γ-observability. For any $i \in [2], h \in [2], b_1, b_2 \in \Delta(\mathcal{S})$, let

$$\begin{aligned}
\|\mathbb{O}_{i,h}^{\top}(b_1 - b_2)\|_1 &= \sum_{o_{i,h}^1 \in [2]} \sum_{o_{i,h}^2 \in \mathcal{S}} \left| \sum_{s_h \in \mathcal{S}} (b_1(s_h) - b_2(s_h)) \mathbb{O}_{i,h}((o_{i,h}^1, o_{i,h}^2) | s_h) \right| \\
&\geq \sum_{o_{i,h}^2 \in \mathcal{S}} \left| \sum_{o_{i,h}^1 \in [2]} \sum_{s_h \in \mathcal{S}} (b_1(s_h) - b_2(s_h)) \mathbb{O}_{i,h}((o_{i,h}^1, o_{i,h}^2) | s_h) \right| \\
&= \sum_{o_{i,h}^2 \in \mathcal{S}} \left| \sum_{s_h \in \mathcal{S}} \sum_{o_{i,h}^1 \in [2]} (b_1(s_h) - b_2(s_h)) \mathbb{1}[o_{i,h}^1 = s_h^{i+2h-2}] (\frac{1-\alpha_2}{16} + \alpha_2 \mathbb{1}[o_{i,h}^2 = s_h]) \right| \\
&= \sum_{o_{i,h}^2 \in \mathcal{S}} \left| \sum_{s_h \in \mathcal{S}} (b_1(s_h) - b_2(s_h)) (\frac{1-\alpha_2}{16} + \alpha_2 \mathbb{1}[o_{i,h}^2 = s_h]) \right| \\
&= \sum_{o_{i,h}^2 \in \mathcal{S}} \left| \frac{1-\alpha_2}{16} (\sum_{s_h \in \mathcal{S}} (b_1(s_h) - b_2(s_h))) + \alpha_2 (b_1(o_{i,h}^2) - b_2(o_{i,h}^2)) \right| \\
&= \sum_{o_{i,h}^2 \in \mathcal{S}} \alpha_2 |b_1(o_{i,h}^2) - b_2(o_{i,h}^2)| = \alpha_2 \|b_1 - b_2\|_1.
\end{aligned}$$

For any $i \in [2], h = 3$, the proof is similar, where we replace $o_{i,h}^1 \in [2]$ with $o_{i,h}^1 \in \widetilde{Y}_i$ for $h = 3$.

- Problem $\mathcal{L}$ satisfies Assumption III.5, because the joint control action histories $a_{1:4}$ do not influence the underlying states, and we restrict the communication and control strategies to not use them as input.

We will show below that computing a team-optimal strategy of $\mathcal{L}$ can give us a team-optimal strategy of $\mathcal{TD}$.

Given $(g_{1:3}^{a,*}, g_{1:3}^{m,*})$ being a team optimal strategy of $\mathcal{L}$, firstly, it will have no additional sharing at timesteps $h = 1, 2$ under $(g_{1:3}^{a,*}, g_{1:3}^{m,*})$, namely, for $h = 1, 2, \mathbb{P}(z_h^a = \{m_{1,h}, m_{2,h}\} | g_{1:3}^{a,*}, g_{1:3}^{m,*}) = 1$. If not, then for any $i \in [2]$, consider $g_{i,1}^{m,'}, g_{i,2}^{m,'}, g_{i,3}^{a,*}$ defined as

$$\forall \tau_{i,1^-}, \tau_{i,2^-}, \tau_{i,3^+}, g_{i,1}^{m,'}(\tau_{i,1^-}) = (0), \quad g_{i,2}^{m,'}(\tau_{i,2^-}) = (0,0), \quad g_{i,3^+}^{a,'} = \begin{cases} o_{3-i,1}^1 & \text{if } o_{3-i,1} \in \tau_{i,3^+} \\ o_{3-i,2}^1 & \text{o.w.} \end{cases},$$

and replace the $g_{1:2,1}^{m,*}, g_{1:2,2}^{m,*}, g_{1:2,3}^{a,*}$ by $g_{1:2,1}^{m,'}, g_{1:2,2}^{m,'}, g_{1:2,3}^{a,'}$ in $g_{1:3}^{a,*}, g_{1:3}^{m,*}$ to get $(g_{1:3}^{a,'}, g_{1:3}^{m,'})$. It is easy to verify that $\mathcal{J}_{\mathcal{L}}(g_{1:3}^{a,*}, g_{1:3}^{m,*}) < \mathcal{J}_{\mathcal{L}}(g_{1:3}^{a,'}, g_{1:3}^{m,'})$, since $g_{1:3}^{a,'}, g_{1:3}^{m,'}$ can guarantee $r_3 = 1$ always holds; if there is no additional sharing under $g_{1:3}^{a,'}, g_{1:3}^{m,'}$ at first two timesteps, these two strategies has the same communication cost, otherwise $(g_{1:3}^{a,'}, g_{1:3}^{m,'})$ has $\sum_{h=1}^3 \kappa_h = \kappa_3 \leq \frac{1}{2}$ but $(g_{1:3}^{a,*}, g_{1:3}^{m,*})$ has $\sum_{h=1}^3 \kappa_h \geq 1$. This leads to the contradiction that $(g_{1:3}^{a,*}, g_{1:3}^{m,*})$ is a team-optimal strategy.

Also, if we replace the $g_{1:2,3}^{a,*}$ by $g_{1:2,3}^{a,'}$, the communication cost does not change and reward can achieve optimal, i.e., $r_3 = 1$ always holds. Thus, without loss of generality, we can assume that $g_{1:2,3}^{a,*} = g_{1:2,3}^{a,'}$, since otherwise we can do the replacement and it is still a team-optimal strategy.

Therefore, $J_{\mathcal{L}}(g_{1:3}^{a,*}, g_{1:3}^{m,*}) = \mathbb{E}[\sum_{h=1}^3 r_h - \kappa_h | g_{1:3}^{a,*}, g_{1:3}^{m,*}] = 1 - \mathbb{E}[\kappa_3 | g_{1:3}^{a,*}, g_{1:3}^{m,*}] = 1 - \frac{1}{\alpha_1}\mathbb{E}[\widetilde{c}(o_{1,3}^1, o_{2,3}^1, m_{1,3}, m_{2,3})]$, where $m_{1,3} = g_{1,3}^{m,*}(\{\tau_{1,3^-}\}), m_{2,3} = g_{2,3}^{m,*}(\{\tau_{2,3^-}\})$, which means $(g_{1:3}^{a,*}, g_{1:3}^{m,*})$ can minimize κ_3 through choosing m_3 properly, if there is no additional sharing at first two timesteps, i.e., $\tau_{i,3^-} = \{o_{i,1}, o_{i,2}, o_{i,3}, m_{1:2}\}$. By construction, κ_3 [kz:this is only a "realization"?] only depends on o_3 and m_3 and is irrelevant of $\{o_{1,1}, o_{1,2}, o_{1,3}^2\}$, and $m_{i,1} = (0), m_{i,2} = (0,0), \forall i \in [2]$

always hold. Hence, $\{o_{1,1}, o_{1,2}, o^2_{1,3}\}$ are useless information for agent 1 to choose $m_{1,3}$ and minimize $\mathbb{E}[\kappa_3]$[kz:but how can we ensure, for ANY realization of κ_h, it is "useless"??]. Therefore, not using them in $g^{m,*}_{1,3}$ does not lose any optimality. Hence, we can consider the $g^{m,*}_{1,3}$ that only has $o^1_{1,3}$ as input. In the same way, we can consider the $g^{m,*}_{2,3}$ that only has $o^1_{2,3}$ as input. Therefore, $J_{\mathcal{L}}(g^{a,*}_{1:3}, g^{m,*}_{1:3}) = 1 - \sum_{o^1_{1,3}, o^2_{1,3}} \frac{1}{\alpha_1} \widetilde{c}(o^1_{1,3}, o^1_{2,3}, g^{m,*}_{1,3}(o^1_{1,3}), g^{m,*}_{2,3}(o^1_{2,3})) \widetilde{p}(o^1_{1,3}, o^1_{2,3})$. Further, we can leverage $\widetilde{\gamma}^*_1 = g^{m,*}_{1,3}, \widetilde{\gamma}^*_2 = g^{m,*}_{2,3}$ to minimize the expected cost $\widetilde{J}$ of the TDP. Therefore, from the NP-hardness of TDPs [31, Corollary 4.1], we complete our proof. □

B-D Proof of Lemma III.6

Proof of Lemma III.6. We prove this result by showing a reduction from the Team Decision problem. Given any TDP $\mathcal{TD} = (\widetilde{Y}_1, \widetilde{Y}_2, \widetilde{U}_1, \widetilde{U}_2, \widetilde{c}, \widetilde{p}, \widetilde{J})$ with $|\widetilde{U}_1| = |\widetilde{U}_2| = 2$, let $\widetilde{U}_1 = \{1, 2\}, \widetilde{U}_2 = \{1, 2\}$, then we can construct an $H = 3$ and 2-agent LTC $\mathcal{L}$ as follows:

- Underlying state: $\mathcal{S} = [2]^2$. For each $s_1 \in \mathcal{S}$, we can split s_1 into 2 parts as $s_1 = (s^1_1, s^2_1)$, where $s^1_1, s^2_1 \in [2]$. Similarly, $s_2, s_3 \in \mathcal{S}$ can be split in the same way.
- Initial state distribution: $\forall s_1 \in \mathcal{S}, \mu_1(s_1) = \frac{1}{4}$.
- Control action space: For any $i \in [2], h = 1$, $\mathcal{A}_{i,1} = \{\emptyset\}$; for $h = 2$, $\mathcal{A}_{i,2} = \{(0, x), (x, 0) | x \in [2]\}$; We can write $a_{i,2} = (a^1_{i,2}, a^2_{i,2}), a^1_{i,2}, a^2_{i,2} \in \{0, 1, 2\}$; for $h = 3, \mathcal{A}_{i,3} = [2]$
- Transition: $\forall s \in \mathcal{S}, a_1 \in \mathcal{A}_1, a_2 \in \mathcal{A}_2, a_3 \in \mathcal{A}_3, \mathbb{T}_1(s | s, a_1) = \mathbb{T}_2(s | s, a_2) = \mathbb{T}_3(s | s, a_3) = 1$. Note that under the transition dynamics above, $s_1 = s_2 = s_3 = s_4$ always holds, for any $s_1 \in \mathcal{S}$.
- Observation space: $\mathcal{O}_{1,1} = \mathcal{O}_{2,1} = [2] \times \mathcal{S}, \mathcal{O}_{1,2} = \widetilde{Y}_1 \times \mathcal{S}, \mathcal{O}_{2,2} = \widetilde{Y}_2 \times \mathcal{S}, \mathcal{O}_{3,1} = \mathcal{O}_{3,2} = \mathcal{S}$. For each $i \in [2], o_{i,1} \in \mathcal{O}_{i,1}$, we can split $o_{i,1}$ into 2 parts as $o_{i,1} = (o^1_{i,1}, o^2_{i,1})$, where $o^1_{i,1} \in [2], o^2_{i,1} \in \mathcal{S}$. For each $i \in [2], o_{i,2} \in \mathcal{O}_{i,2}$, similarly, we can split $o_{i,2}$ into 2 parts as $o_{i,2} = (o^1_{i,2}, o^2_{i,2})$, where $o^1_{i,2} \in \widetilde{Y}_i, o^2_{i,2} \in \mathcal{S}$.
- The baseline sharing is null.
- Communication action space: For $i \in [2], h \in \{1, 2\}, \mathcal{M}_{i,h} = \{0, 1\}^{2h-1}$ and $\phi_{i,h}$ is defined as $\phi_{i,h}(p_{i,h^-}, m_{i,h}) = \{o_{i,k} \in p_{i,h^-} | \forall k \le h, \text{the } (2k-1)\text{-th digit of } m_{i,h} \text{ is } 1\} \cup \{a_{i,k} \in p_{i,h^-} | \forall k \le h-1, \text{the } 2k\text{-th digit of } m_{i,h} \text{ is } 1\} \cup \{m_{i,h}\}$; For $h = 3, \mathcal{M}_{i,3} = \{1, 2\}, \forall p_{i,3^-} \in \mathcal{P}_{i,3^-}$, $\phi_{i,3}(p_{i,3^-}, 1) = \{o_{i,2}, a_{i,2}, m_{i,3}\}$ and $\phi_{i,3}(p_{i,3^-}, 2) = \{o_{i,2}, a_{i,2}, o_{i,3}, m_{i,3}\}$.
- Emission: For $h = 1, \forall s_1 \in \mathcal{S}, o_{i,1} \in \mathcal{O}_{i,1}, \mathbb{O}_1(o_1 | s_1) = \Pi^2_{j=1} \mathbb{O}_{j,1}(o_{j,1} | s_1)$ and $\forall i \in [2], \mathbb{O}_{i,1}(o_{i,1} | s_1)$ is defined as:

$$\mathbb{O}_{i,h}(o_{i,h} | s_h) = \begin{cases} \frac{1-\alpha_2}{4} & o^1_{i,h} = s^{i+2h-2}_h, o^2_{i,h} \neq s_h \\ \frac{1-\alpha_2}{4} + \alpha_2 & o^1_{i,h} = s^{i+2h-2}_h, o^2_{i,h} = s_h \\ 0 & \text{o.w.} \end{cases};$$

  for $h = 2, \forall s_2 \in \mathcal{S}, o_2 \in \mathcal{O}_2, \mathbb{O}_2(o_2 | s_2) = \widetilde{p}(o^1_{1,2}, o^1_{2,2}) \Pi^2_{j=1} \mathbb{O}^2_{j,2}(o^2_{j,2} | s_2)$, and $\forall i \in [2], \mathbb{O}^2_{i,2}(o_{i,2} | s_2)$ is defined as:

$$\mathbb{O}^2_{i,2}(o^2_2 | s_2) = \begin{cases} \frac{1-\alpha_2}{4} & o^2_{i,2} \neq s_2 \\ \frac{1-\alpha_2}{4} + \alpha_2 & o^2_{i,2} = s_2 \\ 0 & \text{o.w.} \end{cases};$$

for $h=3, \forall s_3 \in \mathcal{S}, o_{i,3} \in \mathcal{O}_{i,3}, \mathbb{O}_3(o_3|s_3) = \Pi_{j=1}^2 \mathbb{O}_{j,3}(o_{j,3}|s_3)$, and $\forall i \in [2], \mathbb{O}_{i,3}(o_{i,3}|s_3)$ is defined as:

$$\mathbb{O}_{i,3}(o_{i,3}|s_3) = \begin{cases} \frac{1-\alpha_2}{4} & o_{i,3} \neq s_3 \\ \frac{1-\alpha_2}{4} + \alpha_2 & o_{i,3} = s_3 \\ 0 & \text{o.w.} \end{cases}.$$

- Reward functions:

$$\mathcal{R}_1(s_1, a_1) = \mathcal{R}_2(s_2, a_2) = 0, \quad \forall s_1, s_2 \in \mathcal{S}, a_1 \in \mathcal{A}_1, a_2 \in \mathcal{A}_2;$$
$$\mathcal{R}_3(s_3, a_3) = \begin{cases} 1 & \text{if } a_{1,3} = s_3^2 \text{ and } a_{2,3} = s_3^1 \\ 0 & \text{o.w.} \end{cases}.$$

- Communication cost functions:

$$\forall h \in [2], z_h^a \in \mathcal{Z}_h^a, \mathcal{K}_h(z_h^a) = 1 \qquad \text{if } z_h^a \neq \{m_h\}, \text{ else } 0;$$
$$\mathcal{K}_3(z_3^a) = \begin{cases} \widetilde{c}(o_{1,2}^1, o_{2,2}^1, 1, 1)/\alpha_1 & \text{if } a_{1,2}, a_{2,2} \in z_2^a, a_{1,2}^1 = 0, a_{2,2}^1 = 0 \\ \widetilde{c}(o_{1,2}^1, o_{2,2}^1, 2, 1)/\alpha_1 & \text{if } a_{1,2}, a_{2,2} \in z_3^a, a_{1,2}^2 = 0, a_{2,2}^1 = 0 \\ \widetilde{c}(o_{1,2}^1, o_{2,2}^1, 1, 2)/\alpha_1 & \text{if } a_{1,2}, a_{2,2} \in z_3^a, a_{1,2}^1 = 0, a_{2,2}^2 = 0 \\ \widetilde{c}(o_{1,2}^1, o_{2,2}^1, 2, 2)/\alpha_1 & \text{if } a_{1,2}, a_{2,2} \in z_3^a, a_{1,2}^2 = 0, a_{2,2}^2 = 0 \\ 0 & \text{o.w.} \end{cases}.$$

Let $\alpha_0 = \max_{y_1, y_2, u_1, u_2} \widetilde{c}(y_1, y_2, u_1, u_2)$, which is supposed to be positive without loss of optimality (see a discussion in §B-C). We set $\alpha_1 = 2\alpha_0$, and hence for any $z_4^a = \{m_4, o_{1,3}, o_{2,3}, a_{1,3}, a_{2,3}\}$, $\widetilde{\mathcal{K}}_4(z_4^a) \leq \frac{1}{2}$ always holds. Also, we restrict agents to decide their communication strategies only based on their common information. Under such a construction, $\mathcal{L}$ satisfies the following conditions:

- Problem $\mathcal{L}$ is QC: For any $i_1, i_2 \in [2], h_1, h_2 \in [3]$, agent (i_1, h_1) does not influence (i_2, h_2) because agent (i_1, h_1) cannot influence the observation of agent (i_2, h_2), and the baseline sharing is null.
- Problem $\mathcal{L}$ satisfies Assumptions III.1 and III.7: We prove this by showing that each agent $i \in [2]$

satisfies γ-observability. For any $i \in [2], h = 1, b_1, b_2 \in \Delta(\mathcal{S})$, we have

$$\begin{aligned}
\|\mathbb{O}_{i,h}^{\top}(b_1 - b_2)\|_1 &= \sum_{o_{i,h}^1 \in [2]} \sum_{o_{i,h}^2 \in \mathcal{S}} \left| \sum_{s_h \in \mathcal{S}} (b_1(s_h) - b_2(s_h)) \mathbb{O}_{i,h}((o_{i,h}^1, o_{i,h}^2)|s_h) \right| \\
&\geq \sum_{o_{i,h}^2 \in \mathcal{S}} \left| \sum_{o_{i,h}^1 \in [2]} \sum_{s_h \in \mathcal{S}} (b_1(s_h) - b_2(s_h)) \mathbb{O}_{i,h}((o_{i,h}^1, o_{i,h}^2)|s_h) \right| \\
&= \sum_{o_{i,h}^2 \in \mathcal{S}} \left| \sum_{s_h \in \mathcal{S}} \sum_{o_{i,h}^1 \in [2]} (b_1(s_h) - b_2(s_h)) \mathbb{1}[o_{i,h}^1 = s_h^{i+2h-2}] (\frac{1-\alpha_2}{4} + \alpha_2 \mathbb{1}[o_{i,h}^2 = s_h]) \right| \\
&= \sum_{o_{i,h}^2 \in \mathcal{S}} \left| \sum_{s_h \in \mathcal{S}} (b_1(s_h) - b_2(s_h)) (\frac{1-\alpha_2}{4} + \alpha_2 \mathbb{1}[o_{i,h}^2 = s_h]) \right| \\
&= \sum_{o_{i,h}^2 \in \mathcal{S}} \left| \frac{1-\alpha_2}{4} (\sum_{s_h \in \mathcal{S}} (b_1(s_h) - b_2(s_h))) + \alpha_2 (b_1(o_{i,h}^2) - b_2(o_{i,h}^2)) \right| \\
&= \sum_{o_{i,h}^2 \in \mathcal{S}} \alpha_2 |b_1(o_{i,h}^2) - b_2(o_{i,h}^2)| = \alpha_2 \|b_1 - b_2\|_1.
\end{aligned}$$

For any $i \in [2], h = 2, 3$, the proof is similar, by replacing $o_{i,h}^1 \in [2]$ with $o_{i,h}^1 \in \widetilde{Y}_i$ for $h = 2$ and replacing the space $o_{i,h}^1 \in [2]$ with $\{\emptyset\}$ for $h = 3$.

- Problem $\mathcal{L}$ satisfies Assumption III.4 since we restrict agents to decide their communication actions only based on the common information.

Now, we show that any team-optimal strategy of $\mathcal{L}$ will give us the decision rules γ_1, γ_2 to solve $\mathcal{TD}$.
Given $(g_{1:3}^{a,*}, g_{1:3}^{m,*})$ being a team optimal strategy of $\mathcal{L}$, we can construct $(g_{1:3}^{a,'}, g_{1:3}^{m,'})$ as

$$\begin{aligned}
&\forall i \in [2], \forall \tau_{i,1^-}, \tau_{i,2^-}, \tau_{i,3^-}, g_{i,1}^{m,'}(\tau_{i,1^-}) = (0), \quad g_{i,2}^{m,'}(\tau_{i,2^-}) = (0,0,0), g_{i,3}^{m,'}(\tau_{i,3^-}) = 1 \\
&\forall i \in [2], \forall \tau_{i,2^+}, g_{i,2}^{a,'}(\tau_{i,2^+}) = \begin{cases} (0, o_{i,1}) & \text{if the first term of } [g_{i,2^+}^{a,'}(\tau_{i,2^+})] \text{ is } 1 \\ (o_{i,1}, 0) & \text{o.w.} \end{cases}, \\
&\forall i \in [2], \forall \tau_{i,3^+}, g_{i,3}^{a,'}(\tau_{i,3^+}) = \begin{cases} a_{3-i,2}^2 & \text{if } a_{3-i,2}^1 = 0. \\ a_{3-i,2}^1 & \text{o.w.} \end{cases}.
\end{aligned}$$

One can verify that $\mathcal{J}_{\mathcal{L}}(g_{1:3}^{a,*}, g_{1:3}^{m,*}) \leq \mathcal{J}_{\mathcal{L}}(g_{1:3}^{a,'}, g_{1:3}^{m,'})$. This is because, $(g_{1:3}^{a,'}, g_{1:3}^{m,'})$ can always achieve $r_3 = 1$, which means $\mathbb{E}[\sum_{h=1}^3 r_h | (g_{1:3}^{a,'}, g_{1:3}^{m,'})] \geq \mathbb{E}[\sum_{h=1}^3 r_h | g_{1:3}^{a,*}, g_{1:3}^{m,*}]$. Also, at timesteps $h = 1, 2$, $(g_{1:3}^{a,'}, g_{1:3}^{m,'})$ has no additional sharing; if $(g_{1:3}^{a,*}, g_{1:3}^{m,*})$ has additional sharing, then the communication cost $\sum_{h=1}^3 \kappa_h$ is at least 1; if $(g_{1:3}^{a,*}, g_{1:3}^{m,*})$ has no additional sharing, then $(g_{1:3}^{a,*}, g_{1:3}^{m,*})$ and $(g_{1:3}^{a,'}, g_{1:3}^{m,'})$ has the same communication cost of $\sum_{h=1}^3 \kappa_h = \kappa_3 \leq \frac{1}{2}$. Therefore, $\mathbb{E}[\sum_{h=1}^3 \kappa_h | (g_{1:3}^{a,'}, g_{1:3}^{m,'})] \leq \mathbb{E}[\sum_{h=1}^3 \kappa_h | g_{1:3}^{a,*}, g_{1:3}^{m,*}]$, and thus $\mathcal{J}_{\mathcal{L}}(g_{1:3}^{a,*}, g_{1:3}^{m,*}) \leq \mathcal{J}_{\mathcal{L}}(g_{1:3}^{a,'}, g_{1:3}^{m,'})$.

From the above, we know that $(g_{1:3}^{a,'}, g_{1:3}^{m,'})$ is also a team-optimal strategy. Let $U_1 = f_1(a_{1,2}) := 2 - \mathbb{1}[a_{1,2}^1 = 0], U_2 = f_2(a_{2,2}) := 2 - \mathbb{1}[a_{2,2}^1 = 0]$, then $J_{\mathcal{L}}(g_{1:3}^{a,'}, g_{1:3}^{m,'}) = \mathbb{E}[\sum_{h=1}^3 r_h - \kappa_h | g_{1:3}^{a,'}, g_{1:3}^{m,'}] = 1 - \mathbb{E}[\kappa_3 | g_{1:3}^{a,'}, g_{1:3}^{m,'}] = 1 - \frac{1}{\alpha_1} \mathbb{E}[\widetilde{c}(o_{1,2}^1, o_{2,2}^1, U_1, U_2)]$, where $U_1 = f_1(a_{1,2}) = f_1(g_{1,2}^{a,'}(\tau_{1,2^+})), U_2 = f_2(a_{2,2}) = f_2(g_{2,2}^{a,'}(\tau_{2,2^+}))$, which means that $(g_{1,2}^{a,'}, g_{2,2}^{a,'})$ can minimize κ_3 through properly choosing a_2 (and $U_{1:2}$)

if there is no additional sharing in the first two timesteps, i.e., $\tau_{i,2^+} = \{o_{i,1}, a_{i,1}, o_{i,2}, m_{1:2}\}$. By construction, κ_3 only depends on $o^1_{1,2}, o^1_{2,2}$ and $U_{1:2}$[kz:why these three quantities?], and is irrelevant of $\{o_{1,1}, a_{1,1}, o^2_{1,2}\}$, and $m_{i,1} = (0), m_{i,2} = (0,0,0), \forall i \in [2]$ always hold. Hence, $\{o_{1,1}, o_{1,2}, o^2_{1,3}\}$ are useless information for agent 1 to choose $a_{1,2}$ (and U_1), and to minimize $\mathbb{E}[\kappa_3]$[kz:again, as before, I am not sure if this is true – it is only irrelevant WITHIN certain strategy spaces? not $\mathbb{E}[\kappa_3]$ under ANY strategy? we may need to specify WHICH strategy space we are talking about here.]. Therefore, not using them in $g^{a,'}_{1,2}$ does not lose any optimality. Hence, we can consider the $g^{a,'}_{1,2}$ that only has $o^1_{1,2}$ as input. In the same way, we can consider the $g^{a,'}_{2,2}$ that only has $o^1_{2,2}$ as input. Therefore, $J_{\mathcal{L}}(g^{a,'}_{1:3}, g^{m,'}_{1:3}) = 1 - \sum_{o^1_{1,3}, o^2_{1,3}} \frac{1}{\alpha_1} \widetilde{c}(o^1_{1,2}, o^1_{2,2}, f_1(g^{a,'}_{1,2}(o^1_{1,2})), f_2(g^{a,'}_{2,2}(o^1_{2,2})) \widetilde{p}(o^1_{1,2}, o^1_{2,2})$. Further, we can leverage $\widetilde{\gamma}^*_1 = f_1 \circ g^{a,'}_{1,2}, \widetilde{\gamma}^*_2 = f_2 \circ g^{a,'}_{2,2}$ to minimize the expected cost $\widetilde{J}$ of the TDP, where $\circ$ denotes function composition. Therefore, from the NP-hardness of TDPs [31, Corollary 4.1], we complete our proof. □

B-E Proof of Lemma III.8

Proof. We prove this by showing a reduction from the hardness of finding an ϵ-optimal strategy for POMDPs (Proposition B.3). Given any POMDP $\mathcal{P} = (\mathcal{S}^{\mathcal{P}}, \mathcal{A}^{\mathcal{P}}, \mathcal{O}^{\mathcal{P}}, \{\mathbb{O}^{\mathcal{P}}_h\}_{h\in[H^{\mathcal{P}}]}, \{\mathbb{T}^{\mathcal{P}}_h\}_{h\in[H^{\mathcal{P}}]}, \{\mathcal{R}^{\mathcal{P}}_h\}_{h\in[H^{\mathcal{P}}]}, \mu^{\mathcal{P}}_1)$ with rewards bounded in $[0, \frac{1}{2}]$, we can construct an LTC $\mathcal{L}$ with 2 agents as follows:

- Number of agents: $n = 2$; length of episode: $H = H^{\mathcal{P}}$.
- $\mathcal{S} = \mathcal{S}^{\mathcal{P}} \times [2]$; for any $s \in \mathcal{S}$, we can split state as two parts: $s = (s^1, s^2), s^1 \in \mathcal{S}, s^2 \in [2]$.
- Initial state distribution: $\forall s_1 \in \mathcal{S}, \mu_1(s_1) = \frac{\mu^{\mathcal{P}}_1(s^1_1)}{2}$.
- Control action space: For any $h \in [H]$, $\mathcal{A}_{1,h} = \mathcal{A}^{\mathcal{P}}_h, \mathcal{A}_{2,h} = [2]$.
- Transition: For any $h \in [H]$, $\forall s_h, s_{h+1} \in \mathcal{S}, a_h \in \mathcal{A}_h, \mathbb{T}_h(s_{h+1}|s_h, a_h) = \mathbb{T}^{\mathcal{P}}_h(s^1_{h+1}|s^1_h, a_{1,h}) \mathbb{1}[s^2_{h+1} = a_{2,h}]$.
- Observation space: For any $h \in [H]$, $\mathcal{O}_{1,h} = \mathcal{O}^{\mathcal{P}}, \mathcal{O}_{2,h} = \mathcal{S}$.
- Emission: For any $h \in [H], \forall o_h \in \mathcal{O}_h, s_h \in \mathcal{S}, \mathbb{O}_h(o_h|s_h) = \mathbb{O}^{\mathcal{P}}_h(o_{1,h}|s^1_h) \mathbb{1}[o_{2,h} = s_h]$.
- Reward functions: For any $h \in [H]$, $\forall s_h \in \mathcal{S}, a_h \in \mathcal{A}_h, \mathcal{R}_h(s_h, a_h) = \mathcal{R}^{\mathcal{P}}_h(s^1_h, a_{1,h})/H$.
- Baseline sharing: For any $h \in [H], z^b_h = \{o_{1,h}, a_{1,h-1}\}$.
- Communication action space: For any $h \in [H], \mathcal{M}_{1,h} = \{\emptyset\}, \mathcal{M}_{2,h} = \{0,1\}^{2h-1}$. For any $p_{1,h^-} \in \mathcal{P}_{1,h^-}, p_{2,h^-} \in \mathcal{P}_{2,h^-}, m_h \in \mathcal{M}_h, \phi_{1,h}(p_{1,h^-}, m_{1,h}) = \{m_{1,h}\}, \phi_{2,h}(p_{2,h^-}, m_{2,h}) = \{o_{2,k} | \forall k \le h$, the $2k-1$-th digit of $m_{2,h}$ is 1 and $o_{2,k} \in p_{2,h^-}\} \cup \{a_{2,k} | \forall k < h$, the $2k$-th digit of $m_{2,h}$ is 1 and $a_{2,k} \in p_{2,h^-}\} \cup \{m_{2,h}\}$.
- Communication cost functions: For any $h \in [H], z^a_h \in \mathcal{Z}^a_h, \mathcal{K}_h(z^a_h) = \mathbb{1}[z^a_h \neq \{m_h\}]$, which means that the communication cost is 1 unless there is no additional sharing.
- We restrict the communication strategy to only use c_h as input.

We first verify that $\mathcal{L}$ is QC and satisfies Assumptions III.1, III.4, and III.5.

- $\mathcal{L}$ is QC: For any $\forall h_1 < h_2 \le H$, agent $(2, h_1)$ does not influence agent $(1, h_2)$ under baseline sharing, since agent $(2, h_1)$ does not influence $s^1_h, \forall h \in [H]$, then does not influence $o_{1,h}, \forall h \in [H]$. Also, agent 2 shares nothing via baseline sharing. Therefore, agent $(2, h_1)$ does not influence agent $(1, h_2)$.

For any $h_1 < h_2 \le H$, under baseline sharing, $p_{1,h_1^-} = \emptyset$. Then, we have $\sigma(\tau_{1,h_1^-}) \subseteq \sigma(c_{h_1^-}) \subseteq \sigma(c_{h_2^-}) \subseteq \sigma(\tau_{2,h_2^-})$. [kz:I think the indices are a bit messed up – we started with $(2,h_1)$ and $(1,h_2)$, but now we switched them. can we check here.] [hy:it discusses two cases here.]

- $\mathcal{L}$ satisfies Assumption III.1: For any $h \in [H], b_1, b_2 \in \Delta(\mathcal{S})$, $\mathbb{O}_h$ satisfies that

$$\begin{aligned}
\|\mathbb{O}_h^\top (b_1 - b_2)\|_1 &= \sum_{o_{1,h} \in \mathcal{O}^{\mathcal{P}}} \sum_{o_{2,h} \in \mathcal{S}} \left| \sum_{s_h \in \mathcal{S}} (b_1(s_h) - b_2(s_h)) \mathbb{O}_h((o_{1,h}, o_{2,h}) | s_h) \right| \\
&\ge \sum_{o_{2,h} \in \mathcal{S}} \left| \sum_{o_{1,h} \in \mathcal{O}^{\mathcal{P}}} \sum_{s_h \in \mathcal{S}} (b_1(s_h) - b_2(s_h)) \mathbb{O}_{1,h}(o_{1,h} | s_h) \mathbb{O}_{2,h}(o_{2,h} | s_h) \right| \\
&= \sum_{o_{2,h} \in \mathcal{S}} \left| \sum_{s_h \in \mathcal{S}} (b_1(s_h) - b_2(s_h)) \mathbb{O}_{2,h}(o_{2,h} | s_h) \sum_{o_{1,h} \in \mathcal{O}^{\mathcal{P}}} \mathbb{O}_{1,h}(o_{1,h} | s_h) \right| \\
&= \sum_{o_{2,h} \in \mathcal{S}} \left| \sum_{s_h \in \mathcal{S}} (b_1(s_h) - b_2(s_h)) \mathbb{1}[o_{2,h} = s_h] \right| \\
&= \sum_{o_{2,h} \in \mathcal{S}} |b_1(o_{2,h}) - b_2(o_{2,h})| = \|b_1 - b_2\|_1,
\end{aligned}$$

  showing that γ-observability is satisfied with $\gamma = 1$.

- $\mathcal{L}$ satisfies Assumption III.4: For any $h \in [H]$, we restricted that each agent i decides $m_{i,h}$ based on c_h only.

- $\mathcal{L}$ satisfies Assumption III.5: For any $h \in [H-1], a_{1,h}$ influences s_{h+1}^1, $a_{2,h}$ influences s_{h+1}^2.

Consider the communication strategy $g_{1:H}^{m,'}$ that yields no additional sharing, namely, $\forall h \in [H], \tau_{2,h^-} \in \mathcal{T}_{2,h^-}, g_{2,h}^{m,'}(\tau_{2,h^-}) = \mathbf{0}_h$. For any ϵ-optimal strategy $(g_{1:H}^{a,*}, g_{1:H}^{m,*})$ with $\epsilon \in [0, \frac{1}{4})$, we claim that $(g_{1:H}^{a,*}, g_{1:H}^{m,'})$ is also an ϵ-optimal strategy. This is because, comparing to strategy $(g_{1:H}^{a,*}, g_{1:H}^{m,'})$, for the trajectories where $(g_{1:H}^{a,*}, g_{1:H}^{m,*})$ leads to no additional sharing, these two strategies output the same actions and gain the same total return; for the other trajectories where $(g_{1:H}^{a,*}, g_{1:H}^{m,*})$ leads to some additional sharing, i.e., $z_h \neq \{m_h\}$ for some $h \in [H]$, then it has the return of $\sum_{t=1}^H (r_t - \kappa_t) \le \sum_{t=1}^H (r_t) - \kappa_h \le \frac{1}{2H} \cdot H - 1 < 0$, which is less than when replacing it by $(g_{1:H}^{a,*}, g_{1:H}^{m,'})$ with return $\sum_{t=1}^H (r_t - \kappa_t) = \sum_{t=1}^H r_t \ge 0$. Therefore, $J_{\mathcal{L}}(g_{1:H}^{a,*}, g_{1:H}^{m,'}) \ge J_{\mathcal{L}}(g_{1:H}^{a,*}, g_{1:H}^{m,*})$, and $(g_{1:H}^{a,*}, g_{1:H}^{m,'})$ is also an ϵ-optimal strategy with $\epsilon \in [0, \frac{1}{4})$.

Meanwhile, any $(g_{1:H}^{a,*}, g_{1:H}^{m,'})$ being an $\frac{\epsilon}{H}$-team optimal strategy of $\mathcal{L}$ will directly give an ϵ-team-optimal strategy of $\mathcal{P}$ as $\{g_{1,h}^{a,*}\}_{h \in [H]}$, since when there is no sharing, the decision process is only controlled by agent 1. From Proposition B.3, we complete the proof. □

C. Deferred Details of §IV

C-A Proof of Proposition IV.1

Proof. Given any strategy $(g_{1:H}^a, g_{1:H}^m), g_{1:H}^a \in \mathcal{G}_{1:H}^a, g_{1:H}^m \in \mathcal{G}_{1:H}^m$, we can define $\widetilde{g}_{1:\widetilde{H}} = (g_1^m, g_1^a, \cdots, g_H^m, g_H^a)$. From the construction of $\mathcal{D}_{\mathcal{L}}$, comparing to applying $(g_{1:H}^a, g_{1:H}^m)$ in $\mathcal{L}$, if we apply $\widetilde{g}_{1:\widetilde{H}}$ in $\mathcal{D}_{\mathcal{L}}$, then it is easy to verify that $\forall h \in H, \widetilde{r}_{2h-1} = -\kappa_h, \widetilde{r}_{2h} = r_h$ always holds. Therefore, $\mathcal{J}_{\mathcal{D}_{\mathcal{L}}}(\widetilde{g}_{1:\widetilde{H}}) = \mathcal{J}_{\mathcal{L}}(g_{1:H}^a, g_{1:H}^m)$. In the same way, for any strategy $\widetilde{g}_{1:\widetilde{H}}$, we can define $\widetilde{g}_{1:\widetilde{H}} = (g_1^m, g_1^a, \cdots, g_H^m, g_H^a)$, and verify $\mathcal{J}_{\mathcal{L}}(g_{1:H}^a, g_{1:H}^m) = \mathcal{J}_{\mathcal{D}_{\mathcal{L}}}(\widetilde{g}_{1:\widetilde{H}})$. □

C-B Proof of Theorem IV.2

Proof. Firstly, we prove the QC case. To show $\mathcal{D}_{\mathcal{L}}$ is QC, we need to prove $\forall i_1, i_2 \in [n], h_1, h_2 \in [\widetilde{H}]$, if agent (i_1, h_1) influences agent (i_2, h_2) with $h_1 < h_2$, then $\sigma(\widetilde{\tau}_{i_1,h_1}) \subseteq \sigma(\widetilde{\tau}_{i_2,h_2})$, where we use $\widetilde{\tau}_{i,h}$ to denote the available information of agent (i,h) in $\mathcal{D}_{\mathcal{L}}$. We prove this by considering the following cases:

1. If $h_1 = 2t_1 - 1$ with $t_1 \in [H]$, by the construction of $\mathcal{D}_{\mathcal{L}}$ and Assumption III.4, we have $\widetilde{\tau}_{i_1,h_1} = \widetilde{c}_{h_1} = c_{t_1^-} \subseteq \widetilde{\tau}_{i_2,h_2}$, since common information accumulates over time by definition, and will always be included in the available information $\widetilde{\tau}_{i,h}$ in later steps with $h > h_1$. Thus, $\sigma(\widetilde{\tau}_{i_1,h_1}) \subseteq \sigma(\widetilde{\tau}_{i_2,h_2})$.

2. If $h_1 = 2t_1, h_2 = 2t_2$ with $t_1, t_2 \in [H]$, then $\widetilde{\tau}_{i_1,h_1} = \tau_{i_1,t_1^+} = \tau_{i_1,t_1^-} \cup z^a_{t_1}$ by definition. Consider agent (i_1, t_1) and (i_2, t_2) in $\mathcal{L}$. From Lemma B.2, we know $\sigma(\tau_{i_1,t_1^-}) \subseteq \sigma(\tau_{i_2,t_2^-}) \subseteq \sigma(\tau_{i_2,t_2^+})$. Also, $z^a_{t_1} \subseteq c_{t_1^+} \subseteq c_{t_2^+} \subseteq \tau_{i_2,t_2^+} = \widetilde{\tau}_{i_2,h_2}$ by the accumulation of c_{h^+} over time. Thus, we have $\sigma(\widetilde{\tau}_{i_1,h_1}) \subseteq \sigma(\widetilde{\tau}_{i_2,h_2})$.

3. If $h_1 = 2t_1, h_2 = 2t_2 - 1, t_1, t_2 \in [H]$, then $\widetilde{\tau}_{i_2,h_2} = \widetilde{c}_{h_2}$. Hence, $\exists i_3 \in [n], i_3 \neq i_1$, such that $\widetilde{\tau}_{i_2,h_2} \subseteq \widetilde{c}_{h_2+1} \subseteq \widetilde{\tau}_{i_3,h_2+1}$. Since agent (i_1, h_1) influences agent (i_2, h_2), we thus know that agent (i_1, h_1) also influences agent $(i_3, h_2 + 1)$ in $\mathcal{D}_{\mathcal{L}}$. Since $\mathcal{L}$ is QC, we know $\sigma(\tau_{i_1,t_1^-}) \subseteq \sigma(\tau_{i_3,t_2^-})$ by Lemma B.2. From Assumption II.2 and $i_1 \neq i_3$, we know $\sigma(\tau_{i_1,t_1^-}) \subseteq \sigma(c_{t_2^-}) = \sigma(\widetilde{\tau}_{i_2,h_2})$. Also, it holds that $\tau_{i_1,t_1^+} \backslash \tau_{i_1,t_1^-} \subseteq c_{t_1^+} \subseteq c_{t_2^-}$. Hence, we have $\sigma(\widetilde{\tau}_{i_1,h_1}) = \sigma(\tau_{i_1,h^+}) \subseteq \sigma(c_{t_2^-}) = \sigma(\widetilde{\tau}_{i_2,h_2})$.

Second, we prove the sQC case. In $\mathcal{D}_{\mathcal{L}}$, for any $i_1, i_2 \in [n], h_1, h_2 \in [\widetilde{H}]$, agent (i_1, h_1) influences (i_2, h_2). From the proof above, we know $\sigma(\widetilde{\tau}_{i_1,h_1}) \subseteq \sigma(\widetilde{\tau}_{i_2,h_2})$. We only need to prove $\sigma(\widetilde{a}_{i_1,h_1}) \subseteq \sigma(\widetilde{\tau}_{i_2,h_2})$.

1. If $h_1 = 2t_1 - 1$ with $t_1 \in [H]$, then we know $\widetilde{a}_{i_1,h_1} = m_{i_1,t_1}$. From Assumption II.1, we know that $m_{i_1,t_1} \subseteq z^a_{i_1,t_1}$. Then, we get $\sigma(\widetilde{a}_{i_1,h_1}) \subseteq \sigma(\widetilde{z}_{h_1+1}) \subseteq \sigma(\widetilde{c}_{h_2}) \subseteq \sigma(\widetilde{\tau}_{i_2,h_2})$.

2. If $h_1 = 2t_1, h_2 = 2t_2$ with $t_1, t_2 \in [H]$, then from Lemma B.2, we know that $\sigma(\widetilde{a}_{i_1,h_1}) \subseteq \sigma(\widetilde{\tau}_{i_2,h_2})$.

3. If $h_1 = 2t_1, h_2 = 2t_2 - 1, t_1, t_2 \in [H]$, then $\widetilde{\tau}_{i_2,h_2} = \widetilde{c}_{h_2}$. Hence, $\exists i_3 \in [n], i_3 \neq i_1$, such that $\widetilde{\tau}_{i_2,h_2} \subseteq \widetilde{c}_{h_2+1} \subseteq \widetilde{\tau}_{i_3,h_2+1}$. Since agent (i_1, h_1) influences agent (i_2, h_2), we thus know that agent (i_1, h_1) also influences agent $(i_3, h_2 + 1)$. Since $\mathcal{L}$ is sQC, we know $\sigma(a_{i_1,t_1}) \subseteq \sigma(\tau_{i_3,t_2^-})$ in $\mathcal{D}_{\mathcal{L}}$ by Lemma B.2. From Assumption II.2 and $i_1 \neq i_3$, we know $\sigma(\widetilde{a}_{i_1,h_1}) = \sigma(a_{i_1,t_1}) \subseteq \sigma(c_{t_2^-}) = \sigma(\widetilde{\tau}_{i_2,h_2})$.

This completes the proof. □

C-C Proof of Lemma IV.3

Proof. From the construction of $\mathcal{D}^{\dagger}_{\mathcal{L}}$, since $\mathcal{D}^{\dagger}_{\mathcal{L}}$ requires agents to share more than $\mathcal{D}_{\mathcal{L}}$, it is easy to observe that $\forall h \in [\widetilde{H}], i \in [n], \widetilde{c}_h \subseteq \check{c}_h, \widetilde{\tau}_{i,h} \subseteq \check{\tau}_{i,h}$.
Let $i_1, i_2 \in [n], h_1, h_2 \in [\widetilde{H}], h_1 < h_2$, and agent (i_1, h_1) influences agent (i_2, h_2) in $\mathcal{D}^{\dagger}_{\mathcal{L}}$.

- If $h_1 = 2t_1 - 1$ with $t_1 \in [H]$, then h_1 is a communication step. Hence, $\check{\tau}_{i_1,h_1} = \check{c}_{h_1} \subseteq \check{c}_{h_2}$ and $\widetilde{a}_{i_1,h_1} = m_{i_1,t_1} \subseteq \check{c}_{h_1+1} \subseteq \check{c}_{h_2}$ from Assumption II.1. Therefore, we have $\sigma(\check{\tau}_{i_1,h_1}) \cup \sigma(\check{a}_{i_1,h_1}) \subseteq \sigma(\check{c}_{h_2}) \subseteq \sigma(\check{\tau}_{i_2,h_2})$.

- If $h_1 = 2t_1, h_2 = 2t_2 - 1$ with $t_1, t_2 \in [H]$, then $\check{\tau}_{i_2,h_2} = \check{c}_{h_2}$. If agent (i_1, h_1) does not influence (i_2, h_2) in $\mathcal{D}_{\mathcal{L}}$, but agent (i_1, h_1) influences (i_2, h_2) in $\mathcal{D}^{\dagger}_{\mathcal{L}}$, then it means $\check{a}_{i_1,h_1} \in \check{\tau}_{i_2,h_2}$ but $\widetilde{a}_{i_1,h_1} \notin \widetilde{\tau}_{i_2,h_2}$. This can only happen when $\sigma(\widetilde{\tau}_{i_1,h_1}) \subseteq \sigma(\widetilde{c}_{h_2}) \subseteq \sigma(\check{c}_{h_2})$, and $\widetilde{a}_{i_1,h_1} \subseteq \check{c}_{h_2}$. Also, from the construction of $\mathcal{D}^{\dagger}_{\mathcal{L}}$, we know that $\check{\tau}_{i_1,h_1} \backslash \widetilde{\tau}_{i_1,h_1} \subseteq \check{c}_{h_1}$. Therefore, we have $\sigma(\check{\tau}_{i_1,h_1}) \cup \sigma(\widetilde{a}_{i_1,h_1}) \subseteq \sigma(\check{c}_{h_2}) \subseteq \sigma(\check{\tau}_{i_2,h_2})$.

  If agent (i_1, h_1) influences agent (i_2, h_2) in $\mathcal{D}_{\mathcal{L}}$, then we claim that $\widetilde{a}_{i_1,h_1}$ influences $\widetilde{s}_{h_1+1}$ in $\mathcal{D}_{\mathcal{L}}$, since otherwise, $\widetilde{a}_{i_1,h_1}$ does not influence any $\widetilde{s}_h, \widetilde{o}_h$, and $\widetilde{a}_h$ with $h > h_1$; also, from Assumption

III.5, $\widetilde{a}_{i_1,h_1}$ is removed from $\widetilde{\tau}_h, \forall h > h_1$, and thus agent (i_1,h_1) cannot influence agent (i_2,h_2). Meanwhile, from the QC IS of $\mathcal{D}_{\mathcal{L}}$, we know that $\sigma(\widetilde{\tau}_{i_1,h_1}) \subseteq \sigma(\widetilde{\tau}_{i_2,h_2}) = \sigma(\widetilde{c}_{h_2})$. Then, from the construction of $\mathcal{D}^{\dagger}_{\mathcal{L}}$, we know that $\widetilde{a}_{i_1,h_1}$ is added in $\check{c}_{h_2}$, i.e., $\widetilde{a}_{i_1,h_1} \in \check{c}_{h_2}$. Still, due to $\check{\tau}_{i_1,h_1} \backslash \widetilde{\tau}_{i_1,h_1} \subseteq \check{c}_{h_1}$, we further have $\sigma(\check{\tau}_{i_1,h_1}) \cup \sigma(\widetilde{a}_{i_1,h_1}) \subseteq \sigma(\check{\tau}_{i_2,h_2})$.

- If $h_1 = 2t_1, h_2 = 2t_2$ with $t_1, t_2 \in [H]$. If agent (i_1,h_1) does not influence (i_2,h_2) in $\mathcal{D}_{\mathcal{L}}$, then it means that adding $\widetilde{a}_{i_1,h_1}$ in $\check{c}_{h_2}$ via strict expansion leads to such influence. Then, it must hold that $\sigma(\widetilde{\tau}_{i_1,h_1}) \subseteq \sigma(\widetilde{c}_{h_2}) \subseteq \sigma(\check{c}_{h_2})$, and $\widetilde{a}_{i_1,h_1} \subseteq \check{c}_{h_2}$. We can conclude $\sigma(\check{\tau}_{i_1,h_1}) \cup \sigma(\widetilde{a}_{i_1,h_1}) \subseteq \sigma(\check{c}_{h_2}) \subseteq \sigma(\check{\tau}_{i_2,h_2})$.
  If agent (i_1,h_1) influences (i_2,h_2) in $\mathcal{D}_{\mathcal{L}}$, then we know that $\widetilde{a}_{i_1,h_1}$ influences $\widetilde{s}_{h_1+1}$ in $\mathcal{D}_{\mathcal{L}}$, similarly as shown in the case for $h_2 = 2t_2 - 1$. Therefore, from Assumption III.7, we know that there exists some $i_3 \neq i_1$ such that agent $\widetilde{a}_{i_1,h_1}$ influences agent (i_3, h_1+1) in $\mathcal{D}_{\mathcal{L}}$. Then, from the QC IS of $\mathcal{D}_{\mathcal{L}}$, we know that $\sigma(\widetilde{\tau}_{i_1,h_1}) \subseteq \sigma(\widetilde{\tau}_{i_3,h_1+1})$. Meanwhile, from Assumption II.1 (e), we know that $\widetilde{\tau}_{i_3,h_1+1} \subseteq \widetilde{\tau}_{i_3,h_2}$. Therefore, we can get $\sigma(\widetilde{\tau}_{i_1,h_1}) \subseteq \sigma(\widetilde{\tau}_{i_3,h_2})$ and further $\sigma(\widetilde{\tau}_{i_1,h_1}) \subseteq \sigma(\widetilde{c}_{h_2})$ due to Assumption II.2 and $i_3 \neq i_1$. Then, from the construction of $\mathcal{D}^{\dagger}_{\mathcal{L}}$, we know that $\widetilde{a}_{i_1,h_1}$ is added in $\check{c}_{h_2}$. Together with the fact that $\check{\tau}_{i_1,h_1} \backslash \widetilde{\tau}_{i_1,h_1} \subseteq \check{c}_{h_1} \subseteq \check{c}_{h_2}$, we can conclude that $\sigma(\check{\tau}_{i_1,h_1}) \cup \sigma(\check{a}_{i_1,h_1}) \subseteq \sigma(\check{\tau}_{h_2})$.

This completes the proof. □

C-D Proof of Theorem IV.4

Proof. We claim that given any strategy $\check{g}_{1:\check{H}}$ and $\widetilde{g}_{1:\widetilde{H}} = \varphi(\check{g}_{1:\check{H}}, \mathcal{D}_{\mathcal{L}})$, $J_{\mathcal{D}^{\dagger}_{\mathcal{L}}}(\check{g}_{1:\check{H}}) = J_{\mathcal{D}_{\mathcal{L}}}(\widetilde{g}_{1:\widetilde{H}})$, where the function φ is given by Algorithm 3. In order to prove this statement, we prove that $\widetilde{g}_{i,h}(\widetilde{\tau}_{i,h}) = \check{g}_{i,h}(\check{\tau}_{i,h})$ always holds for any $\widetilde{\tau}_{i,h}$, which is equivalent to prove that for any $i \in [n], h \in [\widetilde{H}]$, and $\widetilde{\tau}_{i,h}$, Algorithm 3 can compute the associated $\check{\tau}_{i,h}$ from the expansion in Equation (IV.2), and use it as the input of $\check{g}_{i,h}$ (Line 11 of Algorithm 3). Let $\check{\tau}'_{i,h}$ be the input of $\check{g}_{i,h}$ used in Algorithm 3, then it holds that $\widetilde{\tau}_{i,h} \subseteq \check{\tau}'_{i,h}$. We now aim to show that $\check{\tau}_{i,h} = \check{\tau}'_{i,h}$:

- For any $j \in [n], t < h$ such that $\widetilde{a}_{j,t} \in \check{\tau}_{i,h} \backslash \widetilde{\tau}_{i,h}$, it must hold that $\sigma(\widetilde{\tau}_{j,t}) \subseteq \sigma(\widetilde{c}_h)$ from Equation (IV.2). From Algorithm 3, we know that $\widetilde{a}_{j,t} \in \check{\tau}'_{i,h}$, and thus $\check{\tau}_{i,h} \subseteq \check{\tau}'_{i,h}$.
- For any $j \in [n], t < h$ such that $\widetilde{a}_{j,t} \in \check{\tau}'_{i,h} \backslash \widetilde{\tau}_{i,h}$, from Algorithm 3, it must hold that $\sigma(\widetilde{\tau}_{j,t}) \subseteq \sigma(\widetilde{c}_h)$. Therefore, from Equation (IV.2), we know that $\widetilde{a}_{j,t} \in \check{\tau}_{i,h}$, and thus $\check{\tau}'_{i,h} \subseteq \check{\tau}_{i,h}$.

Therefore, we conclude that $\check{\tau}_{i,h} = \check{\tau}'_{i,h}$ and prove the statement.

Since $\mathcal{D}^{\dagger}_{\mathcal{L}}$ has larger strategy spaces, i.e., $\max_{\widetilde{g}_{1:\widetilde{H}} \in \widetilde{G}_{1:\widetilde{H}}} J_{\mathcal{D}_{\mathcal{L}}}(\widetilde{g}_{1:\widetilde{H}}) \leq \max_{\check{g}_{1:\check{H}} \in \check{G}_{1:\check{H}}} J_{\mathcal{D}^{\dagger}_{\mathcal{L}}}(\check{g}_{1:\check{H}})$. Let $\check{g}^*_{1:\check{H}}$ be the strategy satisfying $J_{\mathcal{D}^{\dagger}_{\mathcal{L}}}(\check{g}^*_{1:\check{H}}) \geq \max_{\check{g}_{1:\check{H}} \in \check{G}_{1:\check{H}}} J_{\mathcal{D}^{\dagger}_{\mathcal{L}}}(\check{g}_{1:\check{H}}) - \epsilon$. Then, we have $J_{\mathcal{D}_{\mathcal{L}}}(\varphi(\check{g}^*_{1:\check{H}}, \mathcal{D}_{\mathcal{L}})) = J_{\mathcal{D}^{\dagger}_{\mathcal{L}}}(\check{g}^*_{1:\check{H}}) \geq \max_{\check{g}_{1:\check{H}} \in \check{G}_{1:\check{H}}} J_{\mathcal{D}^{\dagger}_{\mathcal{L}}}(\check{g}_{1:\check{H}}) - \epsilon \geq \max_{\widetilde{g}_{1:\widetilde{H}} \in \widetilde{G}_{1:\widetilde{H}}} J_{\mathcal{D}_{\mathcal{L}}}(\widetilde{g}_{1:\widetilde{H}}) - \epsilon$. Thus, $\varphi(\check{g}^*_{1:\check{H}}, \mathcal{D}_{\mathcal{L}})$ is an ϵ-team optimal strategy of $\mathcal{D}_{\mathcal{L}}$. □

[kz:I HAVE COPY PASTED ALL THE "TECHNICAL PROOFS" FROM TAC VERSION UNTIL THIS POINT...]

Lemma C.1. For any given strategy $\check{g}_{1:\check{H}} \in \check{\mathcal{G}}_{1:\check{H}}$, implementing Algorithm 4 in $\mathcal{D}_{\mathcal{L}}$ is equivalent to implementing $\varphi(\check{g}_{1:\check{H}}, \mathcal{D}_{\mathcal{L}})$ in $\mathcal{D}_{\mathcal{L}}$.

Proof. As shown in Theorem IV.4, Algorithm 3 can compute the associated $\check{\tau}_{i,h}$ and use it as the input of $\check{g}_{i,h}$ (Line 12). Therefore, it suffices to prove that Algorithm 4 can also compute the associated $\check{\tau}_{i,h}$ and use it as the input of $\check{g}_{i,h}$ (Line 6), i.e., Algorithm 5 can output the associated $\check{\tau}_{i,h}$ from $\widetilde{\tau}_{i,h}$ and

$\check{g}_{1:h-1}$. We prove this by induction.
Firstly, when $h = 1$, it holds for any $i \in [n]$ such that $\widetilde{\tau}_{i,1} = \check{\tau}_{i,1}$. In Algorithm 5, when $h = 1$, it will never enter the for loop, and thus the output is $\widetilde{\tau}_{i,1} = \check{\tau}_{i,1}$.
Secondly, we assume for any $h < t$, the hypothesis holds. Then for $h = t$, given any $\widetilde{\tau}_{i,t} \in \widetilde{\mathcal{T}}_{i,t}$ and $\check{g}_{1:t-1}$, let $\check{\tau}'_{i,t}$ be the output of Algorithm 5. For any $j \in [n], h' < t$, if it holds that $\sigma(\widetilde{\tau}_{j,h'}) \subseteq \sigma(\widetilde{c}_t)$ in $\mathcal{D}_{\mathcal{L}}$ and $\widetilde{a}_{j,h'} \notin \widetilde{\tau}_{i,t}$, then it can compute the associated $\check{\tau}_{j,h'}$ from induction hypothesis (Line 5-6), compute the exact $\check{a}_{j,h'}$ based on $\check{g}_{j,h'}$ (Line 7), and add it into $\check{\tau}'_{i,h}$ (Line 8). Therefore, we know $\check{\tau}'_{i,t} = \widetilde{\tau}_{i,t} \cup \{\widetilde{a}_{j,h'} \,|\, \sigma(\widetilde{\tau}_{j,h'}) \subseteq \sigma(\widetilde{c}_t) \text{ and } \widetilde{a}_{j,h'} \notin \widetilde{\tau}_{i,t}\} = \check{\tau}_{i,t}$.
From induction, we complete the proof. □

Remark C.2. The difference between Algorithm 3 and 4 lies as follows. Given any $\check{g}_{1:\check{H}}$ and $\mathcal{D}_{\mathcal{L}}$, Algorithm 3 needs to recover the output $\widetilde{a}_{i,h}$ of $\widetilde{g}_{i,h}$ under all possible input $\widetilde{\tau}_{i,h} \in \widetilde{\mathcal{T}}_{i,h}, i \in [n], h \in [\widetilde{H}]$, where the cardinality of $\widetilde{\mathcal{T}}_{i,h}$ could be exponentially large. Thus Algorithm 3 may suffer from computational intractability. However, Algorithm 4 only requires to recover the output $\widetilde{g}_{i,h}$ under the specific $\widetilde{\tau}_{i,h}$ happening in the trajectory, which can be implemented in polynomial time.

C-E Proof of Theorem IV.5

Proof. To prove that $\mathcal{D}^{\dagger}_{\mathcal{L}}$ has SI-CIBs, it suffices to prove that for any $h = 2, \cdots, \check{H}$, fix any $h_1 \in [h-1], i_1 \in [n]$, and for any $\check{g}_{1:h-1} \in \check{\mathcal{G}}_{1:h-1}, \check{g}'_{i_1,h_1} \in \check{\mathcal{G}}_{i_1,h_1}$, let $\check{g}'_{h_1} := (\check{g}_{1,h_1}, \cdots, \check{g}'_{i_1,h_1}, \cdots, \check{g}_{n,h_1})$ and $\check{g}'_{1:h-1} := (\check{g}_1, \cdots, \check{g}'_{h_1}, \cdots, \check{g}_{h-1})$. If $\check{c}_h$ is reachable under both $\check{g}_{1:h-1}$ and $\check{g}'_{1:h-1}$, then the following holds

$$\mathbb{P}(\check{s}_h, \check{p}_h \,|\, \check{c}_h, \check{g}_{1:h-1}) = \mathbb{P}(\check{s}_h, \check{p}_h \,|\, \check{c}_h, \check{g}'_{1:h-1}). \tag{C.1}$$

We prove this result case by case as follows:

1. If there exists some $i_3 \neq i_1$ such that $\sigma(\check{\tau}_{i_1,h_1}) \subseteq \sigma(\check{\tau}_{i_3,h})$ and $\sigma(\check{a}_{i_1,h_1}) \subseteq \sigma(\check{\tau}_{i_3,h})$, then from Assumption II.2, we know that $\sigma(\check{\tau}_{i_1,h_1}) \subseteq \sigma(\check{c}_h), \sigma(\check{a}_{i_1,h_1}) \subseteq \sigma(\check{c}_h)$. Therefore, there exist deterministic measurable functions α_1, α_2 such that $\check{\tau}_{i_1,h_1} = \alpha_1(\check{c}_h), \check{a}_{i_1,h_1} = \alpha_2(\check{c}_h)$, and further it holds that

$$\begin{aligned}
\mathbb{P}(\check{s}_h, \check{p}_h \,|\, \check{c}_h, \check{g}_{1:h-1}) &= \mathbb{P}(\check{s}_h, \check{p}_h \,|\, \alpha_1(\check{c}_h), \alpha_2(\check{c}_h), \check{c}_h, \check{g}_{1:h-1}) \\
&= \mathbb{P}(\check{s}_h, \check{p}_h \,|\, \check{\tau}_{i_1,h_1}, \check{a}_{i_1,h_1}, \check{c}_h, \check{g}_{1:h-1}) \\
&= \mathbb{P}(\check{s}_h, \check{p}_h \,|\, \check{\tau}_{i_1,h_1}, \check{a}_{i_1,h_1}, \check{c}_h, \check{g}'_{1:h-1}).
\end{aligned}$$

   The last equality is due to the fact that both the input and output of $\check{g}_{i_1,h_1}$ are conditioned on.

2. If for any $i_2 \neq i_1$, either $\sigma(\check{\tau}_{i_1,h_1}) \not\subseteq \sigma(\check{\tau}_{i_2,h})$ or $\sigma(\check{a}_{i_1,h_1}) \not\subseteq \sigma(\check{\tau}_{i_2,h})$, then agent (i_1, h_1) does not influence any agent (i_2, h) with $i_2 \neq i_1$ in $\mathcal{D}^{\dagger}_{\mathcal{L}}$, since otherwise, due to the sQC IS of $\mathcal{D}^{\dagger}_{\mathcal{L}}$, it must hold that $\sigma(\check{\tau}_{i_1,h_1}) \subseteq \sigma(\check{\tau}_{i_2,h})$ and $\sigma(\check{a}_{i_1,h_1}) \subseteq \sigma(\check{\tau}_{i_2,h})$. Moreover, we claim that such an h_1 has to be even, since otherwise, the agent will be at a communication step, and we must have $\check{\tau}_{i_1,h_1} = \check{c}_{h_1} \subseteq \check{c}_h \subseteq \check{\tau}_{i_2,h}$ by Assumption III.4 and the reformulation in Equation (IV.1), and $\check{a}_{i_1,h_1} = m_{i_1,\frac{h_1+1}{2}} \in z^a_{\frac{h_1+1}{2}} = \check{z}_{h_1+1} \subseteq \check{c}_h \subseteq \check{\tau}_{i_2,h}$ by Assumption II.1 (b), which violates the premise of this case. Let $k_1 := h_1/2$. Now, we claim that agent (i_1, h_1) does not influence the state $\check{s}_h$ nor the information $\check{\tau}_{i_1,h}$. We prove this case by case as follows:

   (a) Suppose h is odd, then $\check{p}_h = \emptyset$ by Equation (IV.1). If agent (i_1, h_1) influences $\check{s}_h$ in $\mathcal{D}^{\dagger}_{\mathcal{L}}$, then agent (i_1, h_1) influences $\widetilde{s}_h$ in $\mathcal{D}_{\mathcal{L}}$ (because strict expansion does not change system

dynamics). From Assumption III.7, we know that she also influences $\widetilde{o}_{-i_1,h}$, i.e., there must exist some $i_3 \neq i_1$ such that agent (i_1,h_1) influences $\widetilde{o}_{i_3,h}$ in $\mathcal{D}_{\mathcal{L}}$. From Assumption II.1 (e), it holds that $\widetilde{o}_{i_3,h} \in \widetilde{\tau}_{i_3,h+1}$. Therefore, agent (i_1,h_1) influences agent $(i_3,h+1)$ in $\mathcal{D}_{\mathcal{L}}$. From Lemma B.2, we know $\sigma(\tau_{i_1,k_1^-}) \subseteq \sigma(\tau_{i_3,k^-})$ in $\mathcal{L}$, where $k := (h+1)/2$. Furthermore, from Assumption II.2 and $i_3 \neq i_1$, it holds that $\sigma(\tau_{i_1,k_1^-}) \subseteq \sigma(c_{k^-})$. Also, from Equation (IV.1), it holds that $\widetilde{\tau}_{i_1,h_1} = \tau_{i_1,k_1^+} = \tau_{i_1,k_1^-} \cup z^a_{k_1}$ and $z^a_{k_1} = \widetilde{z}_{h_1} \subseteq \widetilde{c}_h$. Then, we have $\sigma(\widetilde{\tau}_{i_1,h_1}) \subseteq \sigma(\widetilde{c}_h) = \sigma(\widetilde{\tau}_{i_3,h})$. Based on the strict expansion from $\mathcal{D}_{\mathcal{L}}$ to $\mathcal{D}^{\dagger}_{\mathcal{L}}$, we can get $\check{\tau}_{i_1,h_1} \backslash \widetilde{\tau}_{i_1,h_1} \subseteq \check{c}_{h_1} \subseteq \check{\tau}_{i_3,h}$ and $\check{a}_{i_1,h_1} \in \check{c}_h$. Then, it holds that $\sigma(\check{\tau}_{i_1,h_1}) \subseteq \sigma(\check{\tau}_{i_3,h}), \sigma(\check{a}_{i_1,h_1}) \subseteq \sigma(\check{\tau}_{i_3,h})$, which leads to a contradiction to the premise that for any $i_2 \neq i_1$, either $\sigma(\check{\tau}_{i_1,h_1}) \nsubseteq \sigma(\check{\tau}_{i_2,h})$ or $\sigma(\check{a}_{i_1,h_1}) \nsubseteq \sigma(\check{\tau}_{i_2,h})$. Hence, we know agent (i_1,h_1) does not influence the state $\check{s}_h$. Additionally, for any $i_2 \neq i_1$, since agent (i_1,h_1) does not influence agent (i_2,h), and $\check{\tau}_{i_1,h} = \check{c}_h = \check{\tau}_{i_2,h}$, then we know that agent (i_1,h_1) does not influence $\check{\tau}_{i_1,h}$.

(b) Suppose h is even. If agent (i_1,h_1) influences $\check{s}_{h_1+1}$, then from Assumption III.7, there exists some $i_3 \neq i_1$ such that agent (i_1,h_1) influences $\check{o}_{i_3,h_1+1}$. However, from Assumption II.1 (e), we know that $\check{o}_{i_3,h_1+1} = o_{i_3,\frac{h_1}{2}+1} \in \tau_{i_3,\frac{h_1}{2}+1} \subseteq \tau_{i_3,\frac{h}{2}} = \widetilde{\tau}_{i_3,h} \subseteq \check{\tau}_{i_3,h}$, which means that agent (i_1,h_1) influences agent (i_3,h), leading to a contradiction. Therefore, we know that agent (i_1,h_1) does not influence $\check{s}_{h_1+1}$, and thus for any $i_2 \in [n]$, it does not influence $\check{o}_{i_2,h_1+1}$. Also, from Assumption III.5, we know that $\check{a}_{i_1,h_1} \notin \check{\tau}_{i_2,h_1+1}$. Therefore, agent (i_1,h_1) does not influence $\check{\tau}_{i_2,h_1+1}$ and $\check{a}_{i_2,h_1+1}$. By recurision, we know that agent (i_1,h_1) does not influence $\check{s}_{h'}$ and $\check{\tau}_{i_2,h'}$ for any $i_2 \in [n], h' > h$.

Combining two cases above, we know agent (i_1,h_1) does not influence $\check{s}_h$, and does not influence $\check{\tau}_{i,h}, \forall i \in [n]$ in $\mathcal{D}^{\dagger}_{\mathcal{L}}$, yielding

$$\begin{aligned}
&\mathbb{P}(\check{s}_h, \check{p}_h \,|\, \check{c}_h, \check{g}_{1:h-1}) = \mathbb{P}(\check{s}_h, \check{p}_h, \check{c}_h \,|\, \check{c}_h, \check{g}_{1:h-1}) \\
&= \mathbb{P}(\check{s}_h, \check{\tau}_h \,|\, \check{c}_h, \check{g}_{1:h-1}) = \mathbb{P}(\check{s}_h, \{\check{\tau}_{i,h}\}_{i\in[n]} \,|\, \check{c}_h, \check{g}_{1:h-1}) \\
&= \mathbb{P}(\check{s}_h, \{\check{\tau}_{i,h}\}_{i\in[n]} \,|\, \check{c}_h, \check{g}'_{1:h-1}) = \mathbb{P}(\check{s}_h, \check{p}_h \,|\, \check{c}_h, \check{g}'_{1:h-1}).
\end{aligned}$$

This completes the proof. □

C-F Proof of Theorem IV.6

Proof. Firstly, from the construction of $\mathcal{D}'_{\mathcal{L}}$ and the strategy space $\overline{\mathcal{G}}_{1:\overline{H}}$, we know that for any $h \in [H], i \in [n], \overline{C}_{2h-1} = \check{C}_{2h-1}, \overline{\mathcal{A}}_{i,2h-1} = \check{\mathcal{A}}_{i,2h-1}, \overline{\mathcal{T}}_{i,2h} = \check{\mathcal{T}}_{i,2h}, \overline{\mathcal{A}}_{i,2h} = \check{\mathcal{A}}_{i,2h}$. Therefore, $\overline{\mathcal{G}}_{1:\overline{H}} = \check{\mathcal{G}}_{1:\check{H}}$, and finding a team optimal strategy of $\mathcal{D}'_{\mathcal{L}}$ in the strategy space $\overline{\mathcal{G}}_{1:\overline{H}}$ is equivalent to finding a team-optimum of $\mathcal{D}^{\dagger}_{\mathcal{L}}$ in the strategy space $\check{\mathcal{G}}_{1:\check{H}}$.

Secondly, we will prove that the Dec-POMDP $\mathcal{D}'_{\mathcal{L}}$ satisfies the information evolution rules. From Assumption II.1, it holds that, for any $i \in [n], h \in [\overline{H}]$, if $h = 2t-1$ with $t \in [H]$

$$\widetilde{z}_h = \chi_t(\widetilde{p}_{h-1}, \widetilde{a}_{h-1}, \widetilde{o}_h), \qquad p_{i,h} = \xi_{i,t}(\widetilde{p}_{i,h-1}, \widetilde{a}_{i,h-1}, \widetilde{o}_{i,h});$$

if $h = 2t$ with $t \in [H]$, then

$$\widetilde{z}_h = \phi_t(p_{h-1}, \widetilde{a}_{h-1}), \qquad \widetilde{p}_{i,h} = p_{i,h-1} \backslash \phi_{i,t}(p_{i,h-1}, \widetilde{a}_{i,h-1}),$$

where $\chi_t, \xi_{i,t}$ are fixed transformations and $\phi_h, \phi_{i,h}$ are additional-sharing functions. Recall that we defined $p_{i,2t-1} = p_{i,t^-}$ for any $i \in [n], t \in [H]$. From the expansion (Equation (IV.2)), we know $\widetilde{c}_h = \check{c}_h$,

and $\check{z}_h = \check{c}_h \backslash \check{c}_{h-1}$. Also, from the refinement, we know $\forall i \in [n], h \in [\overline{H}], \overline{z}_h = \check{z}_h, \overline{c}_h = \check{c}_h, \overline{a}_{i,h} = \check{a}_{i,h} = \widetilde{a}_{i,h}, \overline{o}_{i,h} = \check{o}_{i,h} = \widetilde{o}_{i,h}$, and $\forall t \in [H], \overline{p}_{i,2t-1} = p_{i,t^-}, \overline{p}_{i,2t} = \widetilde{p}_{i,2t}$.

Then, we can construct $\{\overline{\chi}_h\}_{h\in[\overline{H}]}, \{\overline{\xi}_{i,h}\}_{i\in[n],h\in[\overline{H}]}$ accordingly as follows:

- If $h = 2t-1$ with $t \in [H]$, we define $\overline{\chi}_h, \{\overline{\xi}_{i,h}\}_{i\in[n]}$ as

  $\forall i \in [n], \overline{\xi}_{i,h} := \xi_{i,t}, \quad \forall \overline{p}_{h-1} \in \overline{\mathcal{P}}_{h-1}, \overline{a}_{h-1} \in \overline{\mathcal{A}}_{h-1}, \overline{o}_h \in \overline{\mathcal{O}}_h,$
  $\overline{\chi}_h(\overline{p}_{h-1}, \overline{a}_{h-1}, \overline{o}_h) := \chi_t(\overline{p}_{h-1}, \overline{a}_{h-1}, \overline{o}_h) \cup \rho_h^1 \backslash \rho_h^2$, where
  $\rho_h^1 := \{\widetilde{a}_{j,h-1} \mid \forall j \in [n], \sigma(\widetilde{\tau}_{j,h-1}) \subseteq \sigma(\widetilde{c}_h), \widetilde{a}_{j,h-1} \text{ influences } \widetilde{s}_h\} \backslash \chi_t(\overline{p}_{h-1}, \overline{a}_{h-1}, \overline{o}_h)$ if $h > 1$, otherwise $\emptyset$.
  $\rho_h^2 := \{\widetilde{a}_{j,h_0} \mid \forall j \in [n], h_0 < h-1, \sigma(\widetilde{a}_{j,h_0}) \subseteq \sigma(\widetilde{c}_{h-1}), \widetilde{a}_{j,h_0} \text{ influences } \widetilde{s}_{h_0+1}\} \cap \chi_t(\overline{p}_{h-1}, \overline{a}_{h-1}, \overline{o}_h)$, if $h > 2$, otherwise $\emptyset$.

  Note that there exists some functions f_h^1, f_h^2 such that $\rho_h^1 = f_h^1(\overline{p}_{h-1}, \overline{a}_{h-1}, \overline{o}_h), \rho_h^2 = f_h^2(\overline{p}_{h-1}, \overline{a}_{h-1}, \overline{o}_h)$. This is because, $\rho_h^1 \subseteq \{\overline{a}_{i,h-1}\}_{i\in[n]}$; which element $\overline{a}_{j,h-1}$ is in ρ_h^1 is based on whether $\sigma(\widetilde{\tau}_{j,h-1}) \subseteq \sigma(\widetilde{c}_h)$ and whether $\widetilde{a}_{j,h-1}$ influences $\widetilde{s}_h$, which is a property of the problem $\mathcal{D}_{\mathcal{L}}$. Similarly, $\rho_h^2 \subseteq \chi_t(\overline{p}_{h-1}, \overline{a}_{h-1}, \overline{o}_h)$, and which element $\overline{a}_{j,h_0}$ is in ρ_h^2 is based on whether $\sigma(\widetilde{a}_{j,h_0}) \subseteq \sigma(\widetilde{c}_{h-2}), \widetilde{a}_{j,h_0}$ influences $\widetilde{s}_{h_0+1}$. Now we will show that $\overline{\chi}_h, \{\overline{\xi}_{i,h}\}_{i\in[n]}$ satisfy

  $$\overline{c}_h = \overline{c}_{h-1} \cup \overline{z}_h, \ \ \overline{z}_h = \overline{\chi}_h(\overline{p}_{h-1}, \overline{a}_{h-1}, \overline{o}_h), \qquad \text{for each } i \in [n], \ \ \overline{p}_{i,h} = \overline{\xi}_{i,h}(\overline{p}_{i,h-1}, \overline{a}_{i,h-1}, \overline{o}_{i,h}).$$

  For functions $\{\overline{\xi}_{i,h}\}_{i\in[n]}$, based on the construction of $\mathcal{D}'_{\mathcal{L}}$ from $\mathcal{D}_{\mathcal{L}}$, we know $\overline{p}_{i,h-1} = \widetilde{p}_{i,h-1}, \overline{a}_{i,h-1} = \widetilde{a}_{i,h-1}, \overline{o}_{i,h-1} = \widetilde{o}_{i,h-1}$, and $\overline{p}_{i,h} = p_{i,t^-} = p_{i,h}$.

  For function $\overline{\chi}_h$, it is easy to verify that $\overline{z}_h \backslash \widetilde{z}_h = (\overline{c}_h \backslash \widetilde{c}_h) \backslash (\overline{c}_{h-1} \backslash \widetilde{c}_{h-1})$ and $\widetilde{z}_h \backslash \overline{z}_h = (\overline{c}_{h-1} \backslash \widetilde{c}_{h-1}) \backslash (\overline{c}_h \backslash \widetilde{c}_h)$. Together with the fact that $\overline{z}_h = \widetilde{z}_h \cup (\overline{z}_h \backslash \widetilde{z}_h) \backslash (\widetilde{z}_h \backslash \overline{z}_h)$, it suffices to show that $\rho_h^1 = \overline{z}_h \backslash \widetilde{z}_h, \rho_h^2 = \widetilde{z}_h \backslash \overline{z}_h$. From the expansion, we know that for any $h' \in [\overline{H}], \overline{c}_{h'} \backslash \widetilde{c}_{h'}$ only consists of some actions at the even timesteps, since the actions at odd timesteps cannot influence the underlying state. Therefore, $\overline{z}_h \backslash \widetilde{z}_h$ and $\widetilde{z}_h \backslash \overline{z}_h$ only consist of some actions at the even timesteps.

  For any $\widetilde{a}_{i,2t_1}, i \in [n], t_1 < t$, if $\widetilde{a}_{i,2t_1}$ influences $\widetilde{s}_{2t_1+1}$, then from Assumption III.7, there exists $i_2 \neq i$ such that $\widetilde{a}_{i,2t_1}$ influences $\widetilde{o}_{i_2,2t_1+1}$ and thus influences $\widetilde{\tau}_{i_2,2t_1+1}$ due to Assumption II.1 (e). From the QC IS of $\mathcal{D}_{\mathcal{L}}$, we know that $\sigma(\widetilde{\tau}_{i,2t_1}) \subseteq \sigma(\widetilde{\tau}_{i_2,2t_1+1})$, and thus $\sigma(\widetilde{\tau}_{i,2t_1}) \subseteq \sigma(\widetilde{c}_{2t_1+1})$ due to Assumption II.2 and $i_2 \neq i$. Therefore, from Equation (IV.2), we know that $\widetilde{a}_{i,2t_1} \in \check{c}_{2t_1+1}$. This means, if any action $\widetilde{a}_{i,2t_1}$ influences underlying state $\widetilde{s}_{2t_1+1}$, we have $\widetilde{a}_{i,2t_1} \in \check{c}_{2t_1+1} = \overline{c}_{2t_1+1}$, since it will be added in $\check{c}_{2t_1+1}$ via expansion.

  Therefore, if any action $\overline{a}_{i,h_1} = \widetilde{a}_{i,h_1}, i \in [n], h_1 < h$ is in $(\overline{c}_h \backslash \widetilde{c}_h) \backslash (\overline{c}_{h-1} \backslash \widetilde{c}_{h-1})$, it can only happen if $h_1 = h-1$. Also, for any $i \in [n]$, if $\overline{a}_{i,h-1} \in \rho_h^1$, then $\sigma(\widetilde{\tau}_{i,h-1}) \subseteq \sigma(\widetilde{c}_h), \widetilde{a}_{i,h-1}$ influences $\widetilde{s}_h$, and furthermore, $\overline{a}_{i,h-1} \notin \chi_t(\overline{p}_{h-1}, \overline{a}_{h-1}, \overline{o}_h)$. Then it means $\overline{a}_{i,h-1} \notin \widetilde{c}_h$. Therefore, $\overline{a}_{i,h-1} \in (\overline{c}_h \backslash \widetilde{c}_h) \backslash (\overline{c}_{h-1} \backslash \widetilde{c}_{h-1})$ and we proved that $\rho_h^1 \subseteq \overline{z}_h \backslash \widetilde{z}_h$. Also, for any $i \in [n]$, $\overline{a}_{i,h-1} \in \overline{z}_h \backslash \widetilde{z}_h$ only if $\overline{a}_{i,h-1}$ is added via expansion and $\overline{a}_{i,h-1} \notin \chi_t(\overline{p}_{h-1}, \overline{a}_{h-1}, \overline{o}_h)$. Then, this can only happen if $\sigma(\widetilde{\tau}_{i,h-1}) \subseteq \sigma(\widetilde{c}_h)$ and $\widetilde{a}_{i,h-1}$ influences $\widetilde{s}_h$. Therefore, we proved that $\overline{z}_h \backslash \widetilde{z}_h \subseteq \rho_h^1$. Combining the two parts we obtain $\rho_h^1 = \overline{z}_h \backslash \widetilde{z}_h$.

  For any $\overline{a}_{i,h_1}, i \in [n], h_1 < h$ is in $\widetilde{z}_h \backslash \overline{z}_h = (\overline{c}_{h-1} \backslash \widetilde{c}_{h-1}) \backslash (\overline{c}_h \backslash \widetilde{c}_h)$. We will know that $\overline{a}_{i,h_1} \in \overline{c}_{h-1}$, then $h_1 + 1 \leq h-1$ and $\sigma(\overline{a}_{i,h_1}) \subseteq \sigma(\overline{c}_{h-1})$. Also, from the proof above, we know $\overline{a}_{i,h_1}$ influences the state $\overline{s}_{h_1+1}$, then together with $\overline{a}_{i,h_1} \in \widetilde{z}_h$, we have $\overline{a}_{i,h_1} \in \{\widetilde{a}_{j,h_0} \mid \forall j \in [n], h_0 < h-1, \sigma(\widetilde{a}_{j,h_0}) \subseteq \sigma(\widetilde{c}_{h-1}), \widetilde{a}_{j,h_1} \text{ influences } \widetilde{s}_{h_1+1}\} \cap \widetilde{z}_h$. Therefore, $\widetilde{z}_h \backslash \overline{z}_h \subseteq \rho_h^2$. Meanwhile, for any $i \in [n], h_1 < h$, if $\overline{a}_{i,h_1} \in \rho_h^2$, then it holds $\overline{a}_{i,h_1} \in \widetilde{z}_h$, $h_1 < h-1$, and $\overline{a}_{i,h_1}$ influences $\overline{s}_{h_1+1}$. Then, from above, we know that $\overline{a}_{i,h_1} = \check{a}_{i,h_1}$ will be added in $\check{c}_{h_1+1} = \overline{c}_{h_1+1} \subseteq \overline{c}_{h-1}$ since $h_1 < h-1$. Then, $\overline{a}_{i,h_1} \notin \overline{z}_h = \overline{c}_h \backslash \overline{c}_{h-1}$. Therefore, $\overline{a}_{i,h_1} \in \widetilde{z}_h \backslash \overline{z}_h$, and then $\rho_h^2 \subseteq \widetilde{z}_h \backslash \overline{z}_h$. Combining the two parts we obtain $\rho_h^2 = \widetilde{z}_h \backslash \overline{z}_h$.

- If $h = 2t$ with $t \in [H]$, we define $\overline{\chi}_h, \{\overline{\xi}_{i,h}\}_{i\in[n]}$ as

$$\forall i \in [n], \overline{p}_{i,h-1} \in \overline{\mathcal{P}}_{i,h-1}, \overline{a}_{i,h-1} \in \overline{\mathcal{A}}_{i,h-1}, \overline{o}_{i,h} \in \overline{\mathcal{O}}_{i,h}, \overline{\xi}_{i,h}(\overline{p}_{i,h-1}, \overline{a}_{i,h-1}, \overline{o}_{i,h}) = \overline{p}_{i,h-1} \backslash \phi_{i,t}(\overline{p}_{i,h-1}, \overline{a}_{i,h-1})$$
$$\forall \overline{p}_{h-1} \in \overline{\mathcal{P}}_{h-1}, \overline{a}_{h-1} \in \overline{\mathcal{A}}_{h-1}, \overline{o}_h \in \overline{\mathcal{O}}_h, \overline{\chi}_h(\overline{p}_{h-1}, \overline{a}_{h-1}, \overline{o}_h) = \phi_t(\overline{p}_{h-1}, \overline{a}_{h-1}).$$

  Now we will show that $\overline{\chi}_h, \{\overline{\xi}_{i,h}\}_{i\in[n]}$ satisfy

$$\overline{c}_h = \overline{c}_{h-1} \cup \overline{z}_h, \ \ \overline{z}_h = \overline{\chi}_h(\overline{p}_{h-1}, \overline{a}_{h-1}, \overline{o}_h), \qquad \text{for each } i \in [n], \ \ \overline{p}_{i,h} = \overline{\xi}_{i,h}(\overline{p}_{i,h-1}, \overline{a}_{i,h-1}, \overline{o}_{i,h}).$$

  For functions $\{\overline{\xi}_{i,h}\}_{i\in[n]}$, based on the construction of $\mathcal{D}'_{\mathcal{L}}$ from $\mathcal{D}_{\mathcal{L}}$, we know $\overline{p}_{i,h-1} = p_{i,t^-} = p_{i,h-1}, \overline{a}_{i,h-1} = \widetilde{a}_{i,h-1}$, and $\overline{p}_{i,h} = \widetilde{p}_{i,h}$.
  For function $\overline{\chi}_h$, we know that $\overline{p}_{h-1} = p_{t^-} = p_{h-1}, \overline{a}_{h-1} = \widetilde{a}_{h-1}$, so it suffices to show that $\widetilde{z}_h = \overline{z}_h$. As shown above, $\overline{z}_h$ and $\widetilde{z}_h$ only consist of some actions at the even timesteps. Moreover, for any action $\widetilde{a}_{i,2t_1}$ with $i \in [n], t_1 < t$ that influences $\widetilde{s}_{2t_1+1}$, it will be added in $\overline{c}_{2t_1+1}$ via expansion (if $\widetilde{a}_{i,2t_1} \notin \widetilde{c}_{2t_1+1}$); if $\widetilde{a}_{i,2t_1}$ does not influence $\widetilde{s}_{2t_1+1}$, then it will never be added via expansion. Therefore, $\overline{z}_h \backslash \widetilde{z}_h = \emptyset$. Also, if some action $\widetilde{a}_{i,2t_1} \in \widetilde{z}_h \backslash \overline{z}_h$, then we know $\widetilde{a}_{i,2t_1} \in \widetilde{z}_h = \phi_t(\overline{p}_{h-1}, \widetilde{a}_{h-1})$, and further we know $\widetilde{a}_{i,2t_1} \in \overline{p}_{h-1} = p_{i,t^-}$. However, if $\widetilde{a}_{i,2t_1}$ influences $\widetilde{s}_{2t_1+1}$, then it will be added in $\overline{c}_{2t_1+1}$ and cannot lie in $\overline{p}_{h-1}$; if $\widetilde{a}_{i,2t_1}$ does not influence $\widetilde{s}_{2t_1+1}$, then from Assumption III.5, $\widetilde{a}_{i,2t_1} \notin p_{i,t^-}$. Therefore, we know $\widetilde{z}_h \backslash \overline{z}_h = \emptyset$, and get $\widetilde{z}_h = \overline{z}_h$.

Thirdly, we prove that such a Dec-POMDP $\mathcal{D}'_{\mathcal{L}}$ has SI-CIBs with respect to the strategy space $\overline{\mathcal{G}}_{1:\overline{H}}$. This is equivalent to that for any $h \in [2:\overline{H}], \overline{s}_h \in \overline{\mathcal{S}}, \overline{p}_h \in \overline{\mathcal{P}}_h, \overline{c}_h \in \overline{\mathcal{C}}_h, i_1 \in [n], h_1 < h, \overline{g}_{1:h-1}, \overline{g}'_{i_1,h_1} \in \overline{\mathcal{G}}_{i_1:h_1}$, let $\overline{g}'_{h_1} := (\overline{g}_{1,h_1}, \cdots, \overline{g}'_{i_1,h_1}, \cdots, \overline{g}_{n,h_1})$ and $\overline{g}'_{1:h-1} := (\overline{g}_1, \cdots, \overline{g}'_{h_1}, \cdots, \overline{g}_{h-1})$. If $\overline{c}_h$ is reachable from both $\overline{g}_{1:h-1}$ and $\overline{g}'_{1:h-1}$, it holds that

$$\mathbb{P}_h^{\mathcal{D}'_{\mathcal{L}}}(\overline{s}_h, \overline{p}_h | \overline{c}_h, \overline{g}_{1:h-1}) = \mathbb{P}_h^{\mathcal{D}'_{\mathcal{L}}}(\overline{s}_h, \overline{p}_h | \overline{c}_h, \overline{g}'_{1:h-1}). \tag{C.2}$$

We prove this case by case. If $h = 2t$ with $t \in [H]$, then from the result of Theorem IV.5, it holds that

$$\mathbb{P}_h^{\mathcal{D}'_{\mathcal{L}}}(\overline{s}_h, \overline{p}_h | \overline{c}_h, \overline{g}_{1:h-1}) = \mathbb{P}_h^{\mathcal{D}^\dagger_{\mathcal{L}}}(\overline{s}_h, \overline{p}_h | \overline{c}_h, \overline{g}_{1:h-1}) = \mathbb{P}_h^{\mathcal{D}^\dagger_{\mathcal{L}}}(\overline{s}_h, \overline{p}_h | \overline{c}_h, \overline{g}'_{1:h-1}) = \mathbb{P}_h^{\mathcal{D}'_{\mathcal{L}}}(\overline{s}_h, \overline{p}_h | \overline{c}_h, \overline{g}'_{1:h-1}).$$

Therefore, now we consider the case that $h = 2t-1$ with $t \in [H]$.
Suppose h_1 is odd, which means that $\overline{a}_{h_1}$ corresponds to the communication action in $\mathcal{L}$. Then it holds that $\overline{c}_{h_1} \subseteq \overline{c}_h, \overline{a}_{i_1,h_1} = m_{i_1,\frac{h_1+1}{2}} \in \overline{c}_h$, then

$$\begin{aligned}\mathbb{P}_h^{\mathcal{D}'_{\mathcal{L}}}(\overline{s}_h, \overline{p}_h | \overline{c}_h, \overline{g}_{1:h-1}) &= \mathbb{P}_h^{\mathcal{D}'_{\mathcal{L}}}(\overline{s}_h, \overline{p}_h | \overline{c}_{h_1}, \overline{a}_{i_1,h_1}, \overline{c}_h, \overline{g}_{1:h-1}) \\ &= \mathbb{P}_h^{\mathcal{D}'_{\mathcal{L}}}(\overline{s}_h, \overline{p}_h | \overline{c}_{h_1}, \overline{a}_{i_1,h_1}, \overline{c}_h, \overline{g}_{1:h-1} \backslash \overline{g}_{i_1,h_1}) = \mathbb{P}_h^{\mathcal{D}'_{\mathcal{L}}}(\overline{s}_h, \overline{p}_h | \overline{c}_h, \overline{g}'_{1:h-1}),\end{aligned}$$

where the second equality is because the input and output of $\overline{g}_{i_1,h_1}$ are $\overline{c}_{h_1}$ and $\overline{a}_{i_1,h_1}$.
Suppose h_1 is even, which means that h_1 is a control timestep, and let $t_1 = \frac{h_1}{2}$. If $\sigma(\overline{\tau}_{i_1,h_1}) \subseteq \sigma(\overline{c}_h)$ and $\sigma(\overline{a}_{i_1,h_1}) \subseteq \sigma(\overline{c}_h)$, then there exist deterministic measurable functions $\overline{\alpha}_1, \overline{\alpha}_2$ such that $\overline{\tau}_{i_1,h_1} = \overline{\alpha}_1(\overline{c}_h), \overline{a}_{i_1,h_1} = \overline{\alpha}_2(\overline{c}_h)$, and further it holds that

$$\begin{aligned}\mathbb{P}(\overline{s}_h, \overline{p}_h | \overline{c}_h, \overline{g}_{1:h-1}) &= \mathbb{P}(\overline{s}_h, \overline{p}_h | \overline{\alpha}_1(\overline{c}_h), \overline{\alpha}_2(\overline{c}_h), \overline{c}_h, \overline{g}_{1:h-1}) \\ &= \mathbb{P}(\overline{s}_h, \overline{p}_h | \overline{\tau}_{i_1,h_1}, \overline{a}_{i_1,h_1}, \overline{c}_h, \overline{g}_{1:h-1}) = \mathbb{P}(\overline{s}_h, \overline{p}_h | \overline{\tau}_{i_1,h_1}, \overline{a}_{i_1,h_1}, \overline{c}_h, \overline{g}'_{1:h-1}).\end{aligned}$$

If $\sigma(\overline{\tau}_{i_1,h_1}) \not\subseteq \sigma(\overline{c}_h)$ or $\sigma(\overline{a}_{i_1,h_1}) \not\subseteq \sigma(\overline{c}_h)$. Since h_1 is even, $\overline{\tau}_{i_1,h_1} = \check{\tau}_{i_1,h_1}, \overline{a}_{i_1,h_1} = \check{a}_{i_1,h_1}$. Also, we know $\overline{c}_h = \check{c}_h$, then it holds that $\sigma(\check{\tau}_{i_1,h_1}) \not\subseteq \sigma(\check{c}_h)$ or $\sigma(\check{a}_{i_1,h_1}) \not\subseteq \sigma(\check{c}_h)$. Firstly, from the sQC of $\mathcal{D}^\dagger_{\mathcal{L}}$, we know

that agent (i_1,h_1) does not influence agent (i_2,h) in $\mathcal{D}^\dagger_\mathcal{L}$ for any $i_2 \neq i_1$. Then, as shown in the proof of Theorem IV.5, we know that agent (i_1,h_1) does not influence $\check{s}_{h_1+1} = \overline{s}_{h_1+1}$ and does not influence $\check{c}_{h'} = \overline{c}_{h'}$ for any $h' \leq h$. Secondly, from Assumption II.1, we know that for any $i \in [n], p_{i,(t_1+1)^-} = \xi_{i,t_1+1}(p_{i,t_1^+}, a_{i,t_1}, o_{i,t_1+1})$, where ξ_{i,t_1+1} is a fixed transformation. Also, from Assumption III.5, we know that $a_{i,t_1} \notin p_{i,(t_1+1)^-}$. Therefore, we can write $p_{i,(t_1+1)^-} = \xi_{i,t_1+1}(p_{i,t_1^+}, o_{i,t_1+1})$. From the definition of refinement, we know that $\overline{p}_{i,h_1+1} = \xi_{i,t_1+1}(p_{i,t_1^+}, \overline{o}_{i,h_1+1})$. Since agent (i_1,h_1) does not influence $\overline{s}_{h_1+1}$, it does not influence $\overline{o}_{i,h_1+1}$. Also, agent (i_1,h_1) does not influence $\overline{c}_{h_1+1}$ and p_{i,h_1^+} (which happens before choosing $a_{i_1,t_1} = \overline{a}_{i_1,h_1}$). Therefore, agent (i_1,h_1) does not influence $\overline{\tau}_{i,h_1+1}$, and thus does not influence $\overline{a}_{i,h_1+1}$ for any $i \in [n]$. Thirdly, we know that for any $i \in [n], \overline{p}_{i,h_1+2} = \overline{\xi}_{i,h_1+2}(\overline{p}_{i,h_1+1}, \overline{a}_{i,h_1+1}, \overline{o}_{i,h_1+2})$, where $\overline{o}_{i,h_1+2} = \emptyset$. Since agent (i_1,h_1) does not influence $\overline{p}_{i,h_1+1}$ and $\overline{a}_{i,h_1+1}$, it does not influence $\overline{p}_{i,h_1+2}$. Also, we know that it does not influence $\overline{\tau}_{i,h_1+2}$. Therefore, agent (i_1,h_1) does not influence $\overline{\tau}_{i,h_1+2}$, and thus does not influence $\overline{a}_{i,h_1+2}$. In this way, we know that agent (i_1,h_1) does not influence $\overline{\tau}_{i,h'}$ for any $i \in [n], h_1 < h' \leq h$.
Finally, since we proved that agent (i_1,h_1) does not influence $\overline{s}_{h_1+1}$, and does not influence $\overline{\tau}_{i,h'}, \forall i \in [n], h_1 \leq h' < h$, then it does not influence $\overline{a}_{i,h'}$, we have

$$\begin{aligned}
&\mathbb{P}(\overline{s}_h, \overline{p}_h | \overline{c}_h, \overline{g}_{1:h-1}) = \mathbb{P}(\overline{s}_h, \overline{p}_h, \overline{c}_h | \overline{c}_h, \overline{g}_{1:h-1}) = \mathbb{P}(\overline{s}_h, \overline{\tau}_h | \overline{c}_h, \overline{g}_{1:h-1}) \\
&= \mathbb{P}(\overline{s}_h, \{\overline{\tau}_{i,h}\}_{i\in[n]} | \overline{c}_h, \overline{g}_{1:h-1}) = \mathbb{P}(\overline{s}_h, \{\overline{\tau}_{i,h}\}_{i\in[n]} | \overline{c}_h, \overline{g}'_{1:h-1}) = \mathbb{P}(\overline{s}_h, \overline{p}_h | \overline{c}_h, \overline{g}'_{1:h-1}),
\end{aligned}$$

which completes the proof. □

C-G Important Auxiliary Definitions

Definition C.3 (Perfect recall [24])**.** We say that agent i has perfect recall if $\forall h = 2, \cdots, \overline{H}$, it holds that $\overline{\tau}_{i,h-1} \cup \{\overline{a}_{i,h-1}\} \subseteq \overline{\tau}_{i,h}$. If for any $i \in [n]$, agent i has perfect recall, we call that the Dec-POMDP has a perfect recall property.

The following definitions are important in solving the Dec-POMDP constructed from the LTC problem.

[kz:for all the value function definitions (both in $\mathcal{D}$ and in $\mathcal{M}$) below, please define it for $\overline{H}+1$, which is zero, so that u have this notation "defined" later.]

Definition C.4 (Value function)**.** For each $i \in [n]$ and $h \in [\overline{H}]$, given common information $\overline{c}_h$ and strategy $\overline{g}_{1:H} \in \overline{\mathcal{G}}_{1:H}$, the value function conditioned on the common information is defined as:

$$V_h^{\overline{g},\mathcal{D}'_\mathcal{L}}(\overline{c}_h) := \mathbb{E}^{\mathcal{D}'_\mathcal{L}}_{\overline{g}}\left[\sum_{h'=h}^{\overline{H}} \overline{\mathcal{R}}_{h'}(\overline{s}_{h'}, \overline{a}_{h'}, \overline{p}_{h'}) \middle| \overline{c}_h\right], \tag{C.3}$$

where $\overline{\mathcal{R}}_{h'}$ takes $\overline{s}_{h'}, \overline{a}_{h'}, \overline{p}_{h'}$ as input, since after reformulation, the reward may come from communication cost, which is a function of $\overline{p}_{h'}$ and $\overline{a}_{h'}$. For

Definition C.5 (Prescription and Q-Value function)**.** Prescription is an important concept in the common-information-based framework [15, 16]. For any $h \in [\overline{H}], i \in [n]$, the prescription of agent i at the timestep h is defined as $\gamma_{i,h} \in \Gamma_{i,h}$, where $\Gamma_{i,h} := \overline{\mathcal{A}}_{i,h} = \mathcal{M}_{i,\frac{h+1}{2}}$ if $h = 2k-1, k \in [H]$, and $\Gamma_{i,h} := \{\gamma_{i,h} : \overline{\mathcal{P}}_{i,h} \to \overline{\mathcal{A}}_{i,h}\}$ if $h = 2k, k \in [H]$. We use $\gamma_h := (\gamma_{1,h}, \cdots, \gamma_{n,h})$ to denote the joint prescription and Γ_h to denote the joint prescription space. The prescriptions are the marginalization of strategy $\overline{g}_h$, i.e., $\forall h \in [H], \gamma_{i,2h-1} = \overline{g}_{i,2h-1}(\overline{c}_{2h-1}), \gamma_{i,2h}(\cdot) = \overline{g}_{i,2h}(\overline{c}_{2h}, \cdot)$. Then, for any $\overline{g}_{1:\overline{H}}$, we can define the

Q-value function as

$$Q_h^{\overline{g}_{1:\overline{H}},\mathcal{D}'_{\mathcal{L}}}(\overline{c}_h,\gamma_h) := \mathbb{E}_{\overline{g}}^{\mathcal{D}'_{\mathcal{L}}}\left[\sum_{h'=h}^{\overline{H}} \overline{\mathcal{R}}_{h'}(\overline{s}'_h,\overline{a}'_h,\overline{p}'_h)\,\middle|\,\overline{c}_h,\gamma_h\right]. \tag{C.4}$$

Note that for any $i\in[n]$, and for the odd timesteps $2h-1$ with $h\in[H]$, we consider the strategies that only take common information $\overline{c}_h$ as input and thus the prescription $\gamma_{i,2h-1}=\overline{a}_{i,2h-1}$. To unify the notation, we may also write it as $\gamma_{i,2h-1}(\cdot)=\overline{g}_{i,2h-1}(\overline{c}_{2h-1},\cdot)$, where $\cdot$ takes value in $\emptyset$ rather than $\overline{p}_{i,2h-1}$.

Definition C.6 (Expected approximate common information model)**.** We define an *expected approximate common information model* of $\mathcal{D}'_{\mathcal{L}}$ as

$$\mathcal{M} := \left(\{\widehat{\mathcal{C}}_h\}_{h\in[\overline{H}]},\{\widehat{\phi}_h\}_{h\in[\overline{H}]},\{\mathbb{P}_h^{\mathcal{M},z}\}_{h\in[\overline{H}]},\Gamma,\{\widehat{\mathcal{R}}_h^{\mathcal{M}}\}_{h\in[\overline{H}]}\right) \tag{C.5}$$

where $\Gamma=\{\Gamma_h\}_{h\in[\overline{H}]}$ is the joint prescription space, $\widehat{\mathcal{C}}_h$ is the space of approximate common information at timestep h. $\mathbb{P}_h^{\mathcal{M},z}:\widehat{\mathcal{C}}_h\times\Gamma_h\to\Delta(\overline{\mathcal{Z}}_{h+1})$ gives the probability of $\overline{z}_{h+1}$ under $\widehat{c}_h$ and γ_h. $\widehat{\mathcal{R}}_h^{\mathcal{M}}:\widehat{\mathcal{C}}_h\times\Gamma_h\to[0,1]$ gives the reward at timestep h given $\widehat{c}_h$ and γ_h. Then, we call that $\mathcal{M}$ is an $(\epsilon_r(\mathcal{M}),\epsilon_z(\mathcal{M}))$-*expected-approximate common information model* of $\mathcal{D}'_{\mathcal{L}}$, if it has some compression function Compress_h such that $\widehat{c}_h=\text{Compress}_h(\overline{c}_h)$ for each $h\in[\overline{H}]$, and satisfies the following:

- There exists a transformation function $\widehat{\phi}_h$ such that

$$\widehat{c}_h=\widehat{\phi}_h(\widehat{c}_{h-1},\overline{z}_h), \tag{C.6}$$

  where $\overline{z}_h=\overline{c}_h\backslash\overline{c}_{h-1}$.

- For any $\overline{g}_{1:h-1}$ and any prescription $\gamma_h\in\Gamma_h$, it holds that

$$\mathbb{E}^{\mathcal{D}'_{\mathcal{L}}}_{\overline{a}_{1:h-1},\overline{o}_{1:h}\sim\overline{g}_{1:h-1}}\left|\mathbb{E}^{\mathcal{D}'_{\mathcal{L}}}[\overline{\mathcal{R}}_h(\overline{s}_h,\overline{a}_h,\overline{p}_h)\,|\,\overline{c}_h,\gamma_h]-\widehat{\mathcal{R}}_h^{\mathcal{M}}(\widehat{c}_h,\gamma_h)\right|\le\epsilon_r(\mathcal{M}). \tag{C.7}$$

- For any $\overline{g}_{1:h-1}$ and any prescription $\gamma_h\in\Gamma_h$, it holds that

$$\mathbb{E}^{\mathcal{D}'_{\mathcal{L}}}_{\overline{a}_{1:h-1},\overline{o}_{1:h}\sim\overline{g}_{1:h-1}}\left\|\mathbb{P}_h^{\mathcal{D}'_{\mathcal{L}}}(\cdot\,|\,\overline{c}_h,\gamma_h)-\mathbb{P}_h^{\mathcal{M},z}(\cdot\,|\,\widehat{c}_h,\gamma_h)\right\|_1\le\epsilon_z(\mathcal{M}). \tag{C.8}$$

Note that, we can extend the common information to timestep $\overline{H}+1$ as $\overline{c}_{\overline{H}+1}=\overline{c}_{\overline{H}}$, and for any $\gamma_{\overline{H}}\in\Gamma_{\overline{H}}$, $\mathbb{P}_{\overline{H}}^{\mathcal{D}'_{\mathcal{L}}}(\overline{c}_{\overline{H}+1}\,|\,\overline{c}_{\overline{H}},\gamma_h)=\mathbb{1}[\overline{c}_{\overline{H}+1}=\overline{c}_{\overline{H}}]$.

Definition C.7 (Value functions under $\mathcal{M}$)**.** Given a Dec-POMDP $\mathcal{D}'_{\mathcal{L}}$ and its expected approximate common information model $\mathcal{M}$. For any strategy $\overline{g}_{1:\overline{H}}\in\overline{\mathcal{G}}_{1:\overline{H}}$, $h\in[\overline{H}]$, we define the value function and Q-value functions under $\mathcal{M}$ as

$$\begin{aligned}
&V_h^{\overline{g}_{1:\overline{H}},\mathcal{M}}(\overline{c}_h)=\widehat{\mathcal{R}}_h^{\mathcal{M}}(\text{Compress}_h(\overline{c}_h),\{\overline{g}_{j,h}(\overline{c}_h,\cdot)\}_{j\in[n]})+\mathbb{E}^{\mathcal{M}}[V_{h+1}^{\overline{g}_{1:\overline{H}},\mathcal{M}}(\overline{c}_{h+1})\,|\,\text{Compress}_h(\overline{c}_h),\{\overline{g}_{j,h}(\overline{c}_h,\cdot)\}_{j\in[n]}],\\
&Q_h^{\overline{g}_{1:\overline{H}},\mathcal{M}}(\overline{c}_h,\gamma_h)=\widehat{\mathcal{R}}_h^{\mathcal{M}}(\text{Compress}_h(\overline{c}_h),\gamma_h)+\mathbb{E}^{\mathcal{M}}[V_{h+1}^{\overline{g}_{1:\overline{H}},\mathcal{M}}(\overline{c}_{h+1})\,|\,\text{Compress}_h(\overline{c}_h),\gamma_h],\\
&Q_h^{*,\mathcal{M}}(\overline{c}_h,\gamma_h)=\operatorname*{argmax}_{\overline{g}_{1:\overline{H}}\in\overline{\mathcal{G}}_{1:\overline{H}}} Q_h^{\overline{g}_{1:\overline{H}},\mathcal{M}}(\overline{c}_h,\gamma_h).
\end{aligned}$$

Definition C.8 (Model-belief consistency)**.** We say the expected approximate common information model $\mathcal{M}$ of $\mathcal{D}'_{\mathcal{L}}$ is *consistent with* some beliefs $\{\mathbb{P}_h^{\mathcal{M},c}(\bar{s}_h, \bar{p}_h | \widehat{c}_h)\}_{h\in[\overline{H}]}$ if it satisfies the following: for all $i \in [n]$, $h \in H$,

$$\begin{aligned}
\mathbb{P}_{2h-1}^{\mathcal{M},z}(\bar{z}_{2h} | \widehat{c}_{2h-1}, \gamma_{2h-1}) &= \sum_{\substack{\bar{s}_{2h-1}, \bar{p}_{2h-1}: \\ \overline{\chi}_{2h-1}(\bar{p}_{2h-1}, \gamma_{2h-1}, \bar{o}_{2h}=\emptyset)=\bar{z}_{2h}}} \mathbb{P}_{2h-1}^{\mathcal{M},c}(\bar{s}_{2h-1}, \bar{p}_{2h-1} | \widehat{c}_{2h-1}), \\
\mathbb{P}_{2h}^{\mathcal{M},z}(\bar{z}_{2h+1} | \widehat{c}_{2h}, \gamma_{2h}) &= \sum_{\substack{\bar{s}_{2h}, \bar{p}_{2h}, \bar{a}_{2h}, \bar{o}_{2h+1}: \\ \overline{\chi}_{2h}(\bar{p}_{2h}, \bar{a}_{2h}, \bar{o}_{2h+1})=\bar{z}_{2h+1}}} \Big(\mathbb{P}_{2h}^{\mathcal{M},c}(\bar{s}_{2h}, \bar{p}_{2h} | \widehat{c}_{2h}) \mathbb{1}[\bar{a}_{2h} = \gamma_{2h}(\bar{p}_{2h})] \\
&\qquad \sum_{s_{2h+1}} \overline{\mathbb{T}}_{2h}(\bar{s}_{2h+1} | \bar{s}_{2h}, \bar{a}_{2h})] \overline{\mathbb{O}}_{2h+1}(\bar{o}_{2h+1} | \bar{s}_{2h+1}) \Big),
\end{aligned} \tag{C.9}$$

$$\begin{aligned}
\widehat{\mathcal{R}}_{2h-1}^{\mathcal{M}}(\widehat{c}_{2h-1}, \gamma_{2h-1}) &= \sum_{\bar{s}_{2h-1}, \bar{p}_{2h-1}} \mathbb{P}_{2h-1}^{\mathcal{M},c}(\bar{s}_{2h-1}, \bar{p}_{2h-1} | \widehat{c}_{2h-1}) \overline{\mathcal{R}}_{2h-1}(\bar{s}_{2h-1}, \gamma_{2h-1}, \bar{p}_{2h-1}), \\
\widehat{\mathcal{R}}_{2h}^{\mathcal{M}}(\widehat{c}_{2h}, \gamma_{2h}) &= \sum_{\bar{s}_{2h}, \bar{p}_{2h}, \bar{a}_{2h}} \mathbb{P}_h^{\mathcal{M},c}(\bar{s}_{2h}, \bar{p}_{2h} | \widehat{c}_{2h}) \mathbb{1}[\bar{a}_{2h} = \gamma_{2h}(\bar{p}_{2h})] \overline{\mathcal{R}}_{2h}(\bar{s}_{2h}, \bar{a}_{2h}, \bar{p}_{2h}).
\end{aligned} \tag{C.10}$$

Definition C.9 (Strategy-dependent approximate common information model)**.** Given a model $\widetilde{\mathcal{M}}$ (as in Definition C.6) and $\overline{H}$ joint strategies $g^{1:\overline{H}}$, where each $g^h \in \overline{\mathcal{G}}_{1:\overline{H}}$ for $h \in [\overline{H}]$, we say $\widetilde{\mathcal{M}}$ is a *strategy-dependent expected approximate common information model*, denoted as $\widetilde{\mathcal{M}}(g^{1:\overline{H}})$, if it is consistent with the *strategy-dependent* belief $\{\mathbb{P}_h^{g^h, \mathcal{D}'_{\mathcal{L}}}(\bar{s}_h, \bar{p}_h | \widehat{c}_h)\}_{h\in[\overline{H}]}$ (as per Definition C.8).

Definition C.10 (Length of approximate common information)**.** Given the compression functions $\{\text{Compress}_h\}_{h\in[\overline{H}]}$, we define the integer $\widehat{L} > 0$ as the minimum length such that there exists a mapping $\widehat{f}_h : \overline{\mathcal{O}}_{\max\{1,h-\widehat{L}\}:h} \times \overline{\mathcal{A}}_{\max\{1,h-\widehat{L}\}:h-1} \to \widehat{\mathcal{C}}_h$ such that for each $h \in [\overline{H}+1]$ and joint history $\{\bar{o}_{1:h}, \bar{a}_{1:h-1}\}$, we have $\widehat{f}_h(x_h) = \widehat{c}_h$, where $x_h = \{\bar{o}_{\max\{h-\widehat{L},1\}}, \bar{a}_{\max\{h-\widehat{L},1\}}, \bar{o}_{\max\{h-\widehat{L},1\}+1}, \cdots, \bar{a}_{h-1}, \bar{o}_h\}$.

C-H Main Results for Planning in QC LTCs

Theorem C.11 (Full version of Theorem IV.8)**.** Given any QC LTC problem $\mathcal{L}$ satisfying Assumptions III.1, III.4, III.5, and III.7, we can construct a Dec-POMDP problem $\mathcal{D}'_{\mathcal{L}}$ with SI-CIBs such that for any $\epsilon > 0$, solving an ϵ-team optimal strategy in $\mathcal{D}'_{\mathcal{L}}$ can give us an ϵ-team optimal strategy of $\mathcal{L}$, and the following holds. Fix $\epsilon_r, \epsilon_z > 0$ and given any (ϵ_r, ϵ_z)-expected-approximate common information model $\mathcal{M}$ for $\mathcal{D}'_{\mathcal{L}}$ that satisfies Assumption IV.7 and is consistent with some given approximate belief $\{\mathbb{P}_h^{\mathcal{M},c}(\bar{s}_h, \bar{p}_h | \widehat{c}_h)\}_{h\in[\overline{H}]}$, there exists an algorithm that can compute a $(2\overline{H}\epsilon_r + \overline{H}^2\epsilon_z)$-team optimal strategy for the original LTC problem $\mathcal{L}$ with time complexity $\max_{h\in[\overline{H}]} |\widehat{\mathcal{C}}_h| \cdot \text{poly}(|\overline{\mathcal{S}}|, |\overline{\mathcal{A}}_h|, |\overline{\mathcal{P}}_h|, \overline{H})$. In particular, for fixed $\epsilon > 0$, if $\mathcal{L}$ has a baseline sharing protocol as one of the examples in §A, one can construct such am $\mathcal{M}$ and apply Algorithm 1 to compute an ϵ-team optimal strategy for $\mathcal{L}$ with the following complexities:

- **Examples 1, 3, 5, 6:** $\text{poly}(\max_{h\in\overline{H}}(|\overline{\mathcal{O}}_h||\overline{\mathcal{A}}_h|)^{C\gamma^{-4}\log(\frac{|\overline{\mathcal{S}}|}{\epsilon})}, |\overline{\mathcal{S}}|, \overline{H}, \frac{1}{\epsilon})$;
- **Examples 2, 4, 7, 8:** $\text{poly}(\max_{h\in\overline{H}}(|\overline{\mathcal{O}}_h||\overline{\mathcal{A}}_h|)^{C\gamma^{-4}\log(\frac{|\overline{\mathcal{S}}|}{\epsilon})+2d}, |\overline{\mathcal{S}}|, \overline{H}, \frac{1}{\epsilon})$,

for some universal constant $C > 0$. Recall that γ is the constant in Assumption III.1. And d is the delayed step of sharing, which is a constant as stated in §A. Note that, if $d = \text{poly}\log H$, the complexity is still quasi-polynomial.

Proof. We divide the proof into the following three **Parts**.

Part I: Given any QC LTC problem $\mathcal{L}$ satisfying Assumptions III.1, III.4, III.5, and III.7, we can construct a Dec-POMDP problem $\mathcal{D}'_{\mathcal{L}}$ with SI-CIBs such that finding an ϵ-team optimal strategy can give us an ϵ-team optimal strategy of $\mathcal{L}$, as shown in Algorithm 1.

Part II: Given any ϵ-expected-approximate common information model $\mathcal{M}$ of the Dec-POMDP $\mathcal{D}'_{\mathcal{L}}$, there exists an algorithm, Algorithm 6, that can output an ϵ-team optimal strategy of $\mathcal{D}'_{\mathcal{L}}$.
First, we need to prove that solving $\mathcal{M}$ can get the ϵ-team optimal strategy of $\mathcal{D}'_{\mathcal{L}}$. We prove the following 2 lemmas first.

Lemma C.12. For any strategy $\overline{g}_{1:\overline{H}} \in \overline{\mathcal{G}}_{1:\overline{H}}$, and $h \in [\overline{H}]$, we have

$$\mathbb{E}^{\mathcal{D}'_{\mathcal{L}}}_{\overline{g}_{1:\overline{H}}}[|V_h^{\overline{g}_{1:\overline{H}},\mathcal{D}'_{\mathcal{L}}}(\overline{c}_h) - V_h^{\overline{g}_{1:\overline{H}},\mathcal{M}}(\overline{c}_h)|] \le (\overline{H}-h+1)\epsilon_r + \frac{(\overline{H}-h+1)(\overline{H}-h)}{2}\epsilon_z. \tag{C.11}$$

Proof. We prove it by induction. For $h = \overline{H}+1$, we have $V_h^{\overline{g}_{1:\overline{H}},\mathcal{D}'_{\mathcal{L}}}(\overline{c}_h) = V_h^{\overline{g}_{1:\overline{H}},\mathcal{M}}(\overline{c}_h) = 0$. [kz:why "we have"? where did we define it?][kz:pls define the value at $H+1$ at the first time u INTRODUCED value functions?]
For the step $h \le \overline{H}$, we have

$$\begin{aligned}
&\mathbb{E}^{\mathcal{D}'_{\mathcal{L}}}_{\overline{g}_{1:\overline{H}}}[|V_h^{\overline{g}_{1:\overline{H}},\mathcal{D}'_{\mathcal{L}}}(\overline{c}_h) - V_h^{\overline{g}_{1:\overline{H}},\mathcal{M}}(\overline{c}_h)|] \\
\le&\mathbb{E}^{\mathcal{D}'_{\mathcal{L}}}_{\overline{g}_{1:\overline{H}}}\left[|\mathbb{E}^{\mathcal{D}_{\mathcal{L}}}[\overline{\mathcal{R}}_h(\overline{s}_h,\overline{a}_h,\overline{p}_h)|\overline{c}_h,\{\overline{g}_{j,h}(\overline{c}_h,\cdot)\}_{j\in[n]}] - \widehat{\mathcal{R}}_h^{\mathcal{M}}(\widehat{c}_h,\{\overline{g}_{j,h}(\overline{c}_h,\cdot)\}_{j\in[n]})|\right] \\
&+\mathbb{E}^{\mathcal{D}'_{\mathcal{L}}}_{\overline{g}_{1:\overline{H}}}\left[|\mathbb{E}_{\overline{z}_{h+1}\sim\mathbb{P}_h^{\mathcal{D}'_{\mathcal{L}}}(\overline{c}_h,\{\overline{g}_{j,h}(\overline{c}_h,\cdot)\}_{j\in[n]})}[V_h^{\overline{g}_{1:\overline{H}},\mathcal{D}'_{\mathcal{L}}}(\overline{c}_h\cup\overline{z}_{h+1})] - \mathbb{E}_{\overline{z}_{h+1}\sim\mathbb{P}_h^{\mathcal{M},z}(\cdot|\widehat{c}_h,\{\overline{g}_{j,h}(\overline{c}_h,\cdot)\}_{j\in[n]})}[V_h^{\overline{g}_{1:\overline{H}},\mathcal{M}}(\overline{c}_h\cup\overline{z}_{h+1})]|\right] \\
\le&\epsilon_r + (\overline{H}-h)\mathbb{E}^{\mathcal{D}'_{\mathcal{L}}}_{\overline{a}_{1:h-1},\overline{o}_{1:h}\sim\overline{g}_{1:h-1}}\|\mathbb{P}_h^{\mathcal{D}'_{\mathcal{L}}}(\cdot|\overline{c}_h,\gamma_h) - \mathbb{P}_h^{\mathcal{M},z}(\cdot|\widehat{c}_h,\gamma_h)\|_1 + \mathbb{E}^{\mathcal{D}'_{\mathcal{L}}}_{\overline{a}_{1:h-1},\overline{o}_{1:h}\sim\overline{g}_{1:h-1}}\left[|V_{h+1}^{\overline{g}_{1:\overline{H}},\mathcal{D}'_{\mathcal{L}}}(\overline{c}_{h+1}) - V_{h+1}^{\overline{g}_{1:\overline{H}},\mathcal{M}}(\overline{c}_{h+1})|\right] \\
\le&\epsilon_r + (\overline{H}-h)\epsilon_z + (\overline{H}-h)\epsilon_r + \frac{(\overline{H}-h)(\overline{H}-h-1)}{2}\epsilon_z \\
\le&(\overline{H}-h+1)\epsilon_r + \frac{(\overline{H}-h)(\overline{H}-h+1)}{2}\epsilon_z.
\end{aligned}$$

The proof mainly follows from the proof of Lemma 2 in [14]. But the difference is that $\mathcal{D}'_{\mathcal{L}}$ may not satisfy Assumption II.1. In the third line of this proof, we had $\overline{z}_{h+1} \sim \mathbb{P}_h^{\mathcal{D}'_{\mathcal{L}}}(\cdot|\overline{c}_h,\{\overline{g}_{j,h}(\overline{c}_h,\cdot)\}_{j\in[n]})$, where $\overline{z}_{h+1}$ is generated as

$$\begin{aligned}
&\mathbb{P}_h^{\mathcal{D}'_{\mathcal{L}}}(\overline{z}_{h+1}|\overline{c}_h,\gamma_h) \\
=&\sum_{\overline{s}_h\in\overline{\mathcal{S}},\overline{p}_h\in\overline{\mathcal{P}}_h}\mathbb{P}_h^{\mathcal{D}'_{\mathcal{L}}}(\overline{s}_h,\overline{p}_h|\overline{c}_h)\sum_{\overline{s}_{h+1}\in\overline{\mathcal{S}},\overline{o}_{h+1}\in\overline{\mathcal{O}}_{h+1}}\overline{\mathbb{T}}_{h+1}(\overline{s}_{h+1}|\overline{s}_h,\overline{a}_h)\overline{\mathbb{O}}_{h+1}(\overline{o}_{h+1}|\overline{s}_{h+1})\mathbb{1}[\overline{\chi}_{h+1}(\overline{p}_h,\overline{a}_h,\overline{o}_{h+1})],
\end{aligned}$$

with $\gamma_h = \{\overline{g}_{j,h}(\overline{c}_h,\cdot)\}_{j\in[n]}, \overline{a}_h = \gamma_h(\overline{p}_h)$ if $h = 2k$ else $\overline{a}_h = \gamma_h$, for some $k \in [H]$. □

Lemma C.13. Let $\widehat{g}^*_{1:\overline{H}} \in \overline{\mathcal{G}}_{1:\overline{H}}$ be the strategy output by Algorithm 6, then for any $h \in [\overline{H}], \overline{c}_h \in \overline{\mathcal{C}}_h, \overline{g}_{1:\overline{H}} \in \overline{\mathcal{G}}_{1:\overline{H}}$, it holds that

$$V_h^{\overline{g}_{1:\overline{H}},\mathcal{M}}(\overline{c}_h) \le V_h^{\widehat{g}^*_{1:\overline{H}},\mathcal{M}}(\overline{c}_h). \tag{C.12}$$

Proof. We prove it by induction. For $h = \overline{H}+1$ [kz:start from $h=\overline{H}$? check all.], we have $V_h^{\overline{g}_{1:\overline{H}},\mathcal{M}}(\overline{c}_h) = V_h^{\widehat{g}^*_{1:\overline{H}},\mathcal{M}}(\overline{c}_h) = 0$.
For the timestep $h \le \overline{H}$, we have

$$\begin{aligned}
V_h^{\overline{g}_{1:\overline{H}},\mathcal{M}}(\overline{c}_h) &= \mathbb{E}^{\mathcal{M}}[\widehat{r}_h^{\mathcal{M}}(\widehat{c}_h, \{\overline{g}_{j,h}(\overline{c}_h,\cdot)\}_{j\in[n]}) + V_{h+1}^{\overline{g}_{1:\overline{H}},\mathcal{M}}(\overline{c}_{h+1})|\widehat{c}_h, \overline{g}_h] \\
&\le \mathbb{E}^{\mathcal{M}}[\widehat{r}_h^{\mathcal{M}}(\widehat{c}_h, \{\overline{g}_{j,h}(\overline{c}_h,\cdot)\}_{j\in[n]}) + V_{h+1}^{\widehat{g}^*_{1:\overline{H}},\mathcal{M}}(\overline{c}_{h+1})|\widehat{c}_h, \overline{g}_h] \\
&= Q_h^{\widehat{g}^*_{1:\overline{H}},\mathcal{M}}(\overline{c}_h, \{\overline{g}_{j,h}(\overline{c}_h,\cdot)\}_{j\in[n]}) \\
&\le Q_h^{\widehat{g}^*_{1:\overline{H}},\mathcal{M}}(\overline{c}_h, \{\widehat{g}_{j,h}(\overline{c}_h,\cdot)\}_{j\in[n]}) \\
&= V_h^{\widehat{g}^*_{1:\overline{H}},\mathcal{M}}(\overline{c}_h).
\end{aligned}$$

For the first inequality, we use the induction hypothesis. For the second inequality sign, we use the property of argmax in algorithm and $V_h^{\widehat{g}^*_{1:\overline{H}},\mathcal{M}}(\overline{c}_h) = V_h^{\widehat{g}^*_{1:\overline{H}},\mathcal{M}}(\widehat{c}_h)$. By induction, we complete the proof. □

We now go back to the proof of the theorem. Let $\widehat{g}^*_{1:\overline{H}}$ be the solution output by Algorithm 6, then for any $\overline{g}_{1:\overline{H}} \in \overline{\mathcal{G}}_{1:\overline{H}}, h \in [\overline{H}], \overline{c}_h \in \overline{\mathcal{C}}_h$, we have

$$\begin{aligned}
&\mathbb{E}^{\mathcal{D}'_{\mathcal{L}}}_{\overline{g}_{1:\overline{H}}}\left[V_h^{\overline{g}_{1:\overline{H}},\mathcal{D}'_{\mathcal{L}}}(\overline{c}_h) - V_h^{\widehat{g}^*_{1:\overline{H}},\mathcal{D}'_{\mathcal{L}}}(\overline{c}_h)\right] \\
&= \mathbb{E}^{\mathcal{D}'_{\mathcal{L}}}_{\overline{g}_{1:\overline{H}}}\left[\left(V_h^{\overline{g}_{1:\overline{H}},\mathcal{D}'_{\mathcal{L}}}(\overline{c}_h) - V_h^{\widehat{g}^*_{1:\overline{H}},\mathcal{M}}(\overline{c}_h)\right) + \left(V_h^{\widehat{g}^*_{1:\overline{H}},\mathcal{M}}(\overline{c}_h) - V_h^{\widehat{g}^*_{1:\overline{H}},\mathcal{D}'_{\mathcal{L}}}(\overline{c}_h)\right)\right] \\
&\le \mathbb{E}^{\mathcal{D}'_{\mathcal{L}}}_{\overline{g}_{1:\overline{H}}}\left[\left(V_h^{\overline{g}_{1:\overline{H}},\mathcal{D}'_{\mathcal{L}}}(\overline{c}_h) - V_h^{\overline{g}_{1:\overline{H}},\mathcal{M}}(\overline{c}_h)\right) + \left(V_h^{\widehat{g}^*_{1:\overline{H}},\mathcal{M}}(\overline{c}_h) - V_h^{\widehat{g}^*_{1:\overline{H}},\mathcal{D}'_{\mathcal{L}}}(\overline{c}_h)\right)\right] \\
&\le (\overline{H}-h+1)\epsilon_r + \frac{(\overline{H}-h)(\overline{H}-h+1)}{2}\epsilon_z + (\overline{H}-h+1)\epsilon_r + \frac{(\overline{H}-h)(\overline{H}-h+1)}{2}\epsilon_z \\
&= 2(\overline{H}-h+1)\epsilon_r + (\overline{H}-h)(\overline{H}-h+1)\epsilon_z.
\end{aligned} \tag{C.13}$$

For the first inequality, we use Lemma C.13. For the second inequality sign, we use Lemma C.12. Then apply $h=0$, we have $J_{\mathcal{D}'_{\mathcal{L}}}(\overline{g}_{1:\overline{H}}) \le J_{\mathcal{D}'_{\mathcal{L}}}(\widehat{g}^*_{1:\overline{H}}) + 2\overline{H}\epsilon_r + \overline{H}^2\epsilon_z$. This completes the proof of **Part II**.

Part III: If the baseline sharing of $\mathcal{L}$ is one of the 8 cases in §A, we can construct an expected-approximate common information model of $\mathcal{D}'_{\mathcal{L}}$.
We first prove following lemmas: We aim to bound (ϵ_r, ϵ_z) using the following lemma.

Lemma C.14. Given any belief $\{\mathbb{P}_h^{\mathcal{M},c}(\overline{s}_h, \overline{p}_h|\widehat{c}_h)\}_{h\in[\overline{H}]}$ consistent with the expected-approximate-common-information model $\mathcal{M}$, it holds that for any $h \in [\overline{H}], \overline{\mathcal{C}}_h, \gamma_h \in \Gamma_h$:

$$\begin{aligned}
\|\mathbb{P}_h^{\mathcal{D}'_{\mathcal{L}}}(\cdot|\overline{c}_h, \gamma_h) - \mathbb{P}_h^{\mathcal{M},z}(\cdot|\widehat{c}_h, \gamma_h)\|_1 &\le \|\mathbb{P}_h^{\mathcal{D}'_{\mathcal{L}}}(\cdot,\cdot|\overline{c}_h) - \mathbb{P}_h^{\mathcal{M},c}(\cdot,\cdot|\widehat{c}_h)\|_1, \\
|\mathbb{E}^{\mathcal{D}'_{\mathcal{L}}}[\mathcal{R}_h(\overline{s}_h, \overline{a}_h, \overline{p}_h)|\overline{c}_h, \gamma_h] - \widehat{\mathcal{R}}_h^{\mathcal{M}}(\widehat{c}_h, \gamma_h)| &\le \|\mathbb{P}_h^{\mathcal{D}'_{\mathcal{L}}}(\cdot,\cdot|\overline{c}_h) - \mathbb{P}_h^{\mathcal{M},c}(\cdot,\cdot|\widehat{c}_h)\|_1,
\end{aligned}$$

where $\widehat{c}_h$ =Compress$_h(\overline{c}_h)$.

Proof. Adapted from Lemma 4 [kz:Lemma 3 is not about this? pls check all] in [14] by changing the reward function of $r_{i,h}(s_h, a_h)$ to $\mathcal{R}_h(\overline{s}_h, \overline{a}_h, \overline{p}_h)$. Note that the latter can still be evaluated given the common-information-based belief, $\mathbb{P}_h^{\mathcal{D}'_{\mathcal{L}}}(\overline{s}_h, \overline{p}_h|\overline{c}_h)$. □

Then we define the belief states following the notation in [27, 14] as $\overline{\boldsymbol{b}}_1(\emptyset) = \mu_1$, $\overline{\boldsymbol{b}}_h(\overline{o}_{1:h}, \overline{a}_{1:h-1}) = \mathbb{P}(\overline{s}_h = \cdot | \overline{o}_{1:h}, \overline{a}_{1:h-1}), \overline{\boldsymbol{b}}_h(\overline{o}_{1:h-1}, \overline{a}_{1:h-1}) = \mathbb{P}(\overline{s}_h = \cdot | \overline{o}_{1:h-1}, \overline{a}_{1:h-1})$, where $\overline{\boldsymbol{b}} \in \Delta(\mathcal{S})$. Also, for any $L \geq 1$, we define the approximate belief state using the most recent L-step history, that

$$\overline{\boldsymbol{b}}'_h(\overline{o}_{h-L+1:h}, \overline{a}_{h-L:h-1}) = \mathbb{P}(\overline{s}_h = \cdot | \overline{s}_{h-L} \sim \mathrm{Unif}(\mathcal{S}), \overline{o}_{h-L+1:h}, \overline{a}_{h-L:h-1})$$
$$\overline{\boldsymbol{b}}'_h(\overline{o}_{h-L+1:h-1}, \overline{a}_{h-L:h-1}) = \mathbb{P}(\overline{s}_h = \cdot | \overline{s}_{h-L} \sim \mathrm{Unif}(\mathcal{S}), \overline{o}_{h-L+1:h}, \overline{a}_{h-L:h-1}).$$

Also, for any set $N \subseteq [n]$, we define $\overline{o}_{N,h} = \{\overline{o}_{i,h}\}_{i \in N}$, and the same for $\overline{a}_{N,h}$. We can also define the belief of states given historical observations and actions as follows: for any $N \subseteq [n]$,

$$\overline{\boldsymbol{b}}_h(\overline{o}_{1:h-1}, \overline{a}_{1:h-1}, \overline{o}_{N,h}) = \mathbb{P}(\overline{s}_h = \cdot | \overline{o}_{1:h-1}, \overline{a}_{1:h-1}, \overline{o}_{N,h})$$
$$\overline{\boldsymbol{b}}'_h(\overline{o}_{h-L+1:h-1}, \overline{a}_{h-L:h-1}, \overline{o}_{N,h}) = \mathbb{P}_h(\overline{s}_h = \cdot | \overline{s}_{h-L} \sim \mathrm{Unif}(\mathcal{S}), \overline{o}_{h-L+1:h-1}, \overline{a}_{h-L:h-1}, \overline{o}_{N,h}).$$

Let $S = |\mathcal{S}|$ be the cardinality of state space $\mathcal{S}$, we have the following lemma.

Lemma C.15. There is a constant $C \geq 1$ such that the following holds. Given any LTC problem $\mathcal{L}$ satisfying Assumption III.1, and let $\mathcal{D}'_{\mathcal{L}}$ be the Dec-POMDP after reformulation, strict expansion and refinement. Let $\epsilon \geq 0$, fix a strategy $\overline{g}_{1:\overline{H}}$ and indices $1 \leq h - L < h - 1 \leq \overline{H}$. If $L \geq C\gamma^{-4}\log(\frac{|\mathcal{S}|}{\epsilon})$, then the following set of inequalities hold

$$\mathbb{E}_{\overline{o}_{1:h}, \overline{a}_{1:h-1} \sim \overline{g}_{1:\overline{H}}} \|\overline{\boldsymbol{b}}_h(\overline{o}_{1:h}, \overline{a}_{1:h-1}) - \overline{\boldsymbol{b}}'_h(\overline{o}_{h-L+1:h-1}, \overline{a}_{h-L:h})\|_1 \leq \epsilon \tag{C.14}$$
$$\mathbb{E}_{\overline{o}_{1:h}, \overline{a}_{1:h-1} \sim \overline{g}_{1:\overline{H}}} \|\overline{\boldsymbol{b}}_h(\overline{o}_{1:h}, \overline{a}_{1:h-1}) - \overline{\boldsymbol{b}}'_h(\overline{o}_{h-L+1:h-1}, \overline{a}_{h-L:h-1})\|_1 \leq \epsilon \tag{C.15}$$
$$\mathbb{E}_{\overline{o}_{1:h}, \overline{a}_{1:h-1} \sim \overline{g}_{1:\overline{H}}} \|\overline{\boldsymbol{b}}_h(\overline{o}_{1:h-1}, \overline{a}_{1:h-1}, \overline{o}_{N,h}) - \overline{\boldsymbol{b}}'_h(\overline{o}_{h-L+1:h-1}, \overline{a}_{h-L:h-1}, \overline{o}_{N,h})\|_1 \leq \epsilon. \tag{C.16}$$

Proof. Given any LTC problem $\mathcal{L}$, we can construct a Dec-POMDP $\check{\mathcal{D}}$ that the transition and observation functions of $\check{\mathcal{D}}$ are the same as $\mathcal{L}$. And the information of $\check{\mathcal{D}}$ is fully sharing, which means it shares all the $o_{1:h-1}, a_{1:h}$ as common information at timestep h. Since $\mathcal{D}'_{\mathcal{L}}$ is constructed from $\mathcal{L}$ after reformulation, expansion and refinement, for any $h \in [\overline{H}]$, we have

$$\overline{\boldsymbol{b}}_h(\overline{o}_{1:h}, \overline{a}_{1:h-1}) = \boldsymbol{b}_{\lfloor \frac{h+1}{2} \rfloor}(o_{1:\lfloor \frac{h+1}{2} \rfloor}, a_{1:\lfloor \frac{h-1}{2} \rfloor}) = \check{\boldsymbol{b}}_{\lfloor \frac{h+1}{2} \rfloor}(\check{o}_{1:\lfloor \frac{h+1}{2} \rfloor}, \check{a}_{1:\lfloor \frac{h-1}{2} \rfloor})$$
$$\overline{\boldsymbol{b}}_h(\overline{o}_{1:h-1}, \overline{a}_{1:h-1}) = \boldsymbol{b}_{\lfloor \frac{h+1}{2} \rfloor}(o_{1:\lfloor \frac{h}{2} \rfloor}, a_{1:\lfloor \frac{h-1}{2} \rfloor}) = \check{\boldsymbol{b}}_{\lfloor \frac{h+1}{2} \rfloor}(\check{o}_{1:\lfloor \frac{h}{2} \rfloor}, \check{a}_{1:\lfloor \frac{h-1}{2} \rfloor}).$$

And for the approximate belief state, for any $h \in [\overline{H}]$, we have

$$\overline{\boldsymbol{b}}'_h(\overline{o}_{h-L+1:h-1}, \overline{a}_{h-L:h-1}) = \boldsymbol{b}'_{\lfloor \frac{h+1}{2} \rfloor}(o_{\lfloor \frac{h-L+2}{2} \rfloor:\lfloor \frac{h}{2} \rfloor}, a_{\lfloor \frac{h-L}{2} \rfloor:\lfloor \frac{h-1}{2} \rfloor}) = \check{\boldsymbol{b}}'_{\lfloor \frac{h+1}{2} \rfloor}(\check{o}_{\lfloor \frac{h-L+2}{2} \rfloor:\lfloor \frac{h}{2} \rfloor}, \check{a}_{\lfloor \frac{h-L}{2} \rfloor:\lfloor \frac{h-1}{2} \rfloor})$$
$$\overline{\boldsymbol{b}}'_h(\overline{o}_{h-L+1:h}, \overline{a}_{h-L:h-1}) = \boldsymbol{b}'_{\lfloor \frac{h+1}{2} \rfloor}(o_{\lfloor \frac{h-L+2}{2} \rfloor:\lfloor \frac{h+1}{2} \rfloor}, a_{\lfloor \frac{h-L}{2} \rfloor:\lfloor \frac{h-1}{2} \rfloor}) = \check{\boldsymbol{b}}'_{\lfloor \frac{h+1}{2} \rfloor}(\check{o}_{\lfloor \frac{h-L+2}{2} \rfloor:\lfloor \frac{h+1}{2} \rfloor}, \check{a}_{\lfloor \frac{h-L}{2} \rfloor:\lfloor \frac{h-1}{2} \rfloor}).$$

Also, since for any $t \in [H]$, $\overline{a}_{2t-1}$ are communication actions, $\overline{o}_{2t} = \emptyset$ is null, and $\overline{s}_{2t-1} = \overline{s}_{2t}$ always holds. Then we can write Equation (C.14) and Equation (C.15) as

$$\mathbb{E}_{\{\overline{o}_{2t-1}\}_{t=1}^{\lfloor \frac{h+1}{2} \rfloor}, \{\overline{a}_{2t}\}_{t=1}^{\lfloor \frac{h-1}{2} \rfloor} \sim \overline{g}_{1:\overline{H}}} \|\overline{\boldsymbol{b}}_h(\overline{o}_{1:h}, \overline{a}_{1:h-1}) - \overline{\boldsymbol{b}}'_h(\overline{o}_{h-L+1:h}, \overline{a}_{h-L:h-1})\|_1 \leq \epsilon \tag{C.17}$$
$$\mathbb{E}_{\{\overline{o}_{2t-1}\}_{t=1}^{\lfloor \frac{h}{2} \rfloor}, \{\overline{a}_{2t}\}_{t=1}^{\lfloor \frac{h-1}{2} \rfloor} \sim \overline{g}_{1:\overline{H}}} \|\overline{\boldsymbol{b}}_h(\overline{o}_{1:h-1}, \overline{a}_{1:h-1}) - \overline{\boldsymbol{b}}'_h(\overline{o}_{h-L+1:h-1}, \overline{a}_{h-L:h-1})\|_1 \leq \epsilon. \tag{C.18}$$

Since $\check{\mathcal{D}}$ has a fully-sharing IS, for any $i \in [n], h \in [\overline{H}]$ and information $\overline{\tau}_{i,h}, \overline{\tau}_{i,2h}$, we have $\sigma(\overline{\tau}_{i,h}) \subseteq \sigma(\check{\tau}_{i,\lfloor \frac{h+1}{2} \rfloor})$. Therefore, given any strategy $\overline{g}_{1:\overline{H}}$, we can construct a strategy $\check{g}_{1:H}$ such that, for any

$\overline{a}_{1:h-1}, \overline{o}_{1:h}$

$$\mathbb{P}(\{\overline{o}_{2t-1}\}_{t=1}^{\lfloor\frac{h+1}{2}\rfloor}, \{\overline{a}_{2t}\}_{t=1}^{\lfloor\frac{h-1}{2}\rfloor} | \overline{g}_{1:\overline{H}}) = \mathbb{P}(\check{o}_{1:\lfloor\frac{h+1}{2}\rfloor}, \check{a}_{1:\lfloor\frac{h-1}{2}\rfloor} | \check{g}_{1:H}).$$

Since $\check{\mathcal{D}}$ satisfies Assumption III.1, we can apply the Theorem 10 in [14] with $\check{g}_{1:H}$ to get the result that there is a constant $C_0 \geq 1$ such that if $L' \geq C_0 \gamma^{-4} \log(\frac{S}{\epsilon})$, the following holds

$$\mathbb{E}_{\check{o}_{1:\lfloor\frac{h+1}{2}\rfloor}, \check{a}_{1:\lfloor\frac{h-1}{2}\rfloor} \sim \check{g}_{1:H}} \|\check{\boldsymbol{b}}_{\lfloor\frac{h+1}{2}\rfloor}(\check{o}_{1:\lfloor\frac{h+1}{2}\rfloor}, \check{a}_{1:\lfloor\frac{h-1}{2}\rfloor}) - \check{\boldsymbol{b}}'_{\lfloor\frac{h+1}{2}\rfloor}(\check{o}_{\lfloor\frac{h+1}{2}\rfloor - L'+1:\lfloor\frac{h+1}{2}\rfloor}, \check{a}_{\lfloor\frac{h}{2}\rfloor - L':\lfloor\frac{h-1}{2}\rfloor})\|_1 \leq \epsilon \tag{C.19}$$

$$\mathbb{E}_{\check{o}_{1:\lfloor\frac{h+1}{2}\rfloor}, \check{a}_{1:\lfloor\frac{h-1}{2}\rfloor} \sim \check{g}_{1:H}} \|\check{\boldsymbol{b}}_{\lfloor\frac{h+1}{2}\rfloor}(\check{o}_{1:\lfloor\frac{h}{2}\rfloor}, \check{a}_{1:\lfloor\frac{h-1}{2}\rfloor}) - \check{\boldsymbol{b}}'_{\lfloor\frac{h+1}{2}\rfloor}(\check{o}_{\lfloor\frac{h+1}{2}\rfloor - L'+1:\lfloor\frac{h}{2}\rfloor}, \check{a}_{\lfloor\frac{h}{2}\rfloor - L':\lfloor\frac{h-1}{2}\rfloor})\|_1 \leq \epsilon. \tag{C.20}$$

We choose $C = 3C_0, L = 2L'+1$. If $L \geq C\gamma^{-4}\log(\frac{|\mathcal{S}|}{\epsilon})$, we have $L' \geq C_0\gamma^{-4}\log(\frac{|\overline{\mathcal{S}}|}{\epsilon})$. Therefore, we directly get Equation (C.17) and Equation (C.18).
For Equation (C.16), we cannot directly apply Theorem 10 in [14], but we can slightly change Equation (E.11) of Theorem 10 in [14] as

$$\mathbb{E}^{\mathcal{D}'_{\mathcal{L}}}_{o_{1:h}, a_{1:h-1} \sim g_{1:\overline{H}}} \|\overline{\boldsymbol{b}}_h(o_{1:h-1}, a_{1:h-1}, o_{N,h}) - \overline{\boldsymbol{b}}'_h(o_{h-L+1:h-1}, a_{h-L:h-1}, o_{N,h})\|_1 \leq \epsilon. \tag{C.21}$$

It still holds if the posterior update $F^q(P : o_{1,h})$ is changed to $F^q(P : o_{N,h})$, when applying Lemma 12 in the proof of Theorem 10 of [14]. Therefore, we can use the same arguments to prove Equation (C.16) from Equation (C.21) as above, and this completes the proof. □

Then we can compress the common information using a finite-memory truncation. Here, we discuss case by case how to compress it for the 8 examples of QC LTC given in §A. Note that after reformulation, strict expansion, and refinement, **Example 5** and will be the same as **Example 1**, and **Example 7** will be the same as **Example 2**. Hence, we can construct the same expected approximate common information model. Also, after reformulation, strict expansion, and refinement, **Example 5** is similar to **Example 1**, and **Example 8** is similar to **Example 2**, and we can use the construct the same expected approximate common information model, and slightly change the belief which is consistent to the model. We categorize the examples in §A into 6 **Types**.

Type 1: Baseline sharing of $\mathcal{L}$ is one of **Examples 1 and 5** in §A. Then, common information should be that for any $t \in [H], \overline{c}_{2t-1} = \{\overline{o}_{1:2t-2}, \overline{a}_{1:2t-2}\}, \overline{c}_{2t} = \{\overline{o}_{1:2t-2}, \overline{a}_{1:2t-1}, \overline{o}_{N,2t-1}\}, N \subseteq [n]$, where N is the set of agents choose to share their observations through additional sharing, and N can be inferred from $\overline{c}_{2t}$. Then we have that $\mathbb{P}^{\mathcal{D}'_{\mathcal{L}}}_{2t-1}(\overline{s}_{2t-1}, \overline{p}_{2t-1} | \overline{c}_{2t-1}) = \overline{\boldsymbol{b}}_{2t-1}(\overline{o}_{1:2t-2}, \overline{a}_{1:2t-2})(\overline{s}_{2t-1})\overline{\mathbb{O}}_{2t-1}(\overline{o}_{2t-1} | \overline{s}_{2t-1})$. Fix compress length $L > 0$, for timestep $2t-1$, we define the approximate common information as $\widehat{c}_{2t-1} = \{\overline{o}_{2t-L:2t-2}, \overline{a}_{2t-1-L:2t-2}\}$, and the approximate common information conditioned belief as $\mathbb{P}^{\mathcal{M},c}_{2t-1}(\overline{s}_{2t-1}, \overline{p}_{2t-1} | \widehat{c}_{2t-1}) = \overline{\boldsymbol{b}}'_{2t-1}(\overline{o}_{2t-L:2t-2}, \overline{a}_{2t-1-L:2t-2})(\overline{s}_{2t-1})\overline{\mathbb{O}}_{2t-1}(\overline{o}_{2t-1} | \overline{s}_{2t-1})$. Also, we have $\mathbb{P}^{\mathcal{D}'_{\mathcal{L}}}_{2t}(\overline{s}_{2t}, \overline{p}_{2t} | \overline{c}_{2t}) = \overline{\boldsymbol{b}}_{2t-1}(\overline{o}_{1:2t-2}, \overline{a}_{1:2t-2}, \overline{o}_{N,2t-1})(\overline{s}_{2t-1})\mathbb{P}_{2t-1}(\overline{o}_{-N,2t-1} | \overline{s}_{2t-1}, \overline{o}_{N,2t-1})$. For timestep $2t$, we define the approximate common information a $\widehat{c}_{2t} = \{\overline{o}_{2t-L:2t-2}, \overline{a}_{2t-L-1:2t-1}, \overline{o}_{N,2t-1}\}$, and the common information conditioned belief as $\mathbb{P}^{\mathcal{M},c}_{2t}(\overline{s}_{2t}, \overline{p}_{2t} | \widehat{c}_{2t}) = \overline{\boldsymbol{b}}'_{2t-1}(\overline{o}_{2t-L:2t-2}, \overline{a}_{2t-1-L:2t-2}, \overline{o}_{N,2t-1})(\overline{s}_{2t-1})\mathbb{P}_{2t-1}(\overline{o}_{-N,2t-1} | \overline{s}_{2t-1}, \overline{o}_{N,2t-1})$, where $\mathbb{P}_{2t-1}(\overline{o}_{-N,2t-1} | \overline{s}_{2t-1}, \overline{o}_{N,2t-1}) = \frac{\overline{\mathbb{O}}_{2t-1}(\overline{o}_{N,2t-1}, \overline{o}_{-N,2t-1} | \overline{s}_{2t-1})}{\sum_{\overline{o}'_{-N,2t-1}} \overline{\mathbb{O}}_{2t-1}(\overline{o}_{N,2t-1}, \overline{o}'_{-N,2t-1} | \overline{s}_{2t-1})}$.

Now, we need to verify that Definition C.6 is satisfied.

- The $\{\widehat{c}_h\}_{h \in [\overline{H}]}$ satisfied Equation (C.6) since for any $h \in [\overline{H}]$, $\widehat{c}_{h+1} \subseteq \widehat{c}_h \cup \overline{z}_h$.

- Note that for any $\overline{c}_{2t-1}$ and the corresponding $\widehat{c}_{2t-1}$ constructed above:

$$\begin{aligned}
&\|\mathbb{P}^{\mathcal{D}'_{\mathcal{L}}}_{2t-1}(\cdot,\cdot|\overline{c}_h) - \mathbb{P}^{\mathcal{M},c}_{2t-1}(\cdot,\cdot|\widehat{c}_h)\|_1 \\
&\quad = \sum_{\overline{s}_{2t-1},\overline{o}_{2t-1}} |\overline{\boldsymbol{b}}_{2t-1}(\overline{o}_{1:2t-2},\overline{a}_{1:2t-2})(\overline{s}_{2t-1})\overline{\mathbb{O}}_{2t-1}(\overline{o}_{2t-1}|\overline{s}_{2t-1}) \\
&\qquad\qquad - \overline{\boldsymbol{b}}'_{2t-1}(\overline{o}_{2t-L:2t-1},\overline{a}_{2t-1-L:2t-2})(\overline{s}_{2t-1})\overline{\mathbb{O}}_{2t-1}(\overline{o}_{2t-1}|\overline{s}_{2t-1})| \\
&\quad \le \|\overline{\boldsymbol{b}}_{2t-1}(\overline{o}_{1:2t-2},\overline{a}_{1:2t-2}) - \overline{\boldsymbol{b}}'_{2t-1}(\overline{o}_{2t-L:2t-1},\overline{a}_{2t-1-L:2t-2})\|_1.
\end{aligned}$$

  For any $\overline{c}_{2t}$ and the corresponding $\widehat{c}_{2t}$ constructed above:

$$\begin{aligned}
&\|\mathbb{P}^{\mathcal{D}'_{\mathcal{L}}}_{2t}(\cdot,\cdot|\overline{c}_h) - \mathbb{P}^{\mathcal{M},c}_{2t}(\cdot,\cdot|\widehat{c}_h)\| \\
&\quad = \sum_{\overline{s}_{2t-1},\overline{o}_{-N,2t-1}} |\overline{\boldsymbol{b}}_{2t-1}(\overline{o}_{1:2t-2},\overline{a}_{1:2t-2},\overline{o}_{N,2t-1})(\overline{s}_{2t-1})\mathbb{P}_{2t-1}(\overline{o}_{-N,2t-1}|\overline{s}_{2t-1},\overline{o}_{N,2t-1}) \\
&\qquad\qquad - \overline{\boldsymbol{b}}'_{2t-1}(\overline{o}_{2t-L:2t-2},\overline{a}_{2t-1-L:2t-2},\overline{o}_{N,2t-1})(\overline{s}_{2t-1})\mathbb{P}_{2t-1}(\overline{o}_{-N,2t-1}|\overline{s}_{2t-1},\overline{o}_{N,2t-1})| \\
&\quad \le \|\overline{\boldsymbol{b}}_{2t-1}(\overline{o}_{1:2t-2},\overline{a}_{1:2t-2},\overline{o}_{N,2t-1}) - \overline{\boldsymbol{b}}'_{2t-1}(\overline{o}_{2t-L:2t-2},\overline{a}_{2t-1-L:2t-2},\overline{o}_{N,2t-1})\|_1.
\end{aligned}$$

  If we choose $L \ge C\gamma^{-4}\log(\frac{|\overline{\mathcal{S}}|}{\epsilon})$, then we have that for any $h \in [\overline{H}]$

$$\mathbb{E}_{\overline{a}_{1:h-1},o_{1:h}\sim\overline{g}_{1:\overline{H}}}\|\mathbb{P}^{\mathcal{D}'_{\mathcal{L}}}_{h}(\cdot,\cdot|\overline{c}_h) - \mathbb{P}^{\mathcal{M},c}_{h}(\cdot,\cdot|\widehat{c}_h)\|_1 \le \epsilon.$$

Therefore, such a model is an ϵ-expected-approximate common information model.

Type 2: Baseline sharing of $\mathcal{L}$ is **Example 3** in §A. Then, common information should be that for any $t \in [H], \overline{c}_{2t-1} = \{\overline{o}_{1:2t-2},\overline{a}_{1:2t-2},\overline{o}_{1,2t-1}\}, \overline{c}_{2t} = \{\overline{o}_{1:2t-2},\overline{a}_{1:2t-1},\overline{o}_{N,2t-1}\}, N \subseteq [n], 1 \in N$. Here N is the same as defined in **Type 1**, but it must satisfy that $1 \in N$. Similarly as **Type 1**, we construct $\widehat{c}_{2t-1} = \{\overline{o}_{2t-L:2t-2},\overline{a}_{2t-L-1:2t-2},\overline{o}_{1,2t-1}\}, \widehat{c}_{2t} = \{\overline{o}_{2t-L:2t-2},\overline{a}_{2t-L-1:2t-1},\overline{o}_{N,2t-1}\}$, and approximate common information conditioned belief as $\mathbb{P}^{\mathcal{M},c}_{2t-1}(\overline{s}_{2t-1},\overline{p}_{2t-1}|\widehat{c}_{2t-1}) = \overline{\boldsymbol{b}}'_{2t-1}(\overline{o}_{2t-L:2t-2},\overline{a}_{2t-1-L:2t-2},\overline{o}_{1,2t-1})(\overline{s}_{2t-1})\mathbb{P}_{2t-1}(\overline{o}_{-1,2t-1}|\overline{s}_{2t-1},\overline{o}_{1,2t-1}), \mathbb{P}^{\mathcal{M},c}_{2t}(\overline{s}_{2t},\overline{p}_{2t}|\widehat{c}_{2t}) = \overline{\boldsymbol{b}}'_{2t-1}(\overline{o}_{2t-L:2t-2},\overline{a}_{2t-1-L:2t-2},\overline{o}_{N,2t-1})(\overline{s}_{2t-1})\mathbb{P}_{2t-1}(\overline{o}_{-N,2t-1}|\overline{s}_{2t-1},\overline{o}_{N,2t-1})$. Now, we need to verify Definition C.6 is satisfied.

- The $\{\widehat{c}_h\}_{h\in[\overline{H}]}$ satisfies Equation (C.6) since for any $h \in [H]$, $\widehat{c}_{h+1} \subseteq \widehat{c}_h \cup \overline{z}_h$.
- Note that for any $\overline{c}_{2t-1}$ and the corresponding $\widehat{c}_{2t-1}$ constructed above:

$$\begin{aligned}
&\|\mathbb{P}^{\mathcal{D}'_{\mathcal{L}}}_{2t-1}(\cdot,\cdot|\overline{c}_h) - \mathbb{P}^{\mathcal{M},c}_{2t-1}(\cdot,\cdot|\widehat{c}_h)\|_1 \\
&\quad = \sum_{\overline{s}_{2t-1},\overline{o}_{-1,2t-1}} |\overline{\boldsymbol{b}}_{2t-1}(\overline{o}_{1:2t-2},\overline{a}_{1:2t-1},\overline{o}_{1,2t-1})(\overline{s}_{2t-1})\mathbb{P}_{2t-1}(\overline{o}_{-1,2t-1}|\overline{s}_{2t-1},\overline{o}_{1,2t-1}) \\
&\qquad\qquad - \overline{\boldsymbol{b}}'_{2t-1}(\overline{o}_{2t-L:2t-2},\overline{a}_{2t-1-L:2t-2},\overline{o}_{1,2t-1})(\overline{s}_{2t-1})\mathbb{P}_{2t-1}(\overline{o}_{-1,2t-1}|\overline{s}_{2t-1},\overline{o}_{1,2t-1})| \\
&\quad \le \|\overline{\boldsymbol{b}}_{2t-1}(\overline{o}_{1:2t-2},\overline{a}_{1:2t-1},\overline{o}_{1,2t-1}) - \overline{\boldsymbol{b}}'_{2t-1}(\overline{o}_{2t-L:2t-2},\overline{a}_{2t-1-L:2t-2},\overline{o}_{1,2t-1})\|_1.
\end{aligned}$$

For any $\overline{c}_{2t}$ and the corresponding $\widehat{c}_{2t}$ constructed above:

$$\begin{aligned}
&\|\mathbb{P}_{2t}^{\mathcal{D}'_{\mathcal{L}}}(\cdot,\cdot|\overline{c}_h) - \mathbb{P}_{2t}^{\mathcal{M},c}(\cdot,\cdot|\widehat{c}_h)\|_1 \\
&= \sum_{\overline{s}_{2t-1},\overline{o}_{-N,2t-1}} |\overline{\boldsymbol{b}}_{2t-1}(\overline{o}_{1:2t-2},\overline{a}_{1:2t-2},\overline{o}_{N,2t-1})(\overline{s}_{2t-1})\mathbb{P}_{2t-1}(\overline{o}_{-N,2t-1}|\overline{s}_{2t-1},\overline{o}_{N,2t-1}) \\
&\qquad - \overline{\boldsymbol{b}}'_{2t-1}(\overline{o}_{2t-L:2t-2},\overline{a}_{2t-1-L:2t-2},\overline{o}_{N,2t-1})(\overline{s}_{2t-1})\mathbb{P}_{2t-1}(\overline{o}_{-N,2t-1}|\overline{s}_{2t-1},\overline{o}_{N,2t-1})| \\
&\le \|\overline{\boldsymbol{b}}_{2t-1}(\overline{o}_{1:2t-2},\overline{a}_{1:2t-2},\overline{o}_{N,2t-1}) - \overline{\boldsymbol{b}}'_{2t-1}(\overline{o}_{2t-L:2t-2},\overline{a}_{2t-1-L:2t-2},\overline{o}_{N,2t-1})\|_1.
\end{aligned}$$

If we choose $L \ge C\gamma^{-4}\log(\frac{|\overline{\mathcal{S}}|}{\epsilon})$, then from Lemma C.15 we have, for any $h \in [\overline{H}]$

$$\mathbb{E}_{\overline{a}_{1:h-1},o_{1:h}\sim\overline{g}_{1:\overline{H}}}\|\mathbb{P}_h^{\mathcal{D}'_{\mathcal{L}}}(\cdot,\cdot|\overline{c}_h) - \mathbb{P}_h^{\mathcal{M},c}(\cdot,\cdot|\widehat{c}_h)\|_1 \le \epsilon.$$

Therefore, such a model is an ϵ-expected-approximate common information model.

Type 3: Baseline sharing of $\mathcal{L}$ is one of **Examples 2 and 7** in §A. Then the common information should be that, for any $h \in [\overline{H}], \overline{c}_h = \{\overline{o}_{1:h-2d}, \overline{a}_{1,1:h-1}, \{\overline{a}_{-1,2t-1}\}_{t=1}^{\lfloor\frac{h}{2}\rfloor}, \overline{o}_{1,h-2d+1:h}, \overline{o}_M\}$, where $M \subseteq \{(i,t)|1 < i \le n, h-2d+1 \le t \le h\}$, $\overline{o}_M = \{o_{i,t}|(i,t) \in M\}$, and -1 index means all the agents expect agent 1. The corresponding private information is defined as $\overline{p}_h = \{\overline{o}_{i,t}|1 < i \le n, h-2d < t \le h, (i,t) \notin M\}$. Actually, $\overline{o}_M$ are the observations shared by the additional sharing in $\mathcal{L}$. Denote $f_{\tau,h-2d} = \{\overline{a}_{1,1:h-2d-1}, \overline{o}_{h-2d}, \{\overline{a}_{-1,2t-1}\}_{t=1}^{\lfloor\frac{h-2d}{2}\rfloor}\}$, $f_a = \{\overline{a}_{1,h-2d:h-1}, \{\overline{a}_{-1,2t-1}\}_{t=1\lfloor\frac{h-2d+1}{2}\rfloor}^{\frac{h}{2}}\}$, $f_o = \{\overline{o}_{1,h-2d+1:h}, \overline{o}_M\}$. We can compute the common-information-based belief as

$$\begin{aligned}
\mathbb{P}_h^{\mathcal{D}'_{\mathcal{L}}}(\overline{s}_h,\overline{p}_h|\overline{c}_h) &= \sum_{\overline{s}_{h-2d}} \mathbb{P}_h^{\mathcal{D}'_{\mathcal{L}}}(\overline{s}_h,\overline{p}_h|\overline{s}_{h-2d},f_a,f_o)\mathbb{P}_h^{\mathcal{D}'_{\mathcal{L}}}(\overline{s}_{h-2d}|f_{\tau,h-2d},f_a,f_o) \\
&= \sum_{\overline{s}_{h-2d}} \mathbb{P}_h^{\mathcal{D}'_{\mathcal{L}}}(\overline{s}_h,\overline{p}_h|\overline{s}_{h-2d},f_a,f_o)\frac{\mathbb{P}_h^{\mathcal{D}'_{\mathcal{L}}}(\overline{s}_{h-2d},f_a,f_o|f_{\tau,h-2d})}{\sum_{\overline{s}'_{h-2d}}\mathbb{P}_h^{\mathcal{D}'_{\mathcal{L}}}(\overline{s}'_{h-2d},f_a,f_o|f_{\tau,h-2d})}.
\end{aligned}$$

Denote the probability $P_h(f_o|\overline{s}_{h-2d},f_a) := \Pi_{t=1}^{2d}\mathbb{P}_h^{\mathcal{D}'_{\mathcal{L}}}(\overline{o}_{1,h-2d+t},\overline{o}_{M_{h-2d+t}}|\overline{s}_{h-2d},\overline{a}_{1,h-2d:h-2d+t})$, where $M_{h-2d+t} = \{(i,h-2d+t)|(i,h-2d+t) \in M\}$ denotes the set of observations at timestep $h-2d+t$ and shared through additional sharing. With such notation, we have

$$\begin{aligned}
\mathbb{P}_h^{\mathcal{D}'_{\mathcal{L}}}(\overline{s}_{h-2d}|f_{\tau,h-2d},f_a,f_o) &= \frac{\overline{\boldsymbol{b}}_{h-2d}(\overline{o}_{1:h-2d},\overline{a}_{1:h-2d-1})(\overline{s}_{h-2d})P_h(f_o|\overline{s}_{h-2d},f_a)}{\sum_{\overline{s}'_{h-2d}}\overline{\boldsymbol{b}}_{h-2d}(\overline{o}_{1:h-2d},\overline{a}_{1:h-2d-1})(\overline{s}'_{h-2d})P_h(f_o|\overline{s}'_{h-2d},f_a)} \\
&= F^{P_h(\cdot|\cdot,f_a)}(\overline{\boldsymbol{b}}_{h-2d}(\overline{o}_{1:h-2d},\overline{a}_{1:h-2d-1});f_o)(\overline{s}_{h-2d}),
\end{aligned}$$

where $F^{P_h(\cdot|\cdot,f_a)}(\cdot;f_o):\Delta(\mathcal{S})\to\Delta(\mathcal{S})$ is the posterior belief update function. The formal definition is shown in [14, Lemma 12].

Then, we can define the approximate common information as $\widehat{c}_h := \{\overline{o}_{h-2d-L+1:h-2d}, \overline{o}_{1,h-2d+1:h}, \overline{a}_{1,h-2d-L:h-1}, \{\overline{a}_{-1,2t-1}\}_{t=\lfloor\frac{h-2d+1}{2}\rfloor}^{\lfloor\frac{h}{2}\rfloor}, \overline{o}_M\}$ and the corresponding approximate common information conditioned belief as

$$\mathbb{P}_h^{\mathcal{M},c}(\overline{s}_h,\overline{p}_h|\widehat{c}_h) = \sum_{\overline{s}_{h-2d}} \mathbb{P}_h^{\mathcal{D}'_{\mathcal{L}}}(\overline{s}_h,\overline{p}_h|\overline{s}_{h-2d},f_a,f_o)F^{P_h(\cdot|\cdot,f_a)}(\overline{\boldsymbol{b}}'_{h-2d}(\overline{o}_{h-2d-L+1:h-2d},\overline{a}_{h-2d-L:h-2d-1});f_o)(\overline{s}_{h-2d}).$$

Now we verify that Definition C.6 is satisfied.

- Obviously, the $\{\widehat{c}_h\}_{h\in[\overline{H}]}$ satisfies Equation (C.6).
- For any $\overline{c}_h$ and the corresponding $\widehat{c}_h$ constructed above:

$$\begin{aligned}&\|\mathbb{P}_h^{\mathcal{D}'_{\mathcal{L}}}(\cdot,\cdot\,|\,\overline{c}_h)-\mathbb{P}_h^{\mathcal{M},c}(\cdot,\cdot\,|\,\widehat{c}_h)\|_1\\&\quad\le\|F^{P_h(\cdot|\cdot,f_a)}(\overline{\boldsymbol{b}}_{h-2d}(\overline{o}_{1:h-2d},\overline{a}_{1:h-2d-1});f_o)-F^{P_h(\cdot|\cdot,f_a)}(\overline{\boldsymbol{b}}'_{h-2d}(\overline{o}_{h-2d-L+1:h-2d},\overline{a}_{h-2d-L:h-2d-1});f_o)\|_1.\end{aligned}$$

  If we choose $L\ge C\gamma^{-4}\log(\frac{|\overline{\mathcal{S}}|}{\epsilon})$, then for any strategy $\overline{g}_{1:\overline{H}}$, by taking expectations over $f_{\tau,h-2d},f_a,f_o$, from Lemma C.15 and Lemma 12 in [14], we have, for any $h\in[\overline{H}]$

$$\mathbb{E}_{\overline{a}_{1:h-1},o_{1:h}\sim\overline{g}_{1:\overline{H}}}\|\mathbb{P}_h^{\mathcal{D}'_{\mathcal{L}}}(\cdot,\cdot\,|\,\overline{c}_h)-\mathbb{P}_h^{\mathcal{M},c}(\cdot,\cdot\,|\,\widehat{c}_h)\|_1\le\epsilon.$$

Therefore, such a model is an ϵ-expected-approximate common information model.

Type 4: Baseline sharing of $\mathcal{L}$ is **Example 4** in §A. Then, for any $h\in[H]$, the common information should be $\widehat{c}_h:=\{\overline{o}_{1:h-2d},\{\overline{a}_{2t-1}\}_{t=1}^{\lfloor\frac{h}{2}\rfloor},\overline{o}_M\}$, where $M\subseteq\{(i,t)\,|\,i\in[n],h-2d+1\le t\le h\}$. Then, still we denote $f_{\tau,h-2d}=\{\overline{o}_{1:h-2d},\{\overline{a}_{2t-1}\}_{t=1}^{\lfloor\frac{h}{2}\rfloor}\}$, $f_o=\{\overline{o}_M\}$. We can compute the common-information-based belief as

$$\begin{aligned}\mathbb{P}_h^{\mathcal{D}'_{\mathcal{L}}}(\overline{s}_h,\overline{p}_h\,|\,\overline{c}_h)&=\sum_{\overline{s}_{h-2d}}\mathbb{P}_h^{\mathcal{D}'_{\mathcal{L}}}(\overline{s}_h,\overline{p}_h\,|\,\overline{s}_{h-2d},f_o)\mathbb{P}_h^{\mathcal{D}'_{\mathcal{L}}}(\overline{s}_{h-2d}\,|\,f_{\tau,h-2d},f_o)\\&=\sum_{\overline{s}_{h-2d}}\mathbb{P}_h^{\mathcal{D}'_{\mathcal{L}}}(\overline{s}_h,\overline{p}_h\,|\,\overline{s}_{h-2d},f_o)\frac{\mathbb{P}_h^{\mathcal{D}'_{\mathcal{L}}}(\overline{s}_{h-2d},f_o\,|\,f_{\tau,h-2d})}{\sum_{\overline{s}'_{h-2d}}\mathbb{P}_h^{\mathcal{D}'_{\mathcal{L}}}(\overline{s}'_{h-2d},f_o\,|\,f_{\tau,h-2d})}.\end{aligned}$$

Denote the probability $P_h(f_o\,|\,\overline{s}_{h-2d}):=\Pi_{t=1}^{2d}\mathbb{P}_h^{\mathcal{D}'_{\mathcal{L}}}(\overline{o}_{1,h-2d+t},\overline{o}_{M_{h-2d+t}}\,|\,\overline{s}_{h-2d})$, where $M_{h-2d+t}\subseteq\{(i,h-2d+t)\,|\,(i,h-2d+t)\in M\}$ denotes the set of observations at timestep $h-2d+t$ and shared through additional sharing. Since the actions do not influence underlying states, here we use the belief notation $\overline{\boldsymbol{b}}_k(\overline{o}_{1:k}),\overline{\boldsymbol{b}}_k(\overline{o}_{k-L:k})$, $\forall k\in[\overline{H}],L<k$. With such notation, we have

$$\mathbb{P}_h^{\mathcal{D}'_{\mathcal{L}}}(\overline{s}_{h-2d}\,|\,f_{\tau,h-2d},f_o)=\frac{\overline{\boldsymbol{b}}_{h-2d}(\overline{o}_{1:h-2d})(\overline{s}_{h-2d})P_h(f_o\,|\,\overline{s}_{h-2d})}{\sum_{\overline{s}'_{h-2d}}\overline{\boldsymbol{b}}_{h-2d}(\overline{o}_{1:h-2d})(\overline{s}'_{h-2d})P_h(f_o\,|\,\overline{s}'_{h-2d})}=F^{P_h(\cdot|\cdot)}(\overline{\boldsymbol{b}}_{h-2d}(\overline{o}_{1:h-2d});f_o)(\overline{s}_{h-2d}),$$

where $F^{P_h(\cdot|\cdot)}(\cdot;f_o):\Delta(\mathcal{S})\to\Delta(\mathcal{S})$ is the posterior belief update function, the same as discussed in **Type 3**.
Then, we define the approximate common information as $\widehat{c}_h:=\{\overline{o}_{h-2d-L+1:h},\overline{o}_M\}$ and corresponding approximate common information conditioned belief as

$$\mathbb{P}_h^{\mathcal{M},c}(\overline{s}_h,\overline{p}_h\,|\,\widehat{c}_h)=\sum_{\overline{s}_{h-2d}}\mathbb{P}_h^{\mathcal{D}'_{\mathcal{L}}}(\overline{s}_h,\overline{p}_h\,|\,\overline{s}_{h-2d},f_o)F^{P_h(\cdot|\cdot)}(\overline{\boldsymbol{b}}'_{h-2d}(\overline{o}_{h-2d-L+1:h-2d});f_o)(\overline{s}_{h-2d}).$$

Now we verify that Definition C.6 is satisfied.

- Obviously, the $\{\widehat{c}_h\}_{h\in[\overline{H}]}$ satisfies Equation (C.6).

- For any $\bar{c}_h$ and corresponding $\widehat{c}_h$ constructed above:

$$\begin{aligned}&\|\mathbb{P}_h^{\mathcal{D}'_{\mathcal{L}}}(\cdot,\cdot|\bar{c}_h)-\mathbb{P}_h^{\mathcal{M},c}(\cdot,\cdot|\widehat{c}_h)\|_1\\&\quad\le\|F^{P(\cdot|\cdot)}(\bar{\boldsymbol{b}}_{h-2d}(\bar{o}_{1:h-2d});f_o)-F^{P(\cdot|\cdot)}(\bar{\boldsymbol{b}}'_{h-2d}(\bar{o}_{h-2d-L+1:h-2d});f_o)\|_1.\end{aligned}$$

  If we choose $L\ge C\gamma^{-4}\log(\frac{|\bar{\mathcal{S}}|}{\epsilon})$, then for any strategy $\bar{g}_{1:\overline{H}}$, by taking expectations over $f_{\tau,h-2d},f_o$, from Lemma C.15 and Lemma 12 in [14], we have, for any $h\in[\overline{H}]$

$$\mathbb{E}_{\bar{a}_{1:h-1},o_{1:h}\sim\bar{g}_{1:\overline{H}}}\|\mathbb{P}_h^{\mathcal{D}'_{\mathcal{L}}}(\cdot,\cdot|\bar{c}_h)-\mathbb{P}_h^{\mathcal{M},c}(\cdot,\cdot|\widehat{c}_h)\|_1\le\epsilon.$$

Therefore, such a model is an ϵ-expected approximate common information model.

Type 5: Baseline sharing of $\mathcal{L}$ is **Example 6**. Note that, after reformulation, expansion, and refinement, the common information in **Example 6** is the same as that in **Example 1**. The only difference is in the private information part: for any $t\in[H-1]$, $i\in[n]$, $\bar{a}_{i,2t}\in\bar{p}_{i,2t+1}$ always hold, and $\bar{a}_{i,2t}\in\bar{p}_{i,2t+2}$ may happen. Meanwhile, $\bar{a}_{i,2t}\in\bar{c}_{2t+1}\subseteq\bar{c}_{2t+1}$ always holds. Therefore, we can use the same expected approximate common information model as **Type 1** as: $\forall t\in[H],\widehat{c}_{2t-1}=\{\bar{o}_{2t-L:2t-2},\bar{a}_{2t-1-L:2t-2}\},\widehat{c}_{2t}=\{\bar{o}_{2t-L:2t-2},\bar{a}_{2t-1-L:2t-1},\bar{o}_{N,2t-1}\}$. We construct the approximate common information conditioned belief as $\mathbb{P}_{2t-1}^{\mathcal{M},c}(\bar{s}_{2t-1},\bar{p}_{2t-1}|\widehat{c}_{2t-1})=\bar{\boldsymbol{b}}_{2t-1}(\bar{o}_{2t-L-1:2t-2},\bar{a}_{2t-1-L:2t-2})(\bar{s}_{2t-1})\overline{\mathbb{O}}_{2t-1}(\bar{o}_{2t-1}|\bar{s}_{2t-1})\Psi_{2t-1}^1(\bar{p}_{2t-1},\widehat{c}_{2t-1}),\mathbb{P}_{2t}^{\mathcal{M},c}(\bar{s}_{2t},\bar{p}_{2t}|\widehat{c}_{2t})=\bar{\boldsymbol{b}}'_{2t-1}(\bar{o}_{2t-L-1:2t-2},\bar{a}_{2t-1-L:2t-2},\bar{o}_{N,2t-1})(\bar{s}_{2t-1})\mathbb{P}_{2t-1}(\bar{o}_{-N,2t-1}|\bar{s}_{2t-1},\bar{o}_{N,2t-1})\Psi_{2t}^1(\bar{p}_{2t},\widehat{c}_{2t})$. The functions $\{\Psi_h^1\}_{h\in[\overline{H}]}$ are defined as: $\forall t\in[H],\forall\bar{p}_{2t-1},\widehat{c}_{2t-1},\bar{p}_{2t}$ and $\widehat{c}_{2t}$,

$$\Psi_{2t-1}^1(\bar{p}_{2t-1},\widehat{c}_{2t-1})=\begin{cases}1 & \text{the value of } \bar{a}_{2t-1} \text{ in } \bar{p}_{2t-1} \text{ is the same as the value of that in } \widehat{c}_{2t-1}\\ 0 & \text{o.w.}\end{cases}$$

$$\Psi_{2t}^1(\bar{p}_{2t},\widehat{c}_{2t})=\begin{cases}1 & \text{for any } i\in[n] \text{ such that random variable } \bar{a}_{i,2t-2}\in\bar{p}_{2t},\\ & \text{the value of } \bar{a}_{i,2t-2} \text{ in } \bar{p}_{2t} \text{ is the same as that in } \widehat{c}_{2t}\\ 0 & \text{o.w.}\end{cases}.$$

One can verify that Definition C.6 is satisfied as in **Type 1**. This is because, for any $h\in[\overline{H}]$, compared to **Type 1**, the only difference is that there are only some actions in $\bar{p}_h$, and such actions will also appear in $\widehat{c}_h$[kz:how can it not be?][hy:it must be, we do not truncate itndnd]. Therefore, if the value of such actions in $\bar{p}_h$ is consistent with that in $\widehat{c}_h$, which is ensured by the functions $\{\Psi_h^1\}_{h\in[\overline{H}]}$,[kz:is this correct? pls clarify every argument of urs..][hy:yes] then we can leverage the validation in **Type 1**.

Type 6:[kz:pls propagate the previous edits here...] Baseline sharing of $\mathcal{L}$ is **Example 8**. Note that, after reformulation, expansion, and refinement, the common information in **Example 8** is the same as that in **Example 2**. The only difference is in the private information part: for any $t\in[H-1],\bar{a}_{1,2t}\in\bar{p}_{1,2t+1}$ always holds, and $\bar{a}_{1,2t}\in\bar{p}_{1,2t+2}$ may happen. Meanwhile, $\bar{a}_{1,2t}\in\bar{c}_{2t+1}\subseteq\bar{c}_{2t+1}$ always holds. Therefore, we can use the same expected approximate common information model as **Type 3** as: $\forall h\in[\overline{H}],\widehat{c}_h:=\{\bar{o}_{h-2d-L+1:h-2d},\bar{o}_{1,h-2d+1:h},\bar{a}_{1,h-2d-L:h-1},\{\bar{a}_{-1,2t-1}\}_{t=\lfloor\frac{h-2d+1}{2}\rfloor}^{\lfloor\frac{h}{2}\rfloor},\bar{o}_M\}$, and construct beliefs as

$$\mathbb{P}_h^{\mathcal{M},c}(\bar{s}_h,\bar{p}_h|\bar{c}_h)=\sum_{\bar{s}_{h-2d}}\mathbb{P}_h^{\mathcal{D}'_{\mathcal{L}}}(\bar{s}_h,\bar{p}_h|\bar{s}_{h-2d},f_a,f_o)F^{P_h(\cdot|\cdot,f_a)}(\bar{\boldsymbol{b}}'_{h-2d}(\bar{o}_{h-2d-L:h-2d},\bar{a}_{h-2d-L:h-2d-1},);f_o)(\bar{s}_{h-2d})\Psi_h^2(\bar{p}_h,\widehat{c}_h),$$

where the functions $\{\Psi^2_h\}_{h\in\overline{H}}$ are defined as: $\forall t \in [H]$,$\forall \overline{p}_h$, and $\widehat{c}_h$,

$$\Psi^1_{2t-1}(\overline{p}_{2t-1},\widehat{c}_{2t-1}) = \begin{cases} 1 & \text{the value of } \overline{a}_{1,2t-1} \text{ in } \overline{p}_{2t-1} \text{ is the same as the value of that in } \widehat{c}_{2t-1} \\ 0 & \text{o.w.} \end{cases}$$

$$\Psi^1_{2t}(\overline{p}_{2t},\widehat{c}_{2t}) = \begin{cases} 1 & \text{random variable } \overline{a}_{1,2t-2} \notin \overline{p}_{2t} \text{ or the value of } \overline{a}_{1,2t-2} \text{ in } \overline{p}_{2t} \text{ is the same as that in } \widehat{c}_{2t} \\ 0 & \text{o.w.} \end{cases}.$$

One can verify that Definition C.6 is satisfied as in **Type 3**. This is because, for any $h \in [\overline{H}]$, compared to **Type 3**, the only difference is that there are only some actions of agent 1[kz:again, is this edit accurate? I think it should be "the only difference is that XXXXX". pls propagate.] in $\overline{p}_h$, but such actions will also appear in $\widehat{c}_h$. Therefore, if the value of such actions in $\overline{p}_h$ is consistent with that in $\widehat{c}_h$, which is ensured by the functions $\{\Psi^2_h\}_{h\in[\overline{H}]}$,[kz:is this correct? pls clarify every argument of urs..] then we can leverage the validation in **Type 3**.

[kz:question – when u use these functions $\{\Psi^2_h\}_{h\in[\overline{H}]}$ as "indicator" functions, after u "nullify" those events, dont u need to "normalize" to make sure it is a "valid probability" (i.e., sum up to 1)? pls think..][hy:can only one value of $\overline{a}_h$ to make $\Psi = 1$.]

Combining **Parts I, II, III**, we complete the proof. □

Remark C.16. Let $\mathcal{L}$ be an LTC problem satisfying Assumptions III.1, III.4, III.5, and III.7, and $\mathcal{D}'_{\mathcal{L}}$ be the Dec-POMDP after reformulation, strict expansion and refinement. Then, if $\mathcal{L}$ has any one of baseline sharing protocols as one of the examples in §A, and satisfies the conditions as follows, then we can construct an expected-approximate common information model $\mathcal{M}$ of $\mathcal{D}'_{\mathcal{L}}$ that satisfies Assumption IV.7.

- If $\mathcal{L}$ has a baseline sharing protocol as one of **Examples 1, 5, 6** in §A, $\mathcal{L}$ needs to satisfy the **Part (1)** of **Factorized structure** in §G-B.
- If $\mathcal{L}$ has a baseline sharing protocol as one of **Examples 2, 3, 4, 7, 8** in §A, it does not need any additional condition.

[kz:why we need this following paragraph? isnt it already included above?] Under these conditions of $\mathcal{L}$, the $\mathcal{M}$ constructed in the proof of Theorem IV.8 satisfies that: If the baseline sharing protocol of $\mathcal{L}$ is one of **Examples 1, 5, 6**, then $\mathcal{M}$ satisfies the **Factorized structures** condition in §G-B; If the baseline sharing protocol of $\mathcal{L}$ is one of **Examples 2, 4, 7, 8**, then $\mathcal{M}$ satisfies the **Turn-based structures** condition in §G-B; If the baseline sharing protocol of $\mathcal{L}$ is one of **Examples 3**, then $\mathcal{M}$ satisfies the **Nested private information** condition in §G-B. From Lemma G.2, we can conclude that Assumption IV.7 holds, by noticing that in these examples, $\max_{h\in[\overline{H}]}|\overline{\mathcal{P}}_h|$ depends polynomially on the parameters of the original LTC problem $\mathcal{L}$. [kz:it seems that as long as they are one of the examples here, we DO NOT NEED ANY OTHER ASSUMPTION? but I think we need, as shown above..][kz:pls check..][hy:We need additional for one-step delay, others not.]

Finally, we highlight that for the examples with a delay d of sharing, i.e., **Examples 2, 4, 7, 8**, our main result in Theorem IV.8 on the quasi-polynomial time-complexity also applies to the case when $d = \text{poly}\log H$ (beyond being a *constant* as assumed throughout), since the total complexity will contain error bound of order $\mathcal{O}((|\overline{\mathcal{O}}_h||\overline{\mathcal{A}}_h|)^d)$[kz:add overline overall..]. A similar generalization also applies to the main result for learning in LTCs (see Theorem C.17). Such a generalization also resembles that on the quasi-polynomial complexities for the examples in Theorems 7 and 9 in [14].

C-I Main Results for Learning in QC LTCs

Theorem C.17 (Full version of Theorem IV.9)**.** Given any QC LTC problem $\mathcal{L}$ satisfying Assumptions III.1, III.4, III.5, and III.7, we can construct a Dec-POMDP problem $\mathcal{D}'_{\mathcal{L}}$ with SI-CIBs. Moreover, given any compression functions $\{\text{Compress}_h\}_{h\in[H]}$, evolution rules $\{\widehat{\phi}_h\}_{h\in[\overline{H}]}$ of the compressed common information $\{\widehat{c}_h \in \widehat{\mathcal{C}}_h\}_{h\in[\overline{H}]}$, $\epsilon\in(0,1), \delta\in(0,1)$, and $\widehat{L}$ as defined in Definition C.10, we can apply Algorithm 2 with a universal constant C as chosen in [14, Theorem 8]. If the learned $K = 2\overline{H}|\overline{\mathcal{S}}|$ expected-approximate-common-information models $\{\widehat{\mathcal{M}}(\overline{g}^{1:\overline{H},j})\}_{j\in[K]}$ all satisfy Assumption IV.7, then an ϵ_0-team-optimal strategy for $\mathcal{L}$ can be learned with probability $1-\delta$, with time and sample complexities polynomial in the parameters of $\{\widehat{\mathcal{M}}(\overline{g}^{1:\overline{H},j})\}_{j\in[K]}$, where ϵ_0 is defined as

$$\epsilon_0 := \min_{j\in[K]} \ \overline{H}\epsilon_r(\widehat{\mathcal{M}}(\overline{g}^{1:\overline{H},j})) + \overline{H}^2\epsilon_z(\widehat{\mathcal{M}}(\overline{g}^{1:\overline{H},j})) + \frac{6\epsilon}{200(\overline{H}+1)^2}.$$

Specifically, if $\mathcal{L}$ has a baseline sharing protocol as one of the examples in §A, then given any $\epsilon\in(0,1), \delta\in(0,1)$, we can construct compression functions $\{\text{Compress}_h\}_{h\in[H]}$ and evolution rules $\{\widehat{\phi}_h\}_{h\in[\overline{H}]}$, such that Algorithm 2 can learn an ϵ-team optimal strategy of $\mathcal{L}$ with probability $1-\delta$, with the following time and sample complexities:

- **Examples 1, 3, 5, 6:** $\text{poly}\left(\max_{h\in\overline{H}}(|\overline{\mathcal{O}}_h||\overline{\mathcal{A}}_h|)^{C\gamma^{-4}\log(\frac{|\overline{\mathcal{S}}|}{\epsilon})}, |\overline{\mathcal{S}}|, \overline{H}, \frac{1}{\epsilon}, \log(\frac{1}{\delta})\right)$;
- **Examples 2, 4, 7, 8:** $\text{poly}\left(\max_{h\in\overline{H}}(|\overline{\mathcal{O}}_h||\overline{\mathcal{A}}_h|)^{C\gamma^{-4}\log(\frac{|\overline{\mathcal{S}}|}{\epsilon})+2d}, |\overline{\mathcal{S}}|, \overline{H}, \frac{1}{\epsilon}, \log(\frac{1}{\delta})\right)$.

Proof. **Part 1:** Given any LTC problem $\mathcal{L}$ satisfying the assumptions in the theorem, we can construct a $\mathcal{D}'_{\mathcal{L}}$ by the reformulation, strict expansion, and refinement of $\mathcal{L}$. According to Theorem IV.5, we know that $\mathcal{D}'_{\mathcal{L}}$ has SI-CIBs w.r.t. the strategy space $\Gamma_{1:\overline{H}}$.

Part 2: Given any LTC problem $\mathcal{L}$, we can apply Algorithm 2 for learning in such a problem. For any $j\in[K]$, let $\overline{g}^{j,*}_{1:\overline{H}}$ be the output of Algorithm 6 (Line 11 of Algorithm 2), we can guarantee that $\mathcal{J}_{\mathcal{D}'_{\mathcal{L}}}(\overline{g}^{j,*}_{1:\overline{H}}) \geq \max_{\overline{g}_{1:\overline{H}}\in\overline{\mathcal{G}}_{1:\overline{H}}} \mathcal{J}_{\mathcal{D}'_{\mathcal{L}}}(\overline{g}_{1:\overline{H}}) - \overline{H}\epsilon_r(\widehat{\mathcal{M}}(\overline{g}^{1:\overline{H},j})) + \overline{H}^2\epsilon_z(\widehat{\mathcal{M}}(\overline{g}^{1:\overline{H},j}))$. Meanwhile, let $\widehat{j}$ be the index of the strategy selected by Algorithm Pos-Dec (Line 13), i.e., $\overline{g}^{\widehat{j},*}_{1:\overline{H}}$ is the output of Pos-Dec. Then, we can adapt [14, Lemma 20] to the team setting and guarantee that

$$\max_{\overline{g}_{1:\overline{H}}\in\overline{\mathcal{G}}_{1:\overline{H}}} \mathcal{J}_{\mathcal{D}'_{\mathcal{L}}}(\overline{g}_{1:\overline{H}}) - \mathcal{J}_{\mathcal{D}'_{\mathcal{L}}}(\overline{g}^{\widehat{j},*}_{1:\overline{H}}) \leq \min_{j\in[K]}\left(\max_{\overline{g}_{1:\overline{H}}\in\overline{\mathcal{G}}_{1:\overline{H}}} \mathcal{J}_{\mathcal{D}'_{\mathcal{L}}}(\overline{g}_{1:\overline{H}}) - \mathcal{J}_{\mathcal{D}'_{\mathcal{L}}}(\overline{g}^{j,*}_{1:\overline{H}})\right) + \frac{6\epsilon}{200(\overline{H}+1)^2}$$

with probability $1-\delta_3 \geq 1-\delta$. Therefore, we can guarantee that $\max_{\overline{g}_{1:\overline{H}}\in\overline{\mathcal{G}}_{1:\overline{H}}} \mathcal{J}_{\mathcal{D}'_{\mathcal{L}}}(\overline{g}_{1:\overline{H}}) - \mathcal{J}_{\mathcal{D}'_{\mathcal{L}}}(\overline{g}^{\widehat{j},*}_{1:\overline{H}}) \leq \epsilon_0$ with probability $1-\delta$. Also, together with Proposition IV.1 and Theorem IV.4, with probability $1-\delta$, we conclude that the output of Algorithm 2 is an ϵ_0-team-optimal strategy of $\mathcal{L}$.

Part 3: If the baseline sharing of $\mathcal{L}$ is one of the examples in §A, we can construct the compressed common information with length $L \geq 2C\frac{\log(\overline{H}|\overline{\mathcal{S}}|\max_{h\in[\overline{H}]}|\overline{\mathcal{O}}_h|/(\epsilon\gamma)}{\gamma^4}$ as follows.

- **Examples 1, 5, 6:** For any $t\in[H]$, $\widehat{c}_{2h-1} = \{\overline{o}_{2t-L:2t-2}, \overline{a}_{2t-1-L:2t-2}\}, \widehat{c}_{2h} = \{\overline{o}_{2t-L:2t-2}, \overline{a}_{2t-1-L:2t-2}, \overline{o}_{N,2t-1}\}$ with $N\subseteq[n]$. Here $\widehat{L} = L$.
- **Example 3:** For any $t\in[H]$, $\widehat{c}_{2t-1} = \{\overline{o}_{1,2t-L:2t-2}, \overline{a}_{2t-L-1:2t-2}, \overline{o}_{1,2t-1}\}, \widehat{c}_{2t} = \{\overline{o}_{2t-L:2t-2}, \overline{a}_{2t-L-1:2t-1}, \overline{o}_{N,2t-1}\}$ with $N\subseteq[n]$ and $1\in N$. Here $\widehat{L} = L$.

- **Examples 2, 7, 8:** For any $h \in [\overline{H}], \widehat{c}_h := \{\overline{o}_{h-2d-L+1:h-2d}, \overline{o}_{1,h-2d+1:h}, \overline{a}_{1,h-2d-L:h-1}, \{\overline{a}_{-1,2t-1}\}_{t=\lfloor \frac{h-2d+1}{2} \rfloor}^{\lfloor \frac{h}{2} \rfloor}, \overline{o}_M\}$ with $M \subseteq \{(i,t) | 1 < i \le n, h-2d+1 \le t \le h\}$, and recall that $\overline{o}_M = \{o_{i,t} | (i,t) \in M\}$. Here $\widehat{L} = L + 2d$.
- **Example 4:** For any $h \in [\overline{H}], \widehat{c}_h := \{\overline{o}_{h-L-2d+1:h-2d}, \{\overline{a}_{2t-1}\}_{t=\lfloor \frac{h-2d+1}{2} \rfloor}^{\lfloor \frac{h}{2} \rfloor}, \overline{o}_M\}$ with $M \subseteq \{(i,t) | 1 \le i \le n, h-2d+1 \le t \le h\}$, and recall that $\overline{o}_M = \{o_{i,t} | (i,t) \in M\}$. Here $\widehat{L} = L + 2d$.

The compression functions $\{\text{Compress}_h\}_{h \in [\overline{H}]}$ and evolution rules $\{\widehat{\phi}_h\}_{h \in [\overline{H}]}$ of the compressed common information can be constructed correspondingly.

According to Theorems 8 in [14], for any $j \in [K]$, Algorithm LEE (Line 10 of Algorithm 2) can guarantee that the output $\widehat{\mathcal{M}}(\overline{g}^{1:\overline{H},j})$ has an approximation error as

$$\forall \overline{g}_{1:\overline{H}} \in \overline{\mathcal{G}}_{1:\overline{H}}, |V_0^{\overline{g}_{1:\overline{H}}, \mathcal{D}'_{\mathcal{L}}}(\emptyset) - V_0^{\overline{g}_{1:\overline{H}}, \widehat{\mathcal{M}}(\overline{g}^{1:\overline{H},j})}(\emptyset)| \le \overline{H}\epsilon_r(\widetilde{\mathcal{M}}(\overline{g}^{1:\overline{H},j})) + \overline{H}^2\epsilon_z(\widetilde{\mathcal{M}}(\overline{g}^{1:\overline{H},j})) + \epsilon_{apx}^j \tag{C.22}$$

[kz:introdoce what is tilde M?] with probability $1 - \delta_1$, with

$$\epsilon_{apx}^j = \theta_1 + 2\max_{h \in \overline{H}} |\overline{\mathcal{A}}_h| \max_{h \in \overline{H}} |\overline{\mathcal{P}}_h| \frac{\zeta_1}{\zeta_2} + \max_{h \in \overline{H}} |\overline{\mathcal{A}}_h| \max_{h \in \overline{H}} |\overline{\mathcal{P}}_h| \theta_2 + \frac{\max_{h \in \overline{H}} |\overline{\mathcal{A}}_h|^{2\widehat{L}} \max_{h \in \overline{H}} |\overline{\mathcal{O}}_h|^{\widehat{L}}}{\phi}$$
$$+ \max_{h \in [\overline{H}]} \max_{\overline{g} \in \overline{\mathcal{G}}_{1:\overline{H}}} \mathbb{1}[h > \widehat{L}] \cdot 2 \cdot d_{\mathcal{S},h-\widehat{L}}^{\overline{g},\mathcal{D}'_{\mathcal{L}}}(\mathcal{U}_{\phi,h-\widehat{L}}^{\mathcal{D}'_{\mathcal{L}}}(\overline{g}^{h,j})),$$

where for any $\overline{g} \in \overline{\mathcal{G}}_{1:H}, h \in [\overline{H}]$, we define $d_{\mathcal{S},h}^{\overline{g},\mathcal{D}'_{\mathcal{L}}}(\overline{s}) := \mathbb{P}_h^{\overline{g},\mathcal{D}'_{\mathcal{L}}}(\overline{s}_h = \overline{s})$, for any set $X \subseteq \overline{\mathcal{S}}, d_{\mathcal{S},h}^{\overline{g},\mathcal{D}'_{\mathcal{L}}}(X) = \sum_{\overline{s} \in X} d_{\mathcal{S},h}^{\overline{g},\mathcal{D}'_{\mathcal{L}}}(\overline{s}), \mathcal{U}_{\phi,h}^{\mathcal{D}'_{\mathcal{L}}}(\overline{g}) := \{\overline{s} \in \overline{\mathcal{S}} | d_{\mathcal{S},h}^{\overline{g},\mathcal{D}'_{\mathcal{L}}}(\overline{s}) < \phi\}$. Note that $\widetilde{\mathcal{M}}$ here is the expected approximate common-information model with components $\{\widehat{c}_h\}_{h \in \overline{H}}, \{\widehat{\phi}_h\}_{h \in [\overline{H}]}, \Gamma$, and $\forall j \in [K], \widetilde{\mathcal{M}}(g^{1:\overline{H},j})$ are as defined in Definition C.9. Under the parameters specified in Line 6 of Algorithm 2, we have

$$\epsilon_{apx}^j \le 4\epsilon_1 + \max_{h \in [\overline{H}]} \max_{\overline{g} \in \overline{\mathcal{G}}_{1:\overline{H}}} \mathbb{1}[h > \widehat{L}] \cdot 2 \cdot d_{\mathcal{S},h-\widehat{L}}^{\overline{g},\mathcal{D}'_{\mathcal{L}}}(\mathcal{U}_{\phi,h-\widehat{L}}^{\mathcal{D}'_{\mathcal{L}}}(\overline{g}^{h,j})).$$

Meanwhile, we can prove the following lemma.

Lemma C.18. Given any $\widehat{L} > 0$, and parameters $K, \alpha, \beta, \epsilon_1, N_0, N_1, O, S$ specified in Algorithm 2, and let $\{\overline{g}^{1:\overline{H},j}\}_{j \in [K]}$ be the output of the algorithm BaSeCAMP$(\widehat{L}, N_0, N_1, \alpha, \beta, K)$ (Line 8 in Algorithm 2). As long as $\widehat{L} \ge C\frac{\log(\overline{H}SO/(\epsilon_1\gamma))}{\gamma^4}$, where C is a large enough constant as chosen in [14, Theorem 8], then with probability at least $1 - \delta_2$, there exists at least one $j^* \in [K]$ such that for any strategy $\overline{g}_{1:\overline{H}} \in \overline{\mathcal{G}}_{1:\overline{H}}, h \in [\overline{H}], N \subseteq [n]$

$$\mathbb{E}_{\overline{g}_{1:\overline{H}}} \|\overline{\boldsymbol{b}}_h(\overline{o}_{1:h}, \overline{a}_{1:h-1}) - \overline{\boldsymbol{b}}_h^{\overline{g},h,j^*}(\overline{o}_{h-\widehat{L}+1:h}, \overline{a}_{h-\widehat{L}:h-1})\|_1 \le \epsilon_1 + \mathbb{1}[h > \widehat{L}] \cdot 6 \cdot d_{\mathcal{S},h-\widehat{L}}^{\overline{g},\mathcal{D}'_{\mathcal{L}}}(\mathcal{U}_{\phi,h-\widehat{L}}^{\mathcal{D}'_{\mathcal{L}}}(\overline{g}^{h,j^*}))$$
$$\mathbb{E}_{\overline{g}_{1:\overline{H}}} \|\overline{\boldsymbol{b}}_h(\overline{o}_{1:h-1}, \overline{a}_{1:h-1}) - \overline{\boldsymbol{b}}_h^{\overline{g},h,j^*}(\overline{o}_{h-\widehat{L}+1:h-1}, \overline{a}_{h-\widehat{L}:h-1})\|_1 \le \epsilon_1 + \mathbb{1}[h > \widehat{L}] \cdot 6 \cdot d_{\mathcal{S},h-\widehat{L}}^{\overline{g},\mathcal{D}'_{\mathcal{L}}}(\mathcal{U}_{\phi,h-\widehat{L}}^{\mathcal{D}'_{\mathcal{L}}}(\overline{g}^{h,j^*}))$$
$$\mathbb{E}_{\overline{g}_{1:\overline{H}}} \|\overline{\boldsymbol{b}}_h(\overline{o}_{1:h-1}, \overline{a}_{1:h-1}, \overline{o}_{N,h}) - \overline{\boldsymbol{b}}_h^{\overline{g},h,j^*}(\overline{o}_{h-\widehat{L}+1:h-1}, \overline{a}_{h-\widehat{L}:h-1}, \overline{o}_{N,h}\|_1 \le \epsilon_1 + \mathbb{1}[h > \widehat{L}] \cdot 6 \cdot d_{\mathcal{S},h-\widehat{L}}^{\overline{g},\mathcal{D}'_{\mathcal{L}}}(\mathcal{U}_{\phi,h-\widehat{L}}^{\mathcal{D}'_{\mathcal{L}}}(\overline{g}^{h,j^*})),$$
$$d_{\mathcal{S},h-\widehat{L}}^{\overline{g},\mathcal{D}'_{\mathcal{L}}}(\mathcal{U}_{\phi,h-\widehat{L}}^{\mathcal{D}'_{\mathcal{L}}}(\overline{g}^{h,j^*})) < \epsilon_1,$$

where $\overline{\boldsymbol{b}}_h^{\overline{g}}(\cdot) := \overline{\boldsymbol{b}}_h^{apx,\mathcal{D}'_{\mathcal{L}}}(\cdot, d_{\mathcal{S},h-\widehat{L}}^{\overline{g},\mathcal{D}'_{\mathcal{L}}})$ and $\overline{\boldsymbol{b}}_h^{apx,\mathcal{D}'_{\mathcal{L}}}$ is defined as follows:[kz:what is the N below? where did

we introduce?] $\forall D \in \Delta(\overline{\mathcal{S}}), N \subseteq [n]$

$$\overline{\boldsymbol{b}}_h^{apx,\mathcal{D}'_{\mathcal{L}}}(\overline{o}_{h-\widehat{L}+1:h}, \overline{a}_{h-\widehat{L}+1:h-1}, D) = \mathbb{P}(\overline{s}_h = \cdot | \overline{s}_{h-\widehat{L}} \sim D, \overline{o}_{h-\widehat{L}+1:h}, \overline{a}_{h-\widehat{L}:h-1})$$
$$\overline{\boldsymbol{b}}_h^{apx,\mathcal{D}'_{\mathcal{L}}}(\overline{o}_{h-\widehat{L}+1:h}, \overline{a}_{h-\widehat{L}:h}, D) = \mathbb{P}(\overline{s}_h = \cdot | \overline{s}_{h-\widehat{L}} \sim D, \overline{o}_{h-\widehat{L}+1:h}, \overline{a}_{h-\widehat{L}:h})$$
$$\overline{\boldsymbol{b}}_h^{apx,\mathcal{D}'_{\mathcal{L}}}(\overline{o}_{h-\widehat{L}+1:h}, \overline{a}_{h-\widehat{L}:h}, \overline{o}_{N,h}, D) = \mathbb{P}(\overline{s}_h = \cdot | \overline{s}_{h-\widehat{L}} \sim D, \overline{o}_{h-\widehat{L}+1:h}, \overline{a}_{h-\widehat{L}:h}, \overline{o}_{N,h}).$$

Proof. We can adapt Corollary 4 in [14]. Note that $\mathcal{D}'_{\mathcal{L}}$ only has γ-observability at odd steps $h = 2t-1, t \in [H]$. However, for such $h = 2t-1, s_{h+1} = s_h$ and $\overline{o}_{h+1} = \emptyset$ always hold. Therefore, it holds that $\overline{\boldsymbol{b}}_{h+1}(\overline{o}_{1:h+1}, \overline{a}_{1:h}) = \overline{\boldsymbol{b}}_h(\overline{o}_{1:h}, \overline{a}_{1:h-1})$ and $\overline{\boldsymbol{b}}_{h+1}^{\overline{g}^{h,j^*}}(\overline{o}_{h-\widehat{L}+1:h+1}, \overline{a}_{h-\widehat{L}:h}) = \overline{\boldsymbol{b}}_h^{\overline{g}^{h,j^*}}(\overline{o}_{h-\widehat{L}+1:h}, \overline{a}_{h-\widehat{L}:h-1})$, and similarly for the other two types of beliefs. We can thus adapt the corollary. □

Now, we discuss example by example how to validate that under the compressed common information, $\epsilon_r(\widetilde{\mathcal{M}}(\overline{g}^{1:\overline{H},j^*}))$ and $\epsilon_z(\widetilde{\mathcal{M}}(\overline{g}^{1:\overline{H},j^*}))$ in Equation (C.22) can be made small, where j^* is the j selected in Lemma C.18. Since after reformulation, expansion, and refinement, **Examples 1** and **5** are the same, and **Examples 2** and **7** are the same. Also, from Lemma C.14, for any $h \in [\overline{H}]$, we have [kz:same below: should the tilde here all be hat? pls read and check all..][kz:also, $\widetilde{\mathcal{M}}(\overline{g}^{1:\overline{H},j^*})$ below should really all be $\widetilde{\mathcal{M}}(\overline{g}^{1:\overline{H},j^*})$?? check all.]

$$|\mathbb{E}^{\mathcal{D}'_{\mathcal{L}}}[\mathcal{R}_h(\overline{s}_h, \overline{a}_h, \overline{p}_h) | \overline{c}_h, \gamma_h] - \widehat{\mathcal{R}}_h^{\widetilde{\mathcal{M}}(\overline{g}^{1:\overline{H},j^*})}(\widehat{c}_h, \gamma_h)| \le \|\mathbb{P}_h^{\mathcal{D}'_{\mathcal{L}}}(\cdot, \cdot | \overline{c}_h) - \mathbb{P}_h^{\widetilde{\mathcal{M}}(g^{1:H,j^*}),c}(\cdot, \cdot | \widehat{c}_h)\|$$
$$\|\mathbb{P}_h^{\mathcal{D}'_{\mathcal{L}}}(\cdot | \overline{c}_h, \gamma_h) - \mathbb{P}_h^{\widetilde{\mathcal{M}}(\overline{g}^{1:\overline{H},j^*}),z}(\cdot | \widehat{c}_h, \gamma_h)\|_1 \le \|\mathbb{P}_h^{\mathcal{D}'_{\mathcal{L}}}(\cdot, \cdot | \overline{c}_h) - \mathbb{P}_h^{\widetilde{\mathcal{M}}(g^{1:H,j^*}),c}(\cdot, \cdot | \widehat{c}_h)\|.$$

Therefore, all examples in §A can be classified into the following 6 **Types**.

- **Type 1:** The baseline sharing is either **Example 1** or **Example 5**. Consider any $h \in [\overline{H}]$. If $h = 2t-1, t \in [H]$, it holds that [kz:same here, hat or tilde? check all.]

$$\mathbb{P}_h^{\widetilde{\mathcal{M}}(g^{1:H,j^*}),c}(\overline{s}_h, \overline{p}_h | \widehat{c}_h) = \overline{\boldsymbol{b}}_h^{\overline{g}^{h,j^*}}(\overline{o}_{h-\widehat{L}+1:h-1}, \overline{a}_{h-\widehat{L}:h-1})(\overline{s}_h)\overline{\mathbb{O}}_h(\overline{o}_h | \overline{s_h}),$$
$$\mathbb{P}_h^{\mathcal{D}'_{\mathcal{L}}}(\overline{s}_h, \overline{p}_h | \overline{c}_h) = \overline{\boldsymbol{b}}_h(\overline{o}_{1:h-1}, \overline{a}_{1:h-1})(\overline{s}_h)\overline{\mathbb{O}}_h(\overline{o}_h | \overline{s_h}).$$

  Then, we have

$$\max_{\overline{g} \in \overline{\mathcal{G}}_{1:\overline{H}}} \mathbb{E}^{\mathcal{D}'_{\mathcal{L}}}_{\overline{o}_{1:h}, \overline{a}_{1:h-1} \sim \overline{g}} |\mathbb{E}^{\mathcal{D}'_{\mathcal{L}}}[\mathcal{R}_h(\overline{s}_h, \overline{a}_h, \overline{p}_h) | \overline{c}_h, \gamma_h] - \widehat{\mathcal{R}}_h^{\widetilde{\mathcal{M}}(\overline{g}^{1:\overline{H},j^*})}(\widehat{c}_h, \gamma_h)|$$
$$\le \max_{\overline{g} \in \overline{\mathcal{G}}_{1:\overline{H}}} \mathbb{E}^{\mathcal{D}'_{\mathcal{L}}}_{\overline{o}_{1:h-1}, \overline{a}_{1:h-1} \sim \overline{g}} \|\overline{\boldsymbol{b}}_h(\overline{o}_{1:h-1}, \overline{a}_{1:h-1}) - \overline{\boldsymbol{b}}_h^{\overline{g}^{h,j^*}}(\overline{o}_{h-\widehat{L}+1:h-1}, \overline{a}_{h-\widehat{L}:h-1})\|_1 \le 7\epsilon_1.$$

  Similarly, [kz:same here.] $\max_{\overline{g} \in \overline{\mathcal{G}}_{1:\overline{H}}} \mathbb{E}^{\mathcal{D}'_{\mathcal{L}}}_{\overline{o}_{1:h}, \overline{a}_{1:h-1} \sim \overline{g}} \|\mathbb{P}_h^{\mathcal{D}'_{\mathcal{L}}}(\cdot | \overline{c}_h, \gamma_h) - \mathbb{P}_h^{\widetilde{\mathcal{M}}(\overline{g}^{1:\overline{H},j^*}),z}(\cdot | \widehat{c}_h, \gamma_h)\|_1 \le 7\epsilon_1$.[kz:should the $\overline{c}_h$ in the second term be $\widehat{c}_h$?? pls check all and i will not label all of them..]

  If $h = 2t, t \in [H]$, then it holds that,

$$\mathbb{P}_h^{\widetilde{\mathcal{M}}(g^{1:H,j^*}),c}(\overline{s}_h, \overline{p}_h | \widehat{c}_h) = \overline{\boldsymbol{b}}_{h-1}^{\overline{g}^{h,j^*}}(\overline{o}_{h-\widehat{L}+1:h-2}, \overline{a}_{h-\widehat{L}:h-2}, \overline{o}_{N,h-1})(\overline{s}_h)\mathbb{P}_{2t-1}(\overline{o}_{-N,2t-1} | \overline{s}_{2t-1}, \overline{o}_{N,2t-1})$$
$$\mathbb{P}_h^{\mathcal{D}'_{\mathcal{L}}}(\overline{s}_h, \overline{p}_h | \overline{c}_h) = \overline{\boldsymbol{b}}_{h-1}(\overline{o}_{1:h-2}, \overline{a}_{1:h-2}, \overline{o}_{N,h-1})(\overline{s}_h)\mathbb{P}_{2t-1}(\overline{o}_{-N,2t-1} | \overline{s}_{2t-1}, \overline{o}_{N,2t-1}).$$

Note that $N \subseteq [n]$ can be inferred by $\overline{a}_{h-1}$, which lies in $\widehat{c}_h$ and $\overline{c}_h$. Then, we have

$$\max_{\overline{g}\in\overline{\mathcal{G}}_{1:\overline{H}}} \mathbb{E}^{\mathcal{D}'_{\mathcal{L}}}_{\overline{o}_{1:h},\overline{a}_{1:h-1}\sim\overline{g}} |\mathbb{E}^{\mathcal{D}'_{\mathcal{L}}}[\mathcal{R}_h(\overline{s}_h,\overline{a}_h,\overline{p}_h)|\overline{c}_h,\gamma_h] - \widehat{\mathcal{R}}_h^{\widetilde{\mathcal{M}}(\overline{g}^{1:\overline{H},j^*})}(\widehat{c}_h,\gamma_h)|$$
$$\leq \max_{\overline{g}\in\overline{\mathcal{G}}_{1:\overline{H}}} \mathbb{E}^{\mathcal{D}'_{\mathcal{L}}}_{\overline{o}_{1:h-1},\overline{a}_{1:h-2}\sim\overline{g}} \|\overline{\boldsymbol{b}}_{h-1}(\overline{o}_{1:h-2},\overline{a}_{1:h-2},\overline{o}_{N,h-1}) - \overline{\boldsymbol{b}}^{\overline{g}^{h,j^*}}_{h-1}(\overline{o}_{h-\widehat{L}+1:h-2},\overline{a}_{h-\widehat{L}:h-2},\overline{o}_{N,h-1})\|_1 \leq 7\epsilon_1.$$

Similarly, $\max_{\overline{g}\in\overline{\mathcal{G}}_{1:\overline{H}}} \mathbb{E}^{\mathcal{D}'_{\mathcal{L}}}_{\overline{o}_{1:h},\overline{a}_{1:h-1}\sim\overline{g}} \|\mathbb{P}_h^{\mathcal{D}'_{\mathcal{L}}}(\cdot|\overline{c}_h,\gamma_h) - \mathbb{P}_h^{\widetilde{\mathcal{M}}(\overline{g}^{1:\overline{H},j^*}),z}(\cdot|\widehat{c}_h,\gamma_h)\|_1 \leq 7\epsilon_1$.[kz:same.]

- **Type 2:** The baseline sharing is **Example 3**. For any $h \in [\overline{H}]$, if $h = 2t-1, t \in [H]$,

$$\mathbb{P}_h^{\widetilde{\mathcal{M}}(g^{1:H,j^*}),c}(\overline{s}_h,\overline{p}_h|\widehat{c}_h) = \overline{\boldsymbol{b}}_h^{\overline{g}^{h,j^*}}(\overline{o}_{h-\widehat{L}+1:h-1},\overline{a}_{h-\widehat{L}:h-1},\overline{o}_{1,h})(\overline{s}_h)\mathbb{P}_{2t-1}(\overline{o}_{-1,2t}|\overline{s}_{2t-1},\overline{o}_{1,2t-1})$$
$$\mathbb{P}_h^{\mathcal{D}'_{\mathcal{L}}}(\overline{s}_h,\overline{p}_h|\overline{c}_h) = \overline{\boldsymbol{b}}_h(\overline{o}_{1:h-1},\overline{a}_{1:h-1},\overline{o}_{1,h})(\overline{s}_h)\mathbb{P}_{2t-1}(\overline{o}_{-1,2t}|\overline{s}_{2t-1},\overline{o}_{1,2t-1}).$$

Then we have

$$\max_{\overline{g}\in\overline{\mathcal{G}}_{1:\overline{H}}} \mathbb{E}^{\mathcal{D}'_{\mathcal{L}}}_{\overline{o}_{1:h},\overline{a}_{1:h-1}\sim\overline{g}} |\mathbb{E}^{\mathcal{D}'_{\mathcal{L}}}[\mathcal{R}_h(\overline{s}_h,\overline{a}_h,\overline{p}_h)|\overline{c}_h,\gamma_h] - \widehat{\mathcal{R}}_h^{\widetilde{\mathcal{M}}(\overline{g}^{1:\overline{H},j^*})}(\widehat{c}_h,\gamma_h)|$$
$$\leq \max_{\overline{g}\in\overline{\mathcal{G}}_{1:\overline{H}}} \mathbb{E}^{\mathcal{D}'_{\mathcal{L}}}_{\overline{o}_{1:h},\overline{a}_{1:h-1}\sim\overline{g}} \|\overline{\boldsymbol{b}}_h(\overline{o}_{1:h-1},\overline{a}_{1:h-1},\overline{o}_{1,h}) - \overline{\boldsymbol{b}}_h^{\overline{g}^{h,j^*}}(\overline{o}_{h-\widehat{L}+1:h-1},\overline{a}_{h-\widehat{L}:h-1},,\overline{o}_{1,h})\|_1 \leq 7\epsilon_1.$$

Similarly, $\max_{\overline{g}\in\overline{\mathcal{G}}_{1:\overline{H}}} \mathbb{E}^{\mathcal{D}'_{\mathcal{L}}}_{\overline{o}_{1:h},\overline{a}_{1:h-1}\sim\overline{g}} \|\mathbb{P}_h^{\mathcal{D}'_{\mathcal{L}}}(\cdot|\overline{c}_h,\gamma_h) - \mathbb{P}_h^{\widetilde{\mathcal{M}}(\overline{g}^{1:\overline{H},j^*}),z}(\cdot|\widehat{c}_h,\gamma_h)\|_1 \leq 7\epsilon_1$.
If $h = 2t, t \in [H]$, then it holds that,

$$\mathbb{P}_h^{\widetilde{\mathcal{M}}(g^{1:H,j^*}),c}(\overline{s}_h,\overline{p}_h|\widehat{c}_h) = \overline{\boldsymbol{b}}_{h-1}^{\overline{g}^{h,j^*}}(\overline{o}_{h-\widehat{L}+1:h-2},\overline{a}_{h-\widehat{L}:h-2},\overline{o}_{N,h-1})(\overline{s}_h)\mathbb{P}_{2t-1}(\overline{o}_{-N,2t-1}|\overline{s}_{2t-1},\overline{o}_{N,2t-1})$$
$$\mathbb{P}_h^{\mathcal{D}'_{\mathcal{L}}}(\overline{s}_h,\overline{p}_h|\overline{c}_h) = \overline{\boldsymbol{b}}_{h-1}(\overline{o}_{1:h-2},\overline{a}_{1:h-2},\overline{o}_{N,h-1})(\overline{s}_h)\mathbb{P}_{2t-1}(\overline{o}_{-N,2t-1}|\overline{s}_{2t-1},\overline{o}_{N,2t-1}).$$

Note that, it holds that $1 \in N$, and N can be inferred by $\overline{a}_{h-1}$, which lies in $\widehat{c}_h$ and $\overline{c}_h$. Then, we have

$$\max_{\overline{g}\in\overline{\mathcal{G}}_{1:\overline{H}}} \mathbb{E}^{\mathcal{D}'_{\mathcal{L}}}_{\overline{o}_{1:h},\overline{a}_{1:h-1}\sim\overline{g}} |\mathbb{E}^{\mathcal{D}'_{\mathcal{L}}}[\mathcal{R}_h(\overline{s}_h,\overline{a}_h,\overline{p}_h)|\overline{c}_h,\gamma_h] - \widehat{\mathcal{R}}_h^{\widetilde{\mathcal{M}}(\overline{g}^{1:\overline{H},j^*})}(\widehat{c}_h,\gamma_h)|$$
$$\leq \max_{\overline{g}\in\overline{\mathcal{G}}_{1:\overline{H}}} \mathbb{E}^{\mathcal{D}'_{\mathcal{L}}}_{\overline{o}_{1:h-1},\overline{a}_{1:h-2}\sim\overline{g}} \|\overline{\boldsymbol{b}}_{h-1}(\overline{o}_{1:h-2},\overline{a}_{1:h-2},\overline{o}_{N,h-1}) - \overline{\boldsymbol{b}}^{\overline{g}^{h,j^*}}_{h-1}(\overline{o}_{h-\widehat{L}+1:h-2},\overline{a}_{h-\widehat{L}:h-2},\overline{o}_{N,h-1})\|_1 \leq 7\epsilon_1.$$

Similarly, $\max_{\overline{g}\in\overline{\mathcal{G}}_{1:\overline{H}}} \mathbb{E}^{\mathcal{D}'_{\mathcal{L}}}_{\overline{o}_{1:h},\overline{a}_{1:h-1}\sim\overline{g}} \|\mathbb{P}_h^{\mathcal{D}'_{\mathcal{L}}}(\cdot|\overline{c}_h,\gamma_h) - \mathbb{P}_h^{\widetilde{\mathcal{M}}(\overline{g}^{1:\overline{H},j^*}),z}(\cdot|\widehat{c}_h,\gamma_h)\|_1 \leq 7\epsilon_1$.

- **Type 3:** The baseline sharing is either **Example 2** or **Example 7**. [kz:note that i removed all the sentences below for each Type, coz we have said what $\widehat{L}$ is before?] For any $h \in [\overline{H}]$,[kz:should all the P_h below in the superscript be $\mathbb{P}_h$ or not? check and change all..]

$$\mathbb{P}_h^{\mathcal{D}'_{\mathcal{L}}}(\overline{s}_h,\overline{p}_h|\overline{c}_h) = \sum_{\overline{s}_{h-2d}} \mathbb{P}_h^{\mathcal{D}'_{\mathcal{L}}}(\overline{s}_h,\overline{p}_h|\overline{s}_{h-2d},f_a,f_o) F^{P_h(\cdot|\cdot,f_a)}(\overline{\boldsymbol{b}}_{h-2d}(\overline{o}_{1:h-2d},\overline{a}_{1:h-2d-1});f_o)(\overline{s}_{h-2d}),$$

$$\mathbb{P}_h^{\widetilde{\mathcal{M}}(\overline{g}^{1:\overline{H},j^*}),c}(\overline{s}_{h-2d}|\widehat{c}_h) = \sum_{\overline{s}_{h-2d}} \mathbb{P}_h^{\mathcal{D}'_{\mathcal{L}}}(\overline{s}_h,\overline{p}_h|\overline{s}_{h-2d},f_a,f_o) F^{P_h(\cdot|\cdot,f_a)}(\overline{\boldsymbol{b}}_{h-2d}^{\overline{g}^{1:\overline{H},j^*}}(\overline{o}_{h-\widehat{L}+1:h-2d},\overline{a}_{h-\widehat{L}:h-2d-1});f_o)(\overline{s}_{h-2d}),$$

where $f_{\tau,h-2d} = \{\overline{o}_{1:h-2d}, \overline{a}_{1,1:h-2d-1}, \{\overline{a}_{-1,2t-1}\}_{t=1}^{\lfloor\frac{h-2d}{2}\rfloor}\}$, $f_a = \{\overline{a}_{1,h-2d:h-1}, \{\overline{a}_{-1,2t-1}\}_{t=1\lfloor\frac{h-2d+1}{2}\rfloor}^{\frac{h}{2}}\}$, $f_o = \{\overline{o}_{1,h-2d+1:h}, \overline{o}_M\}$, $F^{P_h(\cdot|\cdot,f_a)}(\cdot; f_o) : \Delta(\mathcal{S}) \to \Delta(\mathcal{S})$ is the posterior belief update function (as introduced in **Type 3** part in the proof Theorem C.11). Then, we have

$$\begin{aligned}
&\max_{\overline{g}\in\overline{\mathcal{G}}_{1:\overline{H}}} \mathbb{E}^{\mathcal{D}'_{\mathcal{L}}}_{\overline{o}_{1:h},\overline{a}_{1:h-1}\sim\overline{g}} |\mathbb{E}^{\mathcal{D}'_{\mathcal{L}}}[\mathcal{R}_h(\overline{s}_h,\overline{a}_h,\overline{p}_h)|\overline{c}_h,\gamma_h] - \widehat{\mathcal{R}}_h^{\widetilde{\mathcal{M}}(\overline{g}^{1:\overline{H},j^*})}(\widehat{c}_h,\gamma_h)| \\
&\quad\le \max_{\overline{g}\in\overline{\mathcal{G}}_{1:\overline{H}}} \mathbb{E}^{\mathcal{D}'_{\mathcal{L}}}_{\overline{o}_{1:h},\overline{a}_{1:h-1}\sim\overline{g}} \| \sum_{\overline{s}_{h-2d}} \mathbb{P}_h^{\mathcal{D}'_{\mathcal{L}}}(\overline{s}_h,\overline{p}_h|\overline{s}_{h-2d},f_a,f_o) F^{P_h(\cdot|\cdot,f_a)}(\overline{\boldsymbol{b}}_{h-2d}(\overline{o}_{1:h-2d},\overline{a}_{1:h-2d-1});f_o) \\
&\qquad - \sum_{\overline{s}_{h-2d}} \mathbb{P}_h^{\mathcal{D}'_{\mathcal{L}}}(\overline{s}_h,\overline{p}_h|\overline{s}_{h-2d},f_a,f_o) F^{P_h(\cdot|\cdot,f_a)}(\overline{\boldsymbol{b}}^{\overline{g}^{1:\overline{H},j^*}}_{h-2d}(\overline{o}_{h-\widehat{L}:h-2d},\overline{a}_{h-\widehat{L}:h-2d-1});f_o)(\overline{s}_{h-2d})\|_1 \\
&\quad\le \max_{\overline{g}\in\overline{\mathcal{G}}_{1:\overline{H}}} \mathbb{E}^{\mathcal{D}'_{\mathcal{L}}}_{\overline{o}_{1:h},\overline{a}_{1:h-1}\sim\overline{g}} \|\overline{\boldsymbol{b}}_{h-2d}(\overline{o}_{1:h-2d},\overline{a}_{1:h-1-2d}) - \overline{\boldsymbol{b}}^{\overline{g}^{h,j^*}}_{h-2d}(\overline{o}_{h-\widehat{L}+1:h-2d},\overline{a}_{h-\widehat{L}:h-2d-1})\|_1 \le 7\epsilon_1.
\end{aligned}$$

Similarly, $\max_{\overline{g}\in\overline{\mathcal{G}}_{1:\overline{H}}} \mathbb{E}^{\mathcal{D}'_{\mathcal{L}}}_{\overline{o}_{1:h},\overline{a}_{1:h-1}\sim\overline{g}} \|\mathbb{P}_h^{\mathcal{D}'_{\mathcal{L}}}(\cdot|\overline{c}_h,\gamma_h) - \mathbb{P}_h^{\widetilde{\mathcal{M}}(\overline{g}^{1:\overline{H},j^*}),z}(\cdot|\widehat{c}_h,\gamma_h)\|_1 \le 7\epsilon_1$.

- **Type 4:** The baseline sharing is **Example 4**. For any $h \in [\overline{H}]$,

$$\begin{aligned}
&\mathbb{P}_h^{\mathcal{D}'_{\mathcal{L}}}(\overline{s}_{h-2d}|\overline{c}_h) = \sum_{\overline{s}_{h-2d}} \mathbb{P}_h^{\mathcal{D}'_{\mathcal{L}}}(\overline{s}_h,\overline{p}_h|\overline{s}_{h-2d},f_o) F^{P_h(\cdot|\cdot)}(\overline{\boldsymbol{b}}_{h-2d}(\overline{o}_{1:h-2d},\overline{a}_{1:h-2d-1});f_o)(\overline{s}_{h-2d}), \\
&\mathbb{P}_h^{\widetilde{\mathcal{M}}(\overline{g}^{1:\overline{H},j^*}),c}(\overline{s}_{h-2d}|\widehat{c}_h) = \sum_{\overline{s}_{h-2d}} \mathbb{P}_h^{\mathcal{D}'_{\mathcal{L}}}(\overline{s}_h,\overline{p}_h|\overline{s}_{h-2d},f_o) F^{P_h(\cdot|\cdot)}(\overline{\boldsymbol{b}}^{\overline{g}^{1:\overline{H},j^*}}_{h-2d}(\overline{o}_{h-\widehat{L}+1:h-2d},\overline{a}_{h-\widehat{L}:h-2d-1});f_o)(\overline{s}_{h-2d}),
\end{aligned}$$

where $f_{\tau,h-2d} = \{\overline{o}_{1:h-2d}, \{\overline{a}_{2t-1}\}_{t=1}^{\lfloor\frac{h}{2}\rfloor}\}$, $f_o = \{\overline{o}_M\}$, $F^{P_h(\cdot|\cdot)}(\cdot; f_o) : \Delta(\mathcal{S}) \to \Delta(\mathcal{S})$ is the posterior belief update function (as introduced in **Type 4** part in the proof Theorem C.11)). Recall that $M \subseteq \{(i,t)\,|\,i \in [n], h-2d+1 \le t \le h\}$. Then, we have

$$\begin{aligned}
&\max_{\overline{g}\in\overline{\mathcal{G}}_{1:\overline{H}}} \mathbb{E}^{\mathcal{D}'_{\mathcal{L}}}_{\overline{o}_{1:h},\overline{a}_{1:h-1}\sim\overline{g}} |\mathbb{E}^{\mathcal{D}'_{\mathcal{L}}}[\mathcal{R}_h(\overline{s}_h,\overline{a}_h,\overline{p}_h)|\overline{c}_h,\gamma_h] - \widehat{\mathcal{R}}_h^{\widetilde{\mathcal{M}}(\overline{g}^{1:\overline{H},j^*})}(\widehat{c}_h,\gamma_h)| \\
&\quad\le \max_{\overline{g}\in\overline{\mathcal{G}}_{1:\overline{H}}} \mathbb{E}^{\mathcal{D}'_{\mathcal{L}}}_{\overline{o}_{1:h},\overline{a}_{1:h-1}\sim\overline{g}} \| \sum_{\overline{s}_{h-2d}} \mathbb{P}_h^{\mathcal{D}'_{\mathcal{L}}}(\overline{s}_h,\overline{p}_h|\overline{s}_{h-2d},f_a,f_o) F^{P_h(\cdot|\cdot)}(\overline{\boldsymbol{b}}_{h-2d}(\overline{o}_{1:h-2d},\overline{a}_{1:h-2d-1});f_o) \\
&\qquad - \sum_{\overline{s}_{h-2d}} \mathbb{P}_h^{\mathcal{D}'_{\mathcal{L}}}(\overline{s}_h,\overline{p}_h|\overline{s}_{h-2d},f_a,f_o) F^{P_h(\cdot|\cdot)}(\overline{\boldsymbol{b}}^{\overline{g}^{h,j^*}}_{h-2d}(\overline{o}_{h-\widehat{L}+1:h-2d},\overline{a}_{h-\widehat{L}:h-2d-1});f_o)(\overline{s}_{h-2d})\|_1 \\
&\quad\le \max_{\overline{g}\in\overline{\mathcal{G}}_{1:\overline{H}}} \mathbb{E}^{\mathcal{D}'_{\mathcal{L}}}_{\overline{o}_{1:h},\overline{a}_{1:h-1}\sim\overline{g}} \|\overline{\boldsymbol{b}}_{h-2d}(\overline{o}_{1:h-2d},\overline{a}_{1:h-1-2d}) - \overline{\boldsymbol{b}}^{\overline{g}^{h,j^*}}_{h-2d}(\overline{o}_{h-\widehat{L}+1:h-2d},\overline{a}_{h-\widehat{L}:h-2d-1})\|_1 \le 7\epsilon_1.
\end{aligned}$$

Similarly, $\max_{\overline{g}\in\overline{\mathcal{G}}_{1:\overline{H}}} \mathbb{E}^{\mathcal{D}'_{\mathcal{L}}}_{\overline{o}_{1:h},\overline{a}_{1:h-1}\sim\overline{g}} \|\mathbb{P}_h^{\mathcal{D}'_{\mathcal{L}}}(\cdot|\overline{c}_h,\gamma_h) - \mathbb{P}_h^{\widetilde{\mathcal{M}}(\overline{g}^{1:\overline{H},j^*}),z}(\cdot|\widehat{c}_h,\gamma_h)\|_1 \le 7\epsilon_1$.

- **Type 5:** The baseline sharing is **Example 6**. Consider any $h \in [\overline{H}]$. If $h = 2t-1, t \in [H]$, it holds that

$$\begin{aligned}
&\mathbb{P}_h^{\widetilde{\mathcal{M}}(g^{1:H,j^*}),c}(\overline{s}_h,\overline{p}_h|\widehat{c}_h) = \overline{\boldsymbol{b}}_h^{\overline{g}^{h,j^*}}(\overline{o}_{h-\widehat{L}+1:h-1},\overline{a}_{h-\widehat{L}:h-1})(\overline{s}_h)\overline{\mathbb{O}}_h(\overline{o}_h|\overline{s_h})\Psi^1_{2t-1}(\overline{p}_{2t-1},\widehat{c}_{2t-1}), \\
&\mathbb{P}_h^{\mathcal{D}'_{\mathcal{L}}}(\overline{s}_h,\overline{p}_h|\overline{c}_h) = \overline{\boldsymbol{b}}_h(\overline{o}_{1:h-1},\overline{a}_{1:h-1})(\overline{s}_h)\overline{\mathbb{O}}_h(\overline{o}_h|\overline{s_h})\Psi^1_{2t-1}(\overline{p}_{2t-1},\overline{c}_{2t-1});
\end{aligned}$$

if $h = 2t, t \in [H]$, then it holds that for any $t \in [H]$,

$$\mathbb{P}_h^{\widetilde{\mathcal{M}}(g^{1:H,j^*}),c}(\bar{s}_h, \bar{p}_h | \widehat{c}_h) = \overline{\boldsymbol{b}}_{h-1}^{\overline{g}^{h,j^*}}(\bar{o}_{h-\widehat{L}+1:h-2}, \bar{a}_{h-\widehat{L}:h-2}, \bar{o}_{N,h-1})(\bar{s}_h)\mathbb{P}_{2t-1}(\bar{o}_{-N,2t-1} | \bar{s}_{2t-1}, \bar{o}_{N,2t-1})\Psi_{2t}^1(\bar{p}_{2t}, \widehat{c}_{2t})$$
$$\mathbb{P}_h^{\mathcal{D}'_{\mathcal{L}}}(\bar{s}_h, \bar{p}_h | \bar{c}_h) = \overline{\boldsymbol{b}}_{h-1}(\bar{o}_{1:h-2}, \bar{a}_{1:h-2}, \bar{o}_{N,h-1})(\bar{s}_h)\mathbb{P}_{2t-1}(\bar{o}_{-N,2t-1} | \bar{s}_{2t-1}, \bar{o}_{N,2t-1}).\mathbb{P}_h^{\mathcal{D}'_{\mathcal{L}}}(\bar{s}_h, \bar{p}_h | \bar{c}_h)\Psi_{2t}^1(\bar{p}_{2t}, \bar{c}_{2t}),$$

where recalling $\{\Psi_h^1\}_{h\in[\overline{H}]}$ is defined in the **Type 5** part in the proof of Theorem C.11, and we can extend it as $\Psi_h^1(\bar{p}_h, \bar{c}_h)$ by replacing $\widehat{c}_h$ by $\bar{c}_h, \forall h \in [\overline{H}]$. Then, similar to **Type 1**, we can verify that for any $h \in [\overline{H}]$[kz:parenthesis for the second one is messed up.. check all.]

$$\max_{\overline{g}\in\overline{\mathcal{G}}_{1:\overline{H}}} \mathbb{E}^{\mathcal{D}'_{\mathcal{L}}}_{\bar{o}_{1:h},\bar{a}_{1:h-1}\sim\overline{g}} |\mathbb{E}^{\mathcal{D}'_{\mathcal{L}}}[\mathcal{R}_h(\bar{s}_h, \bar{a}_h, \bar{p}_h) | \bar{c}_h, \gamma_h] - \widehat{\mathcal{R}}_h^{\widetilde{\mathcal{M}}(\overline{g}^{1:\overline{H},j^*})}(\widehat{c}_h, \gamma_h)| \le 7\epsilon_1$$
$$\max_{\overline{g}\in\overline{\mathcal{G}}_{1:\overline{H}}} \mathbb{E}^{\mathcal{D}'_{\mathcal{L}}}_{\bar{o}_{1:h},\bar{a}_{1:h-1}\sim\overline{g}} \|\mathbb{P}_h^{\mathcal{D}'_{\mathcal{L}}}(\cdot | \bar{c}_h, \gamma_h) - \mathbb{P}_h^{\widetilde{\mathcal{M}}(\overline{g}^{1:\overline{H},j^*}),z}(\cdot | \widehat{c}_h, \gamma_h)\|_1 \le 7\epsilon_1.$$

Type 6: The baseline sharing is **Example 8**. For any $h \in [\overline{H}]$,[kz:same edits for $\mathbb{P}_h$.. and all others..]

$$\mathbb{P}_h^{\mathcal{D}'_{\mathcal{L}}}(\bar{s}_{h-2d} | f_{\tau,h-2d}, f_a, f_o) = \sum_{\bar{s}_{h-2d}} \mathbb{P}_h^{\mathcal{D}'_{\mathcal{L}}}(\bar{s}_h, \bar{p}_h | \bar{s}_{h-2d}, f_a, f_o) F^{P_h(\cdot|\cdot,f_a)}(\overline{\boldsymbol{b}}_{h-2d}(\bar{o}_{1:h-2d}, \bar{a}_{1:h-2d-1}); f_o)(\bar{s}_{h-2d})\Psi_h^2(\bar{p}_h, \bar{c}_h),$$
$$\mathbb{P}_h^{\widetilde{\mathcal{M}}(\overline{g}^{1:\overline{H},j^*}),c}(\bar{s}_{h-2d} | f_{\tau,h-2d}, f_a, f_o)$$
$$= \sum_{\bar{s}_{h-2d}} \mathbb{P}_h^{\mathcal{D}'_{\mathcal{L}}}(\bar{s}_h, \bar{p}_h | \bar{s}_{h-2d}, f_a, f_o) F^{P_h(\cdot|\cdot,f_a)}(\overline{\boldsymbol{b}}_{h-2d}^{\overline{g}^{h,j^*}}(\bar{o}_{h-\widehat{L}+1:h-2d}, \bar{a}_{h-\widehat{L}:h-2d-1}); f_o)(\bar{s}_{h-2d})\Psi_h^2(\bar{p}_h, \widehat{c}_h),$$

where $f_a, f_o, f_{\tau,h-2d}, F^{P_h(\cdot|\cdot,f_a)}$ is defined the same as **Type 3**, and recalling that $\{\Psi_h^2\}_{h\in[\overline{H}]}$, and we can extend it as $\Psi_h^2(\bar{p}_h, \bar{c}_h)$ by replacing $\widehat{c}_h$ by $\bar{c}_h, \forall h \in [\overline{H}]$. [kz:should the input for Ψ_h^2 above be $\widehat{c}$ or $\bar{c}$? shouldnt them be consistent? check all?] is defined in the **Type 6** part in the proof of Theorem C.11. Then, similar to **Type 3**, we can verify that for any $h \in [\overline{H}]$[kz:parenthesis for the second one is messed up.. check all similar places.]

$$\max_{\overline{g}\in\overline{\mathcal{G}}_{1:\overline{H}}} \mathbb{E}^{\mathcal{D}'_{\mathcal{L}}}_{\bar{o}_{1:h},\bar{a}_{1:h-1}\sim\overline{g}} |\mathbb{E}^{\mathcal{D}'_{\mathcal{L}}}[\mathcal{R}_h(\bar{s}_h, \bar{a}_h, \bar{p}_h) | \bar{c}_h, \gamma_h] - \widehat{\mathcal{R}}_h^{\widetilde{\mathcal{M}}(\overline{g}^{1:\overline{H},j^*})}(\widehat{c}_h, \gamma_h)| \le 7\epsilon_1$$
$$\max_{\overline{g}\in\overline{\mathcal{G}}_{1:\overline{H}}} \mathbb{E}^{\mathcal{D}'_{\mathcal{L}}}_{\bar{o}_{1:h},\bar{a}_{1:h-1}\sim\overline{g}} \|\mathbb{P}_h^{\mathcal{D}'_{\mathcal{L}}}(\cdot | \bar{c}_h, \gamma_h) - \mathbb{P}_h^{\widetilde{\mathcal{M}}(\overline{g}^{1:\overline{H},j^*}),z}(\cdot | \widehat{c}_h, \gamma_h)\|_1 \le 7\epsilon_1.$$

Therefore, for any example in §A, with probability $1 - \delta_2$[kz:where did this δ_2 appeared in the argument above?][hy:belief contraction?], we know that there exists some $j^* \in [K]$ such that

$$\epsilon_r(\widetilde{\mathcal{M}}(\overline{g}^{1:\overline{H},j^*})) \le 7\epsilon_1, \quad \epsilon_z(\widetilde{\mathcal{M}}(\overline{g}^{1:\overline{H},j^*})) \le 7\epsilon_1, \quad \epsilon_{apx}^{j^*} \le 4\epsilon_1 + \max_{h\in[\overline{H}]} \max_{\overline{g}\in\overline{\mathcal{G}}_{1:\overline{H}}} \mathbb{1}[h > \widehat{L}] \cdot 2 \cdot d_{\mathcal{S},h-\widehat{L}}^{\overline{g},\mathcal{D}'_{\mathcal{L}}}(\mathcal{U}_{\phi,h-\widehat{L}}^{\mathcal{D}'_{\mathcal{L}}}(\overline{g}^{h,j^*})) \le 6\epsilon_1.$$

Then, with probability $1-\delta_1-\delta_2$, we have $\max_{\overline{g}_{1:\overline{H}}\in\overline{\mathcal{G}}_{1:\overline{H}}} \mathcal{J}_{\mathcal{D}'_{\mathcal{L}}}(\overline{g}_{1:\overline{H}}) - \mathcal{J}_{\mathcal{D}'_{\mathcal{L}}}(\overline{g}^{j^*}_{1:\overline{H}}) \le (7\overline{H}^2 + \overline{H} + 6)\epsilon_1$. Hence, with probability $1 - \delta_1 - \delta_2 - \delta_3 \ge 1 - \delta$, we can have $\max_{\overline{g}_{1:\overline{H}}\in\overline{\mathcal{G}}_{1:\overline{H}}} \mathcal{J}_{\mathcal{D}'_{\mathcal{L}}}(\overline{g}_{1:\overline{H}}) - \mathcal{J}_{\mathcal{D}'_{\mathcal{L}}}(\widehat{g}^*_{1:\overline{H}}) \le (7\overline{H}^2 + \overline{H} + 6 + 6)\epsilon_1 \le \epsilon$, where $\widehat{g}^*_{1:\overline{H}}$ is the strategy output by Algorithm 2, which completes the proof. □

Remark C.19. Let $\mathcal{L}$ be an LTC problem satisfying Assumptions III.1, III.4, III.5, and III.7:

- If $\mathcal{L}$ has a baseline sharing protocol as one of **Examples 1, 5, 6** in §A, $\mathcal{L}$ needs to satisfy the **Part (1)** of **Factorized structure** in §G-B;
- If $\mathcal{L}$ has a baseline sharing protocol as one of **Examples 2, 3, 4, 7, 8** in §A, it does not need any additional condition,

then, we can leverage the structures of these examples, and instantiate [14, Algorithm 5] (Line 10 of Algorithm 2) as follows to guarantee that the learned models in $\widehat{\mathcal{M}} = \{\widehat{\mathcal{M}}(\overline{g}^{1:\overline{H},j})\}_{j=1}^{K}$ satisfy Assumption IV.7. This ensures that the conditions in Theorem IV.9 are satisfied, and the quasi-polynomial complexities of the algorithm can be obtained.

- If $\mathcal{L}$ has a baseline sharing protocol as one of **Examples 1, 5, 6**, then we replace the Equation (B.1) of [14, Algorithm 5] by: for any $i \in [n]$

$$\begin{aligned}
&\text{if } h = 2t-1,\ \mathbb{P}_h^{\widehat{\mathcal{M}}(g^{1:\overline{H}}),z}(\overline{z}_{i,h+1}|\widehat{c}_{i,h},\gamma_{i,h}) \leftarrow \sum_{\overline{p}_{i,h}} \mathbb{1}[\overline{z}_{i,h+1} = \overline{\chi}_{i,h+1}(\overline{p}_{i,h},\overline{\gamma}_{i,h},\overline{o}_{i,h+1} = \emptyset)]\mathbb{P}_h^{\widehat{\mathcal{M}}(g^{1:\overline{H}})}(\overline{p}_{i,h}|\widehat{c}_{i,h}),\\
&\text{if } h = 2t,\ \mathbb{P}_h^{\widehat{\mathcal{M}}(g^{1:\overline{H}}),z}(\overline{z}_{i,h+1}|\widehat{c}_{i,h},\gamma_{i,h}) \leftarrow \sum_{\overline{p}_{i,h},\overline{a}_{i,h},\overline{o}_{i,h+1}} \mathbb{1}[\overline{z}_{i,h+1} = \overline{\chi}_{i,h+1}(\overline{p}_{i,h},\overline{a}_{i,h},\overline{o}_{i,h+1})]\mathbb{P}_h^{\widehat{\mathcal{M}}(g^{1:\overline{H}})}(\overline{p}_{i,h}|\widehat{c}_{i,h})\\
&\qquad\qquad \gamma_{i,h}(\overline{a}_{i,h}|\overline{p}_{i,h})\mathbb{P}_h^{\widehat{\mathcal{M}}(g^{1:\overline{H}})}(\overline{o}_{i,h+1}|\widehat{c}_{i,h},\overline{p}_{i,h},\overline{a}_{i,h}),
\end{aligned} \tag{C.23}$$

  and replace Equation (B.2) of [14, Algorithm 5] by: for any $i \in [n]$

$$\begin{aligned}
&\text{if } h = 2t-1,\ \widehat{\mathcal{R}}_{i,h}^{\widehat{\mathcal{M}}(g^{1:\overline{H}})}(\widehat{c}_{i,h},\gamma_{i,h}) \leftarrow \sum_{\overline{p}_{i,h}} \mathbb{P}_h^{\widehat{\mathcal{M}}(g^{1:\overline{H}})}(\overline{p}_{i,h}|\widehat{c}_{i,h})\widehat{\mathcal{R}}_{i,h}^{\widehat{\mathcal{M}}(g^{1:\overline{H}})}(\widehat{c}_{i,h},\overline{p}_{i,h},\gamma_{i,h}),\\
&\text{if } h = 2t,\ \widehat{\mathcal{R}}_{i,h}^{\widehat{\mathcal{M}}(g^{1:\overline{H}})}(\widehat{c}_{i,h},\gamma_{i,h}) \leftarrow \sum_{\overline{p}_{i,h},\overline{a}_{i,h}} \mathbb{P}_h^{\widehat{\mathcal{M}}(g^{1:\overline{H}})}(\overline{p}_{i,h}|\widehat{c}_{i,h})\widehat{\mathcal{R}}_{i,h}^{\widehat{\mathcal{M}}(g^{1:\overline{H}})}(\widehat{c}_{i,h},\overline{p}_{i,h},\overline{a}_{i,h})\gamma_{i,h}(\overline{a}_{i,h}|\overline{p}_{i,h}),
\end{aligned} \tag{C.24}$$

  where $\{\overline{\chi}_{i,h+1}\}_{h\in\overline{H}}$ can be constructed based on $\{\chi_{i,t}\}_{t\in[H]}$ and $\{\phi_{i,t}\}_{t\in[H]}$ similarly as the proof of Theorem IV.6:

  $\forall \overline{p}_{i,h-1} \in \overline{\mathcal{P}}_{i,h-1}, \overline{a}_{i,h-1} \in \overline{\mathcal{A}}_{i,h-1}, \overline{o}_{i,h} \in \overline{\mathcal{O}}_h$, if h is even, then $\overline{\chi}_{i,h}(\overline{p}_{i,h-1},\overline{a}_{i,h-1},\overline{o}_{i,h}) = \phi_{i,\frac{h}{2}}(\overline{p}_{i,h-1},\overline{a}_{i,h-1})$
  if h is even, then $\overline{\chi}_{i,h}(\overline{p}_{i,h-1},\overline{a}_{i,h-1},\overline{o}_{i,h}) = \chi_{i,\frac{h+1}{2}}(\overline{p}_{i,h-1},\overline{a}_{i,h-1},\overline{o}_{i,h}) \cup \varrho_{i,h}^1 \backslash \varrho_{i,h}^2$, where
  $\varrho_{i,h}^1 := \{\widetilde{a}_{i,h-1} | \forall \sigma(\widetilde{\tau}_{i,h-1}) \subseteq \sigma(\widetilde{c}_h), \widetilde{a}_{i,h-1} \text{ influences } \widetilde{s}_h\} \backslash \chi_{i,\frac{h+1}{2}}(\overline{p}_{i,h-1},\overline{a}_{i,h-1},\overline{o}_{i,h})$ if $h > 1$, otherwise $\emptyset$.
  $\varrho_{i,h}^2 := \{\widetilde{a}_{i,h_0} | \forall h_0 < h-1, \sigma(\widetilde{a}_{i,h_0}) \subseteq \sigma(\widetilde{c}_{h-1}), \widetilde{a}_{i,h_0} \text{ influences } \widetilde{s}_{h_0+1}\} \cap \chi_{i,\frac{h+1}{2}}(\overline{p}_{i,h-1},\overline{a}_{i,h-1},\overline{o}_{i,h})$,

- If $\mathcal{L}$ has a baseline sharing protocol as one of **Examples 2, 4, 7, 8**, then we replace the Equation (B.1) of [14, Algorithm 5] by:

$$\begin{aligned}
&\text{if } h = 2t-1, \text{ we do not make any change,}\\
&\text{if } h = 2t,\ \mathbb{P}_h^{\widehat{\mathcal{M}}(g^{1:\overline{H}}),z}(\overline{z}_{h+1}|\widehat{c}_h,\gamma_{ct(h),h}) \leftarrow \sum_{\overline{p}_h,\overline{a}_{ct(h),h},\overline{o}_{h+1}} \mathbb{1}[\overline{z}_{h+1} = \overline{\chi}_{h+1}(\overline{p}_{i,h},\overline{a}_{ct(h),h},\overline{o}_{h+1})]\mathbb{P}_h^{\widehat{\mathcal{M}}(g^{1:\overline{H}})}(\overline{p}_h|\widehat{c}_h)\\
&\qquad\qquad \gamma_{ct(h),h}(\overline{a}_{ct(h),h}|\overline{p}_h)\mathbb{P}_h^{\widehat{\mathcal{M}}(g^{1:\overline{H}})}(\overline{o}_{h+1}|\widehat{c}_h,\overline{p}_h,\overline{a}_{ct(h),h}),
\end{aligned} \tag{C.25}$$

and replace Equation (B.2) of [14, Algorithm 5] by:

if $h = 2t-1$, we do not make any change.

$$\text{if } h = 2t,\ \widehat{\mathcal{R}}_{i,h}^{\widehat{\mathcal{M}}(g^{1:\overline{H}})}(\widehat{c}_h, \gamma_{i,h}) \leftarrow \sum_{\overline{p}_{i,h}, \overline{a}_{i,h}} \mathbb{P}_h^{\widehat{\mathcal{M}}(g^{1:\overline{H}})}(\overline{p}_h|\widehat{c}_h)\widehat{\mathcal{R}}_{i,h}^{\widehat{\mathcal{M}}(g^{1:\overline{H}})}(\widehat{c}_h, \overline{p}_h, \overline{a}_{i,h})\gamma_{i,h}(\overline{a}_{i,h}|\overline{p}_{i,h}). \tag{C.26}$$

- If $\mathcal{L}$ has a baseline sharing protocol as **Example 3**, then we do not make any change.

Then, the $\widehat{\mathcal{M}} = \{\widehat{\mathcal{M}}(g^{1:\overline{H},j})\}_{j=1}^K$ learned in Line 10 of Algorithm 2 satisfies that: for any $j \in [K]$, if the baseline sharing protocol of $\mathcal{L}$ is one of **Examples 1, 5, 6**, then $\widehat{\mathcal{M}}(g^{1:\overline{H},j})$ satisfies the **Factorized structures** condition in §G-B; if the baseline sharing protocol of $\mathcal{L}$ is one of **Examples 2, 4, 7, 8**, then $\widehat{\mathcal{M}}(g^{1:\overline{H},j})$ satisfies the **Turn-based structures** condition in §G-B; if the baseline sharing protocol of $\mathcal{L}$ is **Example 3**, then $\widehat{\mathcal{M}}(g^{1:\overline{H},j})$ satisfies the **Nested private information** condition in §G-B. From Lemma G.2, we can conclude that Assumption IV.7 holds, by noticing that in these examples, $\max_{h\in[\overline{H}]}|\overline{\mathcal{P}}_h|$ depends polynomially on the parameters of the original LTC problem $\mathcal{L}$.

D. Deferred Details of §V

In the following part, we will use $\bar{}$ to denote the elements and random variables in the Dec-POMDP $\mathcal{D}$. We first introduce the notion of *perfect recall* [24]:

Definition D.1 (Perfect recall)**.** We say that agent i has perfect recall if $\forall h = 2, \cdots, \overline{H}$, it holds that $\overline{\tau}_{i,h-1} \cup \{\overline{a}_{i,h-1}\} \subseteq \overline{\tau}_{i,h}$. If for any $i \in [n]$, agent i has perfect recall, we call that the Dec-POMDP has a perfect recall property.

D-A Proof of Theorem V.1

Proof. sQC $\Rightarrow$ SI-CIB:
Let $\mathcal{D}$ be a Dec-POMDP with an sQC information structure, and let $\mathcal{D}$ satisfy Assumptions II.1 (e), II.2, III.5, and III.7. To prove that $\mathcal{D}$ has SI-CIBs, it is sufficient to prove that for any $h = 2, \cdots, \overline{H}$, fix any $h_1 \in [h-1], i_1 \in [n]$, and for any $\overline{g}_{1:h-1} \in \overline{\mathcal{G}}_{1:h-1}, \overline{g}'_{i_1,h_1} \in \overline{\mathcal{G}}_{i_1,h_1}$, let $\overline{g}'_{h_1} := (\overline{g}_{1,h_1}, \cdots, \overline{g}'_{i_1,h_1}, \cdots, \overline{g}_{n,h_1})$ and $\overline{g}'_{1:h-1} := (\overline{g}_1, \cdots, \overline{g}'_{h_1}, \cdots, \overline{g}_{h-1})$. For any $\overline{c}_h \in \overline{\mathcal{C}}_h$, if $\overline{c}_h$ is reachable under both $\overline{g}_{1:h_1}, \overline{g}'_{1:h-1}$, then the following holds

$$\mathbb{P}(\overline{s}_h, \overline{p}_h|\overline{c}_h, \overline{g}_{1:h-1}) = \mathbb{P}(\overline{s}_h, \overline{p}_h|\overline{c}_h, \overline{g}'_{1:h-1}). \tag{D.1}$$

We prove this case by case as follows:

1. If there exists some $i_2 \neq i_1$ such that $\sigma(\overline{\tau}_{i_1,h_1}) \cup \sigma(\overline{a}_{i_1,h_1}) \subseteq \sigma(\overline{\tau}_{i_2,h})$, then from Assumption II.2, we know that $\sigma(\overline{\tau}_{i_1,h_1}) \cup \sigma(\overline{a}_{i_1,h_1}) \subseteq \sigma(\overline{c}_h)$. Therefore, there exist deterministic measurable functions β_1, β_2 such that $\overline{\tau}_{i_1,h_1} = \beta_1(\overline{c}_h), \overline{a}_{i_1,h_1} = \beta_2(\overline{c}_h)$, and further it holds that

$$\begin{aligned}
\mathbb{P}(\overline{s}_h, \overline{p}_h|\overline{c}_h, \overline{g}_{1:h-1}) &= \mathbb{P}(\overline{s}_h, \overline{p}_h|\beta_1(\overline{c}_h), \beta_2(\overline{c}_h), \overline{c}_h, \overline{g}_{1:h-1}) \\
&= \mathbb{P}(\overline{s}_h, \overline{p}_h|\overline{\tau}_{i_1,h_1}, \overline{a}_{i_1,h_1}, \overline{c}_h, \overline{g}_{1:h-1}) \\
&= \mathbb{P}(\overline{s}_h, \overline{p}_h|\overline{\tau}_{i_1,h_1}, \overline{a}_{i_1,h_1}, \overline{c}_h, \overline{g}'_{1:h-1}).
\end{aligned}$$

The last equality is due to the fact that the input and output of $\overline{g}_{i_1,h_1}$ are $\overline{\tau}_{i_1,h_1}$ and $\overline{a}_{i_1,h_1}$, respectively.

2. If there does not exist any $i_2 \neq i_1$ such that $\sigma(\overline{\tau}_{i_1,h_1}) \cup \sigma(\overline{a}_{i_1,h_1}) \subseteq \sigma(\overline{\tau}_{i_2,h})$, i.e., for all $i_2 \neq i_1$, either $\sigma(\overline{\tau}_{i_1,h_1}) \nsubseteq \sigma(\overline{\tau}_{i_2,h})$ or $\sigma(\overline{a}_{i_1,h_1}) \nsubseteq \sigma(\overline{\tau}_{i_2,h})$, then agent (i_1,h_1) does not influence agent (i_2,h) for any $i_2 \neq i_1$, since $\mathcal{D}$ is sQC.

   Firstly, we claim that agent (i_1,h_1) does not influence $\overline{s}_{h_1+1}$: if it influences, from Assumption III.7, there exists some $i_3 \neq i_1$ such that agent (i_1,h_1) influences $\overline{o}_{i_3,h_1+1}$; however, from Assumption II.1 (e), we know $\overline{o}_{i_3,h_1+1} \in \overline{\tau}_{i_3,h_1+1} \subseteq \overline{\tau}_{i_3,h}$; therefore, agent (i_1,h_1) influences agent (i_3,h), contradicting the premise above that the former does not influence (i_2,h) for any $i_2 \neq i_1$. Hence, we further have that agent (i_1,h_1) does not influence $\overline{s}_{h_2}$ for any $h_2 > h_1$.

   Secondly, we claim that agent (i_1,h_1) does not influence $\overline{\tau}_{i_3,h_2}$, for any $i_3 \in [n]$ and $h_2 > h_1$. Since agent (i_1,h_1) does not influence $\overline{s}_{h_1+1}$, then by Assumption III.5, for any $h_2 > h_1$, $\overline{a}_{i_1,h_1} \notin \overline{\tau}_{h_2}$ and agent (i_1,h_1) does not influence $\overline{o}_{i_3,h_1+1}$ for any $i_3 \in [n]$, which implies that agent (i_1,h_1) does not influence any element in $\overline{\tau}_{i_3,h_1+1}$ for any $i_3 \in [n]$, either directly or indirectly. Since $\overline{\tau}_{i_3,h_1+1}$ is the input of agent i_3's strategy at timestep h_1+1 to decide $\overline{a}_{i_3,h_1+1}$, agent (i_1,h_1) thus does not influence $\overline{a}_{i_3,h_1+1}$ for any $i_3 \in [n]$, either, which, together with the fact that it does not influence $\overline{s}_{h_1+2}$ and thus not $\overline{o}_{i_3,h_1+2}$ for any $i_3 \in [n]$, further implies that it does not influence any element in $\overline{\tau}_{i_3,h_1+2}$ for any $i_3 \in [n]$. By recursion[kz:yes, "recursion" is a better word, than "induction" u previously used. change overall.], agent (i_1,h_1) does not influence $\overline{\tau}_{i_3,h_2}$ for any $i_3 \in [n]$ and $h_2 > h_1$.

   Therefore, agent (i_1,h_1) does not influence $\overline{c}_h = \cap_{i_3=1}^n \overline{\tau}_{i_3,h}$ nor $\overline{p}_h = \overline{\tau}_h \backslash \overline{c}_h$, and it does not influence $\overline{s}_h$, either. Then, it means to

$$\mathbb{P}(\overline{s}_h, \overline{p}_h | \overline{c}_h, \overline{g}_{1:h-1}) = \mathbb{P}(\overline{s}_h, \overline{p}_h | \overline{c}_h, \overline{g}'_{1:h-1}).$$

SI-CIB $\Rightarrow$ sQC:
Since $\mathcal{D}$ has perfect recall and has SI-CIBs, i.e., $\forall i \in [n], h \in [\overline{H}], \forall \overline{g}_{1:h-1}, \overline{g}'_{1:h-1} \in \overline{\mathcal{G}}_{1:h-1}, \overline{c}_h \in \overline{\mathcal{C}}_h, \overline{s}_h \in \overline{\mathcal{S}}, \overline{p}_h \in \overline{\mathcal{P}}_h$, if $\overline{c}_h$ is reachable under both $\overline{g}_{1:h-1}, \overline{g}'_{1:h-1}$, then the following holds

$$\mathbb{P}(\overline{s}_h, \overline{p}_h | \overline{c}_h, \overline{g}_{1:h-1}) = \mathbb{P}(\overline{s}_h, \overline{p}_h | \overline{c}_h, \overline{g}'_{1:h-1}).$$

Our goal is to prove that $\mathcal{D}$ is sQC (up to null sets). In particular, we meant to prove that if agent (i_1,h_1) influences agent (i_2,h_2) in the intrinsic model of the Dec-POMDP (see the definition in §F-A), then under any strategy $\overline{g}_{1:\overline{H}} \in \overline{\mathcal{G}}_{1:\overline{H}}$, $\sigma(\overline{\tau}_{i_1,h_1}) \cup \sigma(\overline{a}_{i_1,h_1}) \subseteq \sigma(\overline{\tau}_{i_2,h_2})$ holds. Note that throughout the proof, when it comes to σ-algebra inclusion, we meant it up to the null sets generated by $\overline{g}_{1:\overline{H}}$.

We prove this by contradiction. If this is not true, then there exist $i_1, i_2 \in [n], h_1, h_2 \in [\overline{H}]$, such that agent (i_1,h_1) influences agent (i_2,h_2), but either $\sigma(\overline{\tau}_{i_1,h_1}) \nsubseteq \sigma(\overline{\tau}_{i_2,h_2})$ or $\sigma(\overline{a}_{i_1,h_1}) \nsubseteq \sigma(\overline{\tau}_{i_2,h_2})$. First, we know that $i_2 \neq i_1$, since otherwise it always holds that $\overline{\tau}_{i_1,h_1} \subseteq \overline{\tau}_{i_1,h_2}$ and $\overline{a}_{i_1,h_1} \in \overline{\tau}_{i_1,h_2}$, due to the perfect recall assumption.

Then, we discuss the following different cases.

1. If $\sigma(\overline{a}_{i_1,h_1}) \nsubseteq \sigma(\overline{\tau}_{i_2,h_2})$, then it implies that $\sigma(\overline{a}_{i_1,h_1}) \nsubseteq \sigma(\overline{c}_{h_2})$ because $\overline{c}_{h_2} \subseteq \overline{\tau}_{i_2,h_2}$. This also implies that $\overline{a}_{i_1,h_1} \notin \overline{c}_{h_2}$, and thus $\overline{a}_{i_1,h_1} \in \overline{p}_{i_1,h_2}$ due to perfect recall. Note that there must exists strategy $\overline{g}_{1:h_2-1}$ and some realizations $\overline{c}_{h_2} \in \overline{\mathcal{C}}_{h_2}, \overline{p}_{h_2} \in \overline{\mathcal{P}}_{h_2}, \overline{s}_{h_2} \in \overline{\mathcal{S}}$ such that $\overline{c}_{h_2}$ has non-zero probability under $\overline{g}_{1:h_2-1}$, and $\mathbb{P}(\overline{s}_{h_2}, \overline{p}_{h_2} | \overline{c}_{h_2}, \overline{g}_{1:h_2-1}) \neq 0$. Meanwhile, there must exist another different action realization $\overline{a}'_{i_1,h_1}$ from realizations $\overline{a}_{i_1,h_1}$ such that

$$\mathbb{P}(\overline{s}_{h_2}, \overline{p}_{h_2} \backslash \{\overline{a}_{i_1,h_1}\} \cup \{\overline{a}'_{i_1,h_1}\} | \overline{c}_{h_2}, \overline{g}_{1:h_2-1}) \neq 0, \tag{D.2}$$

since otherwise it holds that $\sigma(\overline{a}_{i_1,h_1}) \subseteq \sigma(\overline{c}_{h_2})$. In fact, this means that there are some non-zero probability trajectories containing $\overline{a}_{i_1,h_1}$ and $\overline{c}_{h_2}$, and some non-zero probability trajectories containing $\overline{a}'_{i_1,h_1}$ and $\overline{c}_{h_2}$. Then, we define another strategy $\overline{g}'_{i_1,h_1}$ as:

$$\forall \overline{\tau}_{i_1,h_1} \in \overline{\mathcal{T}}_{i_1,h_1}, \quad \overline{g}'_{i_1,h_1}(\overline{\tau}_{i_1,h_1}) = \overline{a}'_{i_1,h_1}, \tag{D.3}$$

and we let $\overline{g}'_{h_1} := (\overline{g}_{1,h_1}, \cdots, \overline{g}'_{i_1,h_1}, \cdots, \overline{g}_{n,h_1})$ and $\overline{g}'_{1:h_2-1} := (\overline{g}_1, \cdots, \overline{g}'_{h_1}, \cdots, \overline{g}_{h_2-1})$.
Now we claim that $\overline{c}_{h_2}$ has non-zero probability under $\overline{g}'_{1:h_2-1}$. Since $\overline{c}_{h_2}$ has non-zero probability under $\overline{g}_{1:h_2-1}$, and $\mathbb{P}(\overline{s}_{h_2}, \overline{p}_{h_2} \backslash \{\overline{a}_{i_1,h_1}\} \cup \{\overline{a}'_{i_1,h_1}\} | \overline{c}_{h_2}, \overline{g}_{1:h_2-1}) \neq 0$, we can get $\mathbb{P}(\overline{a}'_{i_1,h_1}, \overline{c}_{h_2} | \overline{g}_{1:h_2-1}) > 0$. Since $\overline{g}'_{1:h_2-1}$ only differs from $\overline{g}_{1:h_2-1}$ in the strategy of agent (i_1, h_1), and $\overline{g}'_{i_1,h_1}$ always chooses $\overline{a}'_{i_1,h_1}$, we have $\mathbb{P}(\overline{a}'_{i_1,h_1}, \overline{c}_{h_2} | \overline{g}'_{1:h_2-1}) \geq \mathbb{P}(\overline{a}'_{i_1,h_1}, \overline{c}_{h_2} | \overline{g}_{1:h_2-1}) > 0$, and thus $\mathbb{P}(\overline{c}_{h_2} | \overline{g}'_{1:h_2-1}) > 0$. This proves our claim.

Meanwhile, due to (D.3), we have

$$\mathbb{P}(\overline{s}_{h_2}, \overline{p}_{h_2} | \overline{c}_{h_2}, \overline{g}'_{1:h_2-1}) = 0 \neq \mathbb{P}(\overline{s}_{h_2}, \overline{p}_{h_2} | \overline{c}_{h_2}, \overline{g}_{1:h_2-1}), \tag{D.4}$$

which leads to a contradiction to the fact that $\mathcal{D}$ has SI-CIBs.

2. If $\sigma(\overline{a}_{i_1,h_1}) \subseteq \sigma(\overline{\tau}_{i_2,h_2})$, then it implies that $\sigma(\overline{\tau}_{i_1,h_1}) \nsubseteq \sigma(\overline{\tau}_{i_2,h_2})$, and further implies that $\sigma(\overline{\tau}_{i_1,h_1}) \nsubseteq \sigma(\overline{c}_{h_2})$ since $\overline{c}_{h_2} \subseteq \overline{\tau}_{i_2,h_2}$. Note that there must exists strategy $\overline{g}_{1:h_2-1}$ and some realizations $\overline{c}_{h_2} \in \overline{\mathcal{C}}_{h_2}, \overline{\tau}_{i_2,h_2} \in \overline{\mathcal{T}}_{i_2,h_2}$ such that $\overline{\tau}_{i_2,h_2}$ has non-zero probability under $\overline{g}_{1:h_2-1}$ and $\overline{c}_{h_2} \subseteq \overline{\tau}_{i_2,h_2}$, and there exist two different realizations $\overline{\tau}_{i_1,h_1}, \overline{\tau}'_{i_1,h_1} \in \overline{\mathcal{T}}_{i_1,h_1}$ such that $\mathbb{P}(\overline{\tau}_{i_1,h_1} | \overline{\tau}_{i_2,h_2}, \overline{g}_{1:h_2-1}) > 0, \mathbb{P}(\overline{\tau}'_{i_1,h_1} | \overline{\tau}_{i_2,h_2}, \overline{g}_{1:h_2-1}) > 0$, since otherwise, it holds that $\sigma(\overline{\tau}_{i_1,h_1}) \subseteq \sigma(\overline{\tau}_{i_2,h_2})$. Furthermore, we know that there exist realizations of $\overline{s}_{h_2} \in \overline{\mathcal{S}}, \overline{p}_{h_2} \in \overline{\mathcal{P}}_{h_2}$ such that $\mathbb{P}(\overline{s}_{h_2}, \overline{p}_{h_2} | \overline{c}_{h_2}, \overline{g}_{1:h_2-1}) > 0$ [kz:why?] and $\overline{\tau}'_{i_2,h_2} \subseteq \overline{c}_{h_2} \cup \overline{p}_{h_2}$ [kz:why?]. Since $\sigma(\overline{a}_{i_1,h_1}) \subseteq \sigma(\overline{\tau}_{i_2,h_2})$, we know that $\mathbb{P}(\overline{a}_{i_1,h_1} | \overline{\tau}_{i_2,h_2}) = 1$ holds under any strategy. Consider another different action $\overline{a}'_{i_1,h_1} \neq \overline{a}_{i_1,h_1}$, we define a new strategy $\overline{g}'_{i_1,h_1}$ as

$$\overline{g}'_{i_1,h_1}(\overline{\tau}_{i_1,h_1}) = \overline{a}'_{i_1,h_1}, \quad \overline{g}'_{i_1,h_1}(\overline{\tau}'_{i_1,h_1}) = \overline{a}_{i_1,h_1}, \tag{D.5}$$

and keep $g'_{i_1,h_1}(\overline{\tau}''_{i_1,h_1})$ the same as $g_{i_1,h_1}(\overline{\tau}''_{i_1,h_1})$ for any other $\overline{\tau}''_{i_1,h_1}$. We denote $\overline{g}'_{h_1} := (\overline{g}_{1,h_1}, \cdots, \overline{g}'_{i_1,h_1}, \cdots, \overline{g}_{n,h_1})$ and $\overline{g}'_{1:h_2-1} := (\overline{g}_1, \cdots, \overline{g}'_{h_1}, \cdots, \overline{g}_{h_2-1})$. Since under $\overline{g}_{1:h_2-1}$, $(\overline{\tau}'_{i_1,h_1}, \overline{\tau}_{i_2,h_2})$ has non-zero probability and $\mathbb{P}(\overline{a}_{i_1,h_1} | \overline{\tau}_{i_2,h_2}) = 1$ [kz:what? not understand..], then we know $(\overline{\tau}'_{i_1,h_1}, \overline{\tau}_{i_2,h_2})$ has non-zero probability under $\overline{g}'_{1:h_2-1}$ [kz:which strategy is this?.. and why?]. Hence, we know that $\overline{c}_{h_2}$ has non-zero probability under $\overline{g}'_{1:h_2-1}$ [kz:same.. and why..]. Meanwhile, it holds that

$$\begin{aligned} \mathbb{P}(\overline{s}_{h_2}, \overline{p}_{h_2} | \overline{c}_{h_2}, \overline{g}'_{1:h_2-1}) &= \frac{\mathbb{P}(\overline{s}_{h_2}, \overline{p}_{h_2}, \overline{c}_{h_2} | \overline{g}'_{1:h_2-1})}{\mathbb{P}(\overline{c}_{h_2} | \overline{g}'_{1:h_2-1})} \\ &= \frac{\mathbb{P}(\overline{s}_{h_2}, \overline{\tau}_{h_2} | \overline{g}'_{1:h_2-1})}{\mathbb{P}(\overline{c}_{h_2} | \overline{g}'_{1:h_2-1})} = 0 \neq \mathbb{P}(\overline{s}_{h_2}, \overline{p}_{h_2} | \overline{c}_{h_2}, \overline{g}_{1:h_2-1}), \end{aligned} \tag{D.6}$$

where the third equality is because $\overline{a}_{i_1,h_1} \in \overline{\tau}_{h_2}, \overline{\tau}_{i_1,h_1} \subseteq \overline{\tau}_{h_2}$ from perfect recall, but $\overline{a}_{i_1,h_1}, \overline{\tau}_{i_1,h_1}$ can never be realized simultaneously under $\overline{g}'_{1:h_2-1}$ due to (D.5). Therefore, (D.6) leads to a contradiction to the fact that $\mathcal{D}$ has SI-CIBs.

This completes the proof. □

E. Collection of Algorithm Pseudocodes

Here we collect both our planning and learning algorithms as pseudocodes in Algorithms 1, 2, 3, 4, 5, and 6.

Algorithm 1 Planning in QC LTC Problems

Require: LTC $\mathcal{L}$.

1: Reformulate $\mathcal{L}$ to $\mathcal{D}_{\mathcal{L}}$ based on Equation (IV.1)
2: Expand $\mathcal{D}_{\mathcal{L}}$ to $\mathcal{D}^{\dagger}_{\mathcal{L}}$ based on Equation (IV.2)
3: Refine $\mathcal{D}^{\dagger}_{\mathcal{L}}$ to $\mathcal{D}'_{\mathcal{L}}$ based on $\mathcal{L}$ and §IV-C
4: Construct an expected approximate common-information model $\mathcal{M}$ from $\mathcal{D}'_{\mathcal{L}}$ (see examples of such constructions in the proof of Theorem IV.8)
5: $\widehat{g}^{*}_{1:\widetilde{H}} \leftarrow$ Algorithm 6($\mathcal{M}$)
6: $\widetilde{g}^{*}_{1:\widetilde{H}} \leftarrow \varphi(\widehat{g}^{*}_{1:\widetilde{H}}, \mathcal{D}_{\mathcal{L}})$
7: $g^{m,*}_{1:H} \leftarrow \{\widetilde{g}^{*}_{1}, \widetilde{g}^{*}_{3}, \cdots, \widetilde{g}^{*}_{2H-1}\}$
8: $g^{a,*}_{1:H} \leftarrow \{\widetilde{g}^{*}_{2}, \widetilde{g}^{*}_{4}, \cdots, \widetilde{g}^{*}_{2H}\}$
9: **return** $(g^{m,*}_{1:H}, g^{a,*}_{1:H})$

F. Decentralized POMDPs (with Information Sharing)

A Dec-POMDP with n agents and potential information sharing can be characterized by a tuple

$$\mathcal{D} = \langle H, \mathcal{S}, \{\mathcal{A}_{i,h}\}_{i\in[n],h\in[H]}, \{\mathcal{O}_{i,h}\}_{i\in[n],h\in[H]}, \{\mathbb{T}_h\}_{h\in[H]}, \{\mathbb{O}_h\}_{h\in[H]}, \mu_1, \{\mathcal{R}_h\}_{h\in[H]}\rangle,$$

where H denotes the length of each episode, $\mathcal{S}$ denotes the state space, and $\mathcal{A}_{i,h}$ denotes the *control action* space of agent i at timestep h. We denote by $s_h \in \mathcal{S}$ the state and by $a_{i,h}$ the control action of agent i at timestep h. We use $a_h := (a_{1,h}, \cdots, a_{n,h}) \in \mathcal{A}_h := \mathcal{A}_{1,h} \times \mathcal{A}_{2,h} \times \cdots \mathcal{A}_{n,h}$ to denote the joint control action for all the n agents at timestep h, with $\mathcal{A}_h$ denoting the joint control action space at timestep h. We denote $\mathbb{T} = \{\mathbb{T}_h\}_{h\in[H]}$ the collection of transition functions, where $\mathbb{T}_h(\cdot|s_h, a_h) \in \Delta(\mathcal{S})$ gives the transition probability to the next state s_{h+1} when taking the joint control action a_h at state s_h. We use $\mu_1 \in \Delta(\mathcal{S})$ to denote the distribution of the initial state s_1. We denote by $\mathcal{O}_{i,h}$ the observation space and by $o_{i,h} \in \mathcal{O}_{i,h}$ the observation of agent i at timestep h. We use $o_h := (o_{1,h}, o_{2,h}, \cdots, o_{n,h}) \in \mathcal{O}_h := \mathcal{O}_{1,h} \times \mathcal{O}_{2,h} \times \cdots \mathcal{O}_{n,h}$ to denote the joint observation of all the n agents at timestep h, with $\mathcal{O}_h$ denoting the joint observation space at timestep h. We use $\{\mathbb{O}_h\}_{h\in[H]}$ to denote the collection of emission matrices, where $o_h \sim \mathbb{O}_h(\cdot|s_h) \in \Delta(\mathcal{O}_h)$ at timestep h under state $s_h \in \mathcal{S}$. For notational convenience, we adopt the matrix convention, where $\mathbb{O}_h$ is a matrix with each row $\mathbb{O}_h(\cdot|s_h)$ for all $s_h \in \mathcal{S}$. Also, we denote by $\mathbb{O}_{i,h}$ the marginalized emission for agent i at timestep h. Finally, $\{\mathcal{R}_h\}_{h\in[H]}$ is a collection of reward functions among all agents, where $\mathcal{R}_h : \mathcal{S} \times \mathcal{A}_h \to [0, 1]$.

At timestep h, each agent i in the Dec-POMDP has access to some information $\tau_{i,h}$, a subset of historical joint observations and actions, namely, $\tau_{i,h} \subseteq \{o_1, a_1, o_2, \cdots, a_{h-1}, o_h\}$, and the collection of all possible such available information is denoted by $\mathcal{T}_{i,h}$. We use τ_h to denote the *joint* available information at timestep h. Meanwhile, agents may *share* part of the history with each other. The *common information* $c_h = \cup_{t=1}^{h} z_t$ at timestep h is thus a subset of the joint history τ_h, where z_h is the information shared at timestep h. We use $\mathcal{C}_h$ to denote the collection of all possible c_h at timestep h, and use $\mathcal{T}_{i,h}$ to denote the collection of all possible $\tau_{i,h}$ of agent i at timestep h. Besides the common information c_h, each agent also has her *private information* $p_{i,h} = \tau_{i,h} \backslash c_h$, where the collection of $p_{i,h}$

Algorithm 2 Learning in QC LTC Problems

Require: Underlying environment LTC $\mathcal{L}$, compression functions $\{\text{Compress}_h\}_{h\in[\overline{H}]}$ and rules $\{\widehat{\phi}_h\}_{h\in[\overline{H}]}$, length $\widehat{L}$, accuracy level ϵ, probability δ, constant C.

Reformulate $\mathcal{L}$ to $\mathcal{D}_{\mathcal{L}}$ based on Equation (IV.1)
Expand $\mathcal{D}_{\mathcal{L}}$ to $\mathcal{D}^{\dagger}_{\mathcal{L}}$ based on Equation (IV.2)
Refine $\mathcal{D}^{\dagger}_{\mathcal{L}}$ to $\mathcal{D}'_{\mathcal{L}}$ based on $\mathcal{L}$ and §IV-C
Construct $\{\widehat{\mathcal{C}}_h\}_{h\in[\overline{H}]}$ as $\forall h\in[\overline{H}], \widehat{\mathcal{C}}_h = \{\text{Compress}_h(\overline{c}_h)\,|\,\overline{c}_h\in\overline{\mathcal{C}}_h\}$
Denote $S = |\overline{\mathcal{S}}|, A = \max_{h\in[\overline{H}]}|\overline{\mathcal{A}}_h|, O = \max_{h\in[\overline{H}]}|\overline{\mathcal{O}}_h|, P = \max_{h\in[\overline{H}]}|\overline{\mathcal{P}}_h|, \widehat{C} = \max_{h\in[\overline{H}]}|\widehat{\mathcal{C}}_h|$, and recall γ is the parameter in Assumption III.1
Define $K = 2\overline{H}S, \alpha = \frac{C\overline{H}^2\epsilon}{200(\overline{H}+1)^2}, \epsilon_1 = \frac{\epsilon}{200(\overline{H}+1)^2}, \theta_1 = \frac{\epsilon}{200(\overline{H}+1)^2O}, \theta_2 = \frac{\epsilon}{200(\overline{H}+1)^2AP}, \phi = \frac{\epsilon_1\gamma^2}{C^2\overline{H}^8S^5O^4}, \zeta_1 = \min\left\{\frac{\epsilon\phi}{200(\overline{H}+1)^2A^{2\widehat{L}}O^{\widehat{L}}}, \frac{\epsilon}{400(\overline{H}+1)^2AP}\right\}, \zeta_2 = \zeta_1^2, \beta = \frac{\delta}{3}, N_0 = \lceil\max\left\{\frac{C(P+\log\frac{4\overline{H}\widehat{C}}{\delta_1}}{\zeta_1\theta_1^2}, \frac{CA(O+\log\frac{4\overline{H}\widehat{C}PA}{\delta_1})}{\zeta_2\theta_2^2}\right\}\rceil, N_1 = \lceil(AO)^{\widehat{L}}\log(\frac{1}{\delta_2})\rceil, N_2 = \lceil C\frac{\overline{H}^2\log\frac{K^2n}{\delta_3}}{\epsilon_1^2}\rceil$
Define $\widehat{\mathcal{M}} := \{\widehat{\mathcal{M}}(\overline{g}^{1:\overline{H},j})\}_{j=1}^{K}$
$\{\overline{g}^{1:\overline{H},j}\}_{j=1}^{K} \leftarrow$ BaSeCAMP$(\widehat{L}, N_0, N_1, \alpha, \beta, K)$ by calling Algorithm 3 of [34] under $\mathcal{D}'_{\mathcal{L}}$
for $j = 1$ to K **do**
    $\widehat{\mathcal{M}}(\overline{g}^{1:\overline{H},j}) \leftarrow$ LEE$(\overline{g}^{1:\overline{H},j}, \{\widehat{\mathcal{C}}_h\}_{h\in[\overline{H}]}, \{\widehat{\phi}_h\}_{h\in[\overline{H}]}, \Gamma, \zeta_1, \zeta_2, \theta_1, \theta_2, \beta)$ by calling Algorithm 5 of [14] under $\mathcal{D}'_{\mathcal{L}}$
    $\overline{g}^{j,*}_{1:\overline{H}} \leftarrow$ Algorithm 6$(\widehat{\mathcal{M}}(\overline{g}^{1:\overline{H},j}))$
end for
$\overline{g}^{*}_{1:\overline{H}} \leftarrow$ Pos-Dec$(\{\overline{g}^{j,*}_{1:\overline{H}}\}_{j=1}^{K}, N_2)$ by calling Algorithm 8 of [14] under $\mathcal{D}'_{\mathcal{L}}$
$\widetilde{g}^{*}_{1:\widetilde{H}} \leftarrow \varphi(\overline{g}^{*}_{1:\overline{H}}, \mathcal{D}_{\mathcal{L}})$
$g^{m,*}_{1:H} \leftarrow \{\widetilde{g}^{*}_1, \widetilde{g}^{*}_3, \cdots, \widetilde{g}^{*}_{2H-1}\}$
$g^{a,*}_{1:H} \leftarrow \{\widetilde{g}_2, \widetilde{g}_4, \cdots, \widetilde{g}_{2H}\}$
return $(g^{m,*}_{1:H}, g^{a,*}_{1:H})$

Algorithm 3 Vanilla Realization of $\varphi(\breve{g}_{1:\breve{H}}, \mathcal{D}_{\mathcal{L}})$

Require: Strategy $\breve{g}_{1:\breve{H}}$, QC Dec-POMDP $\mathcal{D}_{\mathcal{L}}$

$\widetilde{g}_{1:\breve{H}} \leftarrow \emptyset$
for $h_2 = 1$ to $\breve{H}$, $i_2 = 1$ to n, $\widetilde{\tau}_{i_2,h_2} \in \widetilde{\mathcal{T}}_{i_2,h_2}$ **do**
    $\breve{\tau}_{i_2,h_2} \leftarrow \widetilde{\tau}_{i_2,h_2}$
    for $h_1 = 1$ to $h_2 - 1$, $i_1 = 1$ to n **do**
        if $\sigma(\widetilde{\tau}_{i_1,h_1}) \subseteq \sigma(\widetilde{c}_{h_2})^3$ in $\mathcal{D}_{\mathcal{L}}$ **then**
            Obtain the value of $\widetilde{\tau}_{i_1,h_1}$ from that of $\widetilde{c}_{h_2}$ (based on $\widetilde{\tau}_{i_2,h_2}$)
            $\widetilde{a}_{i_1,h_1} \leftarrow \widetilde{g}_{i_1,h_1}(\widetilde{\tau}_{i_1,h_1})$
            $\breve{\tau}_{i_2,h_2} \leftarrow \breve{\tau}_{i_2,h_2} \cup \{\widetilde{a}_{i_1,h_1}\}$
        end if
    end for
    $\widetilde{g}_{i_2,h_2}(\widetilde{\tau}_{i_2,h_2}) \leftarrow \breve{g}_{i_2,h_2}(\breve{\tau}_{i_2,h_2})$
end for
return $\widetilde{g}_{1:\widetilde{H}}$

Algorithm 4 Efficient Implementation of $\varphi(\breve{g}_{1:\breve{H}}, \mathcal{D}_{\mathcal{L}})$

Require: Strategy $\breve{g}_{1:\breve{H}}$, QC Dec-POMDP $\mathcal{D}_{\mathcal{L}}$

for $h = 1$ to $\breve{H}$ **do**
  for $i = 1$ to n **do**
    Agent i receives $\widetilde{\tau}_{i,h}$
    $\breve{\tau}_{i,h} \leftarrow \text{Recover}(\widetilde{\tau}_{i,h}, \breve{g}_{1:h-1}, \mathcal{D}_{\mathcal{L}})$ \\ Recursion of Algorithm 5
    Agent i takes action $\widetilde{a}_{i,h} \leftarrow \breve{g}_{i,h}(\breve{\tau}_{i,h})$
  end for
end for

Algorithm 5 Recover$(\widetilde{\tau}_{i,h}, \breve{g}_{1:h-1}, \mathcal{D}_{\mathcal{L}})$

Require: Information $\widetilde{\tau}_{i,h}$, Strategy $\breve{g}_{1:h-1}$, QC Dec-POMDP $\mathcal{D}_{\mathcal{L}}$

$\breve{\tau}_{i,h} \leftarrow \widetilde{\tau}_{i,h}$
for $j = 1$ to n, $h' = 1$ to $h-1$ **do**
  if $\sigma(\widetilde{\tau}_{j,h'}) \subseteq \sigma(\widetilde{c}_h)$ in $\mathcal{D}_{\mathcal{L}}$ and $\widetilde{a}_{j,h'} \notin \widetilde{\tau}_{i,h}$ **then**
    Obtain the value of $\widetilde{\tau}_{j,h'}$ from that of $\widetilde{c}_h$ (based on $\widetilde{\tau}_{i,h}$)
    $\breve{\tau}_{j,h'} \leftarrow \text{Recover}(\widetilde{\tau}_{j,h'}, \breve{g}_{1:h'-1}, \mathcal{D}_{\mathcal{L}})$
    $\widetilde{a}_{j,h'} \leftarrow \breve{g}_{j,h'}(\breve{\tau}_{j,h'})$
    $\breve{\tau}_{i,h} \leftarrow \breve{\tau}_{i,h} \cup \{\widetilde{a}_{j,h'}\}$
  end if
end for
return $\breve{\tau}_{i,h}$

Algorithm 6 Planning in Dec-POMDPs with expected Approximate Common-information Model

Require: Expected approximate common-information model $\mathcal{M}$

for $i \in [n]$ and $\widehat{c}_{\overline{H}+1} \in \widehat{\mathcal{C}}_{\overline{H}+1}$ **do**
  $V^{*,\mathcal{M}}_{i,\overline{H}+1}(\widehat{c}_{\overline{H}+1}) \leftarrow 0$
end for
for $h = \overline{H}$ to 1 **do**
  for $\widehat{c}_h \in \widehat{\mathcal{C}}_h$ **do**
    Define $Q^{*,\mathcal{M}}_h(\widehat{c}_h, \gamma_{1,h}, \cdots, \gamma_{n,h}) := \widehat{\mathcal{R}}^{\mathcal{M}}_h(\widehat{c}_h, \gamma_h) + \mathbb{E}^{\mathcal{M}}\left[V^{*,\mathcal{M}}_{h+1}(\widehat{c}_{h+1}) | \widehat{c}_h, \gamma_h\right]$

$$\left(\widehat{g}^*_{1,h}(\cdot|\widehat{c}_h, \cdot), \cdots, \widehat{g}^*_{n,h}(\cdot|\widehat{c}_h, \cdot)\right) \leftarrow \underset{\gamma_{1:n,h} \in \Gamma_h}{\operatorname{argmax}} Q^{*,\mathcal{M}}_h(\widehat{c}_h, \gamma_{1,h}, \cdots, \gamma_{n,h}) \tag{E.1}$$

  end for
  $V^{*,\mathcal{M}}_h(\widehat{c}_h) \leftarrow \max_{\gamma_{1:n,h}} Q^{*,\mathcal{M}}_h(\widehat{c}_h, \gamma_{1,h}, \cdots, \gamma_{n,h})$
end for
return $\widehat{g}^*_{1:\overline{H}}$

is denoted by $\mathcal{P}_{i,h}$. We also denote by p_h the *joint* private information, and by $\mathcal{P}_h$ the collection of all possible p_h at timestep h. We refer to the above the *state-space model* of the Dec-POMDP (with information sharing).

Each agent i at timestep h chooses the control action $a_{i,h}$ based on some strategy $g_{i,h} : \mathcal{T}_{i,h} \to \mathcal{A}_{i,h}$. We denote by $g_h := (g_{1,h}, g_{2,h}, \cdots, g_{n,h})$ the joint control strategy of all the agents, and by $g_{1:h} := (g_1, g_2, \cdots, g_h), \forall h \in [H]$ the sequence of joint strategies from timestep 1 to h. We use $\mathcal{G}_{i,h}$ to denote the strategy space of $g_{i,h}$, and use $\mathcal{G}_h, \mathcal{G}_{1:h}$ to denote joint strategy spaces, correspondingly.

Next, we introduce some background on the intrinsic model and information structure of Dec-POMDPs.

F-A Intrinsic Model

In an intrinsic model [25], we regard the agent i at different timesteps as *different agents*, and each agent only acts *once* throughout. Any Dec-POMDP $\mathcal{D}$ with n agents can be formulated within the intrinsic-model framework, and can be characterized by a tuple $\langle (\Omega, \mathscr{F}), N, \{(\mathbb{U}_l, \mathscr{U}_l)\}_{l=1}^N, \{(\mathbb{I}_l, \mathscr{I}_l)\}_{l=1}^N \rangle$ [11], where $(\Omega, \mathcal{F})$ is a measurable space of the environment, $N = n \times H$ is the number of agents in the intrinsic model. By a slight abuse of notation, we write $[N] := [n] \times [H]$, and write $l := (i,h) \in [N]$ for notational convenience. This way, any agent $l \in [N]$ corresponds to an agent $i \in [n]$ at timestep $h \in [H]$ in the state-space model. We denote by $\mathbb{U}_l$ the measurable action space of agent l and by $\mathscr{U}_l$ the σ-algebra over $\mathbb{U}_l$. For $A \subseteq [N]$, let $\mathbb{H}_A := \Omega \times \prod_{l \in A} \mathbb{U}_l$ and $\mathbb{H} := \mathbb{H}_{[N]}$. For any σ-algebra $\mathscr{C}$ over $\mathbb{H}_A$, let $\langle \mathscr{C} \rangle$ denote the cylindrical extension of $\mathscr{C}$ on $\mathbb{H}$. Let $\mathscr{H}_A := \langle \mathscr{F} \otimes (\otimes_{l \in A} \mathscr{U}_l) \rangle$ and $\mathscr{H} = \mathscr{H}_{[N]}$. We denote by $\mathbb{I}_l$ the space of *information available* to agent l, and by $\mathscr{I}_l$ the σ-algebra over $\mathbb{H}$. For $l \in [N]$, we denote by I_l the information of agent l, and U_l the action of agent l. The spaces and random variables of agent $l = (i,h)$ in the intrinsic model are related to those in the state-space model as follows: $\forall l = (i,h) \in [N], \mathbb{U}_l = \mathcal{A}_{i,h}, \mathbb{I}_l = \mathcal{T}_{i,h}, U_l = a_{i,h}, I_l = \tau_{i,h}$.

F-B Information Structures of Dec-POMDPs

An important class of IS is the *quasi-classical* one, which is defined as follows [25, 11, 12].

Definition F.1 (Quasi-classical Dec-POMDPs)**.** We call a Dec-POMDP problem *QC* if each agent in the intrinsic model knows the information available to the agents who influence her, directly or indirectly, i.e., $\forall l_1, l_2 \in [N], l_1 = (i_1, h_1), l_2 = (i_2, h_2), i_1, i_2 \in [n], h_1, h_2 \in [H]$, if agent l_1 influences agent l_2, then $\mathscr{I}_{l_1} \subseteq \mathscr{I}_{l_2}$.

Furthermore, *strictly* quasi-classical IS [25, 26], as a subclass of QC IS, is defined as follows.

Definition F.2 (Strictly quasi-classical Dec-POMDPs)**.** We call a Dec-POMDP problem *sQC* if each agent in the intrinsic model knows the information *and* actions available to the agents who influence her, directly or indirectly. That is, $\forall l_1, l_2 \in [N], l_1 = (i_1, h_1), l_2 = (i_2, h_2), i_1, i_2 \in [n], h_1, h_2 \in [H]$, if agent l_1 influences agent l_2, then $\mathscr{I}_{l_1} \cup \langle \mathscr{U}_{l_1} \rangle \subseteq \mathscr{I}_{l_2}$.

F-C Intrinsic Model of LTC Problems

Given any LTC $\mathcal{L}$ of the state-space-model form defined in §II-A, we define the intrinsic model of $\mathcal{L}$ as a tuple $\langle (\Omega, \mathscr{F}), N, \{(\mathbb{U}_l, \mathscr{U}_l)\}_{l=1}^N, \{(\mathbb{M}_l, \mathscr{M}_l)\}_{l=1}^N, \{(\mathbb{I}_{l^-}, \mathscr{I}_{l^-})\}_{l=1}^N, \{(\mathbb{I}_{l^+}, \mathscr{I}_{l^+})\}_{l=1}^N \rangle$, where $(\Omega, \mathscr{F})$ is the measure space representing all the uncertainty in the system; $N = n \times H$ is the number of agents in the intrinsic model. By a slight abuse of notation, we write $[N] := [n] \times [H]$, and write $l := (i,h) \in [N]$ for convenience. This way, any agent $l \in [N]$ corresponds to an agent $i \in [n]$ at timestep $h \in [H]$ in the state-space model, and we thus define $l^- := (i, h^-)$ and $l^+ := (i, h^+)$ accordingly. We denote

by $\mathbb{U}_l$ and $\mathbb{M}_l$ the measurable control and communication action spaces of agent l, and by $\mathscr{U}_l$ and $\mathscr{M}_l$ the σ-algebra over $\mathbb{U}_l$ and $\mathbb{M}_l$, respectively. For any $A \subseteq [N]$, let $\mathbb{H}_A := \Omega \times \prod_{l\in A}(\mathbb{U}_l \times \mathbb{M}_l)$ and $\mathbb{H} := \mathbb{H}_{[N]}$. For any σ-algebra $\mathscr{C}$ over $\mathbb{H}_A$, let $\langle \mathscr{C} \rangle$ denote the cylindrical extension of $\mathscr{C}$ on $\mathbb{H}$. Let $\mathscr{H}_A := \langle \mathscr{F} \otimes (\otimes_{l\in A}\mathscr{U}_l) \otimes (\otimes_{l\in A}\mathscr{M}_l) \rangle$, $\mathscr{H} = \mathscr{H}_{[N]}$. We denote by $\mathbb{I}_{l^-}$ and $\mathbb{I}_{l^+}$ the spaces of *information available* to agent l *before* and *after* additional sharing, respectively, and by $\mathscr{I}_{l^-} \subseteq \mathscr{H}$ and $\mathscr{I}_{l^+} \subseteq \mathscr{H}$ the associated σ-algebra. The spaces and random variables of agent $l = (i,h)$ in the intrinsic model are related to those in the state-space model as follows: $\forall l = (i,h) \in [N], \mathbb{U}_l = \mathcal{A}_{i,h}, \mathbb{M}_l = \mathcal{M}_{i,h}, \mathbb{I}_{l^-} = \mathcal{T}_{i,h^-}, \mathbb{I}_{l^+} = \mathcal{T}_{i,h^+}, U_l = a_{i,h}, M_l = m_{i,h}, I_{l^-} = \tau_{i,h^-}, I_{l^+} = \tau_{i,h^+}$. For notational convenience, for any random variable B in LTC and the σ-algebra $\mathscr{B}$ generated by B, we overload $\sigma(B)$ to denote the cylindrical extension of $\mathscr{B}$ on $\mathbb{H}$, i.e., $\sigma(B) = \langle \mathscr{B} \rangle$.

G. Other Supplementary Results

G-A Optimality of Deterministic Strategies

We now show a supplementary result that for the formulated LTC problem, it does not lose optimality to consider *deterministic* strategies as introduced in §II. For any LTC problem $\mathcal{L}$, consider generic, stochastic communication and control strategy spaces: $\forall i \in [n],\ h \in [H],\ \mathcal{G}^{m,S}_{i,h} := \{g^{m,S}_{i,h} : \Omega^m_{i,h} \times \mathcal{T}_{i,h^-} \to \mathcal{M}_{i,h}\},\ \mathcal{G}^{a,S}_{i,h} := \{g^{a,S}_{i,h} : \Omega^a_{i,h} \times \mathcal{T}_{i,h^+} \to \mathcal{A}_{i,h}\}$, where $\{\Omega^m_{i,h}\}_{i\in[n],h\in[H]}, \{\Omega^a_{i,h}\}_{i\in[n],h\in[H]}$ are the sets of random seeds which could be correlated to each other across agents and timesteps. Note that these strategy classes include those of the strategies randomized over the action sets, i.e., $\forall i \in [n],\ h \in [H],\ \mathcal{G}^{m,S}_{i,h} := \{g^{m,S}_{i,h} : \mathcal{T}_{i,h^-} \to \Delta(\mathcal{M}_{i,h})\},\ \mathcal{G}^{a,S}_{i,h} := \{g^{a,S}_{i,h} : \mathcal{T}_{i,h^+} \to \Delta(\mathcal{A}_{i,h})\}$. Also, we denote by $\mathcal{G}^{a,S}_h, \mathcal{G}^{m,S}_h$ the *joint* stochastic control and communication spaces at timestep h, respectively. Similarly, we define the objective under the stochastic strategies as

$$\forall g^{a,S}_{1:H} \in \mathcal{G}^{a,S}_{1:H}, g^{m,S}_{1:H} \in \mathcal{G}^{m,S}_{1:H}, \qquad J_{\mathcal{L}}(g^{a,S}_{1:H}, g^{m,S}_{1:H}) := \mathbb{E}_{\mathcal{L}}\left[\sum_{h=1}^{H}(r_h - \kappa_h)\,\middle|\, g^{a,S}_{1:H}, g^{m,S}_{1:H}\right].$$

Lemma G.1. It does not lose optimality to consider deterministic control and communication strategies in LTC. Namely, for any LTC problem $\mathcal{L}$,

$$\max_{g^{a,S}_{1:H}\in\mathcal{G}^{a,S}_{1:H}, g^{m,S}_{1:H}\in\mathcal{G}^{m,S}_{1:H}} J_{\mathcal{L}}(g^{a,S}_{1:H}, g^{m,S}_{1:H}) = \max_{g^{a}_{1:H}\in\mathcal{G}^{a}_{1:H}, g^{m}_{1:H}\in\mathcal{G}^{m}_{1:H}} J_{\mathcal{L}}(g^{a}_{1:H}, g^{m}_{1:H}). \tag{G.1}$$

Proof. For any $i \in [n], h \in [H]$, since space $\mathcal{G}^{m,S}_{i,h}$ covers space $\mathcal{G}^{m}_{i,h}$ and space $\mathcal{G}^{a,S}_{i,h}$ covers space $\mathcal{G}^{a}_{i,h}$, we have that

$$\max_{g^{a,S}_{1:H}\in\mathcal{G}^{a,S}_{1:H}, g^{m,S}_{1:H}\in\mathcal{G}^{m,S}_{1:H}} J_{\mathcal{L}}(g^{a,S}_{1:H}, g^{m,S}_{1:H}) \geq \max_{g^{a}_{1:H}\in\mathcal{G}^{a}_{1:H}, g^{m}_{1:H}\in\mathcal{G}^{m}_{1:H}} J_{\mathcal{L}}(g^{a}_{1:H}, g^{m}_{1:H}). \tag{G.2}$$

In the other direction, from the tower property, for any $g^{a,S}_{1:H} \in \mathcal{G}^{a,S}_{1:H}, g^{m,S}_{1:H} \in \mathcal{G}^{m,S}_{1:H}$

$$
\begin{aligned}
J_{\mathcal{L}}(g^{a,S}_{1:H}, g^{m,S}_{1:H}) &= \mathbb{E}_{\mathcal{L}}\left[\sum_{h=1}^{H}(r_h - \kappa_h)\,\middle|\, g^{a,S}_{1:H}, g^{m,S}_{1:H}\right] \\
&= \mathbb{E}\left[\mathbb{E}_{\mathcal{L}}\Big[\sum_{h=1}^{H}(r_h - \kappa_h)\,\Big|\, g^{a,S}_{1:H}, g^{m,S}_{1:H}, \{\omega^a_{i,h}\}_{i\in[n],h\in[H]}, \{\omega^m_{i,h}\}_{i\in[n],h\in[H]}\Big]\right] \\
&= \mathbb{E}\left[\mathbb{E}_{\mathcal{L}}\Big[\sum_{h=1}^{H}(r_h - \kappa_h)\,\Big|\, \{g^{a,S}_{i,h}[\omega^a_{i,h}]\}_{i\in[n],h\in[H]}, \{g^{m,S}_{i,h}[\omega^m_{i,h}]\}_{i\in[n],h\in[H]}\Big]\right],
\end{aligned}
$$

[kz:all the notation "superscript" should be a,S and m,S, right? there shouldnt be any a,m? check all..] where $\omega^a_{i,h} \in \Omega^a_{i,h}$, $\omega^m_{i,h} \in \Omega^m_{i,h}$ are random seeds, and $g^{a,S}_{i,h}[\omega^a_{i,h}] \in \mathcal{G}^a_{i,h}, g^{a,S}_{i,h}[\omega^m_{i,h}] \in \mathcal{G}^m_{i,h}$[kz:same here..] are deterministic strategies defined [kz:note the "colored" ":=" below] as

$$
\begin{aligned}
&\forall \tau_{i,h^-} \in \mathcal{T}_{i,h^-}, g^{m,S}_{i,h}[\omega^m_{i,h}](\tau_{i,h^-}) := g^{m,S}_{i,h}(\omega^m_{i,h}, \tau_{i,h^-}), \\
&\forall \tau_{i,h^+} \in \mathcal{T}_{i,h^+}, g^{a,S}_{i,h}[\omega^a_{i,h}](\tau_{i,h^+}) := g^{a,S}_{i,h}(\omega^a_{i,h}, \tau_{i,h^+}).
\end{aligned}
$$

Therefore, [kz:same change below as blue above. pls.][kz:no "a,m" below??]

$$
\begin{aligned}
J_{\mathcal{L}}(g^{a,S}_{1:H}, g^{m,S}_{1:H}) &= \mathbb{E}\left[\mathbb{E}_{\mathcal{L}}\Big[\sum_{h=1}^{H}(r_h - \kappa_h)\,\Big|\, \{g^{a,S}_{i,h}[\omega^a_{i,h}]\}, \{g^{m,S}_{i,h}[\omega^m_{i,h}]\}\Big]\right] \\
&\le \max_{\{\omega^a_{i,h}\}_{i\in[n],h\in[H]}, \{\omega^m_{i,h}\}_{i\in[n],h\in[H]}} \mathbb{E}_{\mathcal{L}}\left[\sum_{h=1}^{H}(r_h - \kappa_h)\,\middle|\, \{g^{a,S}_{i,h}[\omega^a_{i,h}]\}, \{g^{m,S}_{i,h}[\omega^m_{i,h}]\}\right] \\
&\le \max_{g^a_{1:H}\in\mathcal{G}^a_{1:H}, g^m_{1:H}\in\mathcal{G}^m_{1:H}} J_{\mathcal{L}}(g^a_{1:H}, g^m_{1:H})
\end{aligned}
$$

holds for any $g^{a,S}_{1:H} \in \mathcal{G}^{a,S}_{1:H}, g^{m,S}_{1:H} \in \mathcal{G}^{m,S}_{1:H}$. Hence, we further get

$$
\max_{g^{a,S}_{1:H}\in\mathcal{G}^{a,S}_{1:H}, g^{m,S}_{1:H}\in\mathcal{G}^{m,S}_{1:H}} J_{\mathcal{L}}(g^{a,S}_{1:H}, g^{m,S}_{1:H}) \le \max_{g^a_{1:H}\in\mathcal{G}^a_{1:H}, g^m_{1:H}\in\mathcal{G}^m_{1:H}} J_{\mathcal{L}}(g^a_{1:H}, g^m_{1:H}) \tag{G.3}
$$

Combining Equation (G.2) and Equation (G.3), we proved Equation (G.1). □

G-B Conditions Leading to Assumption IV.7

As a minimal requirement for computational tractability (for both Dec-POMDPs and LTCs), Assumption IV.7 is needed for the one-step tractability of the team-decision problem involved in the value iteration in Algorithm 6. We now adapt several such structural conditions from [14] to the LTC setting, which lead to this assumption and have been studied in the literature. Note that since we need to do planning in the approximate model $\mathcal{M}$, which is oftentimes constructed based on the original problem $\mathcal{L}$ and the approximate beliefs $\{\mathbb{P}^{\mathcal{M},c}_h(\overline{s}_h, \overline{p}_h | \widehat{c}_h)\}_{h\in[\overline{H}]}$, we necessarily need conditions on these two models $\mathcal{L}$ and $\mathcal{M}$ to ensure Assumption IV.7 holds, for which we refer to as the **Part (1)** and **Part (2)** of the conditions below, respectively.

- **Turn-based structures. Part (1):** At each timestep $h \in [\overline{H}]$, there is only one agent, denoted as $ct(h) \in [n]$, that can affect the state transition. More concretely, the transition dynamics take the forms of $\mathbb{T}_h : \mathcal{S}\times\mathcal{A}_{ct(h)} \to \Delta(\mathcal{S})$. Additionally, we assume the reward function admits an additive

structure such that $\mathcal{R}_h(s_h, a_h) = \sum_{i\in[n]} \mathcal{R}_{i,h}(s_h, a_{i,h})$ for some functions $\{\mathcal{R}_{i,h}\}_{i\in[n]}$. Meanwhile, since only agent $ct(h)$ takes the action, we assume the increment of the common information satisfies $z^b_{h+1} = \chi_{h+1}(p_{h^+}, a_{ct(h),h}, o_{h+1})$. **Part (2):** No additional requirement. Such a structure has been commonly studied in (fully observable) stochastic games and multi-agent RL [35, 36].

- **Nested private information. Part (1):** No additional requirement. **Part (2):** At each timestep $h = 2t, t \in [H]$, all the agents form a *hierarchy* according to the private information they possess, in the sense that $\forall\ i, j \in [n], j < i, \overline{p}_{j,h} = Y^{i,j}_h(\overline{p}_{i,h})$ for some function $Y^{i,j}_h$. More formally, the approximate belief satisfies that $\mathbb{P}^{\mathcal{M},c}_h(\overline{p}_{j,h} = Y^{i,j}_h(\overline{p}_{i,h}) | \overline{p}_{i,h}, \widehat{c}_h) = 1$. Such a structure has been investigated in [37] with heuristic search, and in [14] with finite-time complexity analysis when there is no additional sharing to decide/learn.

- **Factorized structures. Part (1):** At each timestep $h \in [\overline{H}]$, the state s_h can be partitioned into n local states, i.e., $s_h = (s_{1,h}, s_{2,h}, \cdots, s_{n,h})$. Meanwhile, the transition kernel takes the product form of $\mathbb{T}_h(s_{h+1} | s_h, a_h) = \prod_{i=1}^n \mathbb{T}_{i,h}(s_{i,h+1} | s_{i,h}, a_{i,h})$, the emission also takes the product form of $\mathbb{O}_h(o_h | s_h) = \prod_{i=1}^n \mathbb{O}_{i,h}(o_{i,h} | s_{i,h})$, and the communication cost and reward functions can be decoupled into n terms such that $\mathcal{K}_h(z^a_h) = \sum_{i=1}^m \mathcal{K}_{i,h}(z^a_{i,h})$. **Part (2):** At each timestep $h \in [\overline{H}]$, the approximate common information is also factorized so that $\widehat{c}_h = (\widehat{c}_{1,h}, \widehat{c}_{2,h}, \cdots, \widehat{c}_{n,h})$ and its evolution satisfies that $\widehat{c}_{i,h} = \widehat{\phi}_{i,h}(\widehat{c}_{i,h-1}, \overline{z}_{i,h-1})$ for some function $\widehat{\phi}_{i,h}$. Correspondingly, the approximate beliefs need to satisfy that $\mathbb{P}^{\mathcal{M},c}_h(\overline{s}_h, \overline{p}_h | \widehat{c}_h) = \Pi^n_{i=1} \mathbb{P}^{\mathcal{M},c}_{i,h}(\overline{s}_{i,h}, \overline{p}_{i,h} | \widehat{c}_{i,h})$ for some functions $\{\mathbb{P}^{\mathcal{M},c}_{i,h}\}_{i\in[n],h\in[\overline{H}]}$. Such a structure, under general information sharing protocols, can lead to non-classical IS. In this case, it may be viewed as an example of non-classical ISs where the agents have no incentive for signaling [12, §3.8.3].

Lemma G.2. Given any LTC problem $\mathcal{L}$, let $\mathcal{D}'_{\mathcal{L}}$ be the Dec-POMDP after reformulation, expansion, and refinement. For any $\mathcal{M}$ to be the approximate model of $\mathcal{D}'_{\mathcal{L}}$ and $\{\mathbb{P}^{\mathcal{M},c}_h\}_{h\in[\overline{H}]}$ to be the approximate belief, if they satisfy any of the 3 conditions above, then Equation (E.1) in Algorithm 6 can be solved with time complexity $\max_{h\in[\overline{H}]} \text{poly}\,(|\overline{\mathcal{P}}_h|, |\overline{\mathcal{A}}_h|, |\overline{\mathcal{S}}|)$.

Proof. For any $h \in [\overline{H}]$, if $h = 2t-1, t \in [H]$, from Assumption III.4 and the construction of $\mathcal{D}'_{\mathcal{L}}$, since we need to find the optimal strategy of $\mathcal{D}'_{\mathcal{L}}$ in the spaces $\overline{g}_h \in \overline{\mathcal{G}}_h = \{g : \overline{\mathcal{C}}_h \to \overline{\mathcal{A}}_h\}$, where we recall that $\overline{\mathcal{A}}_h = \mathcal{M}_t$ is joint communication action space. Then, $\Gamma_h = M_t$ has cardinality $|\Gamma_h| = |M_t|$, and $\gamma^*_{1:n,h}$ can be computed as

$$\gamma^*_{1:n,h} = \operatorname*{argmax}_{\gamma_{1:n,h}\in\Gamma_h} Q^{*,\mathcal{M}}_h(\widehat{c}_h, \gamma_{1:n,h})$$

by enumerating all possible $\gamma_{1:n,h} \in \Gamma_h$ with complexity $|M_t|$.

If $h = 2t, t \in [H]$, we prove the result case-by-case:

- **Nested private information:** We first define the [kz:can we change all the $\times^i_{j=1}$ to $\prod^i_{j=1}$?] $u_{i,h} \in \mathcal{U}_{i,h} := \{(\times^i_{j=1} \overline{\mathcal{P}}_{j,h}) \times (\times^{i-1}_{j=1} \overline{\mathcal{A}}_{j,h}) \to \overline{\mathcal{A}}_{i,h}\}$ and slightly abuse the notation for $Q^{*,\mathcal{M}}_h$ as follows[kz:all the rewards in the following DPs should include p as input? check all.]

$$Q^{*,\mathcal{M}}_h(\widehat{c}_h, u_{1,h}, \cdots, u_{n,h}) := \sum_{\overline{s}_h, \overline{p}_h, \overline{a}_h, \overline{s}_{h+1}, \overline{o}_{h+1}} \mathbb{P}^{\mathcal{M},c}_h(\overline{s}_h, \overline{p}_h | \widehat{c}_h) \Pi^n_{i=1} \mathbb{1}[\overline{a}_{i,h} = u_{i,h}(\overline{p}_{1:i,h}, \overline{a}_{1:i-1,h})] \overline{\mathbb{T}}_h(\overline{s}_{h+1} | \overline{s}_h, \overline{a}_h)$$
$$\overline{\mathbb{O}}_{h+1}(\overline{o}_{h+1} | \overline{s}_{h+1}) [\overline{\mathcal{R}}_h(\overline{s}_h, \overline{a}_h, \overline{p}_h) + V^{*,\mathcal{M}}_{h+1}(\widehat{c}_{h+1})].$$

Since the space of $\mathcal{U}_{i,h}$ covers the space of $\Gamma_{i,h}$, then for the $u^*_{1:n,h}$ to be an optimal one that maximizes [kz:pls be careful in writing..] the $Q^{*,\mathcal{M}}_h$, we have

$$Q^{*,\mathcal{M}}_h(\widehat{c}_h, u^*_{1,h}, \cdots, u^*_{n,h}) = \max_{\{u_{i,h}\in\mathcal{U}_{i,h}\}_{i\in[n]}} Q^{*,\mathcal{M}}_h(\widehat{c}_h, u_{1,h}, \cdots, u_{n,h}) \geq \max_{\{\gamma_{i,h}\in\Gamma_{i,h}\}_{i\in[n]}} Q^{*,\mathcal{M}}_h(\widehat{c}_h, \gamma_{1,h}, \cdots, \gamma_{n,h}).$$

Meanwhile, due to the nested private information condition, for any $\overline{p}_h \in \overline{\mathcal{P}}_h$, there must exist $\gamma'_{1:n,h}$ such that $\gamma'_{1:n,h}$ output the same actions as $u^*_{1:n,h}$ under $\overline{p}_h$. Therefore, we can conclude that

$$\max_{\{u_{i,h}\in\mathcal{U}_{i,h}\}_{i\in[n]}} Q^{*,\mathcal{M}}_h(\widehat{c}_h, u_{1,h}, \cdots, u_{n,h}) = \max_{\{\gamma_{i,h}\in\Gamma_{i,h}\}_{i\in[n]}} Q^{*,\mathcal{M}}_h(\widehat{c}_h, \gamma_{1,h}, \cdots, \gamma_{n,h}).$$

Therefore, we can solve Equation (E.1) and compute $\gamma^*_{1:n,h}$ from computing $u^*_{1:n,h}$, which can be solved with complexity poly$(|\overline{\mathcal{P}}_h|, |\overline{\mathcal{A}}_h|, |\overline{\mathcal{S}}|)$.

- **Turn-based structures:** For any $\gamma_{ct(h),h} \in \Gamma_{ct(h)}, \gamma_{-ct(h),h}, \gamma'_{-ct(h),h} \in \Gamma_{-ct(h),h}$, where $ct(h)$ is the controller, it holds that for any $\widehat{c}_h$:

$$\begin{aligned}
&Q^{*,\mathcal{M}}_h(\widehat{c}_h, \gamma_{ct(h),h}, \gamma_{-ct(h),h}) \\
=& \sum_{\overline{s}_h,\overline{p}_h,\overline{s}_{h+1},\overline{o}_{h+1}} \mathbb{P}^{\mathcal{M},c}_h(\overline{s}_h, \overline{p}_h|\widehat{c}_h)\overline{\mathbb{T}}_h(\overline{s}_{h+1}|\overline{s}_h, \gamma_{ct(h),h}(\overline{p}_{ct(h),h}), \gamma_{-ct(h),h}(\overline{p}_{-ct(h),h})) \\
&\overline{\mathbb{O}}_{h+1}(\overline{o}_{h+1}|\overline{s}_{h+1})[\overline{\mathcal{R}}_h(\overline{s}_h, \gamma_h(\overline{p}_h)) + V^{*,\mathcal{M}}_{h+1}(\widehat{c}_{h+1})] \\
=& \sum_{\overline{s}_h,\overline{p}_h,\overline{s}_{h+1},\overline{o}_{h+1}} \mathbb{P}^{\mathcal{M},c}_h(\overline{s}_h, \overline{p}_h|\widehat{c}_h)\overline{\mathbb{T}}_h(\overline{s}_{h+1}|\overline{s}_h, \gamma_{ct(h),h}(\overline{p}_{ct(h),h}))\overline{\mathbb{O}}_{h+1}(\overline{o}_{h+1}|\overline{s}_{h+1})[\overline{\mathcal{R}}_h(\overline{s}_h, \gamma_{ct(h),h}(\overline{p}_{ct(h),h})) + V^{*,\mathcal{M}}_{h+1}(\widehat{c}_{h+1})], \\
&\quad + \sum_{i\neq ct(h)} \sum_{\overline{s}_h,\overline{p}_h} \mathbb{P}^{\mathcal{M},c}_h(\overline{s}_h, \overline{p}_h|\widehat{c}_h)\overline{\mathcal{R}}_{i,h}(\overline{s}_h, \gamma_{i,h}(\overline{p}_{i,h})) := \sum_{i\in[n]} U_{i,h}(\widehat{c}_h, \gamma_{i,h}).
\end{aligned}$$

where the last step is due to the fact that $\widehat{c}_{h+1} = \widehat{\phi}_{h+1}(\widehat{c}_h, \overline{z}_{h+1})$. Note that we can write $\overline{\mathcal{R}}_h(\overline{s}_h, \overline{a}_h, \overline{p}_h)$ as $\overline{\mathcal{R}}_h(\overline{s}_h, \overline{a}_h)$ since h is even. Then, we can optimize $Q^{*,\mathcal{M}}_h(\widehat{c}_h, \gamma_{ct(h),h}, \gamma_{-ct(h),h})$ In conclusion, Equation (E.1) can be solved with respect to each individual $\gamma_{i,h}$ with total complexity poly$(|\overline{\mathcal{S}}|, |\overline{\mathcal{P}}_h|, |\overline{\mathcal{A}}_h|)$.

- **Factorized structures:** Note that, for any $h \in [\overline{H}]$, we can write the reward function of $\mathcal{D}'_{\mathcal{L}}$ as $\overline{\mathcal{R}}(\overline{s}_h, \overline{a}_h, \overline{p}_h) = \sum_{i=1}^n \overline{\mathcal{R}}(\overline{s}_{i,h}, \overline{a}_{i,h}, \overline{p}_{i,h}), \forall \overline{s}_h, \overline{a}_h, \overline{p}_h$. Then, for any $h \in [\overline{H}]$, $\widehat{c}_h \in \widehat{\mathcal{C}}_h, \gamma_h \in \Gamma_h$, we use backward induction to prove that, there exist n functions $\{F_{i,h}\}_{i\in[n]}$ such that

$$Q^{*,\mathcal{M}}_h(\widehat{c}_h, \gamma_h) = \sum_{i=1}^n F_{i,h}(\widehat{c}_{i,h}, \gamma_{i,h}).$$

It holds for $h = \overline{H}$ since that

$$Q^{*,\mathcal{M}}_{\overline{H}}(\widehat{c}_{\overline{H}}, \gamma_{\overline{H}}) = \sum_{i=1}^n \sum_{\overline{s}_{i,\overline{H}},\overline{p}_{i,\overline{H}}} \mathbb{P}^{\mathcal{M},c}_{i,\overline{H}}(\overline{s}_{i,\overline{H}}, \overline{p}_{i,\overline{H}}|\widehat{c}_{i,\overline{H}})\overline{\mathcal{R}}_{i,\overline{H}}(\overline{s}_{i,\overline{H}}, \gamma_{i,h}(\overline{p}_{i,\overline{H}}), \overline{p}_{i,\overline{H}}).$$

For any $h \leq \overline{H} - 1$, it holds that[kz:all the rewards in the following DPs should include p as

input? check all.]

$$
\begin{aligned}
&Q_h^{*,\mathcal{M}}(\widehat{c}_h,\gamma_h)\\
&= \sum_{\overline{s}_h,\overline{p}_h,\overline{s}_{h+1},\overline{o}_{h+1}} \mathbb{P}_h^{\mathcal{M},c}(\overline{s}_h,\overline{p}_h|\widehat{c}_h)\overline{\mathbb{T}}_h(\overline{s}_{h+1}|\overline{s}_h,\gamma_h(\overline{p}_h))\overline{\mathbb{O}}_{h+1}(\overline{o}_{h+1}|\overline{s}_{h+1})\\
&\qquad \left[\sum_{i=1}^{n}\overline{\mathcal{R}}_{i,h}(\overline{s}_{i,h},\gamma_{i,h}(\overline{p}_{i,h}),\overline{p}_{i,h}) + F_{i,h+1}(\widehat{c}_{i,h+1},\widehat{g}^*_{i,h+1}(\widehat{c}_{i,h+1}))\right]\\
&= \sum_{i=1}^{n}\sum_{\overline{s}_{i,h},\overline{p}_{i,h},\overline{s}_{i,h+1},\overline{o}_{i,h+1}} \mathbb{P}_{i,h}^{\mathcal{M},c}(\overline{s}_{i,h},\overline{p}_{i,h}|\widehat{c}_{i,h})\overline{\mathbb{T}}_h(\overline{s}_{i,h+1}|\overline{s}_{i,h},\gamma_{i,h}(\overline{p}_{i,h}))\\
&\qquad \overline{\mathbb{O}}_{i,h+1}(\overline{o}_{i,h+1}|\overline{s}_{i,h+1})[\overline{\mathcal{R}}_{i,h}(\overline{s}_{i,h},\gamma_{i,h}(\overline{p}_{i,h}),\overline{p}_{i,h}) + F_{i,h+1}(\widehat{c}_{i,h+1},\widehat{g}^*_{i,h+1}(\widehat{c}_{i,h+1}))]\\
&=: \sum_{i=1}^{n} F_{i,h}(\widehat{c}_{i,h},\gamma_{i,h}).
\end{aligned}
$$

Then, by induction, we know that it holds for any $h \in [\overline{H}]$. We can define $\widehat{g}^*_{i,h}(\widehat{c}_h) \in \operatorname{argmax}_{\gamma_{i,h}\in\Gamma_{i,h}} F_{i,h+1}(\widehat{c}_{i,h+1},\gamma_{i,h})$, and thus solve Equation (E.1) with complexity $\sum_{i=1}^{n}$ poly$(|\overline{\mathcal{S}}_i|,|\overline{\mathcal{A}}_{i,h}|,|\overline{\mathcal{P}}_{i,h}|)$.

This completes the proof. □

Note that, strictly speaking, the time-complexity given in Lemma G.2 does not satisfy Assumption IV.7 yet, since $|\overline{\mathcal{P}}_h|$ may not be necessarily small and polynomial in the LTC parameters $|\mathcal{S}|,|\mathcal{O}_h|,|\mathcal{A}_h|,|\mathcal{M}_h|,H$. For the examples in §A, more specifically, one can show that $|\overline{\mathcal{P}}_h|$ is indeed polynomial in these parameters (when viewing the delay d of sharing, if it exists, as a constant), which led to the final quasi-polynomial complexity guarantees in Theorem IV.8 and Theorem IV.9 (see their proofs for the formal arguments).

H. Examples in the Venn Diagram Figure 2b

Here, we show some examples of the areas ①-⑤ in the Venn diagram in Figure 2b.

- **①: Multi-agent MDP [38] with historical states.** The Dec-POMDPs satisfying that for any $h \in [H], i \in [n], \mathcal{O}_{i,h} = \mathcal{S}, \mathbb{O}_{i,h}(s|s) = 1, c_h = s_{1:h}, p_h = \emptyset$ lie in the area ①.
- **②: Uncontrolled state process without any historical information.** The Dec-POMDPs satisfying that for any $h \in [H], i \in [n], s_h, a_h, a'_h, \mathbb{T}_h(\cdot|s_h,a_h) = \mathbb{T}_h(\cdot|s_h,a'_h), c_h = \emptyset, p_{i,h} = \{o_{i,h}\}$ lie in the area ②.
- **③: Dec-POMDPs with sQC information structure and perfect recall, and satisfying Assumptions III.5 and III.7.** One-step delayed information sharing (Example 1 in A) lies in this area.
- **④: State controlled by one controller with no sharing and only observability of controller.** We consider a Dec-POMDP $\mathcal{D}$. The state dynamics are controller by only one agent (for convenience, agent 1), and only agent 1 has observability, i.e., $\mathbb{T}_h(\cdot|s_h,a_{1,h},a_{-1,h}) = \mathbb{T}_h(\cdot|s_h,a_{1,h},a'_{-1,h})$ for all $s_h,a_{1,h},a_{-1,h},a'_{-1,h}$, and $\mathcal{O}_{-1,h} = \emptyset$. There is no information sharing, i.e., $c_h = \emptyset, p_{1,h} = \{o_{1:h},a_{1:h-1}\}, p_{j,h} = \{a_{j,1:h-1}\}, \forall j \neq 1$. Then $\forall j \neq 1, h_1 < h_2 \in [H]$, agent $(1,h_1)$ does not influence (j,h_2), since $\tau_{j,h_2} = \{a_{j,1:h_2-1}\}$ is not influenced by agent $(1,h_1)$. Therefore, $\mathcal{D}$ is sQC and has

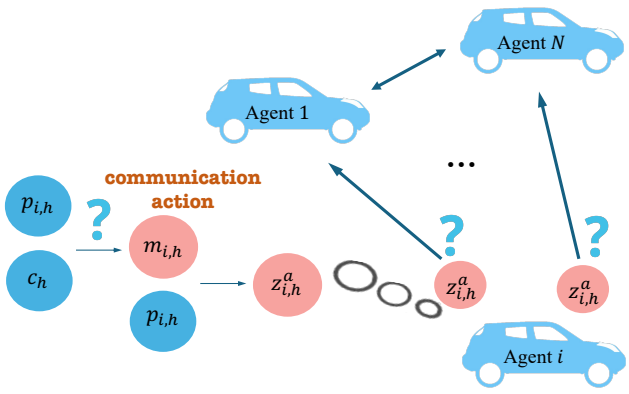


Figure 5: Illustrating the paradigm of the learning-to-communicate problem considered in this paper.

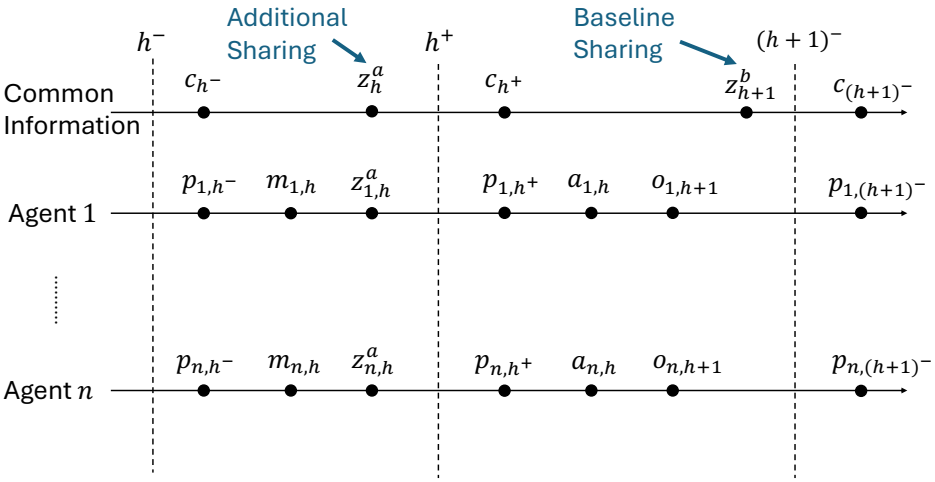


Figure 6: Timeline of the information sharing and evolution protocols in the learning-to-communicate problem considered in this paper.

perfect recall, $\mathcal{D}$ is not SI (underlying state s_h influenced by $g_{1,1:h-1}$). This is because $\mathcal{D}$ does not satisfy Assumption III.7. Then $\mathcal{D}$ lies in the area ④.

- **⑤: One-step delayed observation sharing and two-step delayed action sharing.** The Dec-POMDPs satisfying that for any $h \in [H], i \in [n], c_h = \{o_{1:h-1}, a_{1:h-2}\}, p_{i,h} = \{a_{i,h-1}, o_{i,h}\}$ lie in the area ⑤.

I. Additional Figures

We provide a few figures to better illustrate the paradigms and algorithmic ideas of this paper. Figure 5 and Figure 6 illustrate the paradigm and the timeline of the LTC problems considered in this paper.