Simplified Dynamic Programming for Decentralized POMDPs with Delayed Sharing Information Patterns via Change of Measure

Charalambos D. Charalambous and Photios A. Stavrou

***Abstract*— In this paper, we consider decentralized discrete-time stochastic dynamical optimal control problems with multiple control strategies operating under delayed-sharing information patterns, formulated within the framework of person-by-person (PbP) optimality. We invoke Girsanov's theorem to characterize PbP optimality under a reference probability measure through value functions satisfying simplified dynamic programming (DP) equations, together with corresponding information states that serve as sufficient statistics for the strategies. The value functions and information states retain the fundamental properties of classical partially observable Markov decision problems (POMDPs), namely, both depend on the actions of the minimizing controls, rather than their strategies. The main distinguishing feature of our DP approach is that each control strategy estimates the unobservable state process and the private information components of all other strategies solely from its own private information and the delayed-sharing information components, using information states.**

I. INTRODUCTION

A fundamental model for decentralized discrete-time stochastic dynamical optimal control involves multiple observation posts that collect data and multiple control stations that apply control strategies [2]. At the control stations, the strategies operate under different *information patterns or structures* when applying control actions. The objective of these control strategies is to jointly optimize a common payoff. Such problems are typically formulated as partially observable Markov decision problems (POMDPs).

The characterization of necessary and/or sufficient conditions for optimality in decentralized POMDPs remains a challenging task to date. The available optimization tools are limited, primarily because the *information patterns* of different control strategies are not identical or shared by all decision makers. This stands in contrast to classical centralized POMDPs, which can be successfully addressed using dynamic programming (DP) [3]–[8] and Pontryagin's stochastic maximum principle (SMP) [6], [9], [10].

In recent years, many large-scale and complex systems in science and engineering have been modeled as decentralized POMDPs. Examples include a wide range of scenarios arising in human activities and modern technological systems, all involving multiple decision-making authorities. In such settings, the local data available to any particular strategy is partially shared or communicated to all other strategies, subject to delays. Such data sets are often called *delayed sharing information patterns or structures* [11].

A major challenge in applying the DP approach to decentralized POMDPs is to guarantee that the following two *fundamental properties* of classical centralized POMDPs continue to hold [3].

Property 1. The cost-to-go (value) functions must be defined so that the associated DP equations depend only on the actions of the optimizing controls, and not on their full strategies.

Property 2. The information states (which possess the Markov property) must serve as sufficient statistics for the control strategies, and must likewise depend solely on the actions of the optimizing controls, rather than their strategies.

A. Prior Literature

Over the years, several steps have been taken by many researchers toward developing a systematic framework for addressing optimization issues in general decentralized dynamical optimal control problems and decentralized POMDPs.

Static Team Theory: A major early contribution came from *static team theory*, developed by Marschak and Radner [12], [13], which introduced the notions of *person-by-person (PbP) optimality*, *global or team optimality*, and the relationship between them. However, static team theory is formulated under two restrictive assumptions; namely, the system dynamics are independent of the strategies, and the information patterns available to each strategy are independent of the strategies themselves.

DP Approach for POMDPs with Delayed Sharing Patterns: Another major development was the use of Witsenhausen's Assertion 9 in [2, pp. 1564], and its subsequent variations, to derive DP equations based on information states that serve as sufficient statistics for the strategies [11], [14]–[17]. Although the assertion is correct for 1-step delayed sharing patterns [14]–[16], Varaiya and Walrand [11] demonstrated that *the assertion is incorrect for* 2*-step delayed sharing patterns* by constructing a counterexample. Subsequent studies proposed alternative DP formulations using information states, for example [18], [19]; however, these formulations do not satisfy the fundamental properties 1 and 2 described above.

Decentralized Pontryagin's stochastic maximum principle: Recently, static team theory and Radner's theorem of stationary conditions have been generalized to decentralized stochastic nonlinear dynamical optimal control problems by applying Pontryagin's stochastic maximum principle (SMP) [20]–[23]. The models studied in [20]–[23] are formulated

C. D. Charalambous is with the Department of Electrical and Computer Engineering, University of Cyprus, Cyprus (email: chadcha@ucy.ac.cy). Photios A. Stavrou is with the Department of Communication Systems, EURECOM, France (e-mail: fotios.stavrou@eurecom.fr). The work of P. A. Stavrou is supported by the SNS JU project 6G-GOALS [1] under the EU's Horizon programme (Grant Agreement No. 101139232) and in part by the Huawei France-EURECOM Chair on Future Wireless Networks.

via controlled stochastic differential equations (SDEs), under both the strong and weak (Girsanov's change of measure) formulations of the underlying probability spaces [6], [24].

B. Contribution

This work proposes a new DP framework for decentralized discrete-time POMDPs with delayed sharing information patterns that satisfy the fundamental properties 1 and 2, thereby leading to simplified DP equations. Our approach proceeds in three steps:

(a) **Change of Measure:** We apply a discrete-time version of Girsanov's change-of-measure theorem [24], [25] to reformulate the decentralized POMDP under a reference probability measure in which the observations become statistically independent.

(b) **Construction of Information States:** We identify non-negative conditional measures that act as information states based on the concept of PbP optimality.

(c) **Dynamic Programming and Verification:** We show that PbP optimal strategies satisfy DP equations expressed in terms of these information states. We also establish a verification theorem demonstrating that strategies derived from these DP equations are PbP optimal.

This change-of-measure methodology, combined with PbP optimality, offers significant advantages over previous DP approaches. In particular, it leads to (i) simplified DP equations compared to formulations that do not employ a change of measure, and (ii) information states and DP equations that are independent of the minimizing control strategies, thereby satisfying the classical POMDP properties discussed earlier.

The paper is structured as follows. In Section II, we introduce the model, in Section III we describe the change of measure, and in Section IV, we derive the DP equations and prove a verification theorem for strategies to be PbP optimal.

II. The Decentralized Stochastic Optimal Control Problem

In this section, we introduce the decentralized POMDP.

A. Notation

$\mathbb{R} \triangleq (-\infty,\infty)$, $\mathbb{Z}_+ \triangleq \{1,2,\ldots\}$, $\mathbb{Z}_+^n \triangleq \{1,2,\ldots,n\}$, $n \in \mathbb{Z}_+$. Given a set of elements $s^{(K)} \triangleq \{s^1,\ldots,s^K\}$, $K \in \mathbb{Z}_+$, we define $s^{-k} \triangleq s^{(K)} \setminus \{s^k\} = \{s^1,\ldots,s^{k-1},s^{k+1},\ldots,s^K\}$. $\{(\mathbb{X}_t,\mathscr{B}(\mathbb{X}_t))\big|t \in \mathbb{Z}_+^n\}$ denotes measurable spaces, where $\mathbb{X}_t$ takes values in a complete separable metric space or Borel (Polish) space, and $\mathscr{B}(\mathbb{X}_t)$ is the Borel $\sigma-$algebra of subsets of $\mathbb{X}_t, \forall t \in \mathbb{Z}_+^n$. Points in the product space $\mathbb{X}_{1,n} \triangleq \prod_{t\in\mathbb{Z}_+^n}\mathbb{X}_t$ are denoted by $x_{1,n} \triangleq \{x_1,\ldots,x_n\} \in \mathbb{X}_{1,n}$, and their restrictions for any $(m,n) \in \mathbb{Z}_+ \times \mathbb{Z}_+$ by $x_{m,n} \triangleq \{x_m,\ldots,x_n\} \in \mathbb{X}_{m,n}, n \geq m$. Hence, $\mathscr{B}(\mathbb{X}_{1,n}) \triangleq \otimes_{t\in\mathbb{Z}_+^n}\mathscr{B}(\mathbb{X}_t)$ denotes the $\sigma-$algebra on $\mathbb{X}_{1,n}$ generated by cylinder sets $\{\{x_1,x_2,\ldots,x_n,\ldots\} \in \mathbb{X}_{1,\infty}|x_1 \in A_1,\ldots,x_n \in A_n\}, A_j \in \mathscr{B}(\mathbb{X}_j), \forall j \in \mathbb{Z}_+^n$. We use the convention $X_{k,n} = X_{\max\{k,1\},n}$ and $X_{k,n} = \{\emptyset\}, \forall k > n$.

B. The Decentralized POMDP on the Original Measure

Consider a model that consists of the following elements.

(a) The finite-time horizon set $T_+^n \triangleq \{1,2,\ldots,n\}$.

(b) The complete probability space $(\Omega,\mathscr{F},\mathbb{P}^{U^{(K)}})$, where the measure $\mathbb{P}^{U^{(K)}}$ depends on the controls $U^{(K)} = \{U^1,\ldots,U^K\}$ to be defined shortly.

(c) The unobservable state process, $X_{1,n} \triangleq \{X_1,X_2,\ \ldots,X_n\}$, $X_t : (\Omega,\mathscr{F}) \rightarrow (\mathbb{X}_t,\mathscr{B}(\mathbb{X}_t)), \quad \forall t \in T_+^n$.

(d) The multiple observations processes $Y_{1,n}^{(K)} \triangleq \{Y_{1,n}^1,\ldots,Y_{1,n}^K\}$, $Y_t^k : (\Omega,\mathscr{F}) \rightarrow (\mathbb{Y}_t^k,\mathscr{B}(\mathbb{Y}_t^k)), \quad \forall t \in T_+^n, \quad \forall k \in \mathbb{Z}_+^K$.

(e) The multiple control proc. $U_{1,n}^{(K)} \triangleq \{U_{1,n}^1,\ldots,U_{1,n}^K\}$, $U_t^k : (\Omega,\mathscr{F}) \rightarrow (\mathbb{U}_t^k,\mathscr{B}(\mathbb{U}_t^k)), \forall t \in T_+^n, \forall k \in \mathbb{Z}_+^K$, where $\{\mathbb{U}_t^k|t \in T_+^n\}$ are the *action spaces* of $U_{1,n}^k$ for each $k \in \mathbb{Z}_+^K$.

(f) The $T-$Step delayed sharing patterns. For each $(t,k) \in T_+^n \times \mathbb{Z}_+^K$, control U_t^k is assigned delayed sharing pattern I_t^k,

$$I_t^k \triangleq \{\Delta_t,\Lambda_t^k\} : (\Omega,\mathscr{F}) \rightarrow (\mathbb{I}_t^k,\mathscr{B}(\mathbb{I}_t^k)), \quad \forall (t,k) \in T_+^n \times \mathbb{Z}_+^K$$
$$\Delta_t, : (\Omega,\mathscr{F}) \rightarrow (\mathbb{D}_t,\mathscr{B}(\mathbb{D}_t)) \text{ common comp. to all controls}$$
$$\Lambda_t^k : (\Omega,\mathscr{F}) \rightarrow (\mathbb{L}_t^k,\mathscr{B}(\mathbb{L}_t^k)) \text{ private component of control } k$$

where the two components are specified by

$$\Delta_t = \Big\{Y_{1,t-T}^{(K)},U_{1,t-T}^{(K)}\Big\},\ \Lambda_t^k = \Big\{Y_{t-T+1,t}^k,U_{t-T+1,t-1}^k\Big\} \quad \text{(II.1)}$$

for any delay $T \in T_+^n$.

(g) The strategies of the Controls. Given, $\{I_t^k|t \in T_+^n, k \in \mathbb{Z}_+^K\}$, control U_t^k is generated by a measurable functions $\gamma_t^k(\cdot)$ called strategies of the k-th control at time t, U_t^k, as follows.

$$U_t^k = \gamma_t^k(I_t^k) = \gamma_t^k(\Delta_t,\Lambda_t^k),\ \forall t \in T_+^n,\ \forall k \in \mathbb{Z}_+^K. \quad \text{(II.2)}$$

$\mathscr{U}_{1,n}^k$ the set of admissible strategies of the kth control $U_{1,n}^k$.

Next, we introduce the conditional probability measures (PMs) of the POMDP.

1) The Conditional Probability Measure (PM) of X_{t+1} conditioned on $(X_{1,t},Y_{1,t}^{(K)},U_{1,t}^{(K)})$, $\gamma^{(K)}(\cdot) \in \mathscr{U}_{1,n}^{(K)}$, is Markov,

$$\mathbb{P}^{\gamma^{(K)}}\Big\{X_{t+1} \in dx_{t+1}\big|X_{1,t},Y_{1,t}^{(K)},U_{1,t}^{(K)}\Big\} = \mathbf{P}^{\gamma^{(K)}}_{X_{t+1}|X_t,U_t^{(K)}}$$
$$= S_{t+1}(dx_{t+1}|X_t,U_t^{(K)}),\ \forall t \in T_+^{n-1}. \quad \text{(II.3)}$$

2) The Conditional PM of Y_t^k conditioned on $(X_{1,t},Y_{1,t-1}^{(K)},Y_t^{-k},U_{1,t-1}^{(K)})$, $\gamma^{(K)}(\cdot) \in \mathscr{U}_{1,n}^{(K)}$, satisfies

$$\mathbb{P}^{\gamma^{(K)}}\Big\{Y_t^k \in dy_t^k\big|X_{1,t},Y_{1,t-1}^{(k)},Y_t^{-k},U_{1,t-1}^{(K)}\Big\} = \mathbf{P}^{\gamma^{(K)}}_{Y_t^k|X_t,U_{t-1}^{(K)}}$$
$$= Q_t^k(dy_t^k|X_t,U_{t-1}^{(K)}), \quad \forall t \in T_+^n \quad \forall k \in \mathbb{Z}_+^K. \quad \text{(II.4)}$$

From (II.4) we also deduce,

$$\mathbb{P}^{\gamma^{(K)}}\Big\{Y_t^{(K)} \in dy_t^{(K)}\big|X_{1,t},Y_{1,t-1}^{(K)},U_{1,t-1}^{(K)}\Big\},\ \forall t \in T_+^n$$
$$= Q_t^{(K)}(dy_t^{(K)}|X_t,U_{t-1}^{(K)}) \triangleq \prod_{k=1}^{K} Q_t^k(dy_t^k|X_t,U_{t-1}^{(K)}). \quad \text{(II.5)}$$

3) The Conditional PM of $U_t^{(K)}$ conditioned on $X_{1,t}, Y_{1,t-1}^{(K)}, U_{1,t-1}^{(K)}$ for strategies $\gamma^{(K)}(\cdot) \in \mathscr{U}_{1,n}^{(K)}$ satisfies

$$\mathbb{P}^{\gamma^{(K)}}\Big\{U_t^{(K)} \in du_t^{(K)} \big| X_{1,t}, I_t^{(K)}\Big\}, \ \forall t \in T_+^n \tag{II.6}$$
$$= P_t^{(K)}(du_t^{(K)}|I_t^{(K)}) = \prod_{k=1}^{K} P_t^k(du_t^k|I_t^k) = \prod_{k=1}^{K} \mu_{\gamma_t^k(I_t^k)}(du_t^k)$$

where $\mu_{\gamma_t^k(I_t^k)}(du_t^k)$ is the Dirac measure at $\gamma_t^k(I_t^k)$.

For simplicity we define, $\mathscr{U}_{1,n}^{-k} \triangleq \prod_{j=1, j\neq k}^{K} \mathscr{U}_{1,n}^{j}$,

$\gamma_{1,n}^{-k}(\cdot) \triangleq \Big\{\gamma_{1,n}^1, \ldots, \gamma_{1,n}^{k-1}, \gamma_{1,n}^{k+1}, \ldots, \gamma_{1,n}^{K}\Big\}(\cdot) \in \mathscr{U}_{1,n}^{-k}$,

$\gamma_t^{-k}(I_t^{-k}) \triangleq \Big\{\gamma_t^1(I_t^1), \ldots, \gamma_t^{k-1}(I_t^{k-1}), \gamma_t^{k+1}(I_t^{k+1}), \ldots, \gamma_t^K(I_t^K)\Big\}$

$\equiv \gamma_t^{-k}(\Delta_t, \Lambda_t^{-k})$ for $I_t^k = \{\Delta_t, \Lambda_t^k\}$, $\forall t \in T_+^n$, $\forall k \in \mathbb{Z}_+^K$.

4) The joint PM on $(\Omega, \mathscr{F}, \{\mathscr{F}_t | t \in T_+^n\})$. Let $\{\mathscr{F}_t | t \in T_+^n\}$ denote the complete filtration generated by the $\sigma-$algebra, $\mathscr{F}_t^0 \triangleq \sigma\{(X_{1,t}, Y_{1,t}^{(K)}, U_{1,t-1}^{(K)}), \forall t \in T_+^n\}$. Given $\gamma^{(K)}(\cdot) \in \mathscr{U}_{1,n}^{(K)}$, $\mathbb{P}^{\gamma^{(K)}}\Big\{X_{1,n} \in dx_{1,n}, Y_{1,n}^{(K)} \in dy_{1,n}^{(K)}, U_{1,n}^{(K)} \in du_{1,n}^{(K)}\Big\}$ is uniquely defined using the above PMs.

5) The Payoff Function for any K-tuple $\gamma^{(K)} \triangleq \{\gamma^1, \ldots, \gamma^K\} \in \mathscr{U}_{1,n}^{(K)} \triangleq \times_{k=1}^{K} \mathscr{U}_{1,n}^{k}$ is

$$J_n^{\mathbb{P}^{\gamma^{(K)}}}(\gamma^{(K)}) \triangleq \mathbb{E}^{\mathbb{P}^{\gamma^{(K)}}}\Big\{\sum_{t=1}^{n} \ell(t, X_t, \gamma_t^{(K)})\Big\} \tag{II.7}$$

where $\mathbb{E}^{\mathbb{P}^{\gamma^{(K)}}}$ means the expectation is w.r.t. the PM $\mathbb{P}^{\gamma^{(K)}}$, and $\ell(\cdot)$ is a measurable with finite expectation.

Definition II.1. *(Decentralized PbP Optimality)*
The $K-$tuple of strategies $\gamma^{(K),o} \triangleq \{\gamma^{1,o}, \ldots, \gamma^{K,o}\} \in \mathscr{U}_{1,n}^{(K)}$ is called decentralized PbP optimal, if it satisfies

$$J_n^{\mathbb{P}^{\gamma^{k,o},\gamma^{-k,o}}}(\gamma^{k,o}, \gamma^{-k,o})$$
$$\leq J_n^{\mathbb{P}^{\gamma^{k},\gamma^{-k,o}}}(\gamma^{k}, \gamma^{-k,o}), \ \forall \gamma^k \in \mathscr{U}_{1,n}^{k}, \ \forall k \in \mathbb{Z}^K \tag{II.8}$$

where $J_n^{\mathbb{P}^{\gamma^{k},\gamma^{-k,o}}}(\gamma^{k}, \gamma^{-k,o})$ is the payoff of strategy $\gamma^k(\cdot) \in \mathscr{U}_{1,n-1}^{k}$, when all other strategies are fixed to their optimal responses, $\gamma^{-k,o}(\cdot) \in \mathscr{U}_{1,n-1}^{-k}$.

Assumptions II.1. *(Existence of Prob. Density Function)*
$Q_t^k(\cdot|x_t, u_{t-1}^{(K)})$ and $S_{t+1}(\cdot|x_t, u_t^{(K)})$ are absolutely continuous with respect to (w.r.t.) Lebesgue measures, i.e., there exist density functions $q_t^k(y_t^k|x_t, u_{t-1}^{(K)})$ and $s_{t+1}(x_{t+1}|x_t, u_t^{(K)})$, respectively, $\forall (t,k)$.

III. Change of Measure and Equivalent Decentralized POMDP

In this section, we apply Girsanov's change-of-measure theorem [25] to transform the decentralized POMDP of Section II into an equivalent optimization problem under a reference probability measure $\mathbb{P}$, such that,

1) the observation process $Y_{1,n}^k = \{Y_1^k, Y_2^k, \ldots, Y_n^k\}$ is mutually independent for each $k \in \mathbb{Z}_+^K$, and $Y_{1,n}^k$ is independent of $Y_{1,n}^m$, $\forall k \neq m$, and the state PM is (II.3),

2) the payoff $J_n^{\mathbb{P}^{\gamma^{(K)}}}(\gamma^{(K)})$ is equivalently defined on a reference probability measure $\mathbb{P}$ such that 1) holds.

A. Change of Measure

We consider a reference probability space $(\Omega, \mathscr{F}, \mathbb{P})$ such that the following conditions hold.

1.1) The conditional PM of the state $X_{1,n}$ is given by (II.3).

1.2) The observations $Y_{1,n}^k$ are mutually independent for each $k \in \mathbb{Z}_+^K$, and $Y_{1,n}^k$ is independent of $Y_{1,n}^m$, $\forall k \neq m$, with

$$\mathbf{P}_{Y_{1,n}^k}(dy_{1,n}^k) = \prod_{t=1}^{n} \mathbf{P}_{y_t^k}(dy_t^k) \equiv \prod_{t=1}^{n} \Phi_t^k(dy_t^k), \ \forall k, \tag{III.9}$$

$$\mathbf{P}_{Y_t^{(K)}}(dy_t^{(K)}) = \prod_{k=1}^{K} \mathbf{P}_{Y_t^k}(dy_t^k) \equiv \Phi_t^{(K)}(dy_t^{(K)}), \ \forall t. \tag{III.10}$$

1.3) The control processes $U_{1,n}^{(K)} \triangleq \{U_{1,n}^1, U_{1,n}^2, \ldots, U_{1,n}^K\}$, and their information patterns, and strategies are defined by (II.2), and (II.6) holds.

1.4) Y_t^k is independent of $(X_{1,t}, Y_{1,t-1}^{(K)}, Y_t^{-k}, U_{1,t-1}^{(K)}), \forall (t,k)$.

We introduce σ-algebras, $\mathscr{F}_t^0 \triangleq \sigma\{(X_{1,t}, Y_{1,t}^{(K)}, U_{1,t-1}^{(K)})\}$, $\mathscr{F}^{0,I_t^k} \triangleq \sigma\{I_t^k\}, \forall k$, and denote by $\{\mathscr{F}_t | t \in T_+^n\}$, $\{\mathscr{F}_t^{I^k} | t \in T_+^n\}$, the complete filtrations.

Theorem III.1. *Change from a Reference Measure $\mathbb{P}$ to Original Measure $\mathbb{P}^{\gamma^{(K)}}$.*
Consider the reference probability space $(\Omega, \mathscr{F}, \{\mathscr{F}_t | t \in T_+^n\}, \mathbb{P})$ on which Section III, 1.1)-1.4) hold. Define

$$\theta_s^{U^{(K)}} \triangleq \frac{Q_s^{(K)}(dY_s^{(K)}|X_s, U_{s-1}^{(K)})}{\Phi_s^{(K)}(dY_s^{(K)})} = \prod_{k=1}^{K} \frac{Q_s^k(dY_s^k|X_s, U_{s-1}^{(K)})}{\Phi_s^k(dY_s^k)},$$
$$\Theta_t^{U^{(K)}} \triangleq \prod_{s=1}^{t} \theta_s^{U^{(K)}}, \ \forall t \in T_+^n. \tag{III.11}$$

Assume $Q_s^{(K)}(\cdot|x_s, u_{s-1}^{(K)})$ is absolutely continuous w.r.t $\Phi_s^{(K)}(\cdot)$, denoted by $Q_s^{(K)}(\cdot|x_s, u_s^{(K)}) \ll \Phi_s^{(K)}(\cdot)$-a.a. $(x_s, u_s^{(K)})$, and $\{\Theta_t^{U^{(K)}} | t \in T_+^n\}$ is $\mathbb{P}-$integrable. The following hold.
(1) $\mathbb{E}^{\mathbb{P}}\Big\{\Theta_t^{U^{(K)}}\Big\} = 1, \forall t \in T_+^n$.
(2) Define the Radon-Nikodym derivative (RND),

$$\frac{d\mathbb{P}^{U^{(K)}}}{d\mathbb{P}}\Big|_{\mathscr{F}_t} \triangleq \Theta_t^{U^{(K)}}, \ \ \forall t \in T_+^n. \tag{III.12}$$

Then, $\mathbb{P}^{U^{(K)}} \ll \mathbb{P}$ and

$$\mathbb{P}^{U^{(K)}}(B) = \int_B \Theta_t^{U^{(K)}}(\omega)\mathbb{P}(d\omega), \ \forall B \in \mathscr{F}_t \tag{III.13}$$

is a probability measure.
(3) On the probability space $(\Omega, \mathscr{F}, \{\mathscr{F}_t | t \in T_+^n\}, \mathbb{P}^{U^{(K)}})$, the conditional PMs of $\{X_t | t \in T_+^n\}$, $\{Y_t^{(M)} | t \in T_+^n\}$, $\{U_t^{(K)} | t \in T_+^{n-1}\}$ are given by (II.3)-(II.6).

Proof. The derivation is similar to models described by discrete recursions [24], [26], hence we omit it. ☐

B. Equivalent Decentralized POMDPs on Reference PM

Now, we use Theorem III.1, to equivalently express the payoff $J^{\mathbb{P}^{U^{(K)}}}(U^{(K)})$ given by (II.7) and the decentralized PbP optimality of Definition II.1, on the reference PM $\mathbb{P}$.

Theorem III.2. *(Equivalent Decentralized POMDP)*
Define the payoff on $(\Omega,\mathscr{F},\{\mathscr{F}_t|t\in T_+^n\},\mathbb{P}^{u^{(K)}})$,

$$\mathbb{P}^{U^{(K)}}:\ J_n^{\mathbb{P}^{U^{(K)}}}(U^{(K)})\stackrel{\triangle}{=}\mathbb{E}^{\mathbb{P}^{U^{(K)}}}\Big\{\sum_{t=1}^{n}\ell(t,X_t,U_t^{(K)})\Big\},\quad \text{(III.14)}$$

s.t. $(X_{1,n},Y_{1,n}^{(K)})$*, satisfy (II.3)-(II.6).*

Define the payoff on the reference probability space $(\Omega,\mathscr{F},\{\mathscr{F}_t|t\in T_+^n\},\mathbb{P})$,

$$\mathbb{P}:\quad J_n^{\mathbb{P}}(U^{(K)})\stackrel{\triangle}{=}\mathbb{E}^{\mathbb{P}}\Big\{\sum_{t=1}^{n-1}\Theta_t^{U^{(K)}}\ell(t,X_t,U_t^{(K)})\Big\},\quad \text{(III.15)}$$

$\frac{d\mathbb{P}^{U^{(K)}}}{d\mathbb{P}}\Big|_{\mathscr{F}_t}\stackrel{\triangle}{=}\Theta_t^{U^{(K)}}$ *of Theorem III.1,*

s.t. $(X_{1,n},Y_{1,n}^{(K)})$ *satisfy (II.3), (III.9), (III.10).*

Then $J_n^{\mathbb{P}^{U^{(K)}}}(U^{(K)})=J_n^{\mathbb{P}}(U^{(K)})$*. Moreover, if* $\gamma^{(K),o}$ *is PbP optimal w.r.t. the payoff* $J_n^{\mathbb{P}}(\gamma^{(K)})$ *on the PM* $\mathbb{P}$*, then there exists another* $\gamma^{(K),o}$ *which is PbP optimal w.r.t. the payoff* $J_n^{\mathbb{P}^{\gamma^{(K)}}}(\gamma^{(K)})$ *on the PM* $\mathbb{P}^{\gamma^{(K)}}$*, and vice-versa.*

Proof. This follows directly from Theorem III.1. □

Definition III.1. *On reference PM* $\mathbb{P}$*, for any bounded continuous function,* $\psi(\cdot):\mathbb{X}\times\mathbb{L}^{-k}\to\mathbb{R}$*, and* $\{\gamma^k,\gamma^{-k}\}\in\mathscr{U}_{1,n}^k\times\mathscr{U}_{1,n}^{-k}$*, define* $\forall k\in\mathbb{Z}_+^K,\forall t\in T_+^n$,

$$\mathbb{E}^{\mathbb{P}}\Big\{\psi(X_t,\Lambda_t^{-k})\Theta_t^{\gamma^k,\gamma^{-k}}\big|I_t^k\Big\}\stackrel{\triangle}{=}\int\psi(x_t,\lambda_t^{-k})\mathbf{P}_t^{\gamma^k,\gamma^{-k}}(dx_t,d\lambda_t^{-k}|I_t^k)$$

where $\Xi_t^k[I_t^k](dx_t,d\lambda_t^{-k})\equiv\mathbf{P}_t^{\gamma^k,\gamma^{-k}}(dx_t,d\lambda_t^{-k}|I_t^k)$ *is a non-negative conditional measure (CM) (unnormalized).*

Lemma III.1. *(Payoff in Terms of Non-Negative CM)*
On reference probability space $(\Omega,\mathscr{F},\{\mathscr{F}_t|t\in T_+^n\},\mathbb{P})$*, the payoff of the k-th strategy,* $\gamma^k\in\mathscr{U}_{1,n}^k$ *for* $\gamma^{-k}=\gamma^{-k,o}\in\mathscr{U}_{1,n}^{-k}$*, is expressed using the non-negative CM as*

$$J_n^{\mathbb{P}}(\gamma^k,\gamma^{-k,o})=\mathbb{E}^{\mathbb{P}}\Big\{\sum_{t=1}^{n}\int_{\mathbb{X}_t\times\mathbb{L}_t^{-k}}\ell(t,x_t,\gamma_t^k(I_t^k),\gamma_t^{-k,o}(\Delta_t,\lambda_t^{-k})$$
$$.\mathbf{P}_t^{\gamma^k,\gamma^{-k,o}}(dx_t,d\lambda_t^{-k}|I_t^k)\Big\},\quad\forall k\in\mathbb{Z}_+^K,\quad \text{(III.16)}$$

s.t. $(X_{1,n},Y_{1,n}^{(K)})$ *satisfy (II.3), (III.9), (III.10).*

Proof. This follows from the re-conditioning property of expectation. □

IV. Dynamic Programming and Information States on Reference Measure $\mathbb{P}$

We derive the DP equations on the reference measure $\mathbb{P}$.

Definition IV.1. *(Decentralized PbP Value Processes)*
Consider the payoff $J_n^{\mathbb{P}}(\gamma^k,\gamma^{-k,o})$ *of strategy* $\gamma^k\in\mathscr{U}_{1,n}^k$ *for fixed optimal strategies* $\gamma^{-k}=\gamma^{-k,o}\in\mathscr{U}_{1,n}^{-k}$*,* $\forall k\in\mathbb{Z}_+^K$.
The value process or optimal cost-to-go $\mathscr{V}_t^{\gamma^k,\gamma^{-k,o}}(\cdot):\mathbb{I}_t^k\to[0,\infty)$*, over* $\{t,t+1,\dots,n\}$*, of strategy* $\gamma^k\in\mathscr{U}_{1,n}^k$*, when the optimal strategy* $\gamma^k=\gamma^{k,o}\in\mathscr{U}_{1,t-1}^k$ *is used over* $\{1,2,\dots,t-1\}$*, conditional on* $I_t^k=\{\Delta_t,\Lambda_t^k\}=i_t^k=\{\delta_t,\lambda_t^k\}$ *is defined by*

$$\mathscr{V}_t^{\gamma^k,\gamma^{-k,o}}(i_t^k)\stackrel{\triangle}{=}\inf_{\gamma^k\in\mathscr{U}_{t,n}^k}J_{t,n}^{\gamma^k,\gamma^{-k,o}}(i_t^k),\ \forall t\in T_+^n,\quad \text{(IV.17)}$$

$$J_{t,n}^{\gamma^k,\gamma^{-k,o}}(i_t^k)=\mathbb{E}^{\mathbb{P}}\Big\{\sum_{j=t}^{n}\int_{\mathbb{X}_j\times\mathbb{L}_j^{-k}}\ell(j,x_j,\gamma_j^k(i_j^k),\gamma_j^{-k,o}(\delta_j,\lambda_j^{-k})$$
$$.\mathbf{P}_j^{\gamma^k,\gamma^{-k,o}}(dx_j,d\lambda_j^{-k}|I_j^k=i_j^k)\Big|I_t^k=i_t^k\Big\},\ \forall k\in\mathbb{Z}_+^K.\quad \text{(IV.18)}$$

A. Generalized DP Equations without Information States

Theorem IV.1 gives the first necessary and sufficient conditions for PbP optimality using the generalized DP equations.

Theorem IV.1. *(Decentralized generalized DP Equations-Necessary and Sufficient Conditions)*
Consider the value processes of Definition IV.1, and let $\xi_t^k[i_t^k](dx_t,d\lambda_t^{-k})\stackrel{\triangle}{=}\mathbf{P}_t^{\gamma^k,\gamma^{-k,o}}(dx_t,d\lambda_t^{-k}|I_t^k=i_t^k)$*,* $i_t^k=\{\delta_t,\lambda_t^k\},\forall t\in T_+^n$*. Then, the following necessary and sufficient conditions for PbP optimality hold.*
(1) **Necessity.** *For each* $k\in\mathbb{Z}_+^K$*, suppose a PbP optimal* $\gamma^{k,o}\in\mathscr{U}_{t,n}^k$ *exists.*
(1.1) $\mathscr{V}_t^{\gamma^{k,o},\gamma^{-k,o}}(i_t^k)=\mathscr{V}_t^{\gamma^{-k,o}}(i_t^k),\forall i_t^k,\forall\gamma^{k,o}\in\mathscr{U}_{1,n}^k,\forall t\in T_+^n$ *(i.e., independent of* $\gamma^{k,o}$*) and satisfies the DP equations for all* $I_t^k=i_t^k,\Xi_t^k[I_t^k](dx_t,d\lambda_t^{-k})=\xi_t^k[i_t^k](dx_t,d\lambda_t^{-k})$,

$$\mathscr{V}_n^{\gamma^{-k,o}}(i_n^k)=\inf_{u_n^k\in\mathbb{U}_n^k}\int_{\mathbb{X}_n\times\mathbb{L}_n^{-k}}\ell(n,x_n,u_n^k,\gamma_n^{-k,o}(\delta_n,\lambda_n^{-k}))$$
$$.\xi_n^k[i_t^k](dx_n,d\lambda_n^{-k}),\quad \text{(IV.19)}$$

$$\mathscr{V}_t^{\gamma^{-k,o}}(i_t^k)=\inf_{u_t^k\in\mathbb{U}_t^k}\mathbb{E}^{\mathbb{P}}\Big\{\int_{\mathbb{X}_t\times\mathbb{L}_t^{-k}}\ell(t,x_t,u_t^k,\gamma_t^{-k,o}(\delta_t,\lambda_t^{-k}))$$
$$.\xi_t^k[i_t^k](dx_t,d\lambda_t^{-k})+\mathscr{V}_{t+1}^{\gamma^{-k,o}}(I_{t+1}^k)\Big|i_t^k,u_t^k\Big\},\ \forall t\in T_+^{n-1},k\in\mathbb{Z}_+^K\quad \text{(IV.20)}$$

$$=\inf_{u_t^k\in\mathbb{U}_t^k}\Big\{\int_{\mathbb{X}_t\times\mathbb{L}_t^{-k}}\ell(t,x_t,u_t^k,\gamma_t^{-k,o}(\delta_t,\lambda_t^{-k}))$$
$$.\xi_t^k[i_t^k](dx_t,d\lambda_t^{-k})+\int_{\mathbb{Y}_{t+1}^k\times\mathbb{X}_{t,t+1}\times\mathbb{L}_t^{-k}}\mathscr{V}_{t+1}^{\gamma^{-k,o}}(i_t^k,y_{t+1}^k,$$
$$u_t^k,y_{t-T+1}^{-k},\gamma_{t-T+1}^{-k,o}(\delta_{t-T+1},\lambda_{t-T+1}^{-k}))$$
$$.\mathbf{P}_{t+1}^{\gamma^{-k,o}}(dy_{t+1}^k,d\lambda_t^{-k},dx_t,dx_{t+1}|i_t^k,u_t^k)\Big\},\quad \text{(IV.21)}$$

$$\gamma_{t-T+1}^{-k,o}=\Big\{\gamma_{t-T+1}^{1,o}(\delta_{t-T+1},\lambda_{t-T+1}^1),\dots,$$
$$\gamma_{t-T+1}^{k-1,o}(\delta_{t-T+1},\lambda_{t-T+1}^{k-1}),\gamma_{t-T+1}^{k+1,o}(\delta_{t-T+1},\lambda_{t-T+1}^{k+1}),\dots,$$
$$\gamma_{t-T+1}^{K,o}(\delta_{t-T+1},\lambda_{t-T+1}^K)\Big\}$$

where $\delta_{t-T+1}=\{y_{1,t-2T+1}^{(K)},u_{1,t-2T+1}^{(K)}\}$*,* $\lambda_{t-T+1}^j=\{y_{t-2T+2,t-T+1}^j,u_{t-2T+2,t-T}^j\}$ *are specified by* $\{\delta_t,y_{t-T+1}^j\}$*, and the conditional PM in (IV.21) is*

$$\mathbf{P}_{t+1}^{\gamma^{-k,o}}(dy_{t+1}^k,d\lambda_t^{-k},dx_t,dx_{t+1}|i_t^k,u_t^k)\quad \text{(IV.22)}$$
$$=\mathbf{P}_{t+1}^{\gamma^{-k,o}}(dy_{t+1}^k,d\lambda_t^{-k},dx_t,dx_{t+1}|\xi_t^k,\delta_t,u_t^k).\quad \text{(IV.23)}$$
$$=\Phi_{t+1}^k(dy_{t+1}^k)\quad \text{(IV.24)}$$
$$.S_{t+1}(dx_{t+1}|x_t,u_t^k,\gamma_t^{-k,o}(\delta_t,\lambda_t^{-k}))\xi_t^k[i_t^k](dx_t,d\lambda_t^{-k}).$$

(1.2) For each $k\in\mathbb{Z}_+^K$ the infimum in the DP equations occurs at $u_n^{k,o}=\gamma_n^{k,o}(\xi_n^k[i_n^k],\delta_n)\in\mathbb{U}_n^k, u_t^{k,o}=\gamma_t^{k,o}(\xi_t^k[i_t^k],\delta_t,\lambda_t^k)\in\mathbb{U}_t^k, \forall t\in T_+^{n-1}$, i.e., $\gamma^{k,o}\in\mathscr{U}_{1,n}^k$.

(2) **Verification-Sufficiency.**

(2.1) If for each $k\in\mathbb{Z}_+^K$, the values process $\mathscr{V}_t^{\gamma^{-k,o}}(\cdot), \forall t\in T_+^n$ satisfies the DP equation of part (1), then (almost surely),

$$\mathscr{V}_t^{\gamma^{-k,o}}(I_t^k)\leq J_{t,n}^{\gamma^k,\gamma^{-k,o}}(I_t^k),\ \forall\gamma^k\in\mathscr{U}_{1,n}^k,\ \forall(t,k) \quad \text{(IV.25)}$$

and the resulting $\gamma^{k,o}\in\mathscr{U}_{1,n}^k$, $\forall k\in\mathbb{Z}_+^K$ is PbP optimal.

(2.2) Suppose for each $k\in\mathbb{Z}_+^K$, $\gamma^{k,o}(\cdot)$ is a strategy $u_t^{k,o}=\gamma_t^{k,o}(\xi_t^k[i_t^k],\delta_t,\lambda_t^k)\in\mathbb{U}_t^k$ such that, for all $\{\xi_t^k[i_t^k],\delta_t,\lambda_t^k\}$ achieves the infimum in the DP equations of part (1) for $t=1,\ldots,n$. Then $\gamma^{k,o}\in\mathscr{U}_{1,n}^k$ is PbP optimal and $\mathscr{V}_t^{\gamma^{-k,o}}(I_t^k)=J_{t,n}^{\gamma^{k,o},\gamma^{-k,o}}(I_t^k), \forall(t,k)$ (almost surely).

Proof. We omit the derivation due to space limitations. □

From Theorem IV.1 we directly obtain the next lemma.

Lemma IV.1. *(Value Processes-Private Information State)*
The value process of Definition IV.1 and DP equations of Theorem IV.1 satisfy,

$$\mathscr{V}_t^{\gamma^{-k,o}}(i_t^k)=V_t^{\gamma^{-k,o}}(\xi_t^k,\delta_t,\lambda_t^k),\ \forall(t,k) \quad \text{(IV.26)}$$

i.e, $\mathscr{V}_t^{\gamma^o,\gamma^{-k,o}}(\cdot)$ depends on i_t^k through $(\xi_t^k,\delta_t,\lambda_t^k)$.

Proof. This is obvious from Theorem IV.1. □

B. Generalized DP Equations with Information States

In Theorem IV.3 and Theorem IV.4, we obtain necessary and sufficient conditions for PbP optimality, using simplified generalized DP equations, based on information states and sufficient statistics for the strategies.

First, in Theorem IV.2, Lemma IV.2, Lemma IV.3, we give the recursion of $\mathbf{P}_t^{\gamma^{-k,o}}(dx_t,d\lambda_t^{-k}\big|I_t^k=i_t^k), \forall t\in T_+^n, \forall k$.

Theorem IV.2. *(Limited-Pathwise CMs)*
On the reference probability space $(\{\mathscr{F}_t|t\in T_+^n\},\mathbb{P})$ such that $(X_{1,n},Y_{1,n}^{(K)})$ satisfy (II.3), (III.9), (III.10), and Assumptions II.1 hold, consider the non-negative CMs $\{\Xi_t^k[I_t^k](dx_t,d\lambda_t^{-k})\triangleq\mathbf{P}_t^{\gamma^k,\gamma^{-k}}(dx_t,d\lambda_t^{-k}\big|I_t^k)|t\in T_+^n\}$, $I_t^k\triangleq\{\Delta_t,\Lambda_t^k\}$, and $U_t^k=\gamma_t^k(\Delta_t,\Lambda_t^k), \forall k\in\mathbb{Z}_+^K$, of Definition III.1. For $I_t^k=i_t^k=\{\delta_t,\lambda_t^k\}, \forall t\in T_+^n$ the linear recursions hold.

$$\mathbf{P}_{t+1}^{\gamma^k,\gamma^{-k}}(dx_{t+1},d\lambda_{t+1}^{-k}\big|i_{t+1}^k)\quad\forall t\in\mathbb{Z}_+^{n-1},\ \forall k\in\mathbb{Z}_+^K \quad \text{(IV.27)}$$
$$=\mathbf{T}_{t+1}^k[y_{t+1}^k,u_t^k,\gamma_t^{-k}(\delta_t,\cdot),\mathbf{P}_t^{\gamma^k,\gamma^{-k}}(\cdot\big|i_t^k)](dx_{t+1},d\lambda_{t+1}^{-k}),$$
$$\mathbf{P}_1^{\gamma^k,\gamma^{-k}}(dx_1,d\lambda_1^{-k}|i_1^k)=\mathbf{P}_{X_1,Y_1^{-k}|Y_1^k=y_1^k} \quad \text{(IV.28)}$$
$$\mathbf{T}_{t+1}^k[y_{t+1}^k,u_t^k,\gamma_t^{-k}(\delta_t^k,\cdot),\xi_t^k[i_t^k](\cdot)](dx_{t+1},d\lambda_{t+1}^{-k})$$
$$\triangleq\int_{\mathbb{X}_t}\frac{Q_{t+1}^k(y_{t+1}^k|x_{t+1},u_t^k,\gamma_t^{-k}(\delta_t,\lambda_t^{-k}))}{\Phi_{t+1}^k(dy_{t+1}^k)}$$
$$.Q_{t+1}^{-k}(y_{t+1}^{-k}|x_{t+1},u_t^k,\gamma_t^{-k}(\delta_t,\lambda_t^{-k}))S_{t+1}(x_{t+1}\big|x_t,u_t^k,\gamma_t^{-k}(\delta_t,\lambda_t^{-k}))$$
$$.\prod_{j=1,j\neq k}^K P_t^j(du_t^j\big|\delta_t,\lambda_t^j)\ \xi_t^k[i_t^k](dx_t,d\lambda_t^{-k}) \quad \text{(IV.29)}$$

$P_t^j(du_t^j\big|\delta_t,\lambda_t^j)=\mu_{\gamma_t^j(\delta_t,\lambda_t^j)}(du_t^j)=$ Dirac PM at $\gamma_t^j(\delta_t,\lambda_t^j)$.

Proof. The recursions are derived using Definition III.1 and Theorem III.1, by applying Bayes' theorem. □

Lemma IV.2. *(Property of Non-Negative CMs)*
Consider Theorem IV.2. For each k, the non-negative CM $\mathbf{P}_t^{\gamma^k,\gamma^{-k}}(dx_t,d\lambda_t^{-k}\big|I_t^k), t=1,\ldots,n$ of the kth strategy $\gamma^k(\cdot)\in\mathscr{U}_{1,n}^k$ is independent of $\gamma_t^k(\cdot)$, i.e., for all $k=1,\ldots,K, \forall t\in T_+^n$,

$$\mathbf{P}_t^{\gamma^k,\gamma^{-k}}(dx_t,d\lambda_t^{-k}\big|I_t^k)=\mathbf{P}_t^{\gamma^{-k}}(dx_t,d\lambda_t^{-k}\big|I_t^k),\ \gamma^k\in\mathscr{U}_{1,n}^k.$$

Proof. This follows by induction, using the fact that $\mathbf{P}_1^{\gamma^k,\gamma^{-k}}(dx_1,d\lambda_1^{-k}|I_1^k)=\mathbf{P}_{X_1,Y_1^{-k}|Y_1^k}$ is independent of γ^k. □

Next, we show that $\{\Xi_t^k[i_t^k]\big|t\in T_+^n\}$ is an extended Markov process w.r.t. $\{\Delta_t|t\in T_+^n\}$.

Lemma IV.3. *(Information State and Markov Property of Non-Negative CMs)*
Consider $\Xi_t^k[I_t^k]\triangleq\mathbf{P}_t^{\gamma^k,\gamma^{-k}}(dx_t,d\lambda_t^{-k}\big|I_t^k), \forall t\in T_+^n, \forall k\in\mathbb{Z}_+^K$ of Theorem IV.2.
For each $k\in\mathbb{Z}_+^K$, the CM $\Xi_{1,n}^k\triangleq\{\Xi_t^k[I_t^k]\big|t\in T_+^n\}$ is an extended Markov process w.r.t. $\{\Delta_t|t\in T_+^n\}$, i.e., for bounded continuous $\psi:\mathbb{X}_{t+1}\times\mathbb{L}_{t+1}^{-k}\to\mathbb{R}$,

$$\mathbb{E}^{\mathbb{P}}\Big\{\int_{\mathbb{X}_{t+1}\times\mathbb{L}_{t+1}^{-k}}\psi(x_{t+1},\lambda_{t+1}^{-k})\mathbf{P}_{t+1}^{\gamma^k,\gamma^{-k}}(dx_{t+1},d\lambda_{t+1}^{-k}\big|I_{t+1}^k)\Big|I_t^k,U_t^k\Big\}$$
$$=\mathbb{E}^{\mathbb{P}}\Big\{\int_{\mathbb{X}_{t+1}\times\mathbb{L}_{t+1}^{-k}}\psi(x_{t+1},\lambda_{t+1}^{-k})\mathbf{T}_{t+1}^k[Y_{t+1}^k,U_t^k,\gamma_t^{-k}(\Delta_t,\cdot),$$
$$\mathbf{P}_t^{\gamma^k,\gamma^{-k}}(\cdot\big|I_t^k)](dx_{t+1},d\lambda_{t+1}^{-k})\Big|\Xi_t^k[I_t^k],\Delta_t^k,U_t^k\Big\} \quad \text{(IV.30)}$$
$$=\int_{\mathbb{X}_{t+1}\times\mathbb{Y}_{t+1}^k\times\mathbb{L}_{t+1}^{-k}}\psi(x_{t+1},\lambda_{t+1}^{-k})\Phi_{t+1}^k(dy_{t+1}^k)$$
$$.\Xi_{t+1}^k[I_{t+1}^k](dx_{t+1},d\lambda_{t+1}^{-k}),\ \forall t\in T_+^n,\forall k\in\mathbb{Z}_+^K. \quad \text{(IV.31)}$$

Proof. By Theorem IV.2, we obtain (IV.30), and (IV.31) is due to properties of reference meas. $\mathbb{P}$, (III.9) (III.10). □

Remark IV.1. *Lemma IV.3 and Lemma IV.2 established that the non-negative CM $\xi_t^k=\mathbf{P}_t^{\gamma^k,\gamma^{-k}}(dx_t,d\lambda_t^{-k}\big|I_t^k)$ satisfies Property 2 of classical POMDPs.*

Theorem IV.3. *(Decentralized DP Equations with Private Information State-Necessary Conditions)*
On the reference probability space $(\{\mathscr{F}_t|t\in T_+^n\},\mathbb{P})$ such that $(X_{1,n},Y_{1,n}^{(K)})$ satisfy (II.3), (III.9), (III.10), and Assumptions II.1 hold, consider the value functions $\{V_t^{\gamma^{-k,o}}(\xi_t^k,\delta_t^{(K)},\lambda_t^k)|t\in T_+^n\}$ and the recursions of Theorem IV.2, $\forall k$.
(1) For each $k\in\mathbb{Z}_+^K$ suppose a PbP optimal $\gamma^{k,o}\in\mathscr{U}_{t,n}^k$ exists.

Then $V_t^{\gamma^{-k,o}}(\cdot), \forall t \in T_+^n$ satisfies the DP equations,

$$V_n^{\gamma^{-k,o}}(\xi_n^k, \delta_n^{(K)}, \lambda_n^k) = \inf_{u_n^k \in \mathbb{U}_n^k} \mathbb{E}^{\mathbb{P}}\Big\{ \tag{IV.32}$$

$$\ell(n, X_n, u_n^k, \gamma_n^{-k,o}(\delta_n, \Lambda_n^{-k}))\Theta_n^{u^k,\gamma^{-k,o}} \Big| \xi_n^k, \delta_n, \lambda_n^k, u_n^k\Big\}, \forall k \in \mathbb{Z}_+^K$$

$$= \inf_{u_n^k \in \mathbb{U}_n^k} \int_{\mathbb{X}_n \times \mathbb{L}_n^{-k}} \ell(n, x_n, u_n^k, \gamma_n^{-k,o}(\delta_n, \lambda_n^{-k})) .\xi_n^k[i_n^k](dx_n, d\lambda_n^{-k}), \tag{IV.33}$$

$$V_t^{\gamma^{-k,o}}(\xi_t^k, \delta_t, \lambda_t^k) = \inf_{u_t^k \in \mathbb{U}_t^k} \mathbb{E}^{\mathbb{P}}\Big\{$$

$$\ell(t, X_t, u_t^k, \gamma_t^{-k,o}(\delta_t, \Lambda_t^{-k}))\Theta_t^{u^k,\gamma^{-k,o}} + V_{t+1}^{\gamma^{-k,o}}(\Xi_{t+1}^k, \Delta_{t+1}, \Lambda_{t+1}^k)$$

$$\Big| \xi_t^k, \delta_t, \lambda_t^k, U_t^k\Big\}, \quad \forall t \in T_+^{n-1}, \forall k \in \mathbb{Z}_+^K \tag{IV.34}$$

$$= \inf_{u_t^k \in \mathbb{U}_t^k} \Big\{ \int_{\mathbb{X}_t \times \mathbb{L}_t^{-k}} \ell(t, x_t, u_t^k, \gamma_t^{-k,o}(\delta_t, \lambda_t^{-k}))$$

$$.\xi_t^k(dx_t, d\lambda_t^{-k})[i_t^k] + \int_{\mathbb{Y}_{t+1}^k \times \mathbb{X}_{t,t+1} \times \mathbb{L}_t^{-k}} \Big\{ V_{t+1}^{\gamma^{-k,o}}(\mathbf{T}_{t+1}^k[y_{t+1}^k,$$

$$u_t^k, \gamma_t^{-k}(\delta_t, \cdot), \xi_t^k(\cdot)], \delta_t, \lambda_t^k, y_{t+1}^k, u_t^k, y_{t-T+1}^{-k}, \gamma_{t-T+1}^{-k,o})$$

$$.\mathbf{P}_{t+1}^{\gamma^{-k,o}}(dy_{t+1}^k, d\lambda_t^{-k}, dx_t, dx_{t+1} \Big| \xi_t^k, \delta_t^{(K)}, \lambda_t^k, u_t^k)\Big\}\Big\} \tag{IV.35}$$

where $\gamma_{t-T+1}^{-k,o} = \{\gamma_{t-T+1}^{j,o}(\delta_{t-T+1}^{(K)}, \lambda_{t-T+1}^j)\}_{j=1, j\neq k}^K$ (see Theorem IV.1), and the PM is (IV.22).

(2) For each $k \in \mathbb{Z}_+^K$, if the infimum in the DP equations of part (1) exists, then the optimal strategy is $u_n^{k,o} = \gamma_n^{k,o}(\xi_n^k[i_n^k], \delta_n^{(K)}) \in \mathbb{U}_n^k$, $u_t^{k,o} = \gamma_t^{k,o}(\xi_t^k[i_t^k], \delta_t, \lambda_t^k) \in \mathbb{U}_t^k, \forall t \in T_+^{n-1}$, called semi-separated, denoted by $\gamma^{k,o} \in \mathscr{U}_{1,n}^{k,s-sep}$.

Proof. Follows from Theorem IV.1 and Theorem IV.2. □

Next, we prove the sufficiency, i.e., a *verification theorem*.

Theorem IV.4. *(Decentralized DP Equations with Private Information State-Verification Theorem)*

Suppose the value functions $V_t^{\gamma^{-k,o}}(\cdot), \forall t \in T_+^n, \forall k \in \mathbb{Z}_+^K$ satisfy the DP recursions of Theorem IV.3.

(1) The inequalities hold almost surely (a.s.),

$$V_n^{\gamma^{-k,o}}(\Xi_n^k, \Delta_n, \Lambda_n^k) \leq J_{n,n}^{\gamma^k, \gamma^{-k,o}}(I_n^k), \forall \gamma^k \in \mathscr{U}_{1,n}^k, \forall k, \tag{IV.36}$$

$$V_t^{\gamma^{-k,o}}(\Xi_t^k, \Delta_t, \Lambda_t^k) \leq J_{t,n}^{\gamma^k, \gamma^{-k,o}}(I_t^k), \forall t \in T_+^{n-1}, \forall \gamma^k \in \mathscr{U}_{1,n}^k, \forall k \in \mathbb{Z}_+^K \tag{IV.37}$$

where $J_{t,n}^{\gamma^k,\gamma^{-k,o}}(I_t^k)$ is given by (IV.18).

(2) Given the optimal semi-separated strategies $\gamma_{1,n}^{-k} = \gamma^{-k,o} \in \mathscr{U}_{1,n}^{-k,s-sep}$, let $\gamma^{k,o} \in \mathscr{U}_{1,n}^{k,s-sep}$ be a semi-separated strategy such that for all $\{\xi_t^k, \delta_t, \lambda_t^k\}$, strategy $\gamma_t^{k,o}(\xi_t^k, \delta_t, \lambda_t^k)$ achieves the infimum in the DP equations (IV.32)-(IV.35). Then $\gamma^{k,o}(\xi_t^k, \delta_t, \lambda_t^k)$ is optimal and (almost surely),

$$V_n^{\gamma^{-k,o}}(\Xi_n^k, \Delta_t, \Lambda_t^k) = J_{n,n}^{\gamma^{k,o}, \gamma^{-k,o}}(I_n^k), \forall k, \tag{IV.38}$$

$$V_t^{\gamma^{-k,o}}(\Xi_t^k, \Delta_t, \Lambda_t^k) = J_{t,n}^{\gamma^{k,o}, \gamma^{-k,o}}(I_t^k), \forall t \in T_+^{n-1}, \forall k. \tag{IV.39}$$

where $J_{t,n}^{\gamma^{k,o},\gamma^{-k,o}}(I_t^k) =$ (IV.18) with $\gamma^k = \gamma^{k,o} \in \mathscr{U}_{1,n}^k$.

Proof. This follows using standard DP arguments. □

References

[1] E. C. Strinati et al., "Goal-oriented and semantic communication in 6G AI-native networks: the 6G-GOALS approach," in *Europ. Conf. on Networks and Commun. & 6G Summit*, June 2024.

[2] H. S. Witsenhausen, "Separation of estimation and control for discrete time systems," in *Proc. IEEE*, vol. 59, no. 11, 1971, pp. 1557–1566.

[3] P. R. Kumar and P. Varaiya, *Stochastic Systems: Estimation, Identification, and Adaptive Control*. Prentice Hall, 1986.

[4] N. Ahmed, *Linear and Nonlinear Filtering for Scientists and Engineers*. World Scientific, 1998.

[5] O. Hernandez-Lerma and J. Lasserre, *Discrete-Time Markov Control Processes: Basic Optimality Criteria*, ser. Applications of Mathematics Stochastic Modelling and Applied Probability. Springer Verlag, 1996, no. v. 1.

[6] J. Yong and X. Y. Zhou, *Stochastic Controls, Hamiltonian Systems and HJB Equations*. Springer-Verlag, 1999.

[7] Z. He, C. D. Charalambous, and P. A. Stavrou, "A new finite-horizon dynamic programming analysis of nonanticipative rate-distortion function for Markov sources," in *Proc. European Control Conf.*, 2025, pp. 749–754.

[8] ——, "Rollout-based approximate dynamic programming for MDPs with information-theoretic constraints," in *Proc. IEEE Conf. Decision Control*, 2025, to appear. [Online]. Available: https://arxiv.org/abs/2509.02812

[9] R. Elliott and H. Yang, "Control of partially observed diffusions," *J. Optim. Theory Appl.*, vol. 71, no. 3, pp. 485–501, 1991.

[10] A. Bensoussan, *Stochastic Control of Partially Observable Systems*. Cambridge University Press, 1982.

[11] P. Varaiya and J. Walrand, "On delay sharing patterns," *IEEE Trans. Autom. Control*, vol. 23, no. 3, pp. 443–445, 1978.

[12] R. Radner, "Team decision problems," *The Annals of Mathematical Statistics*, vol. 33, no. 3, pp. 857–881, 1962.

[13] J. Marschak and R. Radner, *Economic Theory of Teams*. New Haven: Yale University Pres, 1972.

[14] B.-Z. Kurtaran and R. Sivan, "Linear-Quadratic-Gaussian control with one-step-delay sharing pattern," *IEEE Trans. Autom. Control*, vol. 19, no. 5, pp. 571–574, 1974.

[15] N. R. Sandell and M. Athans, "Solution of some nonclassical LQG stochastic decision problems," *IEEE Trans. Autom. Control*, vol. 19, no. 2, pp. 108–116, 1974.

[16] T. Yoshikawa, "Dynamic programming approach to decentralized stochastic control problems," *IEEE Trans. Autom. Control*, vol. 20, no. 6, pp. 796–797, 1975.

[17] 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.

[18] B.-Z. Kurtaran, "Corrections and extensions to "decentralized stochastic control with delayed sharing information pattern"," *IEEE Trans. Autom. Control*, vol. 24, no. 4, pp. 656–657, 1979.

[19] A. Nayyar, A. Mahajan, and D. Teneketzis, "Optimal control strategies in delayed sharing information structures," *IEEE Trans. Autom. Control*, vol. 56, no. 7, pp. 1606–1620, 2011.

[20] C. D. Charalambous and N. U. Ahmed, "Centralized versus decentralized optimization of distributed stochastic differential decision systems with different information structures-part I: A general theory," *IEEE Trans. Autom. Control*, vol. 62, no. 3, pp. 1194–1209, March 2017.

[21] ——, "Centralized versus decentralized optimization of distributed stochastic differential decision systems with different information structures—part II: Applications," *IEEE Trans. Autom. Control*, vol. 63, no. 7, pp. 1913–1928, October 2018.

[22] ——, "Team optimality conditions of distributed stochastic differential decision systems with decentralized noisy information structures," *IEEE Trans. Autom. Control*, vol. 62, no. 2, pp. 708–723, 2017.

[23] C. D. Charalambous, "Decentralized optimality conditions of stochastic differential decision problems via Girsanov's measure transformation," *Math. Control Signals Syst.*, vol. 28, no. 3, pp. 1–55, 2016.

[24] R. Elliott, L. Aggoun, and J. Moore, *Hidden Markov Models: estimation and Control*. Springer, 1995.

[25] R. Liptser and A. Shiryayev, *Statistics of Random Processes Vol.1*. Springer-Verlag New York, 1977.

[26] C. D. Charalambous and N. U. Ahmed, "Equivalence of decentralized stochastic dynamic decision systems via girsanov's measure transformation," in *Proc. IEEE Conf. Decision Control*, 2014, pp. 439–444.