DECENTRALIZED STOCHASTIC CONTROL WITH MANY DECISION MAKERS AND THEIR MEAN-FIELD LIMIT

by

SINA SANJARI

A thesis submitted to the
Department of Mathematics and Statistics
in conformity with the requirements for
the degree of Doctor of Philosophy

Queen's University
Kingston, Ontario, Canada
September 2021

Abstract

In this thesis, we study large stochastic team problems (known also as decentralized stochastic control problems) with finite as well as countably infinite number of decision makers.

In the first part of the thesis, we introduce sufficient conditions of optimality and characterize existence and structural properties for globally optimal policies. We first establish sufficient conditions of optimality (in terms of regularity and convexity conditions on the cost function) for static teams with a countably infinite number of decision makers. Then, we focus on static and dynamic team problems and their mean-field limits, where the cost function and dynamics satisfy an exchangeability condition. For this class of team problems, by first imposing convexity, we show that an optimal policy exhibits symmetry and then we establish the convergence of optimal policies for teams with N decision makers to the corresponding optimal policies of mean-field teams. Then, we establish an existence and structural result for convex mean-field teams with a decentralized information structure. Next, we relax the convexity assumption, and we characterize existence and structural properties of an optimal decentralized policy.

Through our analysis of the aforementioned results, it has been shown that static reduction of dynamic stochastic team problems is an effective method for establishing existence and approximation results for optimal policies. With this relation, in the second part, we classify static reductions into three categories: (i) policy-independent, (ii) policy-dependent, and (iii) static measurement with control-sharing reduction. For the first type, we show that there is a bijection between person-by-person optimal (globally optimal) policies of dynamic teams and their policy-independent static reductions. For the second type, although there is a bijection between globally optimal policies of

dynamic teams with partially nested information structures and their static reductions, in general, there is no bijection between person-by-person optimal policies of dynamic teams and their policy-dependent static reductions. We present, however, sufficient conditions under which bijection relationships hold. Under static measurement with control-sharing reduction, the connections between optimality concepts can be established under relaxed conditions. An implication is a convexity characterization of dynamic team problems under static measurement with control-sharing reduction.

Acknowledgments

First, I would like to thank my supervisor Professor Serdar Yüksel for his unlimited motivation and support as well as his enlightening guidance during developing this thesis and my studies at Queen's University. I have been very fortunate to work with such an amazing scholar and supervisor. I am extremely grateful for his insightful guidance in the academic research, career, and in general life. Finishing my PhD would not be possible without his help. Thank you so much, Professor!

I would like to thank my collaborators Professor Tamer Başar and Professor Naci Saldi for their insightful discussions, technical and mathematical writing advises, and careful comments.

I would like to also thank department of Mathematics and Statistics of Queen's university for providing me this wonderful opportunity for learning. In particular, I am very thankful to Professor Fady Alajaji for both accepting to be part of my thesis committee, and also for providing me helpful feedback during my studies. I would like to also thank Professor Bahman Gharesifard for accepting to be part of my supervisory committee as well as for his helpful advise for my PhD and future career. I would like to thank my supervisory and thesis committee members, Professor Nuno Martins, Professor Martin Guay, Professor Yanglei Song, Professor Gregory Smith, Professor Andrew Lewis for their time and effort. I would like to also thank Professor Francesco Cellarosi and Profeeser James Mingo for their excellent probability theory course and real analysis course from which I benefit a lot in my thesis. I would like to thank Professor Ram Murty and Professor Abdolreza Mansouri for their excellent complex analysis course and stochastic calculus course and also for their motivation in studying mathematics. Finally, I thank our graduate secretary Jennifer Read for her enormous help starting my studies at Queen's University to the end.

I want to also thank all of my friends for our great time together in Kingston. I especially like to thank Dak for all fantastic memories together and her support.

I am deeply grateful to my parents Amir and Cita and also my sister Sarah for their unconditional love and unlimited support. I would like to also thank my uncle Taghi for his support and enormous help since my first day in Canada.

Contents

List of Figures

Chapter 1

Introduction: Motivation, Literature Review and Summary of Contributions

A decentralized control problem, or team problem, entails a collection of decision makers (DMs) or agents acting together to optimize a common cost function, but not necessarily sharing all the available information. The term *stochastic teams* refers to the class of team problems where there exist randomness in the initial states, observations, cost realizations, or the evolution of the dynamics. At each time stage, each DM has only partial access to the global information, which is characterized by the *information structure* (IS) of the problem [134]. If there is a pre-defined order in which the decision makers act, then the team is called a *sequential team*. In the following, we formally introduce a general setup of sequential stochastic team problems.

1.1 Intrinsic Model for Stochastic Teams

Hans Witsenhausen's contributions [134, 131, 130, 135, 133] to stochastic control theory, and his characterization of information structures in decentralized stochastic control have been crucial in our modern understanding of decentralized stochastic control and decision theory. In this subsection, we introduce the characterizations as laid out by Witsenhausen, termed as *the Intrinsic Model* [134]; see [139] and [91] for a more comprehensive overview and further characterizations and classifications of information structures (ISs). In this model (described in discrete time), any

action applied at any given time is regarded as applied by an individual DM/agent, who acts only once. One particular advantage of this model, in addition to its generality, is that through it the characterizations regarding ISs can be concisely described. For a recent study on the geometry of ISs, we refer the reader to [111].

Consider decentralized systems where the decision makers act in a pre-defined order. Such systems are called *sequential teams* (for non-sequential teams, we refer the reader to Andersland and Teneketzis [3], [4] and Teneketzis [126], in addition to Witsenhausen [131] and [139, p. 113]). For sequential teams, Witsenhausen's *intrinsic model* [134] consists of the following:

- There exists a collection of *measurable spaces* $\{(\Omega, \mathcal{F}), (\mathbb{U}^i, \mathcal{U}^i), (\mathbb{Y}^i, \mathcal{Y}^i), i \in \mathcal{N}\}$, specifying the system's distinguishable events, and control and measurement spaces. The set $\mathcal{N}$ denotes the collection of decision makers. The set $\mathcal{N}$ can be a finite set $\{1, 2, \ldots, N\}$ or a countable set $\mathbb{N}$ (even though Witsenhausen considered a finite case). The pair $(\Omega, \mathcal{F})$ is a measurable space (on which an underlying probability may be defined). The pair $(\mathbb{U}^i, \mathcal{U}^i)$ denotes the Borel space from which the action u^i of DM^i is selected. The pair $(\mathbb{Y}^i, \mathcal{Y}^i)$ denotes the Borel observation/measurement space.

- There is a *measurement constraint* to establish the relation between the observation variables and the system's distinguishable events. The $\mathbb{Y}^i$-valued observation variables are given by $y^i = h^i(\omega, \underline{u}^{[1,i-1]})$, where $\underline{u}^{[1,i-1]} = \{u^k \mid k \leq i-1\}$ and h^is are measurable functions.

- The set of admissible control laws $\underline{\gamma} = \{\gamma^i\}_{i \in \mathcal{N}}$, also called *designs* or *policies*, are measurable control functions, so that $u^i = \gamma^i(y^i)$. Let Γ^i denote the set of all admissible policies for DM^i and let $\Gamma := \prod_{i \in \mathcal{N}} \Gamma^i$.

- There is a *probability measure* P on $(\Omega, \mathcal{F})$ describing the probability space on which the system is defined.

Under the above intrinsic model, a sequential team problem is *dynamic* if the information available to at least one decision maker (DM) is affected by the action of at least one other DM.

A team problem is *static*, if for every DM the information available is only affected by exogenous disturbances; that is, no other DM can affect the information at any given DM. ISs can also be categorized as *classical*, *quasi-classical (partially-nested)* or *non-classical*. An IS $\{y^i \mid i \in \mathcal{N}\}$ is *classical* if y^i contains all of the information available to DM^k for $k < i$. An IS is *quasi-classical (partially nested)*, if whenever u^k, for some $k < i$, affects y^i through the measurement function h^i, y^i contains y^k (that is $\sigma(y^k) \subset \sigma(y^i)$). An IS which is not partially nested is *non-classical*.

In this thesis, we also sometimes allow for randomized policies, where in addition to y^i, each DM^i has access to, without any loss, a $[0, 1]$-valued independent random variable. This will be made precise later on when randomized policies are utilized.

Under the intrinsic model, every DM acts separately. However, depending on the information structure and the cost function, it may be convenient to consider a collection of DMs as a single DM acting at different time instances (we sometimes refer this collection of DMs as a player (PL) to distinguish the definition of a DM in this setup from the preceding (DM-wise) intrinsic model). In fact, in the classical stochastic control, this is the standard approach. Hence, we may consider a collection of DMs as a single DM ($i \in \mathcal{N}$) acting at different time instances ($t \in \mathcal{T}$), where the set $\mathcal{T}$ denotes the number of times each decision maker can act (the set $\mathcal{T}$ can be a finite set $\{0, 1, \ldots, T-1\}$ or a countable set $\mathbb{N}$ depending on the finite-horizon or infinite-horizon setup). We now define a dynamic (multi-stage) model as:

(i) The observation and action spaces for each DM are standard Borel spaces (that is, Borel subsets of complete, separable and metric spaces) with $\mathbf{Y}^i := \prod_{t=0}^{T-1} \mathbb{Y}_t^i$, $\mathbf{U}^i := \prod_{t=0}^{T-1} \mathbb{U}_t^i$, respectively.

(ii) An admissible policy for DM^i is defined as $\boldsymbol{\gamma}^{\boldsymbol{i}} \in \mathbf{\Gamma^i}$, where $\boldsymbol{\gamma}^{\boldsymbol{i}} := \{\gamma_t^i\}_{t\in\mathcal{T}}$ and $\mathbf{\Gamma^i} = \prod_{t=0}^{T-1} \Gamma_t^i$.

(iii) Observation for each DM^i through time $t \in \mathcal{T}$ is defined as $y_t^i := h_t^i(\underline{\zeta}_{0:t}, \underline{u}_{0:t-1})$, where h_t^is are Borel measurable functions, $\underline{\zeta}_{0:t} := (\zeta_0^i, \ldots, \zeta_t^i)_{i\in\mathcal{N}}$ and $\underline{u}_{0:t-1} := (u_0^i, \ldots, u_{t-1}^i)_{i\in\mathcal{N}}$. Let $\underline{\boldsymbol{\zeta}} := (\boldsymbol{\zeta}^i)_{i\in\mathcal{N}}$, where $\boldsymbol{\zeta}^i := (\zeta_t^i)_{t\in\mathcal{T}}$ denotes all the uncertainty associated with DM^i through times

(which are primitive random variables taking values in standard Borel spaces).

In view of the above formulation for dynamic (multi-stage or multi-level) problems, we will later on define a state space model of stochastic dynamic (multi-stage) teams.

We now introduce a general setup of stochastic team problems.

Problem ($\mathcal{P}'_N$): Stochastic Team Problems with N Decision Makers. Let $\mathcal{N} = \{1, \dots, N\}$. Let $\underline{\gamma}_N := \{\gamma^i\}_{i\in\mathcal{N}} \in \Gamma_N$ where $\Gamma_N = \prod_{i=1}^N \Gamma^i$. Consider a team problem with an expected cost function (to minimize) under a policy $\underline{\gamma}_N$ as

$$J_N(\underline{\gamma}_N) = E^{\underline{\gamma}_N}\Big[c(\omega_0, \underline{u}_N)\Big] := E\Big[c(\omega_0, \gamma^1(y^1), \cdots, \gamma^N(y^N))\Big], \tag{1.1}$$

for some Borel measurable cost function $c : \Omega_0 \times \prod_{i=1}^N \mathbb{U}^i \to \mathbb{R}_+$. We define ω_0 as the Ω_0-valued cost function relevant exogenous random variable, where $\omega_0 : (\Omega, \mathcal{F}, P) \to (\Omega_0, \mathcal{F}_0)$, and Ω_0 is a Borel space with its Borel σ-field $\mathcal{F}_0$. Here, we have the notation $\underline{u}_N := \{u^i \mid i \in \mathcal{N}\}$.

Now, we introduce two notions of optimality for the preceding stochastic team problems:

Global Optimality: For a given stochastic team problem ($\mathcal{P}'_N$) with a given information structure, a policy (strategy) $\underline{\gamma}^*_N := \{\gamma^{i^*}\}_{i\in\mathcal{N}} \in \Gamma_N$ is *(globally) optimal* for ($\mathcal{P}'_N$) if

$$J_N(\underline{\gamma}^*_N) = \inf_{\underline{\gamma}_N \in \Gamma_N} J_N(\underline{\gamma}_N) =: J^*_N.$$

Person-by-Person (PbP) Optimality: For a given stochastic team problem ($\mathcal{P}'_N$) with a given information structure, a person-by-person (pbp) optimal policy is defined as a policy $\underline{\gamma}^*_N := \{\gamma^{i^*}\}_{i\in\mathcal{N}} \in \Gamma_N$ such that for every $i \in \mathcal{N}$

$$J_N(\underline{\gamma}^*_N) = \inf_{\gamma^i \in \Gamma^i} J_N(\gamma^i, \underline{\gamma}^{-i}),$$

where $\underline{\gamma}^{-i} := \{\gamma^1, \dots, \gamma^{i-1}, \gamma^{i+1}, \dots, \gamma^N\}$. This may also be referred to as *Nash equilibrium*.

Before introducing a class of large-scale stochastic team problems, we provide a brief description of static reduction methods, which have been utilized through our analysis in the thesis.

1.2 Static Reductions for Stochastic Dynamic Teams

We now introduce two classes of static reductions which will be used later on in the thesis. Consider teams with finite N number of decision makers (i.e., $\mathcal{N} := \{1, \ldots, N\}$). Since our focus is on teams with N number of decision makers, in this section and later on in Chapter 5, sub/super index N has been eliminated in the problem formulation. We first recall Witsenhausen's static reduction (see [135, 138]), and then, we provide a description of team problems under this static reduction (we refer this as *policy-independent static reduction*).

Consider a stochastic team problem within the intrinsic model with observations of DMs given by

$$y^i = h^i(\omega_0, \omega_i, u^1, \ldots, u^{i-1}, y^1, \ldots, y^{i-1}), \tag{1.2}$$

where ω_i is an exogenous random variable, for $i \in \mathcal{N}$ and the IS of DM^i is $I^i = \{y^i\}$ (or $I^i = \{y^k\}_{k \in K_i}$ for a subset $K_i \subseteq \{1, \ldots, i\}$). Now, we introduce an absolute continuity condition under which a policy-independent static reduction exists (see [138] for further discussions).

An Absolute Continuity Condition of Policy-Independent Static Reduction. If for any DM^i, there exists a probability measure Q^i on $\mathbb{Y}^i$ and a function f^i such that for any Borel set A^i

$$P(y^i \in A^i | \omega_0, u^1, \ldots, u^{i-1}, y^1, \ldots, y^{i-1}) = \int_{A^i} f^i(y^i, \omega_0, u^1, \ldots, u^{i-1}, y^1, \ldots, y^{i-1}) Q^i(dy^i). \tag{1.3}$$

If the preceding absolute continuity condition holds, then

$$P(d\omega_0, du^1, \ldots, du^N, dy^1, \ldots, dy^N)$$

$$= P(d\omega_0) \prod_{i=1}^{N} f^i(y^i, \omega_0, u^1, \ldots, u^{i-1}, y^1, \ldots, y^{i-1}) Q^i(dy^i) 1_{\{\gamma^i(y^i) \in du^i\}}. \tag{1.4}$$

Hence, under the above change of measure (policy-independent static reduction), we can rewrite the expected cost function as follows:

$$J(\underline{\gamma}) := \int \tilde{c}(\omega_0, u^1, \ldots, u^N, y^1, \ldots, y^N) \prod_{i=1}^{N} 1_{\{\gamma^i(y^i) \in du^i\}} Q^i(dy^i) P(d\omega_0), \tag{1.5}$$

with $u^i = \gamma^i(y^i)$ for $i \in \mathcal{N}$, and

$$\tilde{c}(\omega_0, u^1, \ldots, u^N, y^1, \ldots, y^N) := c(\omega_0, u^1, \ldots, u^N) \prod_{i=1}^{N} f^i(y^i, \omega_0, u^1, \ldots, u^{i-1}, y^1, \ldots, y^{i-1}),$$

and under (1.5), measurements are independent. As the analysis above indicates, in this static reduction, under the absolute continuity condition, the probabilistic nature of the problem has been transformed to the cost function by changing the measures of the observations to fixed probability measures. Since the static reduction above is independent of the policies that preceding DMs choose, we refer to this type of static reduction as *policy-independent static reduction*. As Witsenhausen notes, a static reduction always holds when the measurement variables take values from countable sets since a reference measure as in Q^i above can always be constructed on the measurement space $\mathbb{Y}^i$ (e.g., $Q^i(z) = \sum_{j \geq 1} 2^{-j} 1_{\{z = m_j\}}$, where $\mathbb{Y}^i = \{m_j \mid j \in \mathbb{N}\}$) so that the absolute continuity condition always holds. We refer the reader to [44] for relations with the classical continuous-time stochastic control, where the relation with Girsanov's classical measure transformation [61, 24] is recognized, and [139, p. 114] for further discussions. For discrete-time partially observed stochastic control, similar arguments had been presented, e.g. by Borkar in [30], [31].

Now, we briefly introduce Ho and Chu's static reduction [66, 67] (policy-dependent static reduction), and then we provide a description of dynamic team problems under the policy-dependent static reduction.

Stochastic Dynamic Teams with Partially Nested IS. Consider a stochastic dynamic team with

partially nested IS, where observations of DMs are defined as

$$y_i^D := \left\{ y_{\downarrow i}^D, \hat{y}_i^D := g_i(h_i(\zeta), u^{\downarrow i}) \right\}, \tag{1.6}$$

and $\zeta := \{\omega_0, \ldots, \omega_N\}$ denotes the set of all relevant random variables (corresponding to the uncertainty of the team), and g_i and h_i are measurable functions. In the above, $y_{\downarrow i}^D$ is a subset of a collection of observations of precedent DMs, all DM^{j}s for $j \in \mathcal{N}$, such that $\hat{y}_i^D$ is affected by the actions of DM^j. The sub index $\{\downarrow i\}$ denotes

$$\{\downarrow i\} := \{j \mid \hat{y}_i^D \text{ is affected by } u^j\}. \tag{1.7}$$

Let $I_i^D = \{y_i^D\}$.

Now, we introduce a condition under which such a static reduction exists [66, 67].

An Invertibility Condition of Policy-Dependent Static Reduction. For all $i \in \mathcal{N}$ and for every fixed $u^{\downarrow i}$, the function $g_i(\cdot, u^{\downarrow i}) : h_i(\zeta) \mapsto \hat{y}_i^D$ is invertible for all realizations of ζ.

If the preceding invertibility condition holds, following from [66, 67], given a policy $\underline{\gamma}^D$, we can define stochastic dynamic team problems under policy-dependent static reduction as follows:

Stochastic Dynamic Teams with Partially Nested IS under Policy-Dependent static reductions. Consider a stochastic team, where the observations of each DM can be defined as

$$y_i^S = \left\{ y_{\downarrow i}^S, \hat{y}_i^S := h^i(\zeta) \right\}, \tag{1.8}$$

where the IS of DM^i be $I_i^S = \{y_i^S\}$ (we note that given a policy $\underline{\gamma}^D$, under the preceding invertibility condition, we can construct a policy $\underline{\gamma}^S$ through a relation $\gamma_i^S(y_i^S) = \gamma_i^D(y_i^D)$ for P-almost surely for all ζ).

In the preceding static reduction, given policies of DMs, there is a bijection between observations as a function of preceding actions of DMs and primitive random variables, and observations generated under the transformations, where now they are only functions of primitive random variables. We note that this static reduction depends on the policies that preceding DMs choose, and hence, through the thesis, we refer to this static reduction as *policy-dependent static reduction*.

We also introduce a new reduction, by expanding the information structure via control-sharing, where in addition to observations, actions are also shared.

Stochastic Dynamic Teams with Partially Nested with Control-Sharing IS. Consider a stochastic team, where for each DM^i, the observations within partially nested with control-sharing IS is defined as

$$y_i^{D,CS} := \left\{ y_{\downarrow i}^D, u^{\downarrow i}, \hat{y}_i^D \right\}, \tag{1.9}$$

with $I_i^{D,CS} := \{y_i^{D,CS}\}$.

Under the invertibility condition (since there is a bijection between dynamic observations y_i^D and static one y_i^S for each $i \in \mathcal{N}$), we can reduce the preceding dynamic problem with partially nested with control-sharing IS to another dynamic problem where measurements are static.

Stochastic Dynamic Teams with Static Measurements with Control-Sharing IS. Consider a stochastic team, where for each DM^i, the observations can be defined as

$$y_i^{CS} := \left\{ y_{\downarrow i}^S, u^{\downarrow i}, \hat{y}_i^S \right\}, \tag{1.10}$$

with $I_i^{CS} := \{y_i^{CS}\}$.

We emphasize that through the thesis, we refer the above information structures as *partially nested with control-sharing* and *static measurements control-sharing*, respectively. We refer the reduction for control-sharing problems as *static measurements with control-sharing*. This expansion and its reduction turns out to play a pivotal role in the characterization of convexity and

optimality conditions which will be discussed in Chapter 5 in detail.

1.3 Large-Scale Stochastic Teams and their Mean-Field Limit

In this section, we introduce our setup for large-scale stochastic team problems. Since exchangeabiliy plays an important role in our analysis, we first recall the definition of *exchangeability* for random variables, which has been used through the thesis and especially in Chapters 3 and 4.

Definition. *Random vectors (variables)* $x^1, x^2, \dots, x^N$ *defined on a common probability space are* N*-exchangeable if for any permutation* σ *of the set* $\{1, \dots, N\}$*,*

$$\mathcal{L}\Big(x^{\sigma(1)}, x^{\sigma(2)}, \dots, x^{\sigma(N)}\Big) = \mathcal{L}\Big(x^1, x^2, \dots, x^N\Big)$$

where $\mathcal{L}$ *denotes the joint distribution of random vectors. Random vectors* $(x^1, x^2, \dots)$ *is infinitely-exchangeable if finite distributions of* $(x^1, x^2, \dots)$ *and* $(x^{\sigma(1)}, x^{\sigma(2)}, \dots)$ *are identical for any finite permutation (affecting only finitely many elements) of* $\mathbb{N}$*.*

Now, we present a definition for the cost function to be of the exchangeable type:

Definition. *The cost function is exchangeable with respect to actions for all* ω_0 *if for any permutation* σ *of* $\{1, \dots, N\}$*,*

$$c(\omega_0, u^1, \dots, u^N) = c(\omega_0, u^{\sigma(1)}, \dots, u^{\sigma(N)}), \tag{1.11}$$

for all ω_0*.*

A particular case of team problems where the cost function is exchangeable is *mean-field teams*. Let the action space and observation space be identical through DMs and subsets of appropriate Euclidean spaces, that is, $\mathbb{U}^i = \mathbb{U} \subseteq \mathbb{R}^n$ and $\mathbb{Y}^i = \mathbb{Y} \subseteq \mathbb{R}^m$ for all $i \in \mathbb{N}$, where n and m are positive integers. We now introduce an N-DM stochastic static team problem ($\mathcal{P}_N$) (also referred to as a pre-limit stochastic static mean-field team problem) as follows:

Problem ($\mathcal{P}_N$): (Large-Scale) Stochastic (Static) Team Problems with N Decision Makers. Let $\mathcal{N} = \{1, \dots, N\}$. Let $\underline{\gamma}_N \in \Gamma_N$ where $\Gamma_N = \prod_{i=1}^N \Gamma^i$. Given a (decentralized) static information structure, a stochastic static team problem is described with an expected cost function (to minimize) of $\underline{\gamma}_N$ as

$$J_N(\underline{\gamma}_N) = \frac{1}{N} E^{\underline{\gamma}_N}\Big[\sum_{i=1}^N c(\omega_0, u^i, \frac{1}{N}\sum_{p=1}^N u^p)\Big], \tag{1.12}$$

for some Borel measurable cost function $c : \Omega_0 \times \mathbb{U} \times \mathbb{U} \to \mathbb{R}_+$.

Now, we introduce a stochastic static team problem with a countably infinite number of DMs (also referred to as stochastic static mean-field team problem) as follows:

Problem ($\mathcal{P}_\infty$): (Large-Scale) Stochastic (Static) Teams with a Countably Infinite Number of Decision Makers. Consider a stochastic static team with a countably infinite number of decision makers, that is, $\mathcal{N} = \mathbb{N}$. Let $\underline{\gamma} := \{\gamma^i\}_{i\in\mathcal{N}} \in \Gamma$ where $\Gamma = \prod_{i\in\mathbb{N}} \Gamma^i$. Given a (decentralized) static information structure, a stochastic static team problem is described with an expected cost function (to minimize) of a policy $\underline{\gamma}$ as

$$J(\underline{\gamma}) = \limsup_{N\to\infty} \frac{1}{N} E^{\underline{\gamma}}\Big[\sum_{i=1}^N c(\omega_0, u^i, \frac{1}{N}\sum_{p=1}^N u^p)\Big], \tag{1.13}$$

for some Borel measurable cost function $c : \Omega_0 \times \mathbb{U} \times \mathbb{U} \to \mathbb{R}_+$. With a slight abuse of notation, we use the same notation for the cost function c as in (1.12).

One of our main objective here is to characterize the existence and structural properties of globally optimal policies of the above stochastic static team problems $(\mathcal{P}_N)$ and $(\mathcal{P}_\infty)$. We are also interested in the sufficient conditions of optimality for $(\mathcal{P}_\infty)$, and the connections between optimal solutions of team problems $(\mathcal{P}_N)$ and $(\mathcal{P}_\infty)$. Dynamic (multi-stage) version of the above large-scale stochastic team problems is to be introduced and studied later on in Chapters 3 and 4. In particular, in Chapter 3, we focus on convex (multi-stage) dynamic problems, and in Chapter 4, we focus on exchangeable static and dynamic problems, where the convexity condition imposed

in Chapters 2 and 3 is relaxed. Finally, in Chapter 5, we study the connections between optimality concepts in N-DM dynamic problems and their static reductions.

1.4 A General Literature Review and Motivation

For teams with finitely many DMs, Marschak [92] has studied static teams and Radner [102] has established connections between person-by-person (pbp) optimality, stationarity, and global optimality. Radner's results were generalized in [78] by relaxing optimality conditions. The essence of these results is that in the context of static team problems, the convexity of the cost function, subject to minor regularity conditions, suffices for the global optimality of pbp optimal solutions. In the particular case of LQG (Linear Quadratic Gaussian) static teams, this result leads to optimality of linear policies [102]. In addition, under static reductions, the results have been extended to dynamic teams [66, 67].

We also note that static reduction has been utilized for arriving at existence results [62, 140, 138], and formulating universal dynamic programming [138], and characterizing optimality and convexity conditions [140] for stochastic teams with finitely many DMs, and approximating with finite models [112]. We emphasize that in the aforementioned results, the underlying assumption is that the number of decision makers is finite, hence the results above cannot be directly applied for stochastic teams with a countably infinite number of decision makers.

There have been several studies involving decentralized stochastic control with infinitely many decision makers. In particular, when the coupling among the decision makers is only through some aggregate/average effect, such problems can be viewed within the umbrella of mean-field games [70, 69, 86, 19, 40, 19, 40, 110, 43, 39]. There have also been several studies for mean-field games where the limits of sequences of Nash equilibria have been investigated as the number of decision makers $N \to \infty$ (see e.g., [56, 83, 20, 86, 7]). The solution concept in game theory is often Nash equilibrium, and often under various characterizations of it in dynamic Bayesian setups. We however recall that in the context of team problems, these correspond to person-by-person optimal solutions, and hence not necessarily globally optimal solutions. We emphasize that

in team problems, person-by-person optimality does not in general imply global optimality both for teams with a finite number of DMs and teams with a countably infinite number of DMs. In fact, in the context of large-scale stochastic teams, the gap between person-by-person optimality (Nash equilibrium in the game-theoretic context) and global team optimality is significant since a perturbation of finitely many policies fails to deviate the value of the expected cost, which implies that person-by-person optimality is a weak condition for such a setup. Hence, without establishing the uniqueness of the mean-field solution (which may hold for example, under strong monotonicity assumptions [86]), the results presented in the aforementioned papers may be inconclusive regarding global optimality of the limit equilibrium. Nonetheless, on the existence as well as uniqueness and non-uniqueness results on equilibria, there have been several studies for mean-field games [86, 19, 40, 88, 80, 23, 63, 48, 38]. For teams and social optima control problems, the earlier analysis has primarily focused on the LQG model where the centralized performance has been shown to be achieved asymptotically by decentralized controllers (see e.g., [71, 5, 6]). A notable exception is [90] where LQG static teams with a countably infinite number of decision makers have been studied, and sufficient conditions for global optimality have been established.

As a further motivation of our problem in this thesis, we note that for infinite-horizon dynamic team problems, the static reductions lead to static teams with countably many decision makers; thus leading to a different setup where our results in this thesis will be applicable. In particular, the question of whether partially nested dynamic LQG teams admit optimal policies under an expected average cost criterion, in its most general form, has not been conclusively addressed despite the presence of results which impose linearity (and sometimes stabilizability) a priori for the optimal policies under such decentralized information structures [105, 124, 87, 125]. We also establish some results in this direction, and provide sufficient conditions under which an optimal policy is linear.

Now, we provide literature review and motivation for the results in the second part of the thesis, where we study policy-independent and policy-dependent static reductions. As we have discussed above (and can be seen through our analysis of large stochastic teams), static reduction

is an effective method for arriving at existence, structural and approximation results in the context of stochastic teams [62, 140, 109, 138, 68] and for approximations with finite models to [112]. We note that although the purpose of two aforementioned types of static reductions is similar, their impact on the optimality notions (global optimality, person-by-person optimality, and stationarity) as well as convexity and variational analysis is significantly different from one to another and this distinction can be very subtle.

Toward an interpretation and explanation of the subtlety of the above connections, we note that in the language of stochastic control, an interpretation of policies for a dynamic team and its static reduction can be *(path-dependent) feedback* policies (policies that are functions of a subset of the history of states or a noisy observation of states, which may depend on actions of the preceding DMs) versus *noise feedforward* policies (policies that are functions of only disturbances) [24]. We also note that in the game theory literature (see, for example, mean-field games [39]), *closed-loop* policies (where policies are adapted processes to the filtration generated by states or observations, which are functions of preceding actions) can be viewed as policies for dynamic games. On the other hand, *open-loop* policies (where policies are adapted processes to the filtration generated by Brownian motions) can be viewed as policies for dynamic games under a static reduction. These connections are particularly important since the compactness and convexity conditions often can be established under more relaxed assumptions for open-loop policies, and hence, it can simplify the analysis for the convergence and existence for open-loop policies in the context of large-scale games and teams [116, 56, 81, 39, 72, 83]. In view of the above subtlety and interpretation of policies for dynamic team problems and their static reductions, we refer the reader to several papers mostly on games [132, 14, 9, 11, 10, 12, 15, 16, 46, 59, 113, 114]. In the second part of the thesis, we study the above subtleties toward establishing some connections between optimality notions of policies for stochastic dynamic teams and their policy-independent and policy-dependent static reductions.

1.5 Summary of Contributions and Organization of the Thesis

1.5.1 Chapter 2

In this chapter, we study stochastic static teams with a countably infinite number of decision makers, with the goal of obtaining (globally) optimal policies under a decentralized information structure. We present sufficient conditions to connect the concepts of team (global) optimality and person by person optimality for static teams with a countably infinite number of decision makers. We show that under uniform integrability and uniform convergence conditions, an optimal policy for static teams with a countably infinite number of decision makers can be established as the limit of sequences of optimal policies for static teams with N decision makers as $N \to \infty$. Under the presence of a symmetry condition, we relax the conditions and this leads to optimality results for a class of mean-field static team problems, where the existing results have been limited to person-by-person optimality and not global optimality (under strict decentralization). In particular, we establish the optimality of symmetric (i.e., identical) policies for such problems. We consider a number of illustrative examples where the theory is applied to setups with either infinitely many decision makers or an infinite-horizon stochastic control problem reduced to a static team. The results of Chapter 2 have appeared in part in [120, 117].

1.5.2 Chapter 3

This chapter studies convex stochastic dynamic team problems with finite and infinite time horizons under decentralized information structures. First, we introduce two notions called exchangeable teams and symmetric information structures. We show that in convex large-scale team problems an optimal policy exhibits a symmetry structure. We provide a characterization of such symmetrically optimal teams for a general class of convex dynamic team problems under a mild conditional independence condition. In addition, through concentration of measure arguments, we establish the convergence of optimal policies for teams with N decision makers to the corresponding optimal policies for symmetric mean-field teams with infinitely many decision makers. As a

by-product, we present an existence result for convex mean-field teams, where the main contribution of our results is with respect to the information structure in the system when compared with the related results in the literature that have either assumed a classical information structure or a static information structure.

We also apply these results to the important special case of Linear Quadratic Gaussian (LQG) team problems, where while for partially nested LQG team problems with finite time horizons it is known that the optimal policies are linear, for infinite horizon problems the linearity of optimal policies has not been established in full generality. We also study average cost finite and infinite horizon dynamic team problems with a symmetric partially nested information structure and obtain globally optimal solutions, where we establish linearity of optimal policies. The results of Chapter 3 have appeared in [118, 119].

1.5.3 Chapter 4

In this chapter, we study symmetric stochastic team problems and characterize existence and structural properties for (globally) optimal policies. Unlike Chapter 3, in this chapter, convexity of the cost and action spaces is not assumed. We first introduce a suitable topology on policies which then leads to a de Finetti type representation theorem for exchangeable decentralized policies, that is, for the probability measures induced by admissible relaxed control policies under decentralized information structures. This leads to a representation theorem for policies, which admit an infinite exchangeability condition. For a general setup of stochastic team problems with N decision makers, under exchangeability of observations of decision makers and the cost function, we show that without loss of global optimality, the search for optimal policies can be restricted to those that are N-exchangeable. Then, by extending N-exchangeable policies to infinitely exchangeable ones, establishing a convergence argument for the induced costs, and using the presented de Finetti type theorem, we establish the existence of an optimal decentralized policy for static and dynamic teams with a countably infinite number of decision makers, which turns out to be symmetric (i.e.,

identical) and randomized. Finally, we show near optimality of symmetric independently randomized policies for finite N-decision maker team problems and thus establish approximation results for N-decision maker weakly coupled stochastic teams. The results of Chapter 4 are available in [116, 115].

1.5.4 Chapter 5

In this chapter, we study optimality properties under policy-independent, policy-dependent static reductions, and static measurements with control-sharing reduction. Under the first type, we show that there is a bijection between person-by-person optimal (globally optimal, stationary) policies and their policy-independent static reductions. For the second type, although there is a bijection between globally optimal policies of dynamic teams with partially nested information structures and their static reductions, in general there is no bijection between stationary (person-by-person optimal) policies of dynamic teams and their policy-dependent static reductions. We present, however, sufficient conditions under which such a relationship holds and a functional form of the bijection can be constructed. Furthermore, we study the connections between optimality and stationarity concepts under static measurements with control-sharing reduction. An implication is a convexity characterization of dynamic team problems with a partially nested with control-sharing information structure. Finally, we study multi-stage team problems, where we discuss the connections between optimality concepts under the introduced static reductions via introducing two reduction. The results of Chapter 5 are available in [113]. We also note that results on games related to those in Chapter 5 are available in [114] (but these are not included in the thesis).

Chapter 2

Large-Scale Convex Stochastic Static Teams and Their Mean-Field Limit

In this chapter, we study sufficient conditions of (global) optimality, and existence and structural properties of (globally) optimal policies for large-scale stochastic static teams problems with finite as well as a countably infinite number of decision makers. We mainly focus on convex static teams, where the cost function satisfies an exchangeability property. The dynamic case will be studied in Chapter 3.

We first restate the setup for stochastic static teams with finite as well as infinitely many decision makers. Throughout this chapter, we assume that action and observation spaces are subsets of appropriate Euclidean spaces, $\mathbb{U}^i \subseteq \mathbb{R}^n$, and $\mathbb{Y}^i \subseteq \mathbb{R}^m$ for all $i \in \mathbb{N}$, where n and m are positive integers. In view of the intrinsic model described in the introductory chapter, the general stochastic static team problem with a finite number of decision makers (N-DM) is defined as:

Problem ($\mathcal{P}'_N$) Let $\mathcal{N} = \{1, \ldots, N\}$. Let $\underline{\gamma}_N := \{\gamma^i\}_{i\in\mathcal{N}} \in \Gamma_N$ where $\Gamma_N = \prod_{i=1}^N \Gamma^i$. Given a (decentralized) static information structure, a stochastic static team problem is described with an expected cost function (to be minimized) of $\underline{\gamma}_N$ as

$$J_N(\underline{\gamma}_N) := E^{\underline{\gamma}_N}\Big[c(\omega_0, u^1, \ldots, u^N)\Big] := E\Big[c(\omega_0, \gamma^1(y^1), \ldots, \gamma^N(y^N))\Big] \tag{2.1}$$

for some Borel measurable cost function $c : \Omega_0 \times \prod_{k=1}^N \mathbb{U}^k \to \mathbb{R}$. As before, we define ω_0 as the

Ω_0-valued cost function relevant exogenous random variable, where $\omega_0 : (\Omega, \mathcal{F}, P) \to (\Omega_0, \mathcal{F}_0)$, and Ω_0 is a Borel space with its Borel σ-field $\mathcal{F}_0$.

Let the action space and observation space be identical through DMs. Now, under an exchangeability of the cost function, a stochastic static team problems with finite (N) number of decision makers (also referred to as pre-limit mean-field stochastic static team problem) is defined as:

Problem ($\mathcal{P}_N$) Let $\mathcal{N} = \{1, \dots, N\}$. Let $\underline{\gamma}_N := \{\gamma^i\}_{i\in\mathcal{N}} \in \Gamma_N$ where $\Gamma_N = \prod_{i=1}^N \Gamma^i$. Given a (decentralized) static information structure, a stochastic static team problem is described with an expected cost function (to be minimized) of $\underline{\gamma}_N$ as

$$J_N(\underline{\gamma}_N) = \frac{1}{N} E^{\underline{\gamma}_N}\Big[\sum_{i=1}^N c(\omega_0, u^i, \frac{1}{N}\sum_{p=1}^N u^p)\Big], \tag{2.2}$$

for some Borel measurable cost function $c : \Omega_0 \times \mathbb{U} \times \mathbb{U} \to \mathbb{R}_+$.

We note that with slight abuse of notations, we used the same notation c (as (2.1)) for the cost function of (2.2). Now, a stochastic static team problem with a countably infinite number of decision makers (also referred to as a mean-field team problem) is described as:

Problem ($\mathcal{P}_\infty$) Let $\mathcal{N} = \mathbb{N}$. Let $\underline{\gamma} := \{\gamma^i\}_{i\in\mathcal{N}} \in \Gamma$ where $\Gamma = \prod_{i\in\mathcal{N}} \Gamma^i$. Given a (decentralized) static information structure, a stochastic static team problem is described with an expected cost function (to minimize) under a policy $\underline{\gamma}$ as

$$J(\underline{\gamma}) = \limsup_{N\to\infty} \frac{1}{N} E^{\underline{\gamma}}\Big[\sum_{i=1}^N c(\omega_0, u^i, \frac{1}{N}\sum_{p=1}^N u^p)\Big], \tag{2.3}$$

for some Borel measurable cost function $c : \Omega_0 \times \mathbb{U} \times \mathbb{U} \to \mathbb{R}_+$.

The focus of this chapter is to address the following questions:

Problem 1: Optimality conditions for optimal policies of ($\mathcal{P}_\infty$). Assume $c : \Omega_0 \times \mathbb{U} \times \mathbb{U} \to \mathbb{R}_+$ is (jointly) convex and continuously differentiable in its second and third arguments for every

ω_0. Given a policy $\underline{\gamma}^* \in \Gamma$, under what conditions the policy is (globally) optimal for $(\mathcal{P}_\infty)$? In particular, if a policy $\underline{\gamma}^*$ is person-by-person optimal (stationary), under what conditions $\underline{\gamma}^*$ is globally optimal? We note that, using variational analysis, the above question has been addressed for $(\mathcal{P}'_N)$ in [92, 102, 78], here, the extensions of the results to $(\mathcal{P}_\infty)$ are of interest.

Problem 2: Connections between optimal policies of $(\mathcal{P}_N)$ and $(\mathcal{P}_\infty)$. Let a sequence $\{\underline{\gamma}^*_N\}_N$, where for each $N \in \mathcal{N}$, a policy $\underline{\gamma}^*_N \in \Gamma_N$ is a globally optimal policy for a stochastic static team problem $(\mathcal{P}_N)$ be given. Instead of dealing with the variational analysis for the infinite dimensional problem $(\mathcal{P}_\infty)$, is it possible to obtain sufficient conditions of optimality for a policy $\underline{\gamma}^* \in \Gamma$ based on $\{\underline{\gamma}^*_N\}_N$ and its sequence of optimal expected cost $\{J_N(\underline{\gamma}^*_N)\}_N$? In particular, suppose $\{\underline{\gamma}^*_N\}_N$ converges pointwise (almost-surely for every observations) to a limit policy $\underline{\gamma}^*$. Is it the case that $\underline{\gamma}^*$ is (globally) optimal for $(\mathcal{P}_\infty)$? Under what conditions this holds?

Problem 3: Existence and structural properties of optimal policies for $(\mathcal{P}_N)$ and $(\mathcal{P}_\infty)$. The existence of (globally) optimal polices for $(\mathcal{P}'_N)$ has been established under very general setups of stochastic teams [140, 138, 62]. Assume observations of decision makers are identically distributed, and the information structure is decentralized and static. Then, since the cost function in $(\mathcal{P}_N)$ is of an exchangeable type (or in a more general setup, suppose a problem $(\mathcal{P}'_N)$ with an exchangeable cost function), is it the case that (globally) optimal polices for $(\mathcal{P}_N)$ are symmetric (identical) through decision makers? The question that we address is also on the existence and symmetry property of (globally) optimal policy for $(\mathcal{P}_\infty)$. In particular, under what conditions, there exists a symmetric globally (deterministic or randomized) optimal policy for $(\mathcal{P}_\infty)$?

2.1 Literature Review and Contributions

As it has been discussed in the introductory chapter, for teams with finitely many decision makers, the results in [92, 102, 78] showed that convexity of the cost function, subject to minor regularity conditions, may suffice for the global optimality of person-by-person-optimal solutions. However, these results are applicable for static teams with a finite number of decision makers. In this chapter,

the focus is on teams with a countably infinite number of decision makers. for team problems with infinitely many decision makers, a related set of results involves those in mean-field games [70, 69, 86, 19, 40, 19, 40, 110, 43] and in the economic theory literature, [122, 93] (focusing on Cournot-Nash equilibria). There have also been several studies for mean-field games where the limits of sequences of Nash equilibria have been investigated as the number of decision makers tends to infinity (see e.g., [56, 83, 20, 86, 7]). These results, while very useful for establishing equilibria or in the context of team problems, person-by-person-optimal policies, does not guarantee the global optimality among all policies. Compared with [20, 54, 56, 27, 7] where only the convergence of a sequence of Nash equilibria for symmetric games with the mean-field interaction has been studied, we show that, under sufficient conditions, sequences of optimal policies for teams with N number of decision makers as $N \to \infty$ converge to a team optimal policy for static teams with a countably infinite number of decision makers.

Mean-field team and social optima problems have also been studied: Social optima for mean-field LQG control problems under both centralized and a specific decentralized information structure have been considered in [71, 127]. In [6], a setup is considered where decision makers share some information on the mean-field in the system, and through showing that the performance of a corresponding centralized system can be realized under a decentralized information structure, global optimality is established. In this chapter, we follow an approach where optimality for every N is established and also optimality holds as $N \to \infty$ for the limit policy. The papers [74, 73] have studied a continuous-time setup where a major agent is present; by considering the social impact for each individual player, they showed person-by-person optimal policies asymptotically minimize the social cost [71]. By approximating the mean-field term, the authors bound the induced approximation error of order $O(N^{\frac{-1}{2}} + \epsilon_N)$ where ϵ_N goes to zero as the number of players $N \to \infty$ [71]. In [34], mean-field team problems with mixed players have been considered where minor agents act together to minimize a common cost against a major player. Also, for the LQ setup, under the assumption that DMs apply identical policy in addition to some technical assumptions on the cost function and transition probabilities of Markov chains, [5] showed that the

expected cost achieved by a sub-optimal decentralized strategy is ϵ_N close to the optimal cost achieved when mean-field term (empirical distribution of states) has been shared. Such results on mean-field teams either show global optimality through equivalence to the performance of a centralized setup (considering specific sharing patterns on the mean-field model) or typically only assume person-by-person-optimality. In this chapter, we will establish global optimality under a completely decentralized information structure; however, certain technical conditions will be imposed.

A number of studies have focused on the LQG setup (in addition to [71, 127]). A close study is [90] where LQG static teams with a countably infinite number of decision makers have been studied, and sufficient conditions for global optimality have been established. In this chapter, we utilize some of the results from [90], however compared with [90], we propose sufficient conditions of team optimality for average cost problems in a more general setup: except convexity, no specific structure is presumed a priori on the cost function. For our analysis, we do not restrict our setup to the LQG case, where often direct methods can be applied building on [102], [78], and operator theory involving matrix algebra. More importantly, we study the mean-field setting and establish symmetry properties. In fact, for a general setup of static teams, we introduce sufficient conditions (see Theorem 2.2.2 and Theorem 2.2.3) such that the optimal cost and optimal policies of static teams with a countably infinite number of decision makers is obtained as a limit of the optimal cost and optimal policies for static teams with N number of decision makers as $N \to \infty$. In [60], LQG team problems with infinitely many decision makers have been studied for a setup, where the cost function is the expected inner-product of an infinite dimensional vector (and to allow for a Hilbert theoretic formulation, finiteness of the infinite sum of the moments of individual random variables is imposed) and linearity and uniqueness of optimal policies have been established; the finiteness (of the infinite summation) restriction rules out the setup in this chapter. In [99], infinite horizon decentralized stochastic control problems containing a remote controller and a collection of local controllers dealing with linear models have been addressed for a setup where the cost is quadratic and the communication model satisfies a specified sharing pattern of information between local

controller and remote controller. Under the assumed sharing pattern (common information), the connections between the optimal solution and the coupled algebraic Riccati equation for Markov jump linear systems and its convergence to the coupled fixed point equations have been utilized to show the optimality of the solution [99].

Contributions. In view of the above literature review and introduction, this chapter makes the following contributions:

(i) For a general setup of static teams, we show that (see Theorem 2.2.3), under a uniform integrability condition (see Remark 2.2.2), if sequences of team optimal policies of decision makers $i = 1, \ldots, N$ of static teams with N number of decision makers converge uniformly in $i = 1, \ldots, N$ (see (b) in Theorem 2.2.3), then the corresponding limit policies are team optimal for the static team with a countably infinite number of decision makers.

(ii) We establish global optimality results for mean-field teams under strict decentralization of the information structure for both teams with large numbers of players and infinitely many players. Toward this end, we introduce a notion of symmetrically optimal teams (see Definition 2.3.1) to obtain a global optimality result under relaxed sufficient conditions. Under mild conditions on action spaces and observations of decision makers, through concentration of measures arguments, we establish the convergence of optimal policies for symmetric mean-field teams with N decision makers to the corresponding optimal policy of mean-field teams. In addition, we establish an existence result for optimal policies on mean-field teams under relaxed conditions on action spaces and the cost function (see Theorem 2.3.6).

(iii) We apply our results to a number of illustrative examples: We first consider LQG and LQ (non-Gaussian) average cost problems with state coupling (see Section 2.4.1 and Section 2.4.2). We also consider LQG average cost problems with control coupling (see Section 2.4.3). In addition, we show that the team optimal policy of LQG teams with classical information structure (see Section 2.4.5) is obtained using the technique proposed in this chapter. This is important since this result, while well-known in the stochastic control literature, has

not been investigated using static reduction and hence this approach can be viewed as a step to address optimal solutions for infinite-horizon partially nested dynamic LQG problems which can be reduced to a static team with a countably infinite number of decision makers (more on this is to be discussed in Chapter 3, Section 3.2).

The organization of the chapter is as follows. In Section 2.1.1, we review the results on convex static teams. Section 2.2 contains our main results including sufficient conditions for team optimality and asymptotic optimality for a general setup of static teams with a countably infinite number of decision makers. Section 2.3 discusses symmetric and mean-field teams, and applications are presented in Section 2.4. Proofs are provided in Section 2.5.

2.1.1 Convex Static Teams with a Finite Number of Decision Makers

In this subsection, we review the relevant results on convex static teams with a finite number of decision makers [102, 78]. We first recall the definition of stationary policies for team problems.

Definition. *Stationary solution [102].*
A team decision rule $\underline{\gamma}_N(.)$ is stationary if $J(\underline{\gamma}_N) < \infty$, and for all $i = 1, ..., N$, P-almost surely

$$\nabla_{u^i} E\left[c(\omega_0, (\underline{\gamma}_N^{-i}, u^i))\middle| y^i\right]\Big|_{u^i=\gamma^i(y^i)} = 0,$$

where ∇_{u^i} denotes the gradient with respect to u^i.

In this subsection, with slight abuse of notations, we sometimes used γ^i as $\gamma^i(y^i)$. In the following, we review some results for stochastic static teams ($\mathcal{P}'_N$) with N decision makers. The following is known as Radner's theorem [102]. Radner proposed the first result to connect the stationarity concept and global team optimality.

Theorem 2.1.1. *[102] If*

(a) $c(\omega_0, \underline{u}_N)$ is convex and differentiable in $\underline{u}_N$ for P- almost surely;

(b) $\inf_{\underline{\gamma}_N \in \mathbf{\Gamma_N}} J_N(\underline{\gamma}_N) > -\infty$*;*

(c) $J_N(.)$ *is locally finite at* $\underline{\gamma}_N^*$ *[102];*

(d) $\underline{\gamma}_N^*$ *is stationary;*

then $\underline{\gamma}_N^*$ *is globally optimal for* $(\mathcal{P}'_N)$*.*

Radner's theorem fails in some applications because of the restrictive local finiteness assumption. Krainak et al [78] relaxed assumptions and presented sufficient conditions for team optimality on static teams.

Theorem 2.1.2. *[78] Assume that, for every fixed* ω_0*,* $c(\omega_0, \underline{u}_N)$ *is convex differentiable in* $\underline{u}_N$*. Suppose (b) in Theorem 2.1.1 holds. Let* $\underline{\gamma}_N^* \in \mathbf{\Gamma_N}$*, and assume that* $E[c(\omega_0, \underline{\gamma}_N^*(\underline{y}_N))] < \infty$*. If, for all* $\underline{\gamma}_N \in \mathbf{\Gamma_N}$ *with* $E[c(\omega_0, \underline{\gamma}_N(\underline{y}_N))] < \infty$*,*

$$E\bigg[\sum_{i=1}^{N} c_{u^i}(\omega_0, \underline{\gamma}_N^*)(\gamma^i - \gamma^{i*})\bigg] \geq 0, \tag{2.4}$$

where $c_{u^i}(\omega_0, \underline{\gamma}_N^*)$ *is the partial derivative of* $c(\omega_0, \underline{u}_N)$ *with respect to* u^i *valued in* $\underline{u}_N = \underline{\gamma}_N^*$*, then* $\underline{\gamma}_N^*$ *is an optimal team policy for* $(\mathcal{P}'_N)$*. Moreover, if* $c(\omega_0, \underline{u}_N)$ *is strictly convex in* $\underline{u}_N$ P*-almost surely, then* $\underline{\gamma}_N^*$ *is* P*-a.s. unique.*

Since the set of admissible policies is generally uncountable, checking (2.4) is difficult. Krainak et al [78] further developed relaxed conditions under which stationarity of a team decision rule implies its optimality.

Theorem 2.1.3. *[78] Assume, for every fixed* $\omega_0 \in \Omega_0$*, that* $c(\omega_0, \underline{u}_N)$ *is a convex differentiable function of* $\underline{u}_N$ *and suppose (b) in Theorem 2.1.1 holds. Assume that* $\underline{\gamma}_N^* \in \mathbf{\Gamma_N}$ *is a stationary team decision rule. Let, for all* $\underline{\gamma}_N \in \mathbf{\Gamma_N}$ *with* $E[c(\omega_0, \underline{\gamma}_N(\underline{y}_N))] < \infty$*,*

$$E\bigg[c_{u^i}(\omega_0, \underline{\gamma}_N^*)(\gamma^i - \gamma^{i*})\bigg] < \infty \; for \; i = 1, ..., N. \tag{2.5}$$

Then $\underline{\gamma}_N^*$ is a team optimal policy for ($\mathcal{P}'_N$). If $c(\omega_0, \underline{u}_N)$ is strictly convex in $\underline{u}_N$, P-a.s., then $\underline{\gamma}_N^*$ is unique.

Furthermore, (2.5) can be replaced by the following more checkable conditions [139]: Let Γ^i be a Hilbert space consisting of some random variables $\gamma^i(y^i)$ such that for each $i = 1, ..., N$ and $E[c(\omega_0, \underline{\gamma}_N(\underline{y}_N))] < \infty$ for all $\underline{\gamma}_N \in \mathbf{\Gamma_N}$. Moreover, let

$$E\left[c_{u^i}(\omega_0, \underline{\gamma}_N^*)\middle| y^i\right] \in \Gamma^i,\ i = 1, ..., N. \tag{2.6}$$

The above conditions follows directly from (2.5) when Γ^i is a Hilbert space for all $i = 1, 2, \ldots, N$. This condition can be checked for some applications; for example, LQ teams [139].

2.2 Optimal Policies for Teams with Infinitely Many Decision Makers

2.2.1 Sufficient Conditions of Optimality

In the following, we propose sufficient conditions of team optimality for $(\mathcal{P}_\infty)$. We often follow [78], and the result is an extension of [78] to a general setup of static teams with a countably infinite number of decision makers. We also note a related analysis in [90]. We will use the following theorem for LQ static teams with a countably infinite number of decision makers (see Section 2.4.2).

Assumption 2.2.1. *Let*

(A1) $c(\omega_0, u^i, \frac{1}{N}\sum_{p=1}^N u^p)$ *be a $\mathbb{R}_+$-valued jointly convex function of second and third arguments and differentiable in u^i with continuous partial derivatives, for every $\omega_0 \in \Omega_0$.*

(A2) for some $\underline{\gamma}^* \in \mathbf{\Gamma}$,

$$\lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^N E^{\underline{\gamma}^*}\left[c(\omega_0, u^i, \frac{1}{N}\sum_{p=1}^N u^p)\right] < \infty. \tag{2.7}$$

We note that the cost function is differentiable in u^i which means that the cost is totally differentiable in u^i, i.e., $\frac{d}{du^i}c(\omega_0, u^i, \frac{1}{N}\sum_{p=1}^{N} u^p) = \frac{\partial}{\partial u^i}c(\omega_0, u^i, \mu_N) + \frac{1}{N}\frac{\partial}{\partial \mu_N}c(\omega_0, u^i, \mu_N)$ exists.

Theorem 2.2.1. *Assume (A1) holds and (A2) holds for $\underline{\gamma}^* \in \mathbf{\Gamma}$. If for all $\underline{\gamma} \in \mathbf{\Gamma}$ with $J(\underline{\gamma}) < \infty$,*

$$\limsup_{N\to\infty} \frac{1}{N} E\Bigg[\sum_{i=1}^{N}\sum_{k=1}^{N} c_{u^k}(\omega_0, \gamma^{i*}, \mu^*)(\gamma^k - \gamma^{k*})\Bigg] \geq 0, \tag{2.8}$$

where $\mu_N^* = \frac{1}{N}\sum_{p=1}^{N}\gamma^{p*}(y^p)$, then $\underline{\gamma}^*$ is a globally optimal team policy for ($\mathcal{P}_\infty$).

In some applications, (2.8) can be difficult to check since it must be satisfied for all $\underline{\gamma} \in \mathbf{\Gamma}$ with $J(\underline{\gamma}) < \infty$. In the next section, we address this issue by introducing a constructive approach for static teams with a countably infinite number of decision makers as a limit of a sequence of team optimal policies of the corresponding static teams with a finite number of decision makers. In the following, we propose sufficient conditions to approximate the optimal cost and a team optimal policy for static teams with a countably infinite number of decision makers using the optimal cost and an optimal policy for static teams with N decision makers. We note that our first result here is based on [90, Theorem 1], which considered an equality. We denote $\underline{\gamma}|_N \in \mathbf{\Gamma_N}$ as a restriction of $\underline{\gamma} \in \mathbf{\Gamma}$ to the first N components.

Theorem 2.2.2. *Let $\underline{\gamma}_N^* \in \mathbf{\Gamma_N}$ be an optimal policy for ($\mathcal{P}_N$) as* (2.2) *(see [78, 62, 140] for sufficient conditions). If there exists $\underline{\gamma}^* \in \mathbf{\Gamma}$, with $J(\underline{\gamma}^*) < \infty$, satisfying*

$$\limsup_{N\to\infty} J_N(\underline{\gamma}_N^*) \geq J(\underline{\gamma}^*), \tag{2.9}$$

then $\underline{\gamma}^*$ is a globally team optimal policy for ($\mathcal{P}_\infty$).

Remark 2.2.1. *Under (A2), one can replace* (2.9) *with*

$$\limsup_{N\to\infty} \frac{1}{N}\sum_{i=1}^{N}\Bigg\{E^{\underline{\gamma}_N^*}\Big[c(\omega_0, u^i, \mu_N)\Big] - E^{\underline{\gamma}^*}\Big[c(\omega_0, u^i, \mu_N)\Big]\Bigg\} \geq 0. \tag{2.10}$$

The above theorem and remark will be useful for some applications (see for example Section 2.4.5).

2.2.2 Asymptotically Optimal Policies as a Limit of Finite Team Optimal Policies

In the following, we present a sufficient condition for (2.9). The following result also presents a constructive method to obtain optimal policies using asymptotic analysis.

Theorem 2.2.3. *Assume*

(a) for every N, there exist $\underline{\gamma}_N^* \in \mathbf{\Gamma_N}$ *for* $(\mathcal{P}_N)$ *(see* (2.2)*),*

(b) let $\omega \in B$ *for some* $B \in \mathcal{F}$ *event of* P *measure one, for every fixed* $y^i(\omega)$, $\gamma_N^{i*}(y^i)$ *converges to* $\gamma_\infty^{i*}(y^i)$ *uniformly in* $i = 1, 2, \dots, N$*, i.e.,*

$$\lim_{N\to\infty} \sup_{1\le i\le N} ||\gamma_N^{i*}(y^i) - \gamma_\infty^{i*}(y^i)|| = 0 \; P-a.s.,$$

where $||\cdot||$ *is the sup norm,*

(c) there exists a P*-integrable function* $g(\omega_0, \underline{y})$ *such that, for every* N*,*

$$\frac{1}{N}\sum_{i=1}^{N} c\left(\omega_0, \gamma_\infty^{i*}(y^i), \frac{1}{N}\sum_{p=1}^{N}\gamma_\infty^{p*}(y^p)\right) \le g(\omega_0, \underline{y}),$$

where $\underline{y} = (y^1, y^2, \dots)$*, then* $\underline{\gamma}^*$*, a team optimal policy for* $(\mathcal{P}_\infty)$*, is a pointwise limit of* $\underline{\gamma}_N^*$*, an optimal policy for* $(\mathcal{P}_N)$*, i.e.,* $\gamma^{i*}(y^i) = \lim_{N\to\infty} \gamma_N^{i*}(y^i) = \gamma_\infty^{i*}(y^i)$ P*-almost surely.*

Remark 2.2.2. *One can relax conditions in Theorem 2.2.3 as follows:*

(i) relax (a) by considering a sequence of ϵ_N*-optimal policy, where* ϵ_N *are non-negative and converges to zero as* $N \to \infty$*,*

(ii) relax (c) with a uniform integrability condition which is satisfied if the following expression is finite (see [26, Theorem 3.5]),

$$\sup_{N\geq 1} E\left[\left|\frac{1}{N}\sum_{i=1}^{N} c\left(\omega_0, \gamma_\infty^{i*}(y^i), \frac{1}{N}\sum_{i=1}^{N}\gamma_\infty^{i*}(y^i)\right)\right|^{1+\epsilon}\right],$$

for some $\epsilon > 0$. This new condition can be checked in some applications (see Section 2.4). The result follows from [26, Theorem 3.5],

(iii) relax the P-almost sure convergence in (b) by considering convergence in probability, i.e.,

$$\lim_{N\to\infty} P(\sup_{1\leq i\leq N} ||\gamma_N^{i*}(y^i) - \gamma_\infty^{i*}(y^i)||\geq \epsilon) = 0,$$

hence similar to the proof of Theorem 2.2.3, Step 1, using continuous mapping theorem (see for example, [26]), we show that $c(\omega_0, \gamma_N^{i*}(y^i), \mu_N^*)$ converges to $c(\omega_0, \gamma_\infty^{i*}(y^i), \lim_{N\to\infty}\mu_\infty^*)$ in probability. Similarly, the result of Step 2 holds in probability. Using [26, Theorem 3.5], under the uniform integrablity of $X_N := \frac{1}{N}\sum_{i=1}^{N} c\left(\omega_0, \gamma_\infty^{i*}(y^i), \frac{1}{N}\sum_{i=1}^{N}\gamma_\infty^{i*}(y^i)\right)$ and under convergence in probability X_N to $X := \lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N} c\left(\omega_0, \gamma_\infty^{i*}(y^i), \frac{1}{N}\sum_{i=1}^{N}\gamma_\infty^{i*}(y^i)\right)$, we can conclude $E(X_N)\to E(X)$. This relaxation can be useful when the weak law of large numbers can be invoked to check (c), but the strong law of large numbers fails to apply.

We apply the results of this section to two examples in Sections 2.4.1 and 2.4.2.

2.3 Globally Optimal Policies for Mean-Field Teams

2.3.1 Symmetric (Exchangeable) Teams

In the following, we present sufficient conditions for team optimality in symmetric and mean-field teams. The concept of symmetry has been studied in a variety of contexts; see e.g., [96], [47] and many others.

Definition. *(Exchangeable teams)*

An N-DM team is exchangeable if the value of the expected cost function (see (2.1)*) is invariant under every permutation of policies.*

We note that it is also called *totally symmetric* in a game theoretic context (see for example [47]).

Definition. *(Symmetrically optimal teams)*
A team is symmetrically optimal, if for every given policy, there exists an identically symmetric policy (i.e., each DM has the same policy) which performs at least as good as the given policy.

In the following, we characterize the symmetry of the general setup for ($\mathcal{P}'_N$) (see (2.1)). Clearly, the result will also hold for the ($\mathcal{P}_N$) (see (2.2)).

Lemma 2.3.1. *For a fixed N, consider an N-DM team defined as ($\mathcal{P}'_N$) (see* (2.1)*) and let the cost function be a convex function of $\underline{u}_N$ P-almost surely. Assume the cost function is exchangeable P-almost surely with respect to the actions. If the action space $\mathbb{U}$ is convex, and observations of DMs are exchangeable conditioned on ω_0, then the team is symmetrically optimal.*

In the following, we present another characterization of symmetrically optimal teams; this looks to be an immediate result; however, a proof is included for completeness since we could not find an explicit reference (this proof will also be utilized in Chapter 4).

Lemma 2.3.2. *For a fixed N, consider an N-DM team defined as ($\mathcal{P}'_N$) (see* (2.1)*) and let the cost function be a convex function of $\underline{u}_N$ P-almost surely. Assume the action space for each DM is convex. If the expected cost function (see* (2.1)*) is exchangeable with respect to the policies, then the team is symmetrically optimal.*

Proof. Let $\underline{\gamma}^*_N = (\gamma^{1*}, \gamma^{2*}, \ldots, \gamma^{N*})$ be a given policy for ($\mathcal{P}'_N$). According to the definition of exchangeable teams, any permutation of policies, say $\hat{\underline{\gamma}}^*_N = (\gamma^{i_1*}, \gamma^{i_2*}, \ldots, \gamma^{i_N*})$, fails to change the value of the expected cost function, and hence achieve the same expected cost as the one induced by $\underline{\gamma}^*_N$. Consider $\tilde{\underline{\gamma}}_N$ as a uniform randomization among all possible permutations of

optimal policies, since $\mathbb{U}$ is convex then $\tilde{\underline{\gamma}}_N$ is a control policy. By convexity of the cost function, through Jensen's inequality, and the fact that any permutation of optimal policies preserves the value of the cost function, we have $J_N(\tilde{\underline{\gamma}}_N) \leq J_N(\underline{\gamma}^*_N)$. Since $\tilde{\underline{\gamma}}_N$ is also identically symmetric, the proof is completed. □

Now, we characterize symmetrically optimal teams for $(\mathcal{P}_N)$ (see (2.2)).

Theorem 2.3.1. *Consider an N-DM team defined as $(\mathcal{P}_N)$ (see* (2.2)*). Let action spaces be convex and the cost function be convex in the second and third arguments P-almost surely. If observations are exchangeable conditioned on ω_0, then the team is symmetrically optimal.*

Proof. The cost function defined in $(\mathcal{P}_N)$ is exchangeable in actions, hence under convexity of the action spaces and the cost function and following from the hypothesis that observations are exchangeable condition on ω_0, the proof is completed using Lemma 2.3.1. □

Theorem 2.3.1 will be utilized in our analysis to follow.

2.3.2 Optimal Solutions for Mean-field Teams as Limits of Optimal Policies for Finite Symmetric Teams

In the following, we present results for symmetrically optimal static teams. First, we focus on the case that the observations of decision makers are identical and independent, then we deal with non-identical and dependent observations under additional assumptions.

Our next theorem, under the assumption that observations are independent and identically distributed, utilizes a measure concentration argument to establish a convergence result.

Theorem 2.3.2. *Consider a team defined as $(\mathcal{P}_\infty)$ (see* (2.3)*) with the convex cost function in the second and third arguments P-almost surely. Let the action space be compact and convex for each decision maker, and y^is be i.i.d. random variables. If there exists a sequence of optimal policies for $(\mathcal{P}_N)$ (see* (2.2)*), $\{\gamma^*_N\}_N$, which converges (for every decision maker due to the symmetry) pointwise to γ^*_∞ as $N \to \infty$, then γ^*_∞ (which is identically symmetric) is an optimal policy for $(\mathcal{P}_\infty)$.*

Remark 2.3.1. *In view of the analysis in the proof of the preceding Theorem, we can observe that if P-almost surely the sequence $\{Q_N\}_N$ converges weakly to Q, then Theorem 2.3.2 can be generalized to a class of team problems defined as ($\mathcal{P}_\infty$) (see* (2.3)*) which may include ones with a non-convex cost function and/or the ones with conditionally non-exchangeable observations: This relaxation contains a class of problems (see e.g. Example 4 in Section 2.4.4) where one can consider a sequence of N-DM teams which admits asymmetric optimal policies that define ($\mathcal{P}_N$) (see* (2.2)*), but whose limit is identified with ($\mathcal{P}_\infty$) under an optimal sequence of policies.*

In the following, we relax the hypothesis that observations of decision makers are independent.

Proposition 2.3.1. *Consider a team defined as ($\mathcal{P}_\infty$) (see* (2.3)*) with the convex cost function in the second and third arguments P-almost surely. Let the action space be compact and convex for each decision maker, and $y^i = h(x, z^i)$, where z^is are i.i.d. random variables. If there exists a sequence of optimal policies for ($\mathcal{P}_N$), $\{\gamma_N^*\}_N$, which converges pointwise to γ_∞^* as $N \to \infty$, then γ_∞^* (which is identically symmetric) is an optimal policy for ($\mathcal{P}_\infty$).*

Remark 2.3.2. *Existence of optimal policies for ($\mathcal{P}_N$) and dynamic teams satisfying static reduction have been studied in [138] and [62]. In [138, Theorem 4.8], the existence of optimal policies achieved under σ-compactness of each decision maker's action space and under mild conditions on the control law and the cost function. Hence the existence of identically symmetric optimal policies for ($\mathcal{P}_N$) (see* (2.2)*) follows from symmetry and [138, Theorem 4.8]; thus, the existence result for ($\mathcal{P}_\infty$) is obtained under assumptions of Theorem 2.3.2.*

In the following, action spaces need not be compact; this is particularly important for LQG models as we will see in the next section.

Theorem 2.3.3. *Consider a team defined as ($\mathcal{P}_\infty$) (see* (2.3)*) with the convex cost function in the second and third arguments P-almost surely. Let the action spaces be convex for each decision maker. Let y^is be i.i.d. random variables. If there exists a sequence of optimal policies for ($\mathcal{P}_N$) (see* (2.2)*), $\{\gamma_N^*\}_N$, which converges pointwise to γ_∞^* as $N \to \infty$, and*

(A3) for some $\delta > 0$, $\sup_{N \geq 1} E[||\gamma_N^*(y^1)||^{1+\delta}] < \infty$,

then γ_∞^* *(which is identically symmetric) is an optimal policy for* $(\mathcal{P}_\infty)$.

In the following, we present a result for monotone mean-field coupled teams.

Theorem 2.3.4. *Consider a team defined as* $(\mathcal{P}_\infty)$ *(see* (2.3)*) with the convex cost function in the second and third arguments* P*-almost surely. Let the action spaces be convex for each decision maker. Let the cost function be increasing in the last argument, and* y^i*s be i.i.d. random variables. If there exists a sequence of optimal policies for* $(\mathcal{P}_N)$, $\{\gamma_N^*\}_N$ *(see* (2.2)*), which converges pointwise to* γ_∞^* *then* γ_∞^* *as* $N \to \infty$ *(which is identically symmetric) is an optimal policy for* $(\mathcal{P}_\infty)$.

In the following, observations need not be identical or independent.

Theorem 2.3.5. *Consider a team defined as* $(\mathcal{P}_\infty)$ *(see* (2.3)*) with the convex cost function in the second and third arguments* P*-almost surely. Let the action spaces be convex for each decision maker. Let (a), and (c) in Theorem 2.2.3 hold, and let observations be exchangeable conditioned on* ω_0. *Assume there exists a sequence* $\{\gamma_N^*\}_N$ *converges pointwise to* γ_∞^* *as* $N \to \infty$, *and let* P*-a.s.*

$$\left|\left|\gamma_N^*(y^i) - \gamma_\infty^*(y^i)\right|\right|^2 \leq \frac{f(y^i)h(N)}{N}, \tag{2.11}$$

where $\lim_{N\to\infty} N^{-1} \sum_{i=1}^N f(y^i) < \infty$ *and* $\lim_{N\to\infty} h(N) = 0$. *Then, a team optimal policy for* $(\mathcal{P}_\infty)$ *is symmetrically optimal and an optimal policy is identified as a limit of a sequence of team optimal policies for* $(\mathcal{P}_N)$ *(see* (2.2)*) as* $N \to \infty$.

2.3.3 An Existence Theorem on Globally Optimal Policies for Mean-Field Teams

An implication of our analysis is the following existence result on globally optimal policies for mean-field problems. In Theorem 2.3.2, we showed that if a pointwise limit as $N \to \infty$ of a sequence of optimal policies for $(\mathcal{P}_N)$ (see (2.2)) exists, this limit is a globally optimal policy for $(\mathcal{P}_\infty)$, but under the conditions stated in the following theorem, an existence result also can be

established. In the following, we relax the assumption that there exists a pointwise convergence sequence of optimal policies for $(\mathcal{P}_N)$ (see (2.2)). For the following theorem, we do not establish the pointwise convergence; but clearly if a sequence of optimal policies for $(\mathcal{P}_N)$ (see (2.2)) converges pointwise, a global optimal policy exists. Let $Q_N(B) := \frac{1}{N}\sum_{i=1}^{N} \delta_{\zeta_N^i}(B)$, where $B \in \mathcal{Z} := \mathbb{U}\times\mathbb{Y}$, and $\zeta_N^i := (\gamma_N^*(y^i), y^i)$.

Theorem 2.3.6. *Consider $(\mathcal{P}_\infty)$ (see* (2.3)*) with the convex cost function in the second and third arguments P-almost surely. Let the action spaces be convex for each decision maker. Assume further that, without any loss, the optimal control laws can be restricted to those with $E[\phi_i(u^i)] \leq K$ for some finite K, where $\phi_i : \mathbb{U}^i \to \mathbb{R}_+$ is lower semi-continuous. If y^is are i.i.d. random variables, then there exists an optimal policy for $(\mathcal{P}_\infty)$.*

We note that the limit policy is not necessarily deterministic according to the above result; this interesting discussion is left open for further study.

We apply the results of this section in Section 2.4.3.

2.4 Examples

In the following, we present a number of examples to demonstrate results in previous sections. We also investigate dynamic infinite-horizon average cost LQG teams with the classical information structure.

2.4.1 Example 1: Static quadratic Gaussian teams with coupling between states

Consider the following observation scheme,

$$y^i = x^i + z^i, \tag{2.12}$$

where $\{z^i\}_{i\in\mathbb{N}}$ and $\{x^i\}_{i\in\mathbb{N}}$ are i.i.d. zero mean Gaussian random variables. Let $\{z^i\}_{i\in\mathbb{N}}$ be independent of $\{x^i\}_{i\in\mathbb{N}}$. The expected cost function is defined as

$$J(\underline{\gamma}) = \limsup_{N\to\infty} \frac{1}{N} E^{\underline{\gamma}}\Big[\sum_{i=1}^{N} R(u^i)^2 + Q(u^i - x^i - \mu_N)^2\Big], \tag{2.13}$$

where $\mu_N := \frac{1}{N}\sum_{k=1}^{N} x^k$ and let R be a positive number and Q be a non-negative number.

Theorem 2.4.1. *For LQG static teams as formulated above, under the measurement scheme* (2.12)*, $\gamma_\infty^{i*}(y^i)$ is globally optimal for ($\mathcal{P}_\infty$) achieved as the limit $N \to \infty$ of $\gamma_N^{i*}(y^i)$, an optimal solution for ($\mathcal{P}_N$).*

2.4.2 Example 2: Static non-Gaussian teams with coupling between states

Let the observation scheme be (2.12), where $\{z^i\}_{i\in\mathbb{N}}$ and $\{x^i\}_{i\in\mathbb{N}}$ are i.i.d. zero mean random variables with finite variance. Let $\{z^i\}_{i\in\mathbb{N}}$ be independent of $\{x^i\}_{i\in\mathbb{N}}$. The expected cost function is defined as (2.13). Let R be a positive number and Q be a non-negative number.

Theorem 2.4.2. *For LQ static teams as formulated above, under the measurement scheme* (2.12)*, $\gamma_\infty^{k*}(y^k) = (R+Q)^{-1}QE[x^k|y^k]$ is globally optimal for ($\mathcal{P}_\infty$) and is obtained as the limit of $\gamma_N^{k*}(y^k)$ as $N \to \infty$.*

2.4.3 Example 3: LQG symmetric teams with coupling between control actions

Let

$$\underline{y}^i = H^i\underline{x} + \underline{z}^i, \tag{2.14}$$

where $\{\underline{z}^i\}_{i\in\mathbb{N}}$ is independent zero mean Gaussian random vectors also independent of $\underline{x}$, with covariance $\Sigma_{jj} = N^0 > 0$. Define $\omega = (\underline{x}, \underline{z}^1, \underline{z}^2, \dots)$, and $\omega_0 := \underline{x}$ where $\underline{x}$ is a Gaussian random vector with covariance $E[\underline{x}\underline{x}^T] = \Sigma_{00}$. Let

$$J(\underline{\gamma}) = \limsup_{N\to\infty} \frac{1}{N} E^{\underline{\gamma}}\Big[\sum_{i=1}^{N}(u^i)^T R u^i - 2\sum_{i=1}^{N}(u^i)^T D(\underline{x} + \frac{1}{N}\sum_{k=1}^{N} u^k)$$

$$+ (\underline{x} + \frac{1}{N}\sum_{k=1}^{N} u^k)^T Q(\underline{x} + \frac{1}{N}\sum_{k=1}^{N} u^k)\Big], \tag{2.15}$$

where R is an appropriate dimension positive definite matrix and D, and Q are appropriate dimension positive semi-definite matrices, and $R > 2D$. In the following, we follow steps in [139, Theorem 2.6.8] to obtain optimal policies for $(\mathcal{P}_N)$.

Lemma 2.4.1. *Consider an N-DM LQG team formulated above, under the measurement scheme* (2.14)*, the global optimal policy for $(\mathcal{P}_N)$ is linear, i.e., $\gamma_N^{k*}(y^k) = \pi_N^k y^k$. Here, $\pi_N^k \in \mathcal{M}_{n,m}(\mathbb{R})$, $n \times m$ real-valued matrix, is obtained by solving the following parallel update scheme,*

$$\pi_{N,(i)}^k = -L_N\Big[S^k + \frac{1}{N}\sum_{p=1,p\neq k}^{N} \pi_{N,(i)}^p H^p S^k\Big], \tag{2.16}$$

where $L_N := (R + \frac{Q}{N^2} - \frac{2D}{N})^{-1}(\frac{Q}{N} - D)$, $S^k := \Sigma_{00}(H^k)^T(H^k\Sigma_{00}(H^k)^T + \Sigma_{kk})^{-1}$ and the initial points of the iterations are considered as zero functions.

Proof. By Definition 2.1.1, stationary policies satisfy the following equality for $k = 1, \ldots, N$,

$$M\gamma_N^{k*}(y^k) + \left(\frac{Q}{N} - D\right)\left[E[\underline{x}|y^k] + \frac{1}{N}\sum_{p=1,p\neq k}^{N} E\left[\gamma_N^{p*}(y^p)\Big|y^k\right]\right] = 0, \tag{2.17}$$

where $M := R + \frac{Q}{N^2} - \frac{2D}{N}$, and (2.17) can be rewritten as $P\hat{\underline{R}}\underline{\gamma}_N^*(\underline{y}) + Pr(\omega) = 0$, where P is a block diagonal matrix with ii-th block $P_{ii}\beta^i(\omega) := E[\beta^i(\omega)|y^i]$, $\hat{\underline{R}}$ is a matrix where $\hat{R}_{ii} := M$ and $\hat{R}_{ij} := \frac{1}{N}(\frac{Q}{N} - D)$ for every $i, j = 1, ..., N$, $j \neq i$, and $r(\omega) = \underline{x}$. Note that P is a projection operator defined on a Hilbert space whose operator norm is one. Now, we use the successive approximation method [139, Theorem A.6.4]. According to (2.17), we can write for $k = 1, 2, ..., N$

$$\begin{aligned} &M\gamma_{N,(i)}^{k*}(y^k) + \epsilon\gamma_{N,(i)}^{k*}(y^k) - \epsilon\gamma_{N,(i)}^{k*}(y^k) \\ &+ \left(\frac{Q}{N} - D\right)\left[E[\underline{x}|y^k] + \frac{1}{N}\sum_{p=1,p\neq k}^{N} E\left[\gamma_{N,(i)}^{p*}(y^p)\Big|y^k\right]\right] = 0. \end{aligned}$$

Thus, by dividing the expression over ϵ and rearranging it, we have

$$\gamma^{k*}_{N,(i)}(y^k) = (1 - \frac{\hat{R}_{ii}}{\epsilon})\gamma^{k*}_{N,(i)}(y^k)$$
$$-\frac{1}{\epsilon}\left(\frac{Q}{N} - D\right)\left[E[x|y^k] + \frac{1}{N}\sum_{p=1,p\neq k}^{N} E\left[\gamma^{p*}_{N,(i)}(y^p)\middle| y^k\right]\right],$$

where the initial points of the iterations are zero functions. We can write $\underline{\gamma}^*_N(\underline{y}) = P(I - \frac{1}{\epsilon}\hat{\underline{R}})\underline{\gamma}^*_N(\underline{y}) - \frac{1}{\epsilon}Pr(\omega)$. Similar to [139, Theorem 2.6.5], the above sequence converges to the unique fixed point if and only if the spectral radius satisfies the following constraint $\rho\left(P(I - \frac{\hat{\underline{R}}}{\epsilon})\right) = \rho(I - \frac{\hat{\underline{R}}}{\epsilon}) := \lim_{k\to\infty}\sup[||A||^k]^{\frac{1}{k}} < 1$, where $A := I - \frac{\hat{\underline{R}}}{\epsilon}$, $||A||:= \sup_{||x||<1}||Ax||$ and ρ denotes spectral radius. The first equality is true since both P and A maps $\mathbf{\Gamma_N}$ into itself and P has operator norm equal to one. The above constraint can be always satisfied by choosing $\epsilon = \frac{1}{2}(\lambda_{max}(\hat{\underline{R}}) + \lambda_{min}(\hat{\underline{R}}))$. On the other hand, since $(\underline{x}, \underline{z}^1, \ldots, \underline{z}^N)$ are jointly Gaussian, then $\gamma^{k*}_N(y^k) = \pi^k_N y^k$ for $k = 1, \ldots, N$. Hence, $\gamma^{k*}_{N,(i)}(y^k) = \pi^k_{N,(i)}y^k$, and by linearity of the conditional expectation, we have $E[\underline{x}|y^k] = S^k y^k$, and $E[\gamma^{p*}_N(y^p)|y^k] = \pi^p_N H^p S^k y^k$. Hence (2.16) holds. Following from [139], the stationary policy is globally optimal for $(\mathcal{P}_N)$, and this completes the proof. □

Theorem 2.4.3. *Consider* $(\mathcal{P}_\infty)$ *with the expected cost* (2.15)*. Under the following measurement scheme*

$$\underline{y}^i = H\underline{x} + \underline{z}^i, \tag{2.18}$$

where $\underline{z}^i$*s are i.i.d. Gaussian random vectors,* $\gamma^{i*}_\infty(y^i) = \pi^*_\infty y^i$ *is an optimal policy for* $(\mathcal{P}_\infty)$ *and is the pointwise limit of* $\gamma^{i*}_N(y^i) = \pi^*_N y^i$*, an optimal policy for* $(\mathcal{P}_N)$*.*

Proof. In the following, we invoke Proposition 2.3.1 and Theorem 2.3.3 to prove the theorem. Under (2.18), the static team is symmetrically optimal and hence from (2.16), we have $\pi^*_N = L_N[S + N^{-1}(N-1)\pi^*_N HS]$, $\pi^*_\infty = R^{-1}D[S + \pi^*_\infty HS]$, where $L_N := (N^2R - 2DN + Q)^{-1}(N^2D - NQ)$, $S := \Sigma_{00}(H)^T(H\Sigma_{00}(H)^T + \Sigma_{kk})^{-1}$. Since for every N, we have $J_N(\underline{\gamma}^*_N) < \infty$, and since $R > 0$, we have $\sup_{N\geq 1} E[||\gamma^*_N(y^1)||^2_2] < \infty$, which implies (A3). The proof is completed using the results of Proposition 2.3.1 and Theorem 2.3.3. One can also invoke Theorem 2.3.5 to justify the result. □

2.4.4 Example 4: Asymmetric LQG team problems

Here, we consider simple variation of Example 3 considered above to illustrate Remark 2.3.1. Consider the observation scheme (2.18), and let the expected cost function be defined as

$$J(\underline{\gamma})= \limsup_{N\to\infty} \frac{1}{N} E^{\underline{\gamma}}\Big[\sum_{i=1}^{N} (u^i)^T R u^i - 2\sum_{i=1}^{N} (u^i)^T D(\underline{x} + \frac{1}{N}\sum_{k=1}^{N} u^k) \\ +(\underline{x} + \frac{1}{N}\sum_{k=1}^{N} u^k)^T Q(\underline{x} + \frac{1}{N}\sum_{k=1}^{N} u^k) + \frac{1}{N}\sum_{k=1}^{M} (u^k)^T \alpha_k u^k \Big],$$

where $M \in \mathbb{Z}_+$ is independent of N. Clearly, the N-DM team admits asymmetric optimal policies for $(\mathcal{P}_N)$ with the expected cost J_N for every N. However, one can observe that the last term above goes to zero as $N \to \infty$ under a sequence of optimal policies, and hence asymptotically the expected cost would essentially be (2.15) and Theorem 2.4.1 implies γ^*_∞ is an optimal policy since P-almost surely the sequence Q_N converges weakly (the asymmetric term vanishes when $N \to \infty$). That is, the optimal policy designed for the symmetric problem is also a solution for the asymmetric problem since under this policy the additional term (which is a non-negative contribution) vanishes, certifying its optimality.

2.4.5 Example 5: Multivariable Classical LQG Problems: Average Cost Optimality Through Static Reduction

Here, we revisit a well-known problem and a well-known solution, using the technique presented in this chapter. Let

$$x_{t+1} = Ax_t + Bu^t + w^t,$$

where $A \in \mathcal{M}_{n,n}(\mathbb{R})$, $B \in \mathcal{M}_{n,m}(\mathbb{R})$ and w^ts and X_0 are i.i.d. Gaussian random vectors with mean zero and positive variance taking values in $\mathbb{R}^n$. Let (A, B) be controllable and let

$$J(\underline{\gamma}) = \limsup_{T\to\infty} J_T(\underline{\gamma})$$

$$:= \limsup_{T\to\infty} \frac{1}{T} E^{\underline{\gamma}}\Big[\sum_{t=0}^{T-1} x_t^T Q x_t + (u^t)^T R u^t\Big],$$

where $Q \geq 0$ and $R > 0$ are appropriate dimensions real matrices. We can write,

$$J(\underline{\gamma}) = \limsup_{T\to\infty} \frac{1}{T} E^{\underline{\gamma}}\Big[\sum_{t=0}^{T-1} \Big(\sum_{k=1}^{t} A^{t-k} B u^{k-1} + \sum_{k=0}^{t} A^{t-k}\zeta^k\Big)^T \times Q\Big(\sum_{k=1}^{t} A^{t-k} B u^{k-1} + \sum_{k=0}^{t} A^{t-k}\zeta^k\Big) + (u^t)^T R u^t\Big],$$

where $\zeta = (X_0^T, (w^0)^T, (w^1)^T, \dots)^T$. In the following, we consider fully observed classical IS, i.e., $y^t = x^t$, and we can write $y^t = H^t\zeta + \sum_{j=0}^{t-1} D_{tj} u^j$, where H^t and D_{tj} are appropriate dimensional matrices. Using [66, Theorem 1], we can reduce IS to the static one as $v^t = \tilde{H}_t\zeta$. According to [64, Section 3.5], we have $u_T^{t*} = G_T^t x_t$ for $t = 0, 1, \dots, T-1$, where $k_T^T = 0$, and

$$G_T^t = -(R + B^T k_T^{t+1} B)^{-1} B^T k_T^{t+1} A, \tag{2.19}$$

$$k_T^t = Q + A^T k_T^{t+1} A - A^T k_T^{t+1} B (R + B^T k_T^{t+1} B)^{-1} B^T k_T^{t+1} A. \tag{2.20}$$

Theorem 2.4.4. *For LQG teams with the classical information structure as formulated above, $u^{t*} = \lim_{T\to\infty} \gamma_T^{t*}(v^t) = \gamma_\infty^{t*}(v^t)$ is the optimal policy for $J(\underline{\gamma})$, where $\{\gamma_T^{t*}\}_T$ is a sequence of optimal policies for $\{J_T(\underline{\gamma}_T)\}_T$ with the pointwise limit γ_∞^{t*} as $T \to \infty$.*

Although, this result is a classical one in the literature, here, we present a new approach using the static reduction.

Remark 2.4.1. *Similarly, one can show the result for (i) $y^t = Cx^t$, (A, C) is observable and $Q = C^T C$, (ii) the discounted LQG team problems with the classical information structure.*

2.5 Proofs

2.5.1 Proof of Theorem 2.2.1

Under (A1), the required derivatives in (2.8) in the direction of u^i exist and the chain rule of derivatives can be applied since this implies that the cost function is Fréchet differentiable in u^i [57]. Now, we use the convexity property to justify interchanging the expectation and the derivation similar to [78, Theorem 2], then we use (2.7) and (2.8) to establish the global optimality of $\underline{\gamma}^*$ for $(\mathcal{P}_\infty)$. Under (A1), we have for every $\alpha \in (0, 1]$,

$$\sum_{i=1}^{N} c(\omega_0,\gamma^{i*} + \alpha\delta^i, \mu_N^* + \frac{\alpha}{N}\sum_{p=1}^{N}\delta^p) - c(\omega_0, \gamma^{i*}, \mu_N^*) \leq \alpha \sum_{i=1}^{N} c(\omega_0, \gamma^i, \mu_N) - c(\omega_0, \gamma^{i*}, \mu_N^*),$$

where $\mu_N = \frac{1}{N}\sum_{p=1}^{N}\gamma^p(y^p)$ and $\delta^i = \gamma^i - \gamma^{i*}$. Let

$$h_N^\omega(\alpha) := \frac{1}{\alpha}\bigg[\frac{1}{N}\sum_{i=1}^{N} c(\omega_0, \gamma^{i*} + \alpha\delta^i, \mu_N^* + \frac{\alpha}{N}\sum_{p=1}^{N}\delta^p) - c(\omega_0, \gamma^{i*}, \mu_N^*)\bigg].$$

Hence, [50, Proposition 6.3.2] implies that $h_N^\omega(\alpha)$ is a monotone non-increasing function as $\alpha \to 0$ in $\alpha \in [0, 1]$ and bounded from above by $h_N^\omega(1)$. Thus, by [50, Corollary 6.3.3], $h'_{+,N}(\omega, 0) := \lim_{\alpha\to 0} h_N^\omega(\alpha)$ exists. Since $h_N^\omega(\alpha)$ is a monotonic non-increasing function as $\alpha \to 0$ in $\alpha \in [0, 1]$ and bounded above by $h_N^\omega(1)$, and since $J(\underline{\gamma}^*)$ and $J(\underline{\gamma})$ are finite, we can choose N large enough such that $E[h_N^\omega(1)] < \infty$. Now, we can use the monotone convergence theorem (see [64, page. 170]) to interchange the limit and the expectation

$$\lim_{\alpha\to 0} E[h_N^\omega(\alpha)] = E[\lim_{\alpha\to 0} h_N^\omega(\alpha)] = E[h'_{+,N}(\omega, 0)]. \tag{2.21}$$

From [78, Lemma 1], we have $E[h'_{+,N}(\omega, 0)] = \frac{1}{N}E[\sum_{i=1}^{N}\sum_{k=1}^{N} c_{u^k}(\omega_0, \gamma^{i*}, \mu_N^*)\delta^k]$. Define

$$F_{\underline{\gamma}_N}^N(\alpha) := \frac{1}{N}E\bigg[\sum_{i=1}^{N} c(\omega_0, \gamma^{i*} + \alpha\delta^i, \mu_N^* + \frac{\alpha}{N}\sum_{p=1}^{N}\delta^p)\bigg].$$

Note that $F^N_{\underline{\gamma}_N}(\alpha)$ exists for $\alpha \in [0,1]$ since $E[h^\omega_N(\alpha)] \leq E[h^\omega_N(1)] < \infty$, and $E[\frac{1}{N}\sum_{i=1}^N c(\omega_0, \gamma^{i*}, \mu^*_N)] < \infty$. Therefore, one can write $F'^N_{\underline{\gamma}^+_N}(0) = \lim_{\alpha\to 0} E[h^\omega(\alpha)]$, and

$$F'^N_{\underline{\gamma}^+_N}(0) = \frac{1}{N} E\Big[\sum_{i=1}^N \sum_{k=1}^N c_{u^k}(\omega_0, \gamma^{i*}, \mu^*_N)(\gamma^k - \gamma^{k*})\Big]$$

Thus, we can write

$$\begin{aligned} J(\underline{\gamma}) - J(\underline{\gamma}^*) &= \limsup_{N\to\infty} F^N_{\underline{\gamma}_N}(1) - \limsup_{N\to\infty} F^N_{\underline{\gamma}_N}(0) && (2.22)\\ &= \limsup_{N\to\infty} F^N_{\underline{\gamma}_N}(1) - \liminf_{N\to\infty} F^N_{\underline{\gamma}_N}(0) && (2.23)\\ &\geq \limsup_{N\to\infty} \frac{F^N_{\underline{\gamma}_N}(1) - F^N_{\underline{\gamma}_N}(0)}{1} && (2.24)\\ &\geq \limsup_{N\to\infty} F'^N_{\underline{\gamma}^+_N}(0) \geq 0, && (2.25) \end{aligned}$$

where (2.23) follows from (A2), (2.7), and $-\liminf_{N\to\infty} a_N = \limsup_{N\to\infty} -a_N$, $\limsup_{N\to\infty} a_N + \limsup_{N\to\infty} b_N \geq \limsup_{N\to\infty}(a_N + b_N)$ imply (2.24), and (2.25) holds since $F^N_{\underline{\gamma}_N}(.)$ is a convex function and using [50, Corollary 6.3.3], and since $a_N \geq b_N$ then $\limsup_{N\to\infty} a_N \geq \limsup_{N\to\infty} b_N$. Finally, the last inequality follows from (2.8); hence, $J(\underline{\gamma}) - J(\underline{\gamma}^*) \geq 0$, and the proof is completed.

2.5.2 Proof of Theorem 2.2.2

We have

$$\begin{aligned} J(\underline{\gamma}^*) &\leq \limsup_{N\to\infty} \frac{1}{N}\sum_{i=1}^N E^{\underline{\gamma}^*_N}\Big[c(\omega_0, u^i, \mu_N)\Big] && (2.26)\\ &= \limsup_{N\to\infty} \inf_{\underline{\gamma}_N \in \mathbf{\Gamma_N}} \frac{1}{N}\sum_{i=1}^N E^{\underline{\gamma}_N}\Big[c(\omega_0, u^i, \mu_N)\Big] && (2.27)\\ &= \limsup_{N\to\infty} \inf_{\underline{\gamma}\in\mathbf{\Gamma}} \frac{1}{N}\sum_{i=1}^N E^{\underline{\gamma}}\Big[c(\omega_0, u^i, \mu_N)\Big] && (2.28)\\ &\leq \inf_{\underline{\gamma}\in\mathbf{\Gamma}} \limsup_{N\to\infty} \frac{1}{N}\sum_{i=1}^N E^{\underline{\gamma}}\Big[c(\omega_0, u^i, \mu_N)\Big] && (2.29) \end{aligned}$$

$$= \inf_{\underline{\gamma} \in \mathbf{\Gamma}} J(\underline{\gamma}),$$

where $\mu_N := \frac{1}{N} \sum_{p=1}^{N} u^p$ and (2.26) follows from (2.9), and (2.27) is true since γ_N^* is a team optimal policy for $(\mathcal{P}_N)$ (see (2.2)). Furthermore, (2.28) follows from the fact that $[\underline{\gamma}|_N \colon \underline{\gamma} \in \mathbf{\Gamma}] = \mathbf{\Gamma_N}$, where $\underline{\gamma}|_N$ is $\underline{\gamma}$ restricted to the first N components.

2.5.3 Proof of Theorem 2.2.3

According to Theorem 2.2.2, we only need to show that

$$\begin{aligned}
\limsup_{N\to\infty} J_N(\underline{\gamma}_N^*) &\geq \liminf_{N\to\infty} J_N(\underline{\gamma}_N^*) \\
&\geq E\left[\lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} c(\omega_0, \gamma_N^{i*}(y^i), \mu_N^*) \right] \\
&= \lim_{N\to\infty} J_N(\underline{\gamma}_\infty^*),
\end{aligned}$$

where $\mu_N^* = \frac{1}{N} \sum_{p=1}^{N} \gamma_N^{p*}(y^p)$ and the second inequality follows from Fatou's lemma (since the cost function is non-negative). In the following, we justify the equality above. On a set of P measure one, $\omega \in B$ where $B \in \mathcal{F}$, for every fixed $y^i(\omega)$ in this set, define $\underline{y}(\omega) = (y^1(\omega), y^2(\omega), ...)$ and $\underline{y}_N(\omega) = (y^1(\omega), \ldots, y^N(\omega))$. We follow three steps to prove the theorem.

(Step 1): We show that on a set of P measure one, $\omega \in B$ where $B \in \mathcal{F}$, for every fixed $y^i(\omega)$ in this set $\lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} \left(\gamma_N^{i*}(y^i) - \gamma_\infty^{i*}(y^i) \right) = 0$. For a fixed $\underline{y}$, following from (b) for a given $\delta_{\omega,\underline{y}_N} := \sup_{1\leq i\leq N} ||\gamma_N^{i*}(y^i) - \gamma_\infty^{i*}(y^i)|| > 0$ there exists $\hat{N}(\delta_{\omega,\underline{y}_N}) \in \mathbb{N}$ such that for $N > \hat{N}(\delta_{\omega,\underline{y}_N})$, $||\gamma_N^{i*}(y^i) - \gamma_\infty^{i*}(y^i)|| \leq \delta_{\omega,\underline{y}_N}$ for every $i = 1, \ldots, N$, where $\lim_{N\to\infty} \delta_{\omega,\underline{y}_N} = 0$ P-almost surely. We have P-almost surely,

$$\left|\left| \frac{1}{N} \sum_{i=1}^{N} \left(\gamma_N^{i*}(y^i) - \gamma_\infty^{i*}(y^i) \right) \right|\right| < \frac{1}{N} \sum_{i=1}^{N} \delta_{\omega,\underline{y}_N} = \delta_{\omega,\underline{y}_N},$$

and since $\lim_{N\to\infty} \sup_{1\leq i\leq N} ||\gamma_N^{i*}(y^i) - \gamma_\infty^{i*}(y^i)|| = 0$, one can see $\lim_{N\to\infty} \delta_{\omega,\underline{y}_N} = 0$. Hence, we can show

that $\lim_{N\to\infty} \frac{1}{N}\sum_{i=1}^{N} \gamma_N^{i*}(y^i) = \lim_{N\to\infty} \frac{1}{N}\sum_{i=1}^{N} \gamma_\infty^{i*}(y^i)$. Following from continuity, $c(\omega_0, \gamma_N^{i*}(y^i), \mu_N^*)$ converges to $c(\omega_0, \gamma_\infty^{i*}(y^i), \lim_{N\to\infty} \mu_\infty^*)$ P-a.s. for every $i = 1\ldots, N$.

(Step 2): We show that $c(\omega_0, \gamma_N^{i*}(y^i), \mu_N^*)$ converges to $c(\omega_0, \gamma_\infty^{i*}(y^i), \lim_{N\to\infty} \mu_\infty^*)$ uniformly in $i = 1, \ldots, N$ P-almost surely, where $\mu_\infty^* = \frac{1}{N}\sum_{p=1}^{N} \gamma_\infty^{p*}(y^p)$. By continuity of the cost function, we have for a given $\epsilon_{\omega,\underline{y}_N} > 0$, there exists $\delta_{\omega,\underline{y}_N} > 0$ such that $||\gamma_N^{i*}(y^i) - \gamma_\infty^{i*}(y^i)|| < \delta_{\omega,\underline{y}_N}$, and $||\frac{1}{N}\sum_{i=1}^{N} \gamma_N^{i*}(y^i) - \gamma_\infty^{i*}(y^i)|| < \delta_{\omega,\underline{y}_N}$ implies $|c(\omega_0, \gamma_N^{i*}(y^i), \mu_N^*) - c(\omega_0, \gamma_\infty^{i*}(y^i), \mu_\infty^*)| < \epsilon_{\omega,\underline{y}_N}$ P-almost surely for every $i = 1, \ldots, N$. Following from Step 1, we have for $N > \hat{N}(\delta_{\omega,\underline{y}_N}(\epsilon_{\omega,\underline{y}_N}))$, $||\gamma_N^{i*}(y^i) - \gamma_\infty^{i*}(y^i)|| < \delta_{\omega,\underline{y}_N}$, and $||\frac{1}{N}\sum_{i=1}^{N} \gamma_N^{i*}(y^i) - \gamma_\infty^{i*}(y^i)|| < \delta_{\omega,\underline{y}_N}$, hence P-a.s.

$$\left| c(\omega_0, \gamma_N^{i*}(y^i), \mu_N^*) - c(\omega_0, \gamma_\infty^{i*}(y^i), \mu_\infty^*) \right| < \epsilon_{\omega,\underline{y}_N},$$

where $\lim_{N\to\infty} \epsilon_{\omega,\underline{y}_N} = 0$.

(Step 3): In this step, we show that P-a.s.,

$$\lim_{N\to\infty} \frac{1}{N}\sum_{i=1}^{N} \left[c(\omega_0, \gamma_N^{i*}(y^i), \mu_N^*) - c(\omega_0, \gamma_\infty^{i*}(y^i), \mu_\infty^*) \right] = 0.$$

According to Step 2, for $N > \hat{N}(\delta_{\omega,\underline{y}_N}(\epsilon_{\omega,\underline{y}_N})$, we have P-a.s.

$$\left| \frac{1}{N}\sum_{i=1}^{N} c(\omega_0, \gamma_N^{i*}(y^i), \mu_N^*) - c(\omega_0, \gamma_\infty^{i*}(y^i), \mu_\infty^*) \right| < \epsilon_{\omega,\underline{y}_N}.$$

Following from (c), we can interchange the limit and the integral using the dominated convergence theorem, and the proof is completed.

2.5.4 Proof of Lemma 2.3.1

Any permutation of policies does not deviate the value of $J_N(\underline{\gamma}_N)$ since

$$J_N(\underline{\gamma}_N^\sigma) = \int c(\omega_0, u^1, \ldots, u^N) P_N(dy^1, \ldots, dy^N|\omega_0) 1_{\{(\gamma^{\sigma(1)}(y^1), \ldots, \gamma^{\sigma(N)}(y^N))\}}(du^1, \ldots, du^N) P(d\omega_0)$$

$$
\begin{aligned}
&= \int c(\omega_0, u^{\sigma(1)}, \ldots, u^{\sigma(N)}) 1_{\{(\gamma^{\sigma(1)}(y^{\sigma(1)}), \ldots, \gamma^{\sigma(N)}(y^{\sigma(N)}))\}}(du^{\sigma(1)}, \ldots, du^{\sigma(N)}) \\
&\times P_N(dy^{\sigma(1)}, \ldots, dy^{\sigma(N)}|\omega_0) P(d\omega_0) \qquad (2.30) \\
&= \int c(\omega_0, u^1, \ldots, u^N) 1_{\{(\gamma^1(y^1), \ldots, \gamma^N(y^N))\}}(du^1, \ldots, du^N) P_N(dy^1, \ldots, dy^N|\omega_0) P(d\omega_0) \\
&= J_N(\underline{\gamma}_N),
\end{aligned}
$$

where (2.30) follows from the assumption that the cost function is exchangeable with respect to the actions, and the hypothesis that observations of DMs are P-almost surely exchangeable conditioned on the random variable ω_0. Let $\underline{\gamma}^*_N = (\gamma^{1*}, \gamma^{2*}, \ldots, \gamma^{N*})$ be a given policy. Consider $\tilde{\underline{\gamma}}_N$ as a convex combination of all possible permutations of policies by averaging them, $\sigma \in \Sigma$, where Σ is the set of all possible permutation. Since $\mathbb{U}$ is convex, $\tilde{\underline{\gamma}}_N$ is a control policy. Following from convexity of the cost function P-almost surely, we have for $\alpha_\sigma = \frac{1}{|\Sigma|}$ (where $|\Sigma|$ denotes the cardinality of Σ),

$$
\begin{aligned}
J_N(\tilde{\underline{\gamma}}_N) &:= J_N(\sum_{\sigma \in \Sigma} \alpha_\sigma \underline{\gamma}^{*,\sigma}_N) \\
&\leq \sum_{\sigma \in \Sigma} \alpha_\sigma J_N(\underline{\gamma}^{*,\sigma}_N) \\
&= \sum_{\sigma \in \Sigma} \alpha_\sigma J_N(\underline{\gamma}^*_N) \ = J_N(\underline{\gamma}^*_N),
\end{aligned}
$$

where the inequality follows from convexity of the cost function P-almost surely for every fixed realization of observations since we have

$$
\begin{aligned}
&E\bigg[c\bigg(\omega_0, \sum_{\sigma \in \Sigma} \alpha_\sigma (\underline{\gamma}^{*,\sigma}_N)^1(y^1), \ldots, \sum_{\sigma \in \Sigma} \alpha_\sigma (\underline{\gamma}^{*,\sigma}_N)^N(y^N)\bigg)\bigg] \\
&\leq E\bigg[\sum_{\sigma \in \Sigma} \alpha_\sigma c\bigg(\omega_0, (\underline{\gamma}^{*,\sigma}_N)^1(y^1), \ldots, (\underline{\gamma}^{*,\sigma}_N)^N(y^N)\bigg)\bigg] \\
&= \sum_{\sigma \in \Sigma} \alpha_\sigma E\bigg[c\bigg(\omega_0, (\underline{\gamma}^{*,\sigma}_N)^1(y^1), \ldots, (\underline{\gamma}^{*,\sigma}_N)^N(y^N)\bigg)\bigg],
\end{aligned}
$$

where $(\underline{\gamma}_N^{*,\sigma})^j$ denotes the j-the component of $\underline{\gamma}_N^{*,\sigma}$, and the inequality above follows from Jensen's inequality since the cost function is convex P-almost surely. Hence, the team is symmetrically optimal.

2.5.5 Proof of Theorem 2.3.2

Action spaces and the cost function are convex and following from the hypothesis that y^is are i.i.d. random variables (hence they are exchangeable conditioned on ω_0) and the result of Theorem 2.3.1, one can consider a sequence of N-DM teams which are symmetrically optimal that defines $(\mathcal{P}_N)$ (see (2.2)) and whose limit is identified with $(\mathcal{P}_\infty)$. Define empirical measures on actions and observation of $Q_N(B) := \frac{1}{N}\sum_{i=1}^N \delta_{\zeta_N^i}(B)$, and $\tilde{Q}_N(B) := \frac{1}{N}\sum_{i=1}^N \delta_{\zeta_\infty^i}(B)$, where $B \in \mathcal{Z} := \mathbb{U}\times\mathbb{Y}$, $\zeta_N^i := (\gamma_N^*(y^i), y^i)$, $\zeta_\infty^i := (\gamma_\infty^*(y^i), y^i)$, and $\delta_Y(\cdot)$ is the Dirac measure for any random variable Y. In the following, we first show that $\tilde{Q}_N$ converges weakly to $Q = \mathcal{L}(\zeta_\infty^i)$ P-almost surely, then we show (2.9) holds, and we invoke Theorem 2.2.2.

(Step 1): For every $g \in C_b(\mathcal{Z})$, where we denote $C_b(X)$ as the space of continuous and bounded functions in X, we have

$$\begin{aligned}
&\lim_{N\to\infty} P\left(\left|\int g dQ_N - \int g d\tilde{Q}_N\right| \geq \epsilon\right) \\
&\leq \epsilon^{-1} \lim_{N\to\infty} \frac{1}{N}\sum_{i=1}^N E\left[\left|g(\gamma_N^*(y^i), y^i) - g(\gamma_\infty^*(y^i), y^i)\right|\right] &(2.31)\\
&= \epsilon^{-1} \lim_{N\to\infty} E\left[\left|g(\gamma_N^*(y^i), y^i) - g(\gamma_\infty^*(y^i), y^i)\right|\right] &(2.32)\\
&= \epsilon^{-1} E\left[\lim_{N\to\infty}\left|g(\gamma_N^*(y^i), y^i) - g(\gamma_\infty^*(y^i), y^i)\right|\right] = 0, &(2.33)
\end{aligned}$$

where (2.31) follows from Markov's inequality, the triangle inequality and the definition of the empirical measure, and (2.32) follows from the hypothesis that y^is are identical random variables. Since g is bounded and continuous, the dominated convergence theorem implies (2.33). Hence, for every subsequence there exists a subsubsequence such that $|\int g dQ_{N_{k_l}} - \int g d\tilde{Q}_{N_{k_l}}|$ converges to zero P-almost surely as $l \to \infty$. On the other hand, since y^is are i.i.d. random variables, the

strong law of large numbers (SLLN) implies $\tilde{Q}_N$ converges weakly to Q P-almost surely, that is $|\int g d\tilde{Q}_N - \int g dQ|$ converges to zero P-almost surely for every $g \in C_b(\mathcal{Z})$. Hence, through choosing a suitable subsequence, Q_{N_k} converges P-almost sure weakly to Q since for every continuous and bounded function g, we have P-a.s.,

$$\lim_{N\to\infty} \left| \int g dQ_N - \int g dQ \right| \leq \lim_{N\to\infty} \left(\left| \int g dQ_N - \int g d\tilde{Q}_N \right| + \left| \int g d\tilde{Q}_N - \int g dQ \right| \right) = 0. \tag{2.34}$$

(Step 2): Following from [123, Theorem 3.5] and [53, Lemma 1.5], or [85, Theorem 3.1] using the fact that the cost function is non-negative and continuous, we have

$$\begin{aligned}
&\limsup_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} E\left[c\left(\omega_0, \gamma_N^*(y^i), \frac{1}{N} \sum_{i=1}^{N} \gamma_N^*(y^i) \right) \right] \\
&\geq \liminf_{N\to\infty} E\left[E\left[\int_{\mathcal{Z}} c\left(\omega_0, u, \int_{\mathbb{U}} u Q_N(du \times \mathbb{Y}) \right) Q_N(du, dy) \middle| \omega_0 \right] \right] \\
&\geq E\left[E\left[\liminf_{N\to\infty} \int_{\mathcal{Z}} c\left(\omega_0, u, \int_{\mathbb{U}} u Q_N(du \times \mathbb{Y}) \right) Q_N(du, dy) \middle| \omega_0 \right] \right] \\
&\geq E\left[E\left[\int_{\mathcal{Z}} c\left(\omega_0, u, \int_{\mathbb{U}} u Q(du \times \mathbb{Y}) \right) Q(du, dy) \middle| \omega_0 \right] \right],
\end{aligned} \tag{2.35}$$

where the first inequality follows from the definition of Q_N and replacing limsup by liminf. The second inequality follows from Fatou's lemma. In the following, we justify (2.35). Since Q_N converges weakly to Q P-almost surely, using continuous mapping theorem [26, page 20], we have $Q_N(du \times \mathbb{Y})$ converges weakly to $Q(du \times \mathbb{Y})$ P-almost surely, hence the compactness of $\mathbb{U}$ implies $\int_{\mathbb{U}} u Q_N(du \times \mathbb{Y}) \to \int_{\mathbb{U}} u Q(du \times \mathbb{Y})$ P-almost surely, and continuity of the cost function P-almost surely implies $c(\omega_0, u, \int_{\mathbb{U}} u Q_N(du \times \mathbb{Y}))$ converges to $c\left(\omega_0, u, \int_{\mathbb{U}} u Q(du \times \mathbb{Y})\right)$ P-almost surely. Define a non-negative bounded sequence $G_N^M := \min\{M, c\left(\omega_0, u, \int_{\mathbb{U}} u Q_N(du \times \mathbb{Y})\right)\}$, where $G_M^N \uparrow G^N := c\left(\omega_0, u, \int_{\mathbb{U}} u Q_N(du \times \mathbb{Y})\right)$ as $M \to \infty$, then we have P-almost surely

$$\liminf_{N\to\infty} \int_{\mathcal{Z}} c\left(\omega_0, u, \int_{\mathbb{U}} u Q_N(du \times \mathbb{Y}) \right) Q_N(du, dy)$$

$$
\begin{aligned}
&= \lim_{M\to\infty} \liminf_{N\to\infty} \int_{\mathcal{Z}} c\left(\omega_0, u, \int_{\mathbb{U}} uQ_N(du\times\mathbb{Y})\right) Q_N(du, dy) \\
&\geq \lim_{M\to\infty} \liminf_{N\to\infty} \int_{\mathcal{Z}} G_N^M Q_N(du, dy) \\
&= \lim_{M\to\infty} \int_{\mathcal{Z}} G^M Q(du, dy) \\
&= \int_{\mathcal{Z}} c\left(\omega_0, u, \int_{\mathbb{U}} uQ(du\times\mathbb{Y})\right) Q(du, dy),
\end{aligned}
$$

where the first inequality follows from the definition of G_N^M and the second equality is true using [123, Theorem 3.5] since G_N^M is bounded (hence it is uniformly Q_N-integrable) and continuously converges to G^M, and the monotone convergence theorem implies the last equality. Hence, (2.35) holds which implies (2.9), and the proof is completed using Theorem 2.2.2.

2.5.6 Proof of Preposition 2.3.1

Since z^is are i.i.d. random variables, observations, $y^i = h(x, z^i)$, have identical distributions (but are not independent), and similar to the proof of Theorem 2.3.2, using symmetry, one can show (2.33) holds. In the following, we show (2.34) and (2.35) hold.

$$
\begin{aligned}
&\lim_{N\to\infty} P\left(\left|\int g d\tilde{Q}_N - \int g dQ\right| \geq \epsilon\right) \\
&\leq \frac{1}{\epsilon^2} \lim_{N\to\infty} E\left[\left|\frac{1}{N}\sum_{i=1}^{N} g(\gamma_\infty^*(y^i), y^i) - E(g(\gamma_\infty^*(y^1), y^1))\right|^2\right] &(2.36)\\
&= \lim_{N\to\infty} (\epsilon)^{-2} E\left[E\left[\left|\frac{1}{N}\sum_{i=1}^{N} L(\gamma_\infty^*(y^i), y^i)\right|^2 \middle| x\right]\right] &(2.37)\\
&= 0 &(2.38)
\end{aligned}
$$

where $L(\gamma_\infty^*(y^i), y^i) := g(\gamma_\infty^*(y^i), y^i) - E(g(\gamma_\infty^*(y^1), y^1)|x)$, and (2.36) follows from Chebyshev's inequality, and (2.37) follows from the law of iterated expectations. The structure $y^i = h(x, z^i)$ implies conditional independence of y^is given x, hence, using the law of large numbers and since $g \in C_b(\mathcal{Z})$, we have (2.38), and this implies $\tilde{Q}_{N_k}$ converges weakly to $\mathcal{L}(\zeta_\infty^i|x)$ P-almost surely as $k \to \infty$, hence through choosing a suitable subsequence, $Q_{N_{k_l}}$ converges P-almost sure weakly

to $Q = \mathcal{L}(\zeta_\infty^i|x)$ as $l \to \infty$ and the rest of the proof to justify (2.35) is the same as that of Theorem 2.3.2.

2.5.7 Proof of Theorem 2.3.3

In the following, we just show $\int_{\mathbb{U}} uQ_N(du \times \mathbb{Y}) \to \int_{\mathbb{U}} uQ(du \times \mathbb{Y})$ P-almost surely, and the rest of the proof follows from that of Theorem 2.3.2. We have

$$\begin{aligned}
&\lim_{N\to\infty} P\left(\left| \int_{\mathbb{U}} uQ_N(du \times \mathbb{Y}) - \int_{\mathbb{U}} ud\tilde{Q}_N(du \times \mathbb{Y})\right| \geq \epsilon\right) \\
&\qquad \leq \epsilon^{-1} \lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} E\left[\left|\gamma_N^*(y^i) - \gamma_\infty^*(y^i)\right|\right] &(2.39)\\
&\qquad = \epsilon^{-1} \lim_{N\to\infty} E\left[\left|\gamma_N^*(y^1) - \gamma_\infty^*(y^1)\right|\right] &(2.40)\\
&\qquad = \epsilon^{-1} E\left[\lim_{N\to\infty} \left|\gamma_N^*(y^1) - \gamma_\infty^*(y^1)\right|\right] = 0, &(2.41)
\end{aligned}$$

where (2.39) follows from Markov's inequality and the triangle inequality, and (2.40) is true since observations have identical distributions, and (2.41) follows from the uniform integrability assumption (A3) and the pointwise convergence of γ_N^* using [26, Theorem 3.5]. On the other hand, SLLN implies P-almost surely that $\int_{\mathbb{U}} u\tilde{Q}_N(du \times \mathbb{Y}) = \frac{1}{N}\sum_{i=1}^{N} \gamma_\infty^*(y^i) \to \int_{\mathbb{U}} uQ(du \times \mathbb{Y})$, and this completes the proof.

2.5.8 Proof of Theorem 2.3.4

We show (2.9) holds, then we invoke Theorem 2.2.2. We use the same definitions in Theorem 2.3.2 for measures Q_N and Q. We have

$$\begin{aligned}
&E\left[E\left[\liminf_{N\to\infty} \int_{\mathcal{Z}} c\left(\omega_0, u, \int_{\mathbb{U}} uQ_N(du \times \mathbb{Y})\right) Q_N(du, dy)\middle|\omega_0\right]\right] \\
&\geq E\left[E\left[\int_{\mathcal{Z}} \liminf_{N\to\infty} c\left(\omega_0, u, \int_{\mathbb{U}} uQ_N(du \times \mathbb{Y})\right) Q(du, dy)\middle|\omega_0\right]\right] &(2.42)\\
&\geq E\left[E\left[\int_{\mathcal{Z}} c\left(\omega_0, u, \int_{\mathbb{U}} uQ(du \times \mathbb{Y})\right) Q(du, dy)\middle|\omega_0\right]\right], &(2.43)
\end{aligned}$$

where (2.42) follows from a version of Fatou's lemma in [52, Theorem 1.1], and (2.43) is true since from the lower semi-continuity of $\int_{\mathbb{U}} uQ_N(du \times \mathbb{Y})$, we have $\liminf_{N\to\infty} \int_{\mathbb{U}} uQ_N(du \times \mathbb{Y}) \geq \int_{\mathbb{U}} uQ(du \times \mathbb{Y})$, and continuity and the hypothesis that the cost function is increasing in the last argument imply for all $u \in \mathbb{U}$, $\liminf_{N\to\infty} c\left(\omega_0, u, \int_{\mathbb{U}} uQ_N(du \times \mathbb{Y})\right) \geq c\left(\omega_0, u, \int_{\mathbb{U}} uQ(du \times \mathbb{Y})\right)$ P-almost surely, and this completes the proof.

2.5.9 Proof of Lemma 2.3.5

Following from the result of Theorem 2.3.1, one can consider a sequence of N-DM teams which are symmetrically optimal that defines $(\mathcal{P}_N)$ (see (2.2)) and whose limit is identified with $(\mathcal{P}_\infty)$. Equivalent to (b) in Theorem 2.2.3, we can show that $\lim_{N\to\infty} \sup_{1\leq i\leq N} ||\gamma_N^*(y^i) - \gamma_\infty^*(y^i)||^2 = 0$ P-almost surely. We have

$$\begin{aligned}
\lim_{N\to\infty} \sup_{1\leq i\leq N} ||\gamma_N^*(y^i) - \gamma_\infty^*(y^i)||^2 &\leq \lim_{N\to\infty} \sum_{i=1}^{N} ||\gamma_N^*(y^i) - \gamma_\infty^*(y^i)||^2 \\
&\leq \lim_{N\to\infty} h(N) \frac{1}{N} \sum_{i=1}^{N} f(y^i) \\
&= 0,
\end{aligned}$$

where the last inequality follows from (2.11). Hence, thanks to Theorem 2.2.3, a team optimal policy for $(\mathcal{P}_\infty)$ is the limit of a sequence of team optimal policies for $(\mathcal{P}_N)$ (see (2.2)) as $N \to \infty$, and hence a team optimal policy for $(\mathcal{P}_\infty)$ is symmetrically optimal and the proof is completed.

2.5.10 Proof of Theorem 2.3.6

Proof. We first show that $\{Q_N\}_N$ is pre-compact in the product space $(\mathbb{Y} \times \mathbb{U})$ equipped with the weak convergence topology for each component. Then, we show that an induced policy by the limit Q achieves lower expected cost than $\limsup_{N\to\infty} J_N(\underline{\gamma}_N^*)$, and we invoke Theorem 2.2.2 to complete the proof. Action spaces and the cost function are convex and following from the hypothesis that y^is are i.i.d. random variables (hence they are exchangeable conditioned on ω_0) and the result of

Theorem 2.3.1, one can consider a sequence of N-DM teams which are symmetrically optimal that defines $(\mathcal{P}_N)$ (see (2.2)) and whose limit is identified with $(\mathcal{P}_\infty)$.

(Step 1): In the following, we show that for some subsequence $\{Q_n\}_{n\in\mathbb{I}}$ converges weakly to Q P-almost surely, that is, P-a.s., for every continuous and bounded function g,

$$\lim_{n\to\infty}\left|\int g dQ_n - \int g dQ\right| = 0,$$

where $n \in \mathbb{I}$ is the index set of a converging subsequence. We use the fact that observations are i.i.d. and the space of control policies is weakly compact (see e.g., [138, proof of Theorem 4.7]). That is because, we can represent the control policy spaces with the space of all joint measures on $(\mathbb{Y}^i \times \mathbb{U}^i)$ for each DM with a fixed marginal on y^i [140, 29]. Since the team is static, this decouples the policy spaces from the policies of the previous decision makers, and following from the hypothesis on ϕ_i and the fact that $\nu \to \int \nu(dx)g(x)$ is lower semi-continuous for a continuous function g [138, proof of Theorem 4.7], the marginals on $\mathbb{U}^i$ will be weakly compact. If the marginals are weakly compact, then the collection of all measures with these weakly compact marginals are also weakly compact (see e.g., [137, Proof of Theorem 2.4]) and hence the control policy space is weakly compact. Using Tychonoff's theorem, the countably infinite product space is also compact under the product topology which implies compactness of the space of control policies under the product topology. Hence, there exists a subsequence $\{Q_n\}_{n\in\mathbb{I}}$ converges weakly to Q P-almost surely.

(Step 2): Now, we show that (2.9) holds. We have

$$\begin{aligned}
&E\left[E\left[\int_{\mathcal{Z}} c\left(\omega_0, u, \int_{\mathbb{U}} uQ(du\times\mathbb{Y})\right) Q(du,dy)\middle|\omega_0\right]\right]\\
&= \lim_{M\to\infty} E\left[E\left[\int_{\mathcal{Z}} \min\left\{M, c\left(\omega_0, u, \int_{\mathbb{U}} uQ(du\times\mathbb{Y})\right)\right\} Q(du,dy)\middle|\omega_0\right]\right] \qquad (2.44)\\
&= \lim_{M\to\infty} E\left[E\left[\lim_{n\to\infty}\int_{\mathcal{Z}} \min\left\{M, c\left(\omega_0, u, \int_{\mathbb{U}} uQ_n(du\times\mathbb{Y})\right)\right\} Q_n(du,dy)\middle|\omega_0\right]\right] \qquad (2.45)
\end{aligned}$$

$$= \lim_{M\to\infty}\lim_{n\to\infty} E\left[E\left[\int_{\mathcal{Z}} \min\left\{M, c\left(\omega_0, u, \int_{\mathbb{U}} uQ_n(du\times\mathbb{Y})\right)\right\}Q_n(du,dy)\middle|\omega_0\right]\right] \tag{2.46}$$

$$\leq \lim_{M\to\infty}\limsup_{N\to\infty} E\left[E\left[\int_{\mathcal{Z}} \min\left\{M, c\left(\omega_0, u, \int_{\mathbb{U}} uQ_N(du\times\mathbb{Y})\right)\right\}Q_N(du,dy)\middle|\omega_0\right]\right] \tag{2.47}$$

$$\leq \limsup_{N\to\infty} \frac{1}{N}\sum_{i=1}^{N} E\left[c\left(\omega_0, \gamma_N^*(y^i), \frac{1}{N}\sum_{i=1}^{N}\gamma_N^*(y^i)\right)\right], \tag{2.48}$$

where (2.44) follows from the monotone convergence theorem. Since $\{Q_n\}_{n\in I}$ converges weakly to Q P-almost surely, we have by continuous mapping theorem (by considering a projection to the first component) $\int_{\mathbb{U}} uQ_n(du\times\mathbb{Y}) \to \int_{\mathbb{U}} uQ(du\times\mathbb{Y})$ P-almost surely. Following from (Step 1), (2.45) follows from [123, Theorem 3.5]. That is because, the cost function is continuous in actions, and $\min\{M, c\left(\omega_0, u, \int_{\mathbb{U}} uQ_n(du\times\mathbb{Y})\right)\}$ is continuously converges in u, that is,

$$\min\{M, c(\omega_0, u_n, \int_{\mathbb{U}} uQ_n(du\times\mathbb{Y}))\} \to \min\{M, c(\omega_0, u, \int_{\mathbb{U}} uQ(du\times\mathbb{Y}))\}, \tag{2.49}$$

where $u_n \to u$ as $n\to\infty$. Equality (2.46) follows from the dominated convergence theorem since $\min\{M, c\left(\omega_0, u, \int_{\mathbb{U}} uQ_N(du\times\mathbb{Y})\right)\}$ is bounded, and (2.47) is true since limsup is the greatest convergent subsequence limit for a bounded sequence. Finally, (2.48) follows from the definition of empirical measures and since for every M, $\min\{M, c\left(\omega_0, u, \int_{\mathbb{U}} uQ_N(du\times\mathbb{Y})\right)\} \leq c\left(\omega_0, u, \int_{\mathbb{U}} uQ_N(du\times\mathbb{Y})\right)$; hence, following from Theorem 2.2.2, the randomized limit policy through subsequence is a globally optimal for ($\mathcal{P}_\infty$).

□

2.5.11 Proof of Theorem 2.4.1

We invoke Theorem 2.2.3 to prove the theorem. The stationary policy (see Definition 2.1.1) is obtained as

$$\gamma_N^{i*}(y^i) = (R+Q)^{-1}Q(1+\frac{1}{N})E[x^i|y^i],$$

where the equality follows from the assumption that x^is are independent of z^is and x^ks, $k \neq i$ for every $i = 1, 2, ..., N$ and they are mean zero. Following from [78], stationary policies are team optimal for $(\mathcal{P}_N)$ in this formulation, and $\gamma_\infty^{i*}(y^i) = (R+Q)^{-1}QE[x^i|y^i]$. Since y^is are zero mean Gaussian random variables, we have $E[x^i|y^i] = \Sigma_{x^iy^i}\Sigma_{y^iy^i}^{-1}y^i := Ky^i$, where Σ_{XY} is defined as a covariance of two random variables X and Y. We have P-almost surely,

$$\begin{aligned}\sup_{1\leq i\leq N}||\gamma_N^{i*}(y^i) - \gamma_\infty^{i*}(y^i)|| &= \frac{Q}{R+Q}\sup_{1\leq i\leq N}|\frac{1}{N}E[x^i|y^i]| \\ &= \frac{KQ}{R+Q}\sup_{1\leq i\leq N}|\frac{1}{N}y^i| \xrightarrow[N\to\infty]{} 0, \end{aligned} \tag{2.50}$$

where (2.50) follows from

$$\lim_{N\to\infty}\sup_{1\leq i\leq N}\frac{1}{N^2}(y^i)^2 \leq \lim_{N\to\infty}\frac{1}{N^2}\sum_{i=1}^{N}(y^i)^2 = 0 \; P-a.s,$$

where the first inequality is true since $(y^i)^2$s are non-negative, and equality follows from the strong law of large numbers (SLLN) since y^is are i.i.d. and have a finite variance, hence, (b) holds. One can show that the condition in Remark 2.2.2(ii) holds since y^is and x^is are i.i.d. random variables, hence Theorem 2.2.3 completes the proof.

2.5.12 Proof of Theorem 2.4.2

In the following, we use both Theorem 2.2.1 and Theorem 2.2.3. Clearly, (A1) holds, we show that (A2) holds,

$$\begin{aligned}&\limsup_{N\to\infty}\frac{1}{N}E\bigg[\sum_{i=1}^{N}(\gamma_\infty^{i*}(y^i))^2R + Q(\gamma_\infty^{i*}(y^i) - x^i - \mu_N)^2\bigg] \\ &= \limsup_{N\to\infty}\frac{1}{N}E\bigg[\sum_{i=1}^{N}\frac{-Q^2}{Q+R}E^2[x^i|y^i](1+\frac{2}{N})(x^i+\mu_N)^2Q\bigg] \qquad (2.51) \\ &\leq \limsup_{N\to\infty}\frac{1}{N}E\bigg[\sum_{i=1}^{N}\frac{-Q^2}{Q+R}E^2[x^i|y^i]\bigg] + \lim_{N\to\infty}\frac{Q(N+3)\sigma^2}{N} \qquad (2.52)\end{aligned}$$

$$= \frac{-Q^2}{Q+R} E\Big[E^2[x^1|y^1]\Big] + Q\sigma^2, \tag{2.53}$$

where (2.51) follows from

$$E\Big[E[x^i|y^i](x^i+\mu_N)\Big] = E\Big[E\Big[E[x^i|y^i](x^i+\mu_N)\Big|y^i\Big]\Big] = (1+\frac{1}{N})E\Big[E^2[x^i|y^i]\Big], \tag{2.54}$$

and (2.52) is true since x^i an z^i are i.i.d. random variables and $\limsup_{N\to\infty} a_N + \limsup_{N\to\infty} b_N \geq \limsup_{N\to\infty}(a_N + b_N)$. We can justify (2.53) by defining $Y^i := (E[x^i|y^i])^2$, and since y^is are measurable functions of $\{y^i\}_{i\geq 1}$, and y^is and x^is are i.i.d., hence Y^is are i.i.d. random variables. Similarly, one can show the other side direction for liminf. Hence (A2) is satisfied. Now, we check (2.8), for every γ^k_∞ with $J(\underline{\gamma}_\infty) < \infty$,

$$\limsup_{N\to\infty} \frac{1}{N} E\Big[\sum_{i=1}^N \sum_{k=1}^N c_{u^k}(\omega_0, \gamma^{i*}, \mu^*)(m_k)\Big]$$

$$= \limsup_{N\to\infty} \frac{2Q}{N} \sum_{k=1}^N E\Big[E[x^k(m_k)|y^k]\Big] - E\Big[(x^k+\mu_N)(m_k)\Big] \tag{2.55}$$

$$= \limsup_{N\to\infty} \frac{-2Q}{N} \sum_{k=1}^N E\Big[\mu_N m_k\Big] \tag{2.56}$$

$$= 2Q \limsup_{N\to\infty} \frac{1}{N^2} \sum_{k=1}^N E\Big[x^k \gamma^{k*}_\infty(y^k)\Big] - E\Big[x^k \gamma^k_\infty(y^k)\Big] \tag{2.57}$$

$$= -2Q \liminf_{N\to\infty} \frac{1}{N^2} \sum_{k=1}^N E\Big[x^k \gamma^k_\infty(y^k)\Big] \tag{2.58}$$

$$\geq -2Q\sigma \liminf_{N\to\infty} \frac{1}{N^2} \sum_{k=1}^N \sqrt{E\Big[(\gamma^k_\infty(y^k))^2\Big]} \tag{2.59}$$

$$\geq -2Q\sigma \liminf_{N\to\infty} \sup_{1\leq k\leq N} \sqrt{\frac{E\Big[(\gamma^k_\infty(y^k))^2\Big]}{N^2}} = 0, \tag{2.60}$$

where measurability of $m_k := \gamma^k_\infty(y^k) - \gamma^{k*}_\infty(y^k)$ with respect to the σ-field generated by y^k implies (2.55), and (2.56) follows from the iterated expectations property. Since x^ps are mean

zero and independent of y^k for $k \neq p$, we have (2.57), and (2.58) follows from the fact that γ_∞^{k*} is independent of k, and since y^k and x^k are i.i.d. random variables. Moreover, $J(\underline{\gamma}_\infty) < \infty$, so $E[\gamma_\infty^k(y^k)] \leq E[(\gamma_\infty^k(y^k))^2] < \infty$, and Cauchy–Schwarz inequality implies (2.59), and (2.60) follows from

$$\liminf_{N\to\infty} \sup_{1\leq k\leq N} \frac{E\left[(\gamma_\infty^k(y^k))^2\right]}{N^2} \leq \liminf_{N\to\infty} \frac{1}{N^2} \sum_{k=1}^{N} E\left[(\gamma_\infty^k(y^k))^2\right] = 0, \tag{2.61}$$

where (2.61) is true since $E\left[(\gamma_\infty^k(y^k))^2\right] \geq 0$ and $\limsup_{N\to\infty} \frac{1}{N} E\left[\sum_{k=1}^{N} \left(\gamma_\infty^k(y^k)\right)^2 R\right] \leq J(\underline{\gamma}_\infty) < \infty$. Thus, (2.8) is satisfied and Theorem 2.2.1 completes the proof.

One can also invoke Theorem 2.2.3 to complete the proof. One can show that the condition in Remark 2.2.2(ii) holds since y^is and x^is are i.i.d. random variables. We only justify (b). Stationary policy is team optimal for $(\mathcal{P}_N)$ in this formulation [78], hence $\gamma_N^{i*}(y^i) = (R+Q)^{-1}Q(1 + \frac{1}{N})E[x^i|y^i]$, so we need to show that

$$\lim_{N\to\infty} \sup_{1\leq i\leq N} \left|\gamma_N^{i*}(y^i) - \gamma_\infty^{i*}(y^i)\right| = 0 \; P-a.s,$$

Equivalently, we can show that P-a.s

$$\lim_{N\to\infty} \sup_{1\leq i\leq N} \frac{1}{N^2} \left(E[x^i|y^i]\right)^2 \leq \lim_{N\to\infty} \frac{1}{N^2} \sum_{i=1}^{N} \left(E[x^i|y^i]\right)^2 = 0,$$

where the first inequality is true since $\left(E[x^i|y^i]\right)^2$s are non-negative, and equality follows from SLLN since

$$E\left[(E[x^i|y^i])^2\right] = E\left[(x^i)^2\right] - E\left[(x^i - E[x^i|y^i])^2\right] < \infty,$$

and $\left(E[x^i|y^i]\right)^2$ are i.i.d. sequence of random variables since y^is are i.i.d. random variables and the proof is completed.

2.5.13 Proof of Theorem 2.4.4

Since, $k^t_{T+1} = k^{t-1}_T$ for $t = 1, 2, \ldots, T$, one can write (2.20) as

$$k^t_T = Q + A^T k^t_{T-1} A - A^T k^t_{T-1} B(R + B^T k^t_{T-1} B)^{-1} B^T k^t_{T-1} A.$$

We use Theorem 2.2.2 and Remark 2.2.1, to show that $u^{t*}_\infty = G_\infty x_t$ is team optimal, where $G_\infty = -(R + B^T C^* B)^{-1} B^T C^* A$, and following from controllability of (A, B), $C^* = \lim_{\beta \to 1} C_\beta$, a fixed point of the following recursion exists,

$$C_\beta(n) = Q + A^T \beta C_\beta(n-1) A - A^T \beta C_\beta(n-1) B \left(R + B^T \beta C_\beta(n-1) B\right)^{-1} B^T \beta C_\beta(n-1) A$$

By comparing $C^*(n) = \lim_{\beta \to 1} C^*_\beta(n)$ and (2.20), we see that $\lim_{T\to\infty} k^t_T = K = C^* = \lim_{n\to\infty} C^*(n)$. Hence, we have for $t = 0, 1, \ldots, T-1$, $\lim_{T\to\infty} G^t_T = -(R + B^T K B)^{-1} B^T K A = -(R + B^T C^* B)^{-1} B^T C^* A = G_\infty$. In the following, we show (2.9) holds using Remark 2.2.1.

$$\begin{aligned}
&\limsup_{T\to\infty} |J_T(\underline{\gamma}^*_T) - J_T(\underline{\gamma}^*_\infty)| \\
&\leq \limsup_{T\to\infty} \sup_{0\leq t\leq T-1} \Bigg| E\left[\sum_{k=0}^{t} Tr\left[\zeta_k^T \left(L^{t,k}_T\right)^T (H^t_T) L^{t,k}_T \zeta_k\right]\right] \\
&\quad - E\left[\sum_{k=0}^{t} Tr\left[\zeta_k^T \left(L^{t,k}_\infty\right)^T (H^t_\infty) L^{t,k}_\infty \zeta_k\right]\right]\Bigg| && (2.62)\\
&\leq \limsup_{T\to\infty} \sup_{0\leq t\leq T-1} \left| E\left[\sum_{k=0}^{t} Tr[\zeta_k \zeta_k^T (\left(L^{t,k}_T\right)^T (H^t_T) L^{t,k}_T - \left(L^{t,k}_\infty\right)^T (H^t_\infty) L^{t,k}_\infty)]\right]\right| && (2.63)\\
&\leq \Sigma^2 \limsup_{T\to\infty} \sup_{0\leq t\leq T-1} \left| Tr\left[(H^t_T) C^t_T - (H^t_\infty) C^t_\infty\right]\right| && (2.64)\\
&\leq \Sigma^2 \limsup_{T\to\infty} [\sup_{0\leq t\leq T-1} |Tr((G^t_T)^T R G^t_T C^t_T - (G_\infty)^T R G_\infty C^t_\infty)| + \sup_{0\leq t\leq T-1} |Tr(Q e^t_T)|] && (2.65)
\end{aligned}$$

$$\leq \Sigma^2 \limsup_{T\to\infty} [\sup_{0\leq t\leq T-1} |Tr\bigg((G_T^t)^T R G_T^t - (G_\infty)^T R G_\infty\bigg)| \sup_{0\leq t\leq T-1} |Tr(C_T^t)|$$

$$+ \sup_{0\leq t\leq T-1} |Tr(G_\infty^T R G_\infty e_T^t)| + \sup_{0\leq t\leq T-1} |Tr(Q e_T^t)|] \tag{2.66}$$

$$\leq \Sigma^2 \limsup_{T\to\infty} [\sup_{0\leq t\leq T-1} |Tr\bigg[(G_T^t (G_T^t)^T - G_\infty G_\infty^T) R\bigg]| (\sup_{0\leq t\leq T-1} |Tr(e_T^t)| + \sup_{0\leq t\leq T-1} |Tr(C_\infty^t)|)$$

$$+ \sup_{0\leq t\leq T-1} |Tr(G_\infty^T R G_\infty e_T^t)| + \sup_{0\leq t\leq T-1} |Tr(Q e_T^t)|] = 0, \tag{2.67}$$

where $L_T^{t,k} := \prod_{p=k}^{t-1}(A + BG_T^p)$, $L_\infty^{t,k} := \prod_{p=k}^{t-1}(A + BG_\infty)$, $H_T^t = (Q + (G_T^t)^T R G_T^t)$, $H_\infty^t = (Q + (G_\infty)^T R G_\infty)$, $e_T^t := C_T^t - C_\infty^t$, and $C_T^t := \bigg[\sum_{k=0}^t L_T^{t,k}\bigg(L_T^{t,k}\bigg)^T\bigg]$, $C_\infty^t := \bigg[\sum_{k=0}^t L_\infty^{t,k}\bigg(L_\infty^{t,k}\bigg)^T\bigg]$ and $\Sigma^2 := \max(\sigma_{X_0}^2, \sigma_w^2)$, where $\sigma_{X_0}^2$ and σ_w^2 are the variance of each component of X_0 and w_k, respectively. Equality (2.62) follows from the fact that the trace of scalar is the scalar itself and the fact that $\{w_k\}_k$ are i.i.d. and independent from X_0. Inequality (2.63) follows from the trace property that $Tr(ABC) = Tr(CAB)$, and (2.64) follows from the hypothesis that ζ_ks are i.i.d. random vectors and $Tr(ABC) = Tr(BCA)$ and (2.65) follows from linearity of the trace and $\sup f + g \leq \sup f + \sup g$. Inequality (2.66) follows from adding and subtracting $G_\infty^T R G_\infty C_T^t$ in the first term and using $Tr(AB) \leq Tr(A)Tr(B)$ for A and B positive semi-definite matrices since one can observe that (2.20) implies for a fixed T, $\{k_T^t\}_{t=0}^{T-1}$ is a decreasing sequence, i.e., $K > k_T^0 > k_T^1 > \cdots > k_T^{T-1}$, and hence $\{G_T^t (G_T^t)^T\}_{t=0}^{T-1}$ is a decreasing sequence. Also, from (2.19), we have for a fixed T, $\{(A + BG_T^t)(A + BG_T^t)^T\}_{t=0}^{T-1}$ is an increasing sequence, hence one can show $(G_T^t)^T R G_T^t - G_\infty^T R G_\infty$ is positive semi-definite, and (2.67) follows from the definition of e_T^t and the following calculation. First, we show that for a fixed T, $\{Tr(e_T^t)\}_{t=0}^{T-1}$ is an increasing sequence using the aforementioned observations. Hence,

$$\lim_{T\to\infty} \sup_{0\leq t\leq T-1} |Tr(e_T^t)| = \lim_{T\to\infty} |Tr(e_T^{T-1})| = 0.$$

Similarly, $\lim_{T\to\infty} \sup_{0\leq t\leq T-1} |Tr(Q e_T^t)| = 0$. We have,

$$\lim_{T\to\infty} \sup_{0\leq t\leq T-1} |Tr(C_\infty^t)| = \bigg|Tr\bigg[(I - (A + BG_\infty))^{-1}\bigg]\bigg| = 0,$$

where $Y^{(T)}$ denotes the T power of the matrix Y and the result follows from the fact that $||A + BG_\infty|| < 1$ (following from the controllability assumption, the optimal policy stabilizes the dynamics). Finally, we have

$$\lim_{T\to\infty} \sup_{0\leq t\leq T-1} \left| Tr\left[\left(G_T^t (G_T^t)^T - G_\infty G_\infty^T \right) R \right] \right| = 0$$

where the second equality follows from the aforementioned observations and since R is positive definite. Therefore, $\limsup_{T\to\infty} J_T(\underline{\gamma}_T^*) - J(\underline{\gamma}_\infty^*) = 0$, and the proof is completed.

Chapter 3

Large-Scale Convex Stochastic Dynamic Teams and Their Mean-Field Limit

In this chapter, we study convergence, existence and structural properties of (globally) optimal policies for stochastic dynamic team problems with a finite as well as a countably infinite number of decision makers. We mainly focus on finite-horizon convex multi-stage dynamic teams with a symmetric information structure, where the cost function satisfies an exchangeability property. Hence, we extend the results of the previous chapter to the dynamic case.

We first briefly introduce the setup for stochastic (symmetric) dynamic teams with finite as well as infinitely many decision makers. In view of the intrinsic model described in the introductory chapter, we can view a collection of DMs as a single player acting at different time instances.

The set $\mathcal{N}$ denotes the collection of decision makers (which can be a finite set $\{1, 2, \ldots, N\}$ or a countably infinite set $\mathbb{N}$). Consider a collection of DMs as a single DM ($i \in \mathcal{N}$) acting at different time instances ($t \in \mathcal{T}$), where the set $\mathcal{T}$ denotes number of times each decision maker can act (the set $\mathcal{T}$ can be a finite set $\{1, 2, \ldots, T\}$ or a countably infinite set $\mathbb{N}$, depending on the finite-horizon or infinite-horizon setups). Define state dynamics and observations for each decision maker i at time t as

$$x^i_{t+1} = f_t(x^i_t, u^i_t, w^i_t), \tag{3.1}$$

$$y^i_t = h_t(x^i_{0:t}, u^i_{0:t-1}, v^i_{0:t}), \tag{3.2}$$

where functions f_t and h_t are measurable functions. The information structure of DM^i at time t is $I_t^i = \{y_t^i\}$ (or $I_t^i = \{y_k^i\}_{k\in K_t}$ for a subset $K_t \subseteq \{0,\dots,t\}$), and $\zeta_t^i := (w_t^i, v_t^i)$ (with $\zeta_0^i := (x_0^i, w_0^i, v_0^i)$) denotes uncertainty corresponding to dynamics and observations at time t for DM^i which are exogenous random vectors.

In the view of symmetry in the above setup (since functions f_t and h_t are identical through decision makers), we assume that spaces that state, action, observation, disturbances live are symmetric (identical) through players (also through times for simplicity of our analysis) and Borel subsets of appropriate dimension Euclidean space $\mathbb{X} \subseteq \mathbb{R}^m$, $\mathbb{U} \subseteq \mathbb{R}^{m'}$, $\mathbb{Y} \subseteq \mathbb{R}^{m''}$, $\mathbb{W}$, and $\mathbb{V}$, where m, m', and m'' are positive integers. Now, a dynamic stochastic team problem with finite (N) number of decision makers (also referred to as a pre-limit mean-field stochastic team problem) is defined as:

Problem ($\mathcal{P}_T^{N,\mathbf{MF}}$): Given a (decentralized) information structure, an N-DM finite-horizon stochastic team problem is described with an expected cost function (to minimize) of $\underline{\gamma}_T^{1:N}$ as

$$J_T^N(\underline{\gamma}_T^N) = \frac{1}{N}\sum_{t=0}^{T-1}\sum_{i=1}^{N} E^{\underline{\gamma}_T^{1:N}}\left[c\left(\omega_0, x_t^i, u_t^i, \frac{1}{N}\sum_{p=1}^{N} u_t^p, \frac{1}{N}\sum_{p=1}^{N} x_t^p\right)\right], \tag{3.3}$$

for some Borel measurable cost function $c : \Omega_0 \times \mathbb{X} \times \mathbb{U} \times \mathbb{U} \times \mathbb{X} \to \mathbb{R}_+$, where ω_0 is an exogenous random vector and $\underline{\gamma}_T^{1:N} = \gamma_{0:T-1}^{1:N}$.

We note that to distinguish the finite-horizon problems from infinite-horizon ones (Section 3.4.2), the sub-index T has been introduced. Also the super-index MF (mean-field) has been considered in the problem description to distinguish the mean-field setup from symmetric LQG problems (which will be defined later on in Section 3.4). In this chapter, for some of the results, we consider a more general setup of N-DM (symmetric) stochastic dynamic team problems, where the cost function is of an exchangeable type, and the information structure is symmetric. This will be defined precisely later on. We now introduce finite-horizon stochastic team problem with a countably infinite number of decision makers (also referred to as a mean-field team problem).

Problem ($\mathcal{P}_T^{\infty,\mathbf{MF}}$): Given a (decentralized) information structure, a mean-field team problem

is described with the expected cost function (to minimize) of $\underline{\gamma}_T$ as

$$J_T^{\infty}(\underline{\gamma}_T) = \limsup_{N\to\infty} \frac{1}{N} \sum_{t=0}^{T-1} \sum_{i=1}^{N} E^{\underline{\gamma}_T} \left[c\left(\omega_0, x_t^i, u_t^i, \frac{1}{N}\sum_{p=1}^{N} u_t^p, \frac{1}{N}\sum_{p=1}^{N} x_t^p \right) \right], \tag{3.4}$$

for some Borel measurable cost function $c : \Omega_0 \times \mathbb{X} \times \mathbb{U} \times \mathbb{U} \times \mathbb{X} \to \mathbb{R}_+$, where $\underline{\gamma}_T^i = \gamma_{0:T-1}^i$ for $i \in \mathbb{N}$ and $\underline{\gamma}_T = \{\underline{\gamma}_T^i\}_{i\in\mathbb{N}}$.

The global optimality can be defined similar to that of static problems in the preceding chapters. The main focus of this chapter is to address the following questions:

Problem 1: Connections between optimal policies of ($\mathcal{P}_T^{N,\mathbf{MF}}$) and ($\mathcal{P}_T^{\infty,\mathbf{MF}}$). Let a sequence of policies $\{\underline{\gamma}_T^{N*}\}_N$, where for each $N \in \mathcal{N}$ a policy $\underline{\gamma}_T^{N*} \in \Gamma_N$ is a globally optimal policy for N-DM stochastic dynamic team problem ($\mathcal{P}_T^{N,\mathrm{MF}}$) be given. Suppose $\{\underline{\gamma}_T^{N*}\}_N$ converges pointwise (almost-surely for every observation) to a limit policy $\underline{\gamma}_T^*$. Is $\underline{\gamma}_T^*$ (globally) optimal for ($\mathcal{P}_T^{\infty,\mathrm{MF}}$)?

Problem 2: Existence and structural properties of optimal policies of ($\mathcal{P}_T^{N,\mathbf{MF}}$) and ($\mathcal{P}_T^{\infty,\mathbf{MF}}$). The existence of (globally) optimal polices for ($\mathcal{P}_N'$) has been established under very general setups of team problems [140, 138, 62]. Since the cost function in ($\mathcal{P}_T^{N,\mathrm{MF}}$) are of exchangeable type, is it the case that a (globally) optimal policies for ($\mathcal{P}_T^{N,\mathrm{MF}}$) is symmetric (identical) through decision makers? The question that we also address is toward the existence and symmetry property of (globally) optimal policy for ($\mathcal{P}_T^{\infty,\mathrm{MF}}$). In particular, under what conditions, does there exist a symmetric globally (deterministic or randomized) optimal policy for ($\mathcal{P}_T^{\infty,\mathrm{MF}}$)?

Problem 3: Structural properties of optimal policies for finite-horizon and infinite-horizon dynamic team problems with a specific symmetric network topology. Consider a general setup of N-DM stochastic dynamic teams (not necessarily pre-limit mean-field model), where the cost function is of exchangeable type. Assume the information structure is symmetric for N-DM stochastic team problems, that is, functions h_t^i, inducing the observations of decision makers, are identical through decision makers (this holds for the mean-field problems formulated above). Now,

for these problems, can we show the symmetry of globally optimal policies through decision makers? Now, let us focus on N-DM LQG dynamic teams with a specific partially nested symmetric information structure (for example, a tree information structure). Linearity of a (globally) optimal policy for the finite-horizon case has been established in [66]. Can we establish the linearity of a (globally) optimal policy for the infinite-horizon case? In particular case, let a sequence $\{\underline{\gamma}_T^{N*}\}_T$, where for each $T \in \mathcal{T}$, policy $\underline{\gamma}_T^{N*}$ is a globally optimal policy for an N-DM problem with a finite horizon T, be given. Suppose $\{\underline{\gamma}_T^{N*}\}_T$ converges pointwise (almost-surely for every observation) to a limit policy $\underline{\gamma}^{N*}$ as $T \to \infty$. Under the above convergence, is $\underline{\gamma}^{N*}$ (globally) optimal for the infinite-horizon case?

We note that obtaining structural results in team problems is important towards establishing both existence and computational/approximation methods for optimal policies. In this chapter, we define the notion of exchangeable teams and symmetric information structures (as a refinement of those in the preceding chapter), and we show that, for convex exchangeable dynamic teams with finite horizons, optimal policies exhibit a symmetry structure. For any number of DMs, this symmetry structure is more relaxed when compared with the previous chapter, which focused on problems under a static information structure, and is applicable for dynamic teams which may not admit a static reduction, as long as convexity in policies holds for the team problem.

3.1 Literature Review and Contributions

Some relevant studies on the existence and convergence of equilibria from the mean-field games literature are the following: In [36], for one-shot mean-field games, under regularity assumptions on the cost function, it has been shown that mixed Nash strategies of N-player symmetric games converge through a subsequence to a limit (which is a weak-solution of the mean-field limit). In [56], through a concentration of measures argument, it has been shown that a subsequence of symmetric local approximate Nash equilibria for N player games converge to a solution for the mean-field game under the assumption that the normalized occupational measures converges

weakly to a deterministic measure. Furthermore, using a similar method in [81], assumptions on equilibrium policies of the large population mean-field symmetric stochastic differential games have been presented to allow the convergence of asymmetric approximate Nash equilibria to a weak solution of the mean-field game [81, Theorem 2.6] in the presence of common randomness. Using martingale methods and relaxed controls (see also [56, 81, 80, 40]), an existence result and a limit theory have been established for controlled McKean-Vlasov dynamics [82]. We note that in [81, 82, 80, 40], it has been assumed that each player has full access to the information available to all players, i.e., the controls are functions of all initial states, Wiener processes of all players, and common randomness. Thus, the information structure is centralized.

We further note that the existence results for open-loop equilibria have been established in [81, 40, 39, 56]. We note that in our setup under these strategies the information structure corresponds to the static problems. The equilibria with respect to *closed-loop* can be completely different since the deviating player can still influence the information of other players and hence it can influence the average of states or actions substantially. Under a convexity condition (which has been introduced in [55] and also considered in [82, 80]), and under the classical information structure (i.e., what would be a centralized problem in the team theoretic setup), convergence of Nash equilibria induced by closed-loop controllers (both path-dependent and Markovian) to a weak semi-Markov mean-field equilibrium has been established in [83] for finite horizon mean-field game problems. Recently, in [35], both a convergence result for all correlated equilibrium solutions of discrete finite state mean-field games as limits of exchangeable correlated equilibria restricted to Markov open-loop strategies and an approximation result for N-players correlated equilibria have been established.

We also note a result in [37] for the convergence of the closed-loop equilibria, where an infinite-dimensional PDE, the *master equation*, has been considered and its smooth solution has been used to show the convergence of the empirical measure to a mean-field game equilibrium. We note that, the approach requires uniqueness of the mean-field equilibrium [37]. For infinite horizon problems, in [38], an example of ergodic differential games with mean-field coupling has been constructed

such that limits of sequences of expected costs induced by symmetric Nash-equilibrium of N-player games capture expected costs induced by many more Nash-equilibruim policies including a mean-field equilibrium and social optima. In [83], the classical information structure (a centralized problem) has been considered, where in [38] it has been assumed that players have access to all the history of states of all players but not controls (we note that in the team problem setup with the classical information structure through using a classical result of Blackwell [28] in the case where each decision maker knows all the history of states of all decision makers, optimal policies can be realized as one in the centralized problem where just the global state is a sufficient statistic). As we see, information structure aspects lead to subtle differences in analysis and conclusions.

The above highlights the intricacies due to the information structure aspects: different from the aforementioned studies above, we consider information structures that are not necessarily static or classical. Also, in this chapter, we work with global optimality and not only person-by-person (or mean-field) equilibria and we establish existence and convergence results for a globally optimal policy for mean-field team problems. On the other hand, here since we work under the convexity assumption, the information structure does not allow for the mean-field coupling in the dynamics. This assumption will be relaxed in the next chapter, where non-convex problems are to be studied. In this chapter, we also obtain existence results on optimal policies for the setups considered. Compared to the results on the existence of a globally optimal policy in team problems where (finite) N-DM team problems have been considered [138, 62, 140, 108], we study convex team problems with a countably infinite number of decision makers.

Parts of our results in this chapter correspond to LQG teams. Under static reductions, dynamic teams with a partially nested information structure can be reduced to a static one ([66, 135]) where Radner's theorem concludes global optimality of linear policies for LQG team problems [102]. However, for average cost infinite horizon, partially nested, LQG dynamic team problems so far there has been no universal result establishing that a globally optimal policy is linear, time-invariant, and stabilizing, and this has been often imposed a priori: In [105], the problem of

designing a linear, time-invariant, stabilizing, state feedback optimal policy for decentralized $\mathcal{H}_2$-optimization problems, which satisfy the quadratically invariance property, has been addressed by reparametrizing the problem as a convex problem (via Youla parameterization). In [104], it has been shown that for sequential team problems involving linear systems, quadratic invariance and the partially nested property are equivalent. For a class of partially ordered (POSET) systems, state space techniques have been utilized to obtain optimal, linear, time-invariant, state feedback controllers for $\mathcal{H}_2$-optimization problems with sparsity constraints [124]. A similar result has been established in [125] where linearity and time invariance have been imposed a priori. In [106], the problem of designing linear time-invariant stabilizing controllers, belonging to a pre-selected quadratically invariant subspace has been addressed. In [107], a convex characterization of the set of all stable time-invariant closed-loop linear systems which result from applying linear time-invariant stabilizing controllers has been studied under a strong quadratic invariance condition. In [87], $\mathcal{H}_2$-optimization output feedback problems with two-players have been considered and optimality results have been established when the optimal policies are restricted to linear, time invariant, stabilizing policies. However, the results in [87, 105, 124, 125, 106, 107] may be inconclusive regarding global optimality. Our contribution here is to consider average cost infinite horizon dynamic team problems without restricting the set of policies to those that are linear, time-invariant, and stabilizing unlike the results in [87, 105, 124, 125, 106, 107]. We note again that the optimality of linear policies for infinite horizon LQG team problems is an open problem in its generality and we provide positive results for a class of such problems.

Contributions. In view of the discussion above, this chapter makes the following contributions.

(i) We define a notion of exchangeable teams and symmetric information structures, and we show that, for convex exchangeable dynamic teams with finite horizons, optimal policies exhibit a symmetry structure (Theorem 3.2.2). For any number of DMs, this symmetry structure is more relaxed when compared with the symmetry results developed in [120, 117] and is applicable for dynamic teams which may not admit a static reduction, as long as

convexity in policies holds for the team problem.

(ii) For convex mean-field teams with a symmetric information structure, through concentration of measure arguments, we establish the convergence of optimal policies for mean-field teams with N decision makers to the corresponding optimal policies for mean-field teams (see Theorem 3.3.1).

(iii) We establish an existence result for the class of convex mean-field teams with a symmetric information structure (see Theorem 3.3.3) for finite horizon problems, where, as noted in the literature review, related results assumed more restrictive information structures which are either static or classical.

(iv) We also apply our results to LQG dynamic teams for finite horizon problems (see Section 3.4). For LQG dynamic teams with a symmetric partially information pattern, we obtain an optimal policy for finite horizon problems (see Section 3.4.1). We also apply convex mean-field results to LQG mean-field teams with a symmetric partially nested information structure (see Section 3.4.1) and obtain a globally optimal policy. Building on the result above, we also obtain a globally optimal policy for average cost LQG team problems.

The organization of the chapter is as follows: we study convex exchangeable dynamic teams with finite horizons in Section 3.2, and we study mean-field teams in Section 3.3. We obtain globally optimal solutions for finite horizon problems with a symmetric partially nested information structure and LQG mean-field teams in Section 3.4.1, and we discuss average cost LQG team problems with a symmetric information structure in Section 3.4.2, respectively. Proofs are provided in Section 3.5.

3.2 Finite Horizon Convex Dynamic Team Problems with a Symmetric Information Structure

In this section, we characterize symmetry in dynamic team problems. Consider *sequential systems* and assume the action and measurement spaces are standard Borel spaces. According to the intrinsic model for (multi-stage) dynamic team problems, by considering a collection of DMs as a single DM ($i = 1, \dots, N$) acting at different time instances ($t = 0, \dots, T-1$), we define a team problem with (NT)-DMs as a team with N-DMs:

(i) Let the observation and action spaces be identical for each DM ($i = 1, \dots, N$) with $\mathbb{Y}_i := \mathbf{Y} = \prod_{t=0}^{T-1} \mathbb{Y}^t$, $\mathbb{U}_i := \mathbf{U} = \prod_{t=0}^{T-1} \mathbb{U}^t$, respectively. The sets of all admissible policies are denoted by $\mathbf{\Gamma} = \prod_{i=1}^{N} \Gamma_i = \prod_{i=1}^{N} \prod_{t=0}^{T-1} \Gamma^t$.

(ii) For $i = 1, \dots, N$, $y_t^i := h_t^i(x_0^{1:N}, \zeta_{0:t}^{1:N}, u_{0:t-1}^{1:N})$ represents the observation of DM^i at time t (h_t^is are Borel measurable functions).

(iii) Let $(\underline{\zeta}^{1:N}) := (\underline{\zeta}^1, \dots, \underline{\zeta}^N)$ where $\underline{\zeta}^i := (x_0^i, \zeta_{0:T-1}^i)$ denotes all the uncertainty associated with DM^i including his/her initial states. We assume that $(\underline{\zeta}^i)$ takes values in Ω_ζ (where at each time instances t, it takes value in Ω_{ζ_t}). Let μ denote the law of $\underline{\zeta}^{1:N}$.

(iv) Define the expected cost function of $\underline{\gamma}^{1:N}$ as $J_T(\underline{\gamma}^{1:N}) = E^{\underline{\gamma}^{1:N}}[c(\underline{\zeta}^{1:N}, \underline{u}^{1:N})]$, for some Borel measurable cost function $c : \prod_{i=1}^{N}(\Omega_\zeta \times \mathbf{U}) \to \mathbb{R}_+$, where $\underline{\gamma}^{1:N} = (\underline{\gamma}^1, \underline{\gamma}^2, \dots, \underline{\gamma}^N)$ and $\underline{\gamma}^i = \gamma_{0:T-1}^i$ for $i = 1, \dots, N$.

Now, we present the definition of symmetric information structures (note that symmetric information structures can be classical, partially nested, or non-classical).

Definition. *Let the information of DM^i acting at time t be described as $I_t^i := \{y_t^i\}$ (or $I_t^i = \{y_k^i\}_{k \in K_t}$ for a subset $K_t \subseteq \{0, \dots, t\}$). The information structure of a sequential N-DM team problem is symmetric if*

(i) $y_t^i = h_t(x_0^i, x_0^{-i}, \zeta_{0:t}^i, \zeta_{0:t}^{-i}, u_{0:t-1}^i, u_{0:t-1}^{-i})$ where h_t is identical for all $i = 1, \dots, N$ (note that function's arguments depend on i).

We note that the above definition can be generalized to be applicable for teams with countably infinite DMs and infinite horizon problems.

The symmetric information structure can also be interpreted and defined as a graph, which has often been the common method to describe information structures in control theory, relating DMs and their information through directed edges. Consider $G(V,\mu)$ as a directed graph with $V=\{1,\dots,NT\}$ nodes and where $\mu\subset V\times V$ determines the directed edges between nodes; this represents the dependency notation in the information of nodes, i.e., (i,j) denotes a directed edge from i to j, $i\to j$, it means u^i affects y^j through the relation $y^j=h^j(\omega,\underline{u}^{1:j-1})$ defined in the intrinsic model. We denote by $\downarrow j$ as the set of nodes i such that $i\to j$ (ancestors), and $\downdownarrows j:=\{\downarrow j\}\cup\{j\}$. Similarly, we can define descendants by $\uparrow j$. We can define a collection of DMs as a single DM ($i=1,\dots,N$) acting at different time instances ($t=0,\dots,T-1$) on a graph with a symmetric information structure (two examples are shown in Fig. (3.1), and Fig. (3.1)). Assume

(i) there exists a node $\{i\}$ (root node), ω_0. Each sub-graph represents a single DM acting at time instances $t=0,\dots,T-1$, and there exists a finite number of sub-graphs $G_p(\hat{V},\hat{\mu})$ such that $\cup_{p=1}^{N}G_p\cup\{i\}=G$, where G_ps are isomorphic (see e.g., [128]) for all $p=1,\dots,N$, i.e., for every node with directed edges in each sub-graph there exists a unique node with identical directed edges in the corresponding sub-graphs, where $\hat{V}=\{0,\dots,T-1\}$, and G_p^k refers to a node k in G_p for all $p=1,\dots,N$ and $k=0,\dots,T-1$,

(ii) sharing of the information is symmetric across sub-graphs, i.e., for $p,s=1,\dots,N$, and $k,j=0,\dots,T-1$, and for every edge from a node G_p^k to a node G_s^j, there exists an edge from a node G_p^k to nodes G_{-p}^j , where G_{-p}^j denotes $(G_1^j,\dots,G_{p-1}^j,G_{p+1}^j,\dots,G_N^j)$, and also there exist edges from nodes G_{-p}^k to a node G_p^j.

Assumption 3.2.1. *For any permutation σ of the set $\{1,\dots,N\}$, we have for all ω_0*

$$c(\omega_0,(\underline{\zeta}^{\sigma})^{1:N},(\underline{u}^{\sigma})^{1:N})=c(\omega_0,\underline{\zeta}^{1:N},\underline{u}^{1:N}),\tag{3.5}$$

where $(\underline{\zeta}^{\sigma})^{1:N} := (\underline{\zeta}^{\sigma(1)}, \dots, \underline{\zeta}^{\sigma(N)})$ *and* $(\underline{u}^{\sigma})^{1:N} := (\underline{u}^{\sigma(1)}, \dots, \underline{u}^{\sigma(N)})$.

Here, we recall some definitions and results from [140, Section 3.3] on convexity of static and dynamic team problems required to follow the result in this chapter.

Definition. *[140, Section 3.3] An N-DM team problem (static or dynamic) is convex in policies if for any two team policies $\underline{\gamma}_T^{1:N}$ and $\tilde{\underline{\gamma}}_T^{1:N}$ in the set $\{\underline{\gamma}_T^{1:N} \in \mathbf{\Gamma} : J_T(\underline{\gamma}_T^{1:N}) < \infty\}$, and for any $\alpha \in (0,1)$, we have*

$$J_T(\alpha \underline{\gamma}_T^{1:N} + (1-\alpha)\tilde{\underline{\gamma}}_T^{1:N}) \leq \alpha J_T(\underline{\gamma}_T^{1:N}) + (1-\alpha) J_T(\tilde{\underline{\gamma}}_T^{1:N}).$$

The above definition can also be applied to infinite-horizon and/or teams with a countably infinite number of DMs. We recall sufficient conditions for convexity of static and dynamic team problems following [140, Section 3.3].

Theorem 3.2.1. *[140, Section 3.3] Consider a sequential team problems, and assume action spaces are convex, and the expected cost function is finite for all admissible policies $\underline{\gamma} \in \mathbf{\Gamma}$ (or alternatively, restrict the set to those leading to the finite cost). Then*

(i) for static team problems convexity of the cost function in actions is sufficient for convexity of the team problem in policies,

(ii) for dynamic team problems with a static reduction, convexity of the team problem in policies is equivalent to the convexity of its static reduction.

(iii) in particular, for partially nested dynamic teams with a static reduction (more generally, for stochastically partially nested team problems [140, Section 3.3]) if the cost function is convex

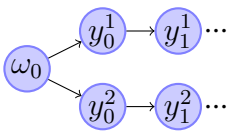

Figure 3.1: A tree structure of a symmetric dynamic team.

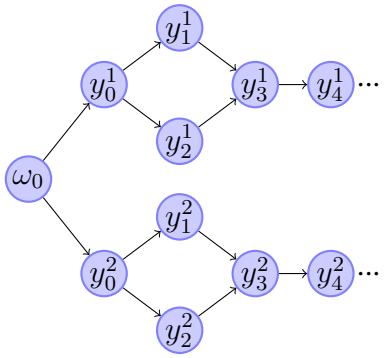

Figure 3.2: An example of a graph structure of a symmetric dynamic team.

in actions then for the reduced team problem with an equivalent information structure (see Chapter 5 for refinement via control sharing according to the partial nested information structure) is convex on $\mathbf{\Gamma}$.

The conditions above, are only sufficient conditions [140, Example 1]. We note however that as a Corollary for (ii) above, for the LQG setup, under partial nestedness, convexity in policies hold as a consequence of Ho-Chu's theorem; we will study this case in Section 3.4. On the other hand, not all LQG problems are convex: the celebrated counterexample of Witsenhausen [130] demonstrates that under non-classical information structures, even LQG problems may not be convex and optimal policies may not be linear (see [140, p. 22] for a proof of non-convexity).

3.2.1 Optimality of Symmetric Policies for Convex Dynamic Teams with A Symmetric Information Structure

In the following, we define notions of exchangeable and symmetrically optimal teams analogous to Section 2.3 of Chapter 2 for dynamic teams.

Definition. *(Exchangeable teams)*

An N-DM team is exchangeable if the value of the expected cost function is invariant under every permutation of policies of DMs, i.e., $J_T(\underline{\gamma}_T^1, \underline{\gamma}_T^2, \ldots, \underline{\gamma}_T^N) = J_T((\underline{\gamma}_T^\sigma)^1, \ldots, (\underline{\gamma}_T^\sigma)^N)$.

Definition. *(Symmetrically optimal teams)*

A team is symmetrically optimal, if for every given policy $\underline{\gamma}_T = (\underline{\gamma}_T^1, \ldots, \underline{\gamma}_T^N)$, *there exists an*

identically symmetric policy (i.e., each DM has the same policy, $\underline{\tilde{\gamma}}_T = (\underline{\tilde{\gamma}}^1_T, \dots, \underline{\tilde{\gamma}}^N_T)$, and $\underline{\tilde{\gamma}}^i_T = \underline{\tilde{\gamma}}^j_T$ for all $i, j = 1, \dots, N$) which performs at least as good as the given policy.

Remark 3.2.1. *The concepts of exchangeable and symmetrically optimal dynamic teams in this chapter are generalizations of those for static teams in preceding chapter. However, here, the value of the expected cost function may not be invariant under exchanging γ^i_t with γ^j_k for $k \neq t$, $k, t = 0, \dots, T-1$, and for $i, j = 1, \dots, N$.*

Here, we give a characterization for exchangeable and symmetrically optimal dynamic teams.

Theorem 3.2.2. *Consider dynamic team problems with a symmetric information structure under Assumption 3.2.1. If*

(a) action spaces $\mathbb{U}^t$ are convex for all $t = 0, \dots, T-1$,

(b) $(\underline{\zeta}^1, \dots, \underline{\zeta}^N)$ are exchangeable,

(c) for all policies $\underline{\gamma} \in \mathbf{\Gamma}$, and for all $A = A^1 \times \ldots \times A^N$ where $A^i \in \mathcal{Y}^i$,

$$\prod_{t=0}^{T-1} P\Big(y_t^{1:N} \in A \Big| x_0^{1:N}, \zeta_{0:t-1}^{1:N}, y_{0:t-1}^{1:N}, \gamma_0^1(y_0^1), \dots, \gamma_{t-1}^1(y_{t-1}^1), \gamma_0^2(y_0^2), \dots, \gamma_{t-1}^N(y_{t-1}^N) \Big)$$
$$= \prod_{t=0}^{T-1} \prod_{i=1}^{N} P\Big(y_t^i \in A^i \Big| x_0^i, \zeta_{0:t-1}^i, y_{\downarrow t}^{\downarrow\downarrow i}, \gamma_{\downarrow t}^{\downarrow\downarrow i}(y_{\downarrow t}^{\downarrow\downarrow i}) \Big), \tag{3.6}$$

where $y_{\downarrow t}^{\downarrow\downarrow i} := \{y_p^j | u_p^j$ affects $y_t^i \; \forall \; p = 0, \dots, t-1$ and $\forall j = 1, \dots, N\}$ and $(\gamma_{\downarrow t}^{\downarrow\downarrow i}(y_{\downarrow t}^{\downarrow\downarrow i}))$ can be defined similarly,

(i) then, the team problem is exchangeable.

(ii) Furthermore, if the team problem is convex in policies (see Theorem 3.2.1), then the team is symmetrically optimal.

Examples will be given in Section 3.3 and Section 3.4.1 where Theorem 3.2.2 can be applied. Here, we present the result for a class of problems that admit a static reduction (see [139, Section 3.7], [140, Section 1.2], [67, 135]).

Lemma 3.2.1. *Consider a dynamic team problem with a symmetric partially nested information structure (see Definition 3.2) which admits a static reduction. Under Assumption 3.2.1, and Assumptions (a), (b), (c) of Theorem 3.2.2, if the cost function is jointly convex in $\underline{u}^1, \dots, \underline{u}^N$ P-almost surely, then the team is symmetrically optimal.*

Proof. The proof follows from Theorem 3.2.1(iii) and Theorem 3.2.2 since the team is convex on $\mathbf{\Gamma}$ under the static reduction which is equivalent to the dynamic problem. □

Hence, it follows that if a static reduction of an exchangeable, symmetrically optimal, dynamic team exists, then it is exchangeable and symmetrically optimal.

3.3 Convex Mean-Field Teams with a Symmetric Information Structure

Consider the dynamic team problems ($\mathcal{P}_T^{N,\mathrm{MF}}$) and ($\mathcal{P}_T^{\infty,\mathrm{MF}}$), where the cost function and dynamics satisfies the following regularity conditions.

Assumption 3.3.1. *Assume*

(a) function $f_t : \mathbb{X} \times \mathbb{U} \times \mathbb{W} \to \mathbb{X}$ is continuous in its first and second arguments for all w_t^i and for each $i \in \mathbb{N}$ and uniformly bounded,

(b) function $h_t : \prod_{k=0}^{t} \mathbb{X} \times \prod_{k=0}^{t-1} \mathbb{U} \times \prod_{k=0}^{t} \mathbb{V} \to \mathbb{Y}$ is continuous in states and actions for all $v_{0:t}^i$ and for each $i \in \mathbb{N}$, and

(c) the cost function in (3.3)*, $c : \Omega_0 \times \mathbb{X} \times \mathbb{U} \times \mathbb{U} \times \mathbb{X} \to \mathbb{R}_+$, is continuous in its second, third, fourth, and fifth arguments for all ω_0.*

In the following, we first establish global optimality results under Assumption 3.3.2 (see Theorem 3.3.1), then we establish the result under a more relaxed assumption, Assumption 3.3.3 (see Theorem 3.3.2):

Assumption 3.3.2. *Assume*

(i) $(x_0^1, x_0^2, \dots)$ *are i.i.d. random vectors conditioned on* ω_0*,*

(ii) for $t = 0, \dots, T-1$, $\{w_t^i\}_{i\in\mathbb{N}}$ *are i.i.d. random vectors, and for* $i \in \mathbb{N}$, $\{w_t^i\}_{t=0}^{T-1}$ *are mutually independent, and independent of* ω_0 *and* $(x_0^1, x_0^2, \dots)$*. For* $t = 0, \dots, T-1$, $\{v_t^i\}_{i\in\mathbb{N}}$ *are i.i.d. random vectors, and for* $i \in \mathbb{N}$, $\{v_t^i\}_{t=0}^{T-1}$ *are mutually independent, and independent of* ω_0*,* $(x_0^1, x_0^2, \dots)$*, and* w_t^i*s for* $i \in \mathbb{N}$ *and* $t = 0, \dots, T-1$*.*

Assumption 3.3.3. *Assume that conditioned on* ω_0*,* $(x_0^1, x_0^2, \dots)$ *are exchangeable random vectors.*

Later on we will establish an existence theorem under Assumption 3.3.2, and we note that the proof under Assumption 3.3.2 is more direct. This is why two separate theorems will be presented, and the proof of the latter will be built on that of the former.

Lemma 3.3.1. *Consider a team defined as* $(\mathcal{P}_T^{N,MF})$ *with a symmetric information structure. Assume the team problem is convex in policies. Let the action space be compact and convex for each decision makers. Under Assumption 3.3.1 and Assumption 3.3.2, the team is symmetrically optimal.*

Proof. The proof follows from Theorem 3.2.2. □

Theorem 3.3.1. *Consider a team defined as* $(\mathcal{P}_T^{\infty,MF})$ *with* $(\mathcal{P}_T^{N,MF})$ *having a symmetric information structure for every* N*. Assume for every* N *the team problem is convex in policies. Let the action space be compact and convex for each DM. Under Assumption 3.3.1, and Assumption 3.3.2, if there exists a sequence of optimal policies for* $(\mathcal{P}_T^{N,MF})$*,* $\{\underline{\gamma}_T^{*,N}\}_N$*, which converges (for every DM due to the symmetry) pointwise to* $\underline{\gamma}_T^{*,\infty}$ *as* $N \to \infty$*, then* $\underline{\gamma}_T^{*,\infty}$ *(which is identically symmetric) is an optimal policy for* $(\mathcal{P}_T^{\infty,MF})$*.*

Remark 3.3.1. *On the connection between finitely many DMs and infinitely many DMs, we note a closely related work on mean-field games by Fischer [56] where the information structure is assumed to be static since the policy of each player is assumed to be adapted to the filtration generated by his/her initial states and Wiener process (also called in the mean-field games' literature,*

somewhat non-standard in the control literature, as open-loop distributed controllers [81],[40, pages 72-76]). This means that the information of each DM is not affected by any of the actions of the other DMs. For dynamic teams, there are two difficulties: (1) obtaining variational equations is challenging since fixing policies of DMs and perturbing only DM's policies, perturbs the observation of other DMs and hence the controls $u^{-i*} = (\underline{\gamma}^{1*}(\underline{y}^1), \ldots, \underline{\gamma}^{i-1*}(\underline{y}^{i-1}), \underline{\gamma}^{i+1*}(\underline{y}^{i+1}), \ldots, \underline{\gamma}^{N*}(\underline{y}^N))$*; (2) solutions of variational equations which give person-by-person optimal policies are inconclusive for global optimality due to the lack of convexity in general.*

Remark 3.3.2. *We also note additional related works by Lacker [82, 83] where either convergence of open-loop controllers, or convergence of Nash equilibria induced by closed-loop controllers (where controls are measurable path-dependent functions of states,* $u_t^i = \phi(t, \boldsymbol{x}_{0:t})$*, where* $\boldsymbol{x}_{0:t} = (x_{0:t}^1, \ldots, x_{0:t}^N)$ *and* ϕ *is a measurable function) or Markovian controllers (*$u_t^i = \phi(t, \boldsymbol{x}_t)$*, where* $\boldsymbol{x}_t = (x_t^1, \ldots, x_t^N)$*) have been considered. In [83], the information structure is classical (a centralized problem) since players have access to all the information available to previous DMs).*

Remark 3.3.3. *In Lemma 3.3.1 and Theorem 3.3.1, we considered a non-classical information structure for teams defined as (*$\mathcal{P}_T^{\infty,MF}$*) with a convex expected cost in policies. For teams defined as (*$\mathcal{P}_T^{\infty,MF}$*) with a symmetric partially nested information structure which admit static reduction, the above result holds and similar to the proof of Theorem 3.3.1, it can be proven under the assumption that the cost functions is convex in actions (since convexity of the cost function in actions is a sufficient condition for convexity of the expected cost function in policies for this class of problems [140, Theorem 3.7]).*

Remark 3.3.4. *Assumptions that action spaces are compact and* f_t*s are bounded can be relaxed by assuming that*

(A1) $\sup_{N \geq 1} E[||\underline{\gamma}_T^{*,N}(\underline{y}^1)||^{1+\delta}] < \infty$ *for some* $\delta > 0$*,*

(A2) $\sup_{N \geq 1} E[||\Lambda(\underline{\gamma}_T^{*,N}(\underline{y}^1), \underline{\zeta}^1)||^{1+\tilde{\delta}}] < \infty$ *for some* $\tilde{\delta} > 0$*.*

That is because, following from the pointwise convergence of $\underline{\gamma}_T^{*,N}$ and continuity of Λ in actions, the above uniform integrability assumption justifies exchanging the limit and the expectation required to establish the convergence in (3.40) *and* (3.41) *using a similar analysis of* (3.31) *and an argument of* (3.32) *based on the strong law of large numbers. This result is particularly important for LQG models (we use this remark in Section 3.4).*

Theorem 3.3.2. *Consider a team defined as ($\mathcal{P}_T^{\infty,MF}$) with ($\mathcal{P}_T^{N,MF}$) having a symmetric information structure for every N. Assume for every N the team problem is convex in policies. Let the action space be compact and convex for each DM, and assume Assumption 3.3.1, Assumption 3.3.2(ii), and Assumption 3.3.3 hold. If there exists a sequence of optimal policies for ($\mathcal{P}_T^{N,MF}$), $\{\underline{\gamma}_T^{*,N}\}_N$, which converges (for every DM due to the symmetry) pointwise to $\underline{\gamma}_T^{*,\infty}$ as $N \to \infty$, then $\underline{\gamma}_T^{*,\infty}$ (which is identically symmetric) is an optimal policy for ($\mathcal{P}_T^{\infty,MF}$).*

3.3.1 An Existence Theorem on Globally Optimal Policies for Dynamic Mean-Field Team Problems with a Symmetric Information Structure

An implication of Theorem 3.3.1 is the following existence result on globally optimal policies for mean-field team problems. In particular, we will establish the existence of a converging subsequence, in an appropriate sense, for a sequence of optimal policies for N-DM teams with an increasing number of DMs. For the following theorem, we do not establish the pointwise convergence; but by Theorem 3.3.1, if a sequence of optimal policies for ($\mathcal{P}_T^{N,MF}$), $\{\underline{\gamma}_T^{*,N}\}_N$, converges pointwise, a global optimal policy exists. To this end, we allow decision makers to apply randomized policies. For each decision maker (DM^i for $i \in \mathbb{N}$), a probability measure $P \in \mathcal{P}(\Omega_0 \times \mathbb{X} \times \prod_{t=0}^{T-1}(\mathbb{W} \times \mathbb{V}) \times \prod_{t=0}^{T-1}(\mathbb{U} \times \mathbb{Y}))$ is a policy induced by a randomized policy if and only if for every $t = 0, \ldots, T-1$ and for all continuous and bounded function g

$$\begin{aligned} &\int g(\omega_0, x_0^i, \zeta_{0:t-1}^i, y_{0:t}^i, u_{0:t}^i) P(d\omega_0, dx_0^i, d\zeta_{0:t-1}^i, dy_{0:t}^i, du_{0:t}^i) \\ &= \int g(\omega_0, x_0^i, \zeta_{0:t-1}^i, y_{0:t}^i, u_{0:t}^i) \mu^i(dx_0^i, d\zeta_{0:t-1}^i | \omega_0) P(d\omega_0) \end{aligned} \tag{3.7}$$

$$\times \prod_{k=0}^{t} \Pi_k^i(du_k^i|y_k^i) p_k^i(dy_k^i|\omega_0, x_0^i, \zeta_{0:k-1}^i, y_{0:k-1}^i, u_{0:k-1}^i),$$

for a stochastic kernel Π_k^i on $\mathbb{U}$ given $\mathbb{Y}$, where p_k^i is the transition kernel characterizing the observations of DM^i at time t,

$$\begin{aligned} &p_k^i\left(y_k^i \in \cdot \middle| \omega_0, x_0^i, \zeta_{0:k-1}^i, y_{0:k-1}^i, u_{0:k-1}^i\right) \\ &\quad := P\left(h_k(x_{0:k}^i, u_{0:k-1}^i, v_{0:k}^i) \in \cdot \middle| \omega_0, x_0^i, \zeta_{0:k-1}^i, y_{0:k-1}^i, u_{0:k-1}^i\right), \end{aligned}$$

and μ^i is a fixed probability measure on initial states and disturbances of DM^i conditioned on ω_0. This equivalency follows from the fact that continuous and bounded functions form a separating class [26, page 12] and [137, Theorem 2.2].

First, we present an absolute continuity assumption on observations of DMs.

Assumption 3.3.4. *For every DM^i and $t = 0, \ldots, T-1$, there exists a function $\psi_t^i : \mathbb{Y} \times \Omega_0 \times \mathbb{X} \times \prod_{k=0}^{t-1}(\mathbb{W} \times \mathbb{V}) \times \prod_{k=0}^{t-1}(\mathbb{Y} \times \mathbb{U}) \to \mathbb{R}_+$ continuous in actions, and a probability measure ν_t^i on $\mathbb{Y}$ such that for all Borel sets $A = A^1 \times \ldots \times A^N$,*

$$P(y_t^{1:N} \in A|\omega_0, x_0^{1:N}, \zeta_{0:t-1}^{1:N}, y_{0:t-1}^{1:N}, u_{0:t-1}^{1:N}) = \prod_{i=1}^{N} \int_{A^i} \psi_t^i(y_t^i, \omega_0, x_0^i, \zeta_{0:t-1}^i, y_{0:t-1}^i, u_{0:t-1}^i) \nu_t^i(dy_t^i).$$

This assumption allows us to obtain an independent measurements reduction (see [138, Section 2.2]). For example, if v_t^i for all $i \in \mathbb{N}$ and $t = 0, \ldots, T-1$ are i.i.d with a probability measure admitting a density function so that the observation of each DM^i at time t is $y_t^i = \tilde{h}_t(x_t^i, u_{\downarrow t}^i) + v_t^i$, where $\tilde{h}_t$ is continuous, then Assumption 3.3.4 holds [62, Lemma 5.1].

Theorem 3.3.3. *Consider a team defined as ($\mathcal{P}_T^{\infty,MF}$) with ($\mathcal{P}_T^{N,MF}$) having a symmetric information structure for every N. Assume for every N, the team problem is convex in policies and the action space is convex. Assume further that without any loss, the optimal policies can be restricted to those with $E(\phi_i(u^i)) \leq K$ for some finite K, where $\phi_i : \mathbb{U} \to \mathbb{R}_+$ is lower semi-continuous (moment condition). Under Assumption 3.3.1 and Assumption 3.3.2 if either*

(i) Assumption 3.3.4 holds (with no further assumptions on the information structure of each DM^i for $i \in \mathbb{N}$ through time $t = 0, \dots, T-1$), or

(ii) for each DM^i for $i \in \mathbb{N}$ through time $t = 0, \dots, T-1$, there exists a static reduction with the classical information structure (i.e., under a static reduction, the information structure is expanding such that $\sigma(y_t^i) \subset \sigma(y_{t+1}^i)$ where σ denotes the σ-field),

then there exists an optimal policy for ($\mathcal{P}_T^{\infty,MF}$).

Since the space of policies that are deterministic (where Π_k^i in (3.7) are indicator functions) are not closed under the weak convergence topology (e.g., as an implication of [140, Theorem 2.7]), we allow for randomization in the policies and therefore the limit policy is not necessarily deterministic according to the above result; however, it is identical for each DM.

Remark 3.3.5. *For the existence result, to show that the set of policies induced by independently randomized policies for each DM through time $t = 0, \dots, T-1$ (see* (3.7)*) is closed under the weak convergence topology, we utilized the result in [138, Section 5.2] which are more general than those in [62, 140]. We note that the extension of the existence results in [138, Section 5.2] to our setup is not immediate since the conclusion of (Step 3) can not be established rigorously without considering the technical steps involving infinite dimensions and limit arguments.*

3.4 Symmetric LQG Dynamic Teams

In the section, we consider LQG setup where the results of Section 3.2 and Section 3.3 can be applied. We first consider N-DM LQG problems where we use Theorem 3.2.2 to show that the globally optimal policies are symmetric. Then, based on symmetry, we calculate N-DM optimal policies for such problems. Next, using Theorem 3.3.1 and Theorem 3.3.2, we show the convergence of N-DM optimal policies to optimal policies of LQG mean-field teams with a countably infinite number of DMs. Finally, we consider infinite horizon problems where we use symmetry and convergence results to obtain global optimal policies for such problems.

3.4.1 Symmetric Partially Nested LQG Dynamic Teams on a Graph

In the following, we consider decentralized problems where Theorem 3.2.2 can be utilized and the optimal policy can be obtained. First, we formulate LQG problems with a symmetric partially nested information structure. Consider the following dynamics. Let $i = 1, 2$, and

$$x^i_{t+1} = Ax^i_t + Bu^i_t + w^i_t. \tag{3.8}$$

Problem ($\mathcal{P}_T$): Consider the expected cost function of $(\underline{\gamma}^1_T, \underline{\gamma}^2_T)$ as

$$J_T(\underline{\gamma}^1_T, \underline{\gamma}^2_T) = E^{(\underline{\gamma}^1_T, \underline{\gamma}^2_T)}\left[\frac{1}{T}\sum_{t=0}^{T-1}\sum_{i=1}^{2}(x^i_t)^T Q x^i_t + (u^i_t)^T R u^i_t + (u^1_t)^T \tilde{R} u^2_t + (u^2_t)^T \tilde{R} u^1_t\right], \tag{3.9}$$

where $\underline{\gamma}^i_T = (\gamma^i_{0:T-1})$, and $R, \tilde{R} > 0$ and $Q \geq 0$. Let

$$y^i_t = H_t \zeta^i_{0:t-1} + \sum_{j=0}^{t-1} D_{tj} u^i_j, \tag{3.10}$$

where $\zeta^i_t = (w^i_t, v^i_t)$ with $\zeta^i_0 = (x^i_0, w^i_0, v^i_0)$. Let $\mathcal{M}_{r,q}$ denote the space of $r \times q$ matrices. Let $n, m, s \in \mathbb{N}$ and $\mathbb{X} = \mathbb{R}^n$, $\mathbb{Y} = \mathbb{R}^s$, $\mathbb{U} = \mathbb{R}^m$, $w^i_t \in \mathbb{R}^n$, $v^i_t \in \mathbb{R}^n$, $A \in \mathcal{M}_{n,n}$, $B \in \mathcal{M}_{n,m}$, $R \in \mathcal{M}_{m,m}$, $Q \in \mathcal{M}_{n,n}$, $\tilde{R} \in \mathcal{M}_{m,m}$, $H_t \in \mathcal{M}_{s,n(2t+1)}$, and $D_{tj} \in \mathcal{M}_{s,m}$. Let the information structure of each DM be $I^i_t = \{y^i_t, y^i_{\downarrow t}\}$.

In the following, we show that the above dynamic teams are symmetrically optimal under sufficient conditions on the observations and initial states.

Corollary 3.4.1. *For a fixed T, consider a finite horizon team problem defined above as ($\mathcal{P}_T$). If x^1_0 and x^2_0 are exchangeable zero mean Gaussian random vectors and w^i_ts and v^i_ts are i.i.d. Gaussian random vectors for $i = 1, 2$ and independent for all $t = 0, \ldots, T-1$ and also independent of initial states, then the dynamic team is symmetrically optimal.*

Proof. Since the dynamic team is LQG with a partially nested information structure, a static reduction exists and the expected cost is convex in policies under static reductions (see [66] and Theorem 3.2.1(iii)). Assumption 3.2.1 is satisfied following from (3.9). We need to show assumptions of Theorem 3.2.2 hold. Following from the hypothesis on disturbances and initial states, Assumption (b) holds. Assumption (c) holds following from Assumption 3.2.2 and since given (x_0^1, x_0^2), $(y_{0:T-1}^1, y_{0:T-1}^2)$ are independent. Hence, Theorem 3.2.2 completes the proof. □

Here, we consider a class of LQG dynamic teams with a tree information structure where we utilize Corollary 3.4.1 and we obtain an explicit recursion for the optimal policy.

Problem ($\mathcal{P}_T^{\text{tree}}$): Consider a finite horizon expected cost (3.9) with $I_t^i = \{x_{[0:t]}^i, u_{[0:t-1]}^i\}$.

We note that this problem is a special case of ($\mathcal{P}_T$) since we assumed that $y_t^i = x_t^i$ for all DM^i and $t = 0, \dots, T-1$ and also all DMs have a total recall ($I_t^i = \{x_{[0:t]}^i, u_{[0:t-1]}^i\}$). For this problem, we calculate an explicit recursion for optimal policies using the symmetry established in Corollary 3.4.1.

Theorem 3.4.1. *For a fixed T, consider a finite horizon team problem defined as ($\mathcal{P}_T^{tree}$). If (x_0^1, x_0^2) are exchangeable with an identical zero mean Gaussian distribution and w_t^is are i.i.d. zero mean Gaussian random vectors for $i = 1, 2$ and independent for all $t = 0, \dots, T-1$ and independent of initial states, then*

$$u_t^{1*,(T)} = K_t^{(T)} x_t^1 + L_t^{(T)} E[x_0^2 | x_0^1], \tag{3.11}$$

$$u_t^{2*,(T)} = K_t^{(T)} x_t^2 + L_t^{(T)} E[x_0^1 | x_0^2], \tag{3.12}$$

where

$$K_t^{(T)} = -(R + B^T P_{t+1}^{(T)} B)^{-1} B^T P_{t+1}^{(T)} A, \tag{3.13}$$

$$P_t^{(T)} = -A^T P_{t+1}^{(T)} B^T (R + B^T P_{t+1}^{(T)} B)^{-1} B^T P_{t+1}^{(T)} A + Q + A^T P_{t+1}^{(T)} A, \tag{3.14}$$

$$L_t^{(T)} = -\left(R + B^T P_{t+1}^{(T)} B\right)^{-1} \left[\tilde{R} K_t^{(T)} G_t^{(T)} + \tilde{R} L_t^{(T)} \Sigma + \sum_{s=t+1}^{T-1} B^T (A^T)^{s-t} P_{s+1}^{(T)} B L_s^{(T)}\right], \quad (3.15)$$

$$G_t^{(T)} = \prod_{s=0}^{t-1} (A + B K_s^{(T)}) + \sum_{s=1}^{t} \prod_{j=s}^{t-1} (A + B K_j^{(T)}) B L_{s-1}^{(T)} \Sigma,$$

where $\Sigma = E[x_0^1 (x_0^2)^T](E[x_0^2 (x_0^2)^T])^{-1}$, $P_T^{(T)} = 0$, $G_0^{(T)} = I$. *Moreover, the optimal cost is*

$$\begin{aligned} J_T(\underline{\gamma}_T^*) = \frac{2}{T} \Bigg\{ & E\left[(x_0^1)^T P_0^{(T)} x_0^1\right] + \sum_{t=0}^{T-1} E\left[(w_t^1)^T P_t^{(T)} w_t^1\right] \\ & + \sum_{t=0}^{T-1} E\left[\left(E[x_0^2|x_0^1]\right)^T \left((L_t^{(T)})^T B^T P_{t+1}^{(T)} B L_t^{(T)}\right) E[x_0^2|x_0^1]\right] \\ & + \sum_{t=1}^{T-1} E\left[(x_0^1)^T (A^T)^t P_{t+1}^{(T)} B L_t^{(T)} E[x_0^2|x_0^1]\right] \Bigg\}. \end{aligned} \quad (3.16)$$

Proof. Following from [66] and Radner's theorem [102], person-by-person optimality implies global optimality due to the uniqueness of the person-by-person optimal policy. That is because the information structure is partially nested, and LQG dynamic teams can be reduced to a static one using Ho-Chu's static reduction [66]. Hence, we only need to show that the policy satisfying (3.11) and (3.12) is person-by-person optimal. We show that for DM^1, $J(\underline{\gamma}_T^*, \underline{\gamma}_T^*) \leq J((\underline{\gamma}_T^{-t*}, \beta), \underline{\gamma}_T^*)$ for all $\beta \in \Gamma^t$ where $(\underline{\gamma}_T^{-t*}, \beta) = (\gamma_{0:t-1}^*, \beta, \gamma_{t+1:T-1}^*)$. This implies that $(\underline{\gamma}_T^*, \underline{\gamma}_T^*)$ is person-by-person optimal thanks to Corollary 3.4.1 since the dynamic team is symmetrically optimal (by exchanging policies $(\underline{\gamma}_T^{-t*}, \beta)$ with $\underline{\gamma}_T^*$ which implies $J(\underline{\gamma}_T^*, \underline{\gamma}_T^*) \leq J(\underline{\gamma}_T^*, (\underline{\gamma}_T^{-t*}, \beta))$ for all $\beta \in \Gamma^t$ and this implies that $(\underline{\gamma}_T^*, \underline{\gamma}_T^*)$ is the fixed point of the equation). The proof is completed by induction. Due to space constraints, we have removed the calculation. □

Remark 3.4.1. *The optimal policies* (3.11) *and* (3.12) *contain two parts which can be interpreted as follows: the first part,* $k_t^{(T)} x_t^i$*, is equivalent to the optimal policy of the branch (DM) by ignoring the other branch in the optimization problem (in this case, this is equivalent to the centralized policies since the information structure of each branch (DM) is centralized). The second part corresponds to the correlation term between branches (DMs).*

In the following, we generalize the result of Theorem 3.4.1 to N-DM LQG dynamic teams. Assume that the dynamics for $i = 1, 2, ..., N$ are defined as (3.8).

Problem ($\mathcal{P}_T^{N,\text{tree}}$): Consider the expected cost function of $\underline{\gamma}_T^{1:N}$ as

$$J_T^N(\underline{\gamma}_T^{1:N}) = \frac{1}{T}\sum_{t=0}^{T-1}\sum_{i=1}^{N} E^{\underline{\gamma}_T^{1:N}}\Big[(x_t^i)^T Q x_t^i + (u_t^i)^T R u_t^i + \sum_{j=1, j\neq i}^{N} (u_t^i)^T \tilde{R} u_t^j + (u_t^j)^T \tilde{R} u_t^i\Big], \quad (3.17)$$

where $\underline{\gamma}_T^i = \gamma_{0:T-1}^i$ for $i = 1, \ldots, N$ and $R, \tilde{R} > 0$ and $Q \geq 0$. Let $I_t^i = \{x_{[0:t]}^i, u_{[0:t-1]}^i\}$.

Corollary 3.4.2. *For a fixed T and N, consider a finite horizon team problem defined as ($\mathcal{P}_T^{N,tree}$). If $(x_0^{1:N})$ are exchangeable with an identical zero mean Gaussian distribution, and w_t^is for $i = 1, \ldots, N$ are i.i.d. zero mean Gaussian random vectors, independent for $t = 0, \ldots, T-1$, and independent of initial states, then*

$$u_t^{i*,(T),(N)} = K_t^{(T)} x_t^i + L_t^{(N),(T)} \sum_{j=1, j\neq i}^{N} E[x_0^j | x_0^i],$$

where $K_t^{(T)}$ and $P_t^{(T)}$ satisfy (3.13) *and* (3.14)*, and $L_t^{(N),(T)}$ is a function of $K_{0:t}^{(T)}$.*

Proof. The proof is similar to the one of Theorem 3.4.1. □

Now we consider more general setup where using Corollary 3.4.1, we establish a structural result for the case where the information structure of each decision maker over time satisfies a structure which is identical for all DMs and is partially nested. An example of such a graph structure has been depicted in Fig. (3.2).

Problem ($\mathcal{P}_T^N$): Consider a finite horizon expected cost of $\underline{\gamma}_T^{1:N}$ as (3.17) with the information structure $I_t^i = \{y_t^i, y_{\downarrow t}^i\}$ where y_t^i is defined in (3.10) and dynamics is defined in (3.8).

Theorem 3.4.2. *For a fixed T and N, consider a finite horizon team problem defined as ($\mathcal{P}_T^N$). If $(x_0^1, \ldots, x_0^N)$ are exchangeable with an identical zero mean Gaussian distribution and w_t^is for*

$i = 1, \dots, N$ are i.i.d. zero mean Gaussian random vectors, independent for all $t = 0, \dots, T-1$ and independent of initial states, then

$$u_t^{i*,(T),(N)} = K_t^{(T)} y_t^i + L_t^{(N),(T)} \sum_{j=1, j\neq i}^{N} E[x_0^j | y_t^i],$$

where $K_t^{(T)}$ are obtained by considering only one DM and ignoring other DMs.

Proof. The proof is similar to that of Theorem 3.4.1 by [66] and Corollary 3.4.1. □

Remark 3.4.2. *A related work is [97], where structural results for optimal policy have been obtained for finite horizon LQG problems on graphs. In our analysis above, the structural result for the optimal policy is obtained without assuming that decision makers who have no common ancestors and no common descendants have either uncorrelated noise or are decoupled through the cost function. Instead, exchangeable partially nested LQG teams with correlated initial states and disturbances are considered. Moreover, here, the graph structures may not be trees in general, as opposed to [97] where a multi-tree structure has been imposed on a graph.*

Now, we present results for LQG teams with a mean-field coupling through the cost function. First, using Corollary 3.4.2, we obtain globally optimal policies for N-DM teams with a mean-field coupling and correlated initial states and disturbances. Next, as an implication of Theorem 3.3.1, we show the convergence of optimal policies for LQG N-DM mean-field teams on a tree to the corresponding optimal policy of mean-field teams. Let $I_t^i = \{x^i_{[0:t]}, u^i_{[0:t-1]}\}$ for $i \in \mathbb{N}$, and dynamics be as (3.8).

Problem ($\mathcal{P}^{N,\mathbf{MF}}_{T,\mathbf{LQG}}$): Consider the expected cost function of $\underline{\gamma}_T^N$ as

$$J_T^N(\underline{\gamma}_T^N) = \frac{1}{T} \sum_{t=0}^{T-1} \sum_{i=1}^{N} E^{\underline{\gamma}_T^{1:N}} \Big[(x_t^i)^T Q x_t^i + (u_t^i)^T R u_t^i \tag{3.18}$$
$$+ \frac{1}{N-1} \sum_{j=1, j\neq i}^{N} (u_t^i)^T \tilde{R} u_t^j + (u_t^j)^T \tilde{R} u_t^i + \frac{1}{N-1} \sum_{j=1, j\neq i}^{N} (x_t^i)^T \tilde{Q} x_t^j + (x_t^j)^T \tilde{Q} x_t^i \Big],$$

where $R, \tilde{R} > 0$ and $Q, \tilde{Q} \geq 0$.

Problem ($\mathcal{P}^{\infty,\mathbf{MF}}_{T,\mathbf{LQG}}$): Consider the expected cost function of $\underline{\gamma}_T$ as

$$J^{\infty}_T(\underline{\gamma}_T) = \limsup_{N\to\infty} J^N_T(\underline{\gamma}_T). \tag{3.19}$$

Corollary 3.4.3. *For a fixed T and N, consider a finite horizon team problem defined as ($\mathcal{P}^{N,MF}_{T,LQG}$). If $(x^{1:N}_0)$ are exchangeable zero mean Gaussian random vectors, and w^i_ts are i.i.d. zero mean Gaussian random vectors for $i = 1, \dots, N$, independent for $t = 0, \dots, T-1$, and independent of initial states, then*

$$u^{i*,(T),(N)}_t = K^{(T)}_t x^i_t + \frac{L^{(N),(T)}_t}{N-1} \sum_{j=1, j\neq i}^{N} E[x^j_0 | x^i_0], \tag{3.20}$$

where $K^{(T)}_t$ and $P^{(T)}_t$ satisfy (3.13) *and* (3.14)*, and $L^{(N),(T)}_t$ is a function of $K^{(T)}_{0:t}$.*

Proof. The proof is similar to the one in Theorem 3.4.1. □

Corollary 3.4.4. *For a fixed T, consider a finite horizon team problem defined as ($\mathcal{P}^{\infty,MF}_{T,LQG}$). Assume $\{x^i_0\}_{i\in\mathbb{N}}$ are exchangeable random vectors with zero mean Gaussian distribution, and w^i_ts are i.i.d. zero mean Gaussian random vectors for $i \in \mathbb{N}$, independent for $t = 0, \dots, T-1$, and independent of initial states. If $L^{(N),(T)}_t$ in* (3.20) *converges pointwise as $N \to \infty$ to $L^{(\infty),(T)}_t$, then*

$$u^{i*,(T),(\infty)}_t = K^{(T)}_t x^i_t + L^{(\infty),(T)}_t \Sigma x^i_0,$$

where $K^{(T)}_t$ and $P^{(T)}_t$ satisfy (3.13) *and* (3.14)*, and $\Sigma = E[x^1_0 (x^2_0)^T](E[x^2_0 (x^2_0)^T])^{-1}$.*

Proof. Following from [1, page 9], since $\{x^i_0\}_{i\in\mathbb{N}}$ are exchangeable Gaussian random vectors, we can describe them explicitly as $x^i_0 = \omega_0 + \theta^i$ where $(\theta^1, \theta^2, \dots)$ are i.i.d. mean zero Gaussian and independent of mean zero Gaussian random vector ω_0. Now, we invoke Theorem 3.3.1 (or Theorem 3.3.2) and Corollary 3.4.3 and we use Remark 3.3.4 to complete the proof. □

3.4.2 Average Cost Infinite Horizons Problems for Partially Nested Dynamic Teams

In the following, we consider average cost problems with a symmetric partially nested information structure as well as LQG dynamic teams with the one step delay sharing pattern and sparsity constraints. We note that the optimality of linear policies for infinite horizon LQG problems is an open problem in its generality. In this subsection, we provide a positive result for a class of such problems.

Average Cost Infinite Horizons Problems for Partially Nested Dynamic Teams

Now, we consider an infinite horizon team problem and we use the result in Section 3.4.1.

Problem ($\mathcal{P}_\infty^{\text{tree}}$): Consider the expected cost function of $(\underline{\gamma}^1, \underline{\gamma}^2)$ as

$$J(\underline{\gamma}^{1:2}) = \limsup_{T\to\infty} E^{(\underline{\gamma}^1,\underline{\gamma}^2)}[c(x^{1:2}_{0:T-1}, u^{1:2}_{0:T-1})], \tag{3.21}$$

where the cost function is defined as (3.9) and $I_t^i = \{x^i_{[0:t]}, u^i_{[0:t-1]}\}$.

First, we introduce a lemma essential for Theorem 3.4.3.

Lemma 3.4.1. *Consider the sequence $\{a_T^i\}_{i=1}^T$. Assume $\lim_{T\to\infty} a_T^i = a$ for $i = 0, \dots, T-1$. If for every fixed $T \in \mathbb{N}$, $a_T^l = a_{T+1}^{l+1}$ for all $l = 0 \dots, T-1$, then $\lim_{T\to\infty} \frac{1}{T}\sum_{i=1}^T a_T^i = a$.*

Proof. We have

$$\lim_{T\to\infty} \frac{1}{T}\sum_{i=1}^{T} a_T^i = \lim_{T\to\infty} \frac{1}{T}\sum_{i=0}^{T-1} a_{T-i}^1 = \lim_{T\to\infty} \frac{1}{T}\sum_{k=1}^{T} a_k^1 = a,$$

where the second equality follows from $a_T^l = a_{T+1}^{l+1}$ and the last equality follows from the Cesáro mean argument. □

Theorem 3.4.3. *Consider average cost infinite horizon team problems defined as ($\mathcal{P}_\infty^{tree}$). Assume (A, B) are stabilizable and $(A, Q^{\frac{1}{2}})$ are detectable. Assume x_0^1 and x_0^2 are exchangeable with an*

identical zero mean Gaussian distribution and w_t^is are i.i.d. zero mean Gaussian random variables for $i = 1, 2$ and for all $t = 0, \ldots, T-1$ and independent of initial states. If $L_t^{(T)}$ in (3.15) *converges pointwise to $L_t^{(\infty)}$ as $T \to \infty$, then the pointwise limit of the sequence of optimal policies for ($\mathcal{P}_T^{tree}$) is team optimal for ($\mathcal{P}_\infty^{tree}$) and stabilizes the closed-loop system,*

$$u_t^{1*,(\infty)} = Kx_t^1 + L_t^{(\infty)} E[x_0^2|x_0^1],$$
$$u_t^{2*,(\infty)} = Kx_t^2 + L_t^{(\infty)} E[x_0^1|x_0^2],$$

where $K, P, L_t^{(\infty)}$ and $G_t^{(\infty)}$ are the pointwise limit of the ones for ($\mathcal{P}_T^{tree}$) as $T \to \infty$.

Average Cost Team Problems with One-Step-Delayed Observation Sharing Pattern and Sparsity Constraints

In [84], the problem of finding a steady state solution for state feedback average cost LQG team problems with a delay and sparsity has been considered. Under the independence of disturbances assumption, convergence of the solution to the stabilizing solutions for finite horizon of the corresponding algebraic Riccati equation has been shown [84]; however, global optimality of the limit solution has not been established. We establish global optimality of the limit solution. Consider the following dynamics for $i = 1, \ldots, N$,

$$x_{t+1}^i = \sum_j^N A^{ij} x_t^j + B^{ij} u_t^j + \omega_t^i,$$

where for a fixed $i = 1, \ldots, N$, A_t^{ij} and B_t^{ij} are zero if sum of the delays along the directed path from j to i with the shortest path, D_{ij}, is more than one, i.e., $D_{ij} > 1$. Assume x_0^is and ω_t^is are mutually independent Gaussian random vectors for all $i = 1, \ldots, N$ and ω_t^is are independent for $t = 0, \ldots, T-1$ with covariance matrix Σ_ω^i. The information structure is defined as $I_t^i = \{x_k^j | j = 1, \ldots, N, \text{ and } k = 0, \ldots, t - D_{ij}\}$. This along with the assumption on the dynamics of the system implies that the information structure is partially nested. Since the disturbances and initial states are

mutually independent, the information structure can be decomposed into the independent sets [84]. To this end, an information graph, $\hat{G}(\mathcal{M}, \mathcal{F})$, has been defined for a graph [84]. In the following, we follow the notation in [84]. Define s_k^j as a set of all nodes reachable from node j within k steps. Define $\mathcal{M} := \{s_k^j | j = 1, \ldots, N \text{ and } k \geq 0\}$ and $\mathcal{F} := \{(s_k^j, s_{k+1}^j) | j = 1, \ldots, N \text{ and } k \geq 0\}$. The following theorem is from [84], and gives an optimal policy for finite horizon problems.

Assume that the delay of sharing between each decision maker is not greater than one. Assume there are no directed cycles with a total delay of zero.

Theorem 3.4.4. *[84] Consider the LQG team problem with*

$$J_T(\underline{\gamma}_T) := \frac{1}{T}\mathbb{E}^{\underline{\gamma}_T}\left[\sum_{t=0}^{T-1}\begin{bmatrix} x_t^T & u_t^T \end{bmatrix}\begin{bmatrix} Q & S \\ S^T & R \end{bmatrix}\begin{bmatrix} x_t \\ u_t \end{bmatrix}\right],$$

where $R > 0$, $Q \geq 0$, *and* $\begin{bmatrix} Q & S \\ S^T & R \end{bmatrix} \geq 0$, $u_t = (u_t^1, \ldots, u_t^N)$. *Then, the optimal controller is* $u_{t*}^T = \sum_{r \ni i} I^{\{i\},r} K_t^r \zeta_t^r$, *where* $r \in \mathcal{M}$, $I^{\{i\},r}$ *is the identity matrix partition according to* $\{i\}$ *and* r, *and*

$$\begin{aligned} K_t^r &= -(R^{rr} + (B^{sr})^T X_{t+1}^s B^{sr})^{-1}(S^{rr} + (A^{sr})^T X_{t+1}^s B^{sr}), \\ X_t^r &= Q^{rr} + (A^{sr})^T X_{t+1}^s A^{sr} - (K_t^r)^T (R^{rr} + (B^{sr})^T X_{t+1}^s B^{sr}) K_t^r, \end{aligned} \tag{3.22}$$

with $X_T^r = Q^{rr}$, *and* $\zeta_{t+1}^s = \sum_{r \to s}(A^{sr} + B^{sr} K_t^r)\zeta_t^r + \sum_{\omega^i \to s} I^{s,\{i\}} \omega_t^i$, *with* $\zeta_0^s = \sum_{\omega^i \to s} I^{s,\{i\}} x_0^i$, *where* $s \in \mathcal{M}$ *is the unique node such that there is an edge between* r *and* s, *i.e.,* $r \to s$. *Moreover, the optimal cost is obtained as*

$$J_T(\underline{\gamma}_T^*) = \frac{1}{T}[\sum_{i=1, \omega^i \to s}^{N} Tr((X_0^s)^{\{i\},\{i\}} \mathbb{E}(x_0^i (x_0^i)^T)) + \sum_{t=0}^{T-1} \sum_{i=1, \omega^i \to s}^{N} Tr((X_{t+1}^s)^{\{i\},\{i\}} \Sigma_\omega^i)], \tag{3.23}$$

where $(X_{t+1}^s)^{\{i\},\{i\}}$ *denotes the sub-matrix* (X_{t+1}^s) *corresponds to the* ii*-th array.*

In the following, we refine a related result in [84]. We show the global optimality of the limit policy similar to the method used for the convergence results in the preceding chapters (by showing that the sequence of optimal expected costs converges to the expected cost of the limit policy).

Theorem 3.4.5. *Consider the LQG team problem with*

$$J(\underline{\gamma}) := \limsup_{T\to\infty} \frac{1}{T}\mathbb{E}^{\underline{\gamma}}\Big[\sum_{t=0}^{T-1} \begin{bmatrix} x_t^T & u_t^T \end{bmatrix} \begin{bmatrix} Q & S \\ S^T & R \end{bmatrix} \begin{bmatrix} x_t \\ u_t \end{bmatrix}\Big].$$

Assume for self loops $s \to s$ in the information graph, (A^{ss}, B^{ss}) is stabilizable and $\begin{bmatrix} A^{ss} - e^{i\theta} & B^{ss} \\ C^{ss} & R^{ss} \end{bmatrix}$ *has a full column rank for every $\theta \in [0, 2\pi]$, where C^{ss} and D^{ss} are the matrix of the form* $\begin{bmatrix} Q^{ss} & S^{ss} \\ (S^{ss})^T & R^{ss} \end{bmatrix} = \begin{bmatrix} C^{ss} & D^{ss} \end{bmatrix}^T \begin{bmatrix} C^{ss} & D^{ss} \end{bmatrix}$*. Then, $u_{t*}^{\infty} = \sum_{r\ni i} I^{\{i\},r} K^{r,(\infty)} \zeta_t^{r,(\infty)}$, where $K^{r,(\infty)} = \lim_{T\to\infty} K_t^{r,(T)}$, $X^{r,(\infty)} = \lim_{T\to\infty} X_t^{r,(T)}$ for all t and $\zeta_t^{r,(\infty)}$ is obtained by replacing $K_t^{r,(T)}$ with $K^{r,(\infty)}$.*

3.5 Proofs

3.5.1 Proof of Theorem 3.2.2

We first show that for any permutation $\sigma \in S$, $J_T((\underline{\gamma}_T^{\sigma})^1, \ldots, (\underline{\gamma}_T^{\sigma})^N) = J_T(\underline{\gamma}_T^1, \ldots, \underline{\gamma}_T^N)$, i.e., the team is exchangeable. We have,

$$\begin{aligned}
&J_T\left((\underline{\gamma}_T^{\sigma})^1, \ldots, (\underline{\gamma}_T^{\sigma})^N\right) \\
&= \int c\left(\underline{\zeta}^{1:N}, (\underline{\gamma}_T^{\sigma})^1(\underline{y}^1), \ldots, (\underline{\gamma}_T^{\sigma})^N(\underline{y}^N)\right) \mu(dx_0^{1:N}, d\zeta_{0:T-1}^{1:N}) \\
&\quad \times \prod_{t=0}^{T-1}\prod_{i=1}^{N} P\left(dy_t^i \Big| x_0^i, \zeta_{0:t-1}^i, y_{\downarrow t}^{\downarrow\downarrow i}, (\gamma^{\sigma})_{\downarrow t}^{\downarrow\downarrow i}(y_{\downarrow t}^{\downarrow\downarrow i})\right) \qquad (3.24) \\
&= \int c\left((\underline{\zeta}^{\sigma})^{1:N}, (\underline{\gamma}_T^{\sigma})^1((\underline{y}^{\sigma})^1), \ldots, (\underline{\gamma}_T^{\sigma})^N((\underline{y}^{\sigma})^N)\right) \mu(d(x_0^{\sigma})^{1:N}, d(\zeta_{0:T-1}^{\sigma})^{1:N}) \\
&\quad \times \prod_{t=0}^{T-1}\prod_{i=1}^{N} P\left(d(y^{\sigma})_t^i \Big| (x_0^{\sigma})^i, (\zeta_{0:t-1}^{\sigma})^i, (y^{\sigma})_{\downarrow t}^{\downarrow\downarrow i}, (\gamma^{\sigma})_{\downarrow t}^{\downarrow\downarrow i}((y^{\sigma})_{\downarrow t}^{\downarrow\downarrow i})\right) \qquad (3.25)
\end{aligned}$$

$$= \int c\left(\underline{\zeta}^{1:N}, \underline{\gamma}_T^1(\underline{y}^1), \ldots, \underline{\gamma}_T^N(\underline{y}^N)\right) \mu(dx_0^{1:N}, d\zeta_{0:T-1}^{1:N}) \prod_{t=0}^{T-1} \prod_{i=1}^{N} P\left(dy_t^i \middle| x_0^i, \zeta_{0:t-1}^i, y_{\downarrow t}^{\downarrow i}, \gamma_{\downarrow t}^{\downarrow i}(y_{\downarrow t}^{\downarrow i})\right) \tag{3.26}$$

$$= J_T(\underline{\gamma}_T^1, \ldots, \underline{\gamma}_T^N),$$

where (3.24) follows from condition (c). Equality (3.25) follows from exchanging $\underline{y}^i$, $\underline{\zeta}^i$ with $(\underline{y}^\sigma)^i$, $(\underline{\zeta}^\sigma)^i$ by relabeling them, respectively. Since the information structure is symmetric, (3.5) and condition (b) imply (3.26). Hence, the team is exchangeable. Let $\underline{\gamma}_T^* = (\underline{\gamma}_T^{1*}, \ldots, \underline{\gamma}_T^{N*})$ be a given policy. Consider $\tilde{\underline{\gamma}}_T$ as a convex combination of all possible permutations of policies by averaging them. Since action spaces are convex by condition (a), $\tilde{\underline{\gamma}}_T$ is a control policy. Following from convexity of the cost function in policies, we have

$$J_T(\tilde{\underline{\gamma}}_T) := J_T(\sum_{\sigma \in S} \frac{1}{|S|} \underline{\gamma}_T^{*,\sigma}) \leq \sum_{\sigma \in S} \frac{1}{|S|} J_T(\underline{\gamma}_T^{*,\sigma}) = J_T(\underline{\gamma}_T^*),$$

where $|S|$ denotes the cardinality of the set S and the inequality above follows from the hypothesis that the team problem is convex on $\mathbf{\Gamma}$ and the last equality follows from exchangeability of the team problem. This implies that the team is symmetrically optimal and completes the proof.

3.5.2 Proof of Theorem 3.3.1

Following from Lemma 3.3.1, one can consider a sequence of N-DM teams which are symmetrically optimal that defines $(\mathcal{P}_T^{N,\mathrm{MF}})$ and whose limit is identified with $(\mathcal{P}_T^{\infty,\mathrm{MF}})$. Define

$$\begin{aligned} Q_N(B) &:= \frac{1}{N} \sum_{i=1}^{N} \delta_{\beta_N^i}(B) \quad \text{where} \quad \beta_N^i := (\underline{\gamma}_T^{*,N}(\underline{y}^i), \underline{y}^i, \underline{\zeta}^i), \\ \tilde{Q}_N(B) &:= \frac{1}{N} \sum_{i=1}^{N} \delta_{\beta_\infty^i}(B) \quad \text{where} \quad \beta_\infty^i := (\underline{\gamma}_T^{*,\infty}(\underline{y}^i), \underline{y}^i, \underline{\zeta}^i), \end{aligned} \tag{3.27}$$

where $\delta_Y(\cdot)$ denotes the Dirac measure for any random vector Y, and $B \in \mathcal{Z} := \mathbf{U} \times \mathbf{Y} \times \mathbf{S}$, $\mathbf{U} := (\prod_{t=0}^{T-1} \mathbb{U})$, $\mathbf{Y} := (\prod_{t=0}^{T-1} \mathbb{Y})$, $\mathbf{S} := (\prod_{t=0}^{T-1} \mathbb{S}) = \mathbb{X} \times (\prod_{t=0}^{T-1} \mathbb{W} \times \mathbb{V})$, $\mathbf{X} = (\prod_{t=0}^{T-1} \mathbb{X})$,

$\underline{y}^i = (y_0^i, \dots, y_{T-1}^i)$, and $\underline{\zeta}^i := (\zeta_0^i, \dots, \zeta_{T-1}^i)$.

In the following, first, we show that conditioned on ω_0, Q_N converges P-almost surely to $Q = \mathcal{L}(\beta_\infty^1|\omega_0)$ in w-s topology (coarsest topology on $\mathcal{P}(\mathbf{U}\times\mathbf{Y}\times\mathbf{S})$ under which $\int f(u,y,\zeta)Q_N(du,dy,d\zeta) : \mathcal{P}(\mathbf{U}\times\mathbf{Y}\times\mathbf{S}) \to \mathbb{R}$ is continuous for every measurable and bounded function f which is continuous in u and y but need not to be continuous in ζ (see e.g., [121] and [138, Theorem 5.6])). Then, we show that

$$\limsup_{N\to\infty} J_T^N(\tilde{\underline{\gamma}}_T^{*,N}) = J_T^\infty(\tilde{\underline{\gamma}}_T^{*,\infty}),$$

where $\tilde{\underline{\gamma}}_T^{*,N} := (\underline{\gamma}_T^{*,N}, \underline{\gamma}_T^{*,N}, \dots, \underline{\gamma}_T^{*,N})$ and $\tilde{\underline{\gamma}}_T^{*,\infty} := (\underline{\gamma}_T^{*,\infty}, \underline{\gamma}_T^{*,\infty}, \dots)$.

(Step 1): In this step, we show that conditioned on ω_0, Q_N converges P-almost surely to Q in w-s topology. First, we show that for every continuous and bounded function g in actions and observations, for every ω_0 on a set of P-measure one,

$$P\left(\left\{\omega\in\Omega\middle|\lim_{N\to\infty}\left(\frac{1}{N}\sum_{i=1}^N\left[g(\underline{\gamma}_T^{*,N}(\underline{y}^i),\underline{y}^i,\underline{\zeta}^i) - g(\underline{\gamma}_T^{*,\infty}(\underline{y}^i),\underline{y}^i,\underline{\zeta}^i)\right]\right)=0\right\}\middle|\omega_0\right)=1. \quad (3.28)$$

Following from symmetry of the information structure and Lemma 3.3.1, every DM applies an identical optimal policy $\underline{\gamma}_T^{*,N}$ and since functions f_t and h_t are identical for each DM, conditioned on ω_0, $(\underline{\gamma}_T^{*,N}(\underline{y}^i),\underline{y}^i,\underline{\zeta}^i)$ and $(\underline{\gamma}_T^{*,\infty}(\underline{y}^i),\underline{y}^i,\underline{\zeta}^i)$ are i.i.d. random vectors. For every $\epsilon>0$ and for every function g continuous and bounded in actions and observations, we have P-almost surely

$$\begin{aligned}
&\lim_{N\to\infty} P\left(\left|\int g dQ_N - \int g d\tilde{Q}_N\right| \geq \epsilon\middle|\omega_0\right)\\
&\leq \epsilon^{-1}\lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^N E\left[\left|g(\underline{\gamma}_T^{*,N}(\underline{y}^i),\underline{y}^i,\underline{\zeta}^i) - g(\underline{\gamma}_T^{*,\infty}(\underline{y}^i),\underline{y}^i,\underline{\zeta}^i)\right|\middle|\omega_0\right] \quad (3.29)\\
&= \epsilon^{-1}\lim_{N\to\infty} E\left[\left|g(\underline{\gamma}_T^{*,N}(\underline{y}^i),\underline{y}^i,\underline{\zeta}^i) - g(\underline{\gamma}_T^{*,\infty}(\underline{y}^i),\underline{y}^i,\underline{\zeta}^i)\right|\middle|\omega_0\right] \quad (3.30)\\
&= \epsilon^{-1} E\left[\lim_{N\to\infty}\left|g(\underline{\gamma}_T^{*,N}(\underline{y}^i),\underline{y}^i,\underline{\zeta}^i) - g(\underline{\gamma}_T^{*,\infty}(\underline{y}^i),\underline{y}^i,\underline{\zeta}^i)\right|\middle|\omega_0\right] = 0, \quad (3.31)
\end{aligned}$$

where (3.29) follows from Markov's inequality, the triangle inequality and the definition of the

empirical measure, and (3.30) follows from the fact that $(\underline{\gamma}_T^{*,N}(\underline{y}^i), \underline{y}^i, \underline{\zeta}^i)$ and $(\underline{\gamma}_T^{*,\infty}(\underline{y}^i), \underline{y}^i, \underline{\zeta}^i)$ are i.i.d. random vectors. Since g is bounded and continuous, the dominated convergence theorem implies (3.31). Hence, for every subsequence, there exists a subsubsequence such that P-almost surely $P\left(\{\omega \in \Omega | \lim_{N\to\infty} \left(\int g dQ_N - \int g d\tilde{Q}_N\right) = 0\} | \omega_0\right) = 1$.

Now, we show that conditioned on ω_0, $\{\tilde{Q}_N\}_N$ converges weakly to Q P-almost surely. Since conditioned on ω_0, $(\underline{\gamma}_T^{*,\infty}(\underline{y}^i), \underline{y}^i, \underline{\zeta}^i)$ are i.i.d. random vectors, the strong law of large numbers implies that P-almost surely

$$P\left(\left\{\omega \in \Omega \middle| \lim_{N\to\infty} \left(\frac{1}{N}\sum_{i=1}^{N} g\left(\underline{\gamma}_T^{*,\infty}(\underline{y}^i), \underline{y}^i, \underline{\zeta}^i\right) - E\left[g(\underline{\gamma}_T^{*,\infty}(\underline{y}^1), \underline{y}^1, \underline{\zeta}^1)\middle|\omega_0\right]\right) = 0\right\}\middle|\omega_0\right) = 1, \quad (3.32)$$

hence, $P\left(\{\omega \in \Omega | \lim_{N\to\infty} \left(\int g d\tilde{Q}_N - \int g dQ\right) = 0\} | \omega_0\right) = 1$ P-almost surely.

Hence, through choosing a suitable subsequence, for every $\omega_0 \in \Omega_0$ on a set of P-measure one, for every function g continous and bounded in actions and observations and measurable and bounded in uncertainty and initial states

$$\lim_{N\to\infty} \left|\int g dQ_N - \int g dQ\right| \leq \lim_{N\to\infty} \left(\left|\int g dQ_N - \int g d\tilde{Q}_N\right| + \left|\int g d\tilde{Q}_N - \int g dQ\right| \right) = 0,$$

hence, conditioned on ω_0, Q_N converges weakly to Q P-almost surely. We note that the convergence is in the weak convergence topology, but since $\underline{\zeta}^i$s are exogenous with a fixed probability measure, the convergence is also in the w-s topology.

(Step 2): Following from (3.1) and (3.2), we have

$$x_t^i = f_{t-1}(f_{t-2}(\dots f_0(x_0^i, u_0^i, w_0^i)), u_{t-1}^i, w_{t-1}^i) = \tilde{f}_{t-1}(\underline{\zeta}^i, u_{0:t-1}^i), \quad (3.33)$$

$$y_t^i = h_t(x_0^i, \tilde{f}_0(\underline{\zeta}^i, u_0^i), \dots, \tilde{f}_{t-1}(\underline{\zeta}^i, u_{0:t-1}^i), u_{0:t-1}^i, v_{0:t}^i) = \tilde{h}_t(\underline{\zeta}^i, u_{0:t-1}^i), \quad (3.34)$$

where following from Assumption 3.3.1, $\tilde{f}_{t-1}$ and $\tilde{h}_t$ are continuous in actions. Hence, under

Assumption 3.3.1(c), we have

$$\frac{1}{N}\sum_{i=1}^{N}\sum_{t=0}^{T-1} E^{\underline{\gamma}_T^{*,1:N}}\left[c\left(\omega_0, x_t^i, u_t^i, \frac{1}{N}\sum_{p=1}^{N} u_t^p, \frac{1}{N}\sum_{p=1}^{N} x_t^p\right)\right]$$
$$= \frac{1}{N}\sum_{i=1}^{N} E\left[\tilde{c}\left(\omega_0, \underline{\zeta}^i, \underline{\gamma}_T^{*,N}(\underline{y}^i), \frac{1}{N}\sum_{p=1}^{N}\underline{\gamma}_T^{*,N}(\underline{y}^p), \frac{1}{N}\sum_{p=1}^{N}\Lambda(\underline{\gamma}_T^{*,N}(\underline{y}^p), \underline{\zeta}^p)\right)\right], \tag{3.35}$$

where (3.35) is true following from (3.33) for some functions $\tilde{c} : \Omega_0 \times \mathbf{S} \times \mathbf{U} \times \mathbf{U} \times \mathbf{X} \to \mathbb{R}_+$ which are continuous in its last three arguments and a function $\Lambda : \mathbf{U}\times\mathbf{S} \to \mathbf{X}$ which is continuous in actions. Hence, by induction and rewriting observations as a functions of policies of the past DMs ($\gamma_{\downarrow t}^{*,N}(y_{\downarrow t}^i)$) since $\underline{\gamma}_T^{*,N}$ converges to $\underline{\gamma}_T^{*,\infty}$, the induced cost by $\underline{\gamma}_T^{*,N}$ also converges to the cost induced by $\underline{\gamma}_T^{*,\infty}$ P-almost surely.

(Step 3): We have

$$\limsup_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}\sum_{t=0}^{T-1} E^{\underline{\gamma}_T^{*,1:N}}\left[c\left(\omega_0, x_t^i, u_t^i, \frac{1}{N}\sum_{p=1}^{N} u_t^p, \frac{1}{N}\sum_{p=1}^{N} x_t^p\right)\right]$$
$$= \limsup_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N} E\left[\tilde{c}\left(\omega_0, \underline{\zeta}^i, \underline{\gamma}_T^{*,N}(\underline{y}^i), \frac{1}{N}\sum_{p=1}^{N}\underline{\gamma}_T^{*,N}(\underline{y}^p), \frac{1}{N}\sum_{p=1}^{N}\Lambda(\underline{\gamma}_T^{*,N}(\underline{y}^p), \underline{\zeta}^p)\right)\right] \tag{3.36}$$
$$\geq \liminf_{N\to\infty} E\Bigg[E\Bigg[\int_{\mathcal{Z}}\tilde{c}\left(\omega_0, \zeta, u, \int_{\mathbf{U}} uQ_N(du\times\mathbf{Y}\times\mathbf{S}), \int_{\mathbf{U}\times\mathbf{S}}\Lambda Q_N(du\times\mathbf{Y}\times d\zeta)\right)$$
$$\times Q_N(du, dy, d\zeta)\Big|\omega_0\Bigg]\Bigg] \tag{3.37}$$
$$\geq E\Bigg[E\Bigg[\liminf_{N\to\infty}\int_{\mathcal{Z}}\tilde{c}\left(\omega_0, \zeta, u, \int_{\mathbf{U}} uQ_N(du\times\mathbf{Y}\times\mathbf{S}), \int_{\mathbf{U}\times\mathbf{S}}\Lambda Q_N(du\times\mathbf{Y}\times d\zeta)\right)$$
$$\times Q_N(du, dy, d\zeta)\Big|\omega_0\Bigg]\Bigg] \tag{3.38}$$
$$\geq E\Bigg[E\Bigg[\int_{\mathcal{Z}}\tilde{c}\left(\omega_0, \zeta, u, \int_{\mathbf{U}} uQ(du\times\mathbf{Y}\times\mathbf{S}), \int_{\mathbf{U}\times\mathbf{S}}\Lambda Q(du\times\mathbf{Y}\times d\zeta)\right)Q(du, dy, d\zeta)\Big|\omega_0\Bigg]\Bigg], \tag{3.39}$$

where (3.36) follows from (3.35). Inequality (3.37) follows from (3.27) and replacing limsup with liminf, and (3.38) follows from Fatou's lemma. In the following, we justify (3.39). Since

conditioned on ω_0, Q_N converges weakly to Q P-almost surely, we have $Q_N(du\times\mathbf{Y}\times\mathbf{S})$ converges weakly to $Q(du\times\mathbf{Y}\times\mathbf{S})$ P-almost surely conditioned on ω_0, hence, the compactness of $\mathbf{U}$ implies that conditioned on ω_0, P-almost surely

$$\frac{1}{N}\sum_{i=1}^{N}\underline{\gamma}_T^{*,N}(\underline{y}^i)=\int_{\mathbf{U}}uQ_N(du\times\mathbf{Y}\times\mathbf{S})\xrightarrow{N\to\infty}\int_{\mathbf{U}}uQ(du\times\mathbf{Y}\times\mathbf{S})=E[\underline{\gamma}_T^{*,\infty}(\underline{y}^1)\Big|\omega_0]. \tag{3.40}$$

Since conditioned on ω_0, Q_N converges weakly to Q P-almost surely, we have $Q_N(du\times\mathbf{Y}\times d\zeta)$ converges P-almost surely to $Q(du\times\mathbf{Y}\times d\zeta)$ in w-s topology conditioned on ω_0. Following from (3.33), since f_ts are bounded and continuous in actions, Λ is bounded and continuous in actions, hence, this implies that conditioned on ω_0, P-almost surely

$$\int_{\mathbf{U}\times\mathbf{S}}\Lambda Q_N(du\times\mathbf{Y}\times d\zeta)\xrightarrow{N\to\infty}\int_{\mathbf{U}\times\mathbf{S}}\Lambda Q(du\times\mathbf{Y}\times d\zeta). \tag{3.41}$$

Since the cost function $\tilde{c}$ is continuous in its last three arguments, P-almost surely

$$\begin{aligned}&\lim_{N\to\infty}\tilde{c}\left(\omega_0,\zeta,u,\int_{\mathbf{U}}uQ_N(du\times\mathbf{Y}\times\mathbf{S}),\int_{\mathbf{U}\times\mathbf{S}}\Lambda Q_N(du\times\mathbf{Y}\times d\zeta)\right)\\&=\tilde{c}\left(\omega_0,\zeta,u,\int_{\mathbf{U}}uQ(du\times\mathbf{Y}\times\mathbf{S}),\int_{\mathbf{U}\times\mathbf{S}}\Lambda Q(du\times\mathbf{Y}\times d\zeta)\right).\end{aligned}$$

Define a non-negative bounded functions

$$\begin{aligned}G_N^M&:=\min\left\{M,\tilde{c}\left(\omega_0,\zeta,u,\int_{\mathbf{U}}uQ_N(du\times\mathbf{Y}\times\mathbf{S}),\int_{\mathbf{U}\times\mathbf{S}}\Lambda Q_N(du\times\mathbf{Y}\times d\zeta)\right)\right\},\\G^M&:=\min\left\{M,\tilde{c}\left(\omega_0,\zeta,u,\int_{\mathbf{U}}uQ(du\times\mathbf{Y}\times\mathbf{S}),\int_{\mathbf{U}\times\mathbf{S}}\Lambda Q(du\times\mathbf{Y}\times d\zeta)\right)\right\},\end{aligned}$$

where the sequence $\{G^M\}_M$ converges as $M\to\infty$ P-almost surely to

$$G:=\tilde{c}\left(\omega_0,\zeta,u,\int_{\mathbf{U}}uQ(du\times\mathbf{Y}\times\mathbf{S}),\int_{\mathbf{U}\times\mathbf{S}}\Lambda Q(du\times\mathbf{Y}\times d\zeta)\right).$$

We have P-almost surely

$$\begin{aligned}
\liminf_{N\to\infty} \int_{\mathcal{Z}} \tilde{c}\left(\omega_0, \zeta, u, \int_{\mathbf{U}} uQ_N(du \times \mathbf{Y} \times \mathbf{S}), \int_{\mathbf{U}\times\mathbf{S}} \Lambda Q_N(du \times \mathbf{Y} \times d\zeta)\right) Q_N(du, dy, d\zeta) \\
&\geq \lim_{M\to\infty} \liminf_{N\to\infty} \int_{\mathcal{Z}} G_N^M Q_N(du, dy.d\zeta) \qquad (3.42)\\
&= \lim_{M\to\infty} \int_{\mathcal{Z}} G^M Q(du, dy, d\zeta) \qquad (3.43)\\
&= \int_{\mathcal{Z}} \tilde{c}\left(\omega_0, \zeta, u, \int_{\mathbf{U}} uQ(du \times \mathbf{Y} \times \mathbf{S}), \int_{\mathbf{U}\times\mathbf{S}} \Lambda Q(du \times \mathbf{Y} \times d\zeta)\right) Q(du, dy, d\zeta), \qquad (3.44)
\end{aligned}$$

where (3.42) is true since

$$\tilde{c}\left(\omega_0, \zeta, u, \int_{\mathbf{U}} uQ_N(du \times \mathbf{Y} \times \mathbf{S}), \int_{\mathbf{U}\times\mathbf{S}} \Lambda Q_N(du \times \mathbf{Y} \times d\zeta)\right) \geq G_N^M.$$

Equality (3.43) follows from the generalized convergence theorem in [123, Theorem 3.5] since G_N^M is bounded and continuously converges to G^M, i.e., P-almost surely

$$\begin{aligned}
\lim_{N\to\infty} \min\left\{M,\ \tilde{c}\left(\omega_0, \zeta, u_N, \int_{\mathbf{U}} uQ_N(du \times \mathbf{Y} \times \mathbf{S}), \int_{\mathbf{U}\times\mathbf{S}} \Lambda Q_N(du \times \mathbf{Y} \times d\zeta)\right)\right\} \\
= \min\left\{M, \tilde{c}\left(\omega_0, \zeta, u, \int_{\mathbf{U}} uQ(du \times \mathbf{Y} \times \mathbf{S}), \int_{\mathbf{U}\times\mathbf{S}} \Lambda Q(du \times \mathbf{Y} \times d\zeta)\right)\right\}, \qquad (3.45)
\end{aligned}$$

when $u_N \to u$ as $N \to \infty$. The monotone convergence theorem implies (3.44). Hence, (3.39) holds which implies $\limsup_{N\to\infty} J_T^N(\underline{\tilde{\gamma}}_T^{*,N}) = J_T^\infty(\underline{\tilde{\gamma}}_T^{*,\infty})$, and this completes the proof following from [120, Theorem 5]. Here, for completeness we present the proof which is similar to the analysis of the proof [120, Theorem 5] for dynamic teams,

$$\begin{aligned}
\inf_{\underline{\gamma}_T} J_T^\infty(\underline{\gamma}_T) &\leq \limsup_{N\to\infty} J_T^N(\underline{\tilde{\gamma}}_T^{*,\infty}) \\
&= \limsup_{N\to\infty} J_T^N(\underline{\gamma}_T^{*,N}) = \limsup_{N\to\infty} \inf_{\underline{\gamma}_T^{1:N}} J_T^N(\underline{\gamma}_T^{1:N}) \\
&= \limsup_{N\to\infty} \inf_{\underline{\gamma}_T} J_T^N(\underline{\gamma}_T) \qquad (3.46)
\end{aligned}$$

$$\leq \inf_{\underline{\gamma}_T} \limsup_{N\to\infty} J_T^N(\underline{\gamma}_T) = \inf_{\underline{\gamma}_T} J^\infty(\underline{\gamma}_T),$$

where (3.46) is true since the restriction $\underline{\gamma}_T$ to the first N components is $\underline{\gamma}_T^{1:N}$. This implies that $\tilde{\underline{\gamma}}_T^{*,\infty}$ is globally optimal.

3.5.3 Proof of Theorem 3.3.2

Under Assumption 3.3.2(ii) and Assumption 3.3.3, for every $A^i \in \mathcal{B}(\mathbf{S})$, and $A^i = B^i \times \prod_{t=0}^{T-1}(D_t^i \times E_t^i)$ (where $B^i \in \mathcal{B}(\mathbb{X})$, $D_t^i \in \mathcal{B}(\mathbb{W})$, and $E_t^i \in \mathcal{B}(\mathbb{V})$), for all $N \in \mathbb{N}$, and permutations σ, we have P-almost surely

$$\begin{aligned}
&P(\underline{\zeta}^1 \in A^1, \ldots, \underline{\zeta}^N \in A^N | \omega_0) \\
&= P(x_0^1 \in B^1, \ldots, x_0^N \in B^N | \omega_0) \prod_{i=1}^{N} \prod_{t=0}^{T-1} P(w_t^i \in D_t^i) P(v_t^i \in E_t^i) \qquad (3.47) \\
&= P((x_0^\sigma)^1 \in B^1, \ldots, (x_0^\sigma)^N \in B^N | \omega_0) \prod_{i=1}^{N} \prod_{t=0}^{T-1} P((w_t^\sigma)^i \in D_t^i) P((v_t^\sigma)^i \in E_t^i) \qquad (3.48) \\
&= P((\underline{\zeta}^\sigma)^1 \in A^1, \ldots, (\underline{\zeta}^\sigma)^N \in A^N | \omega_0),
\end{aligned}$$

where (3.47) follows from Assumption 3.3.2(ii), and (3.48) follows from Assumption 3.3.3. Hence, $(\underline{\zeta}^1, \underline{\zeta}^2, \ldots)$ are exchangeable conditioned on ω_0.

Hence, following from a similar argument as the proof of Theorem 3.2.2 (by considering ω_0 in the cost function and the law of total expectation (by first conditioning on ω_0), under Assumption 3.3.2(ii) and Assumption 3.3.3, one can consider a sequence of N-DM teams which are symmetrically optimal that defines $(\mathcal{P}_T^{N,\mathrm{MF}})$ and whose limit is identified with $(\mathcal{P}_T^{\infty,\mathrm{MF}})$. Since initial states are not necessarily independent conditioned on ω_0, we can not establish that $(\underline{\gamma}_T^{*,\infty}(\underline{y}^i), \underline{y}^i, \underline{\zeta}^i)$ are i.i.d. random vectors conditioned on ω_0 which has been used in (3.32) in (Step 1) of the proof of Theorem 3.3.1 to show that Q_N converges weakly to Q P-almost surely.

However, we note that since $(\underline{\zeta}^1, \underline{\zeta}^2, \ldots)$ are exchangeable conditioned on ω_0, for every $A^i \in$

$\mathcal{B}(\mathbf{S})$ and $C \in \mathcal{B}(\Omega_0)$, and for all $N \in \mathbb{N}$, and permutations σ, we have

$$P\Big(\underline{\zeta}^1 \in A^1, \ldots, \underline{\zeta}^N \in A^N, \omega_0 \in C\Big) = P\Big((\underline{\zeta}^\sigma)^1 \in A^1, \ldots, (\underline{\zeta}^\sigma)^N \in A^N, \omega_0 \in C\Big). \tag{3.49}$$

Let $\alpha^i := (\omega_0, \underline{\zeta}^i)$. Hence, (3.49) implies that $(\alpha^1, \alpha^2, \ldots)$ is exchangeable. Following from [1, Proposition 3.8(a)], there exists a random vector $z \in [0,1]$ such that, $(\underline{\zeta}^1, \underline{\zeta}^2, \ldots)$ are i.i.d. random vectors conditioned on (ω_0, z).

Let $\tilde{\omega}_0 := (\omega_0, z)$. Hence, under Assumption 3.3.2(ii) and Assumption 3.3.3, conditioned on $\tilde{\omega}_0$, $(\underline{\zeta}^1, \underline{\zeta}^2, \ldots)$ are i.i.d. random vectors. Following from standard stochastic realization results [29, Lemma 3.1], we can represent any stochastic kernel in a functional form, with almost sure equivalence, $\underline{\zeta}^i = g(\tilde{\omega}_0, \theta^i)$ for some independent θ^i and measurable g (note that following from exchangeability, g is identical for all $i \in \mathbb{N}$ and $(\theta^1, \theta^2, \ldots)$ are i.i.d. random vectors).

Since conditioned on $\tilde{\omega}_0$, $(\underline{\zeta}^1, \underline{\zeta}^2, \ldots)$ are i.i.d. random vectors, $(\underline{\gamma}_T^{*,\infty}(\underline{y}^i), \underline{y}^i, \underline{\zeta}^i)$ are i.i.d. random vectors conditioned on $\tilde{\omega}_0$, hence for every $\tilde{\omega}_0$ on a set of P-measure one, we have for every continuous and bounded function g in actions and observations, by the strong law of large numbers,

$$P\left(\left\{\omega \in \Omega \,\middle|\, \lim_{N\to\infty} \left| \frac{1}{N} \sum_{i=1}^{N} g\Big(\underline{\gamma}_T^{*,\infty}(\underline{y}^i), \underline{y}^i, \underline{\zeta}^i\Big) - E\Big[g(\underline{\gamma}_T^{*,\infty}(\underline{y}^1), \underline{y}^1, \underline{\zeta}^1)\Big|\tilde{\omega}_0\Big]\right| = 0\right\}\middle|\tilde{\omega}_0\right) = 1.$$

Hence, following from an identical analysis as that of (Step 1) of the proof of Theorem 3.3.1, conditioned on $\tilde{\omega}_0$, Q_N converges weakly to Q, for every $\tilde{\omega}_0$ on a set of P-measure one.

Following from the representation $\underline{\zeta}^i = g(\tilde{\omega}_0, \theta^i)$, we have

$$\begin{aligned}
&\limsup_{N\to\infty} E\Big[E\Big[\int_{\mathcal{Z}} \tilde{c}\left(\omega_0, \zeta, u, \int_{\mathbf{U}} uQ_N(du \times \mathbf{Y} \times \mathbf{S}), \int_{\mathbf{U}\times\mathbf{S}} \Lambda Q_N(du \times \mathbf{Y} \times d\zeta)\right) Q_N(du, dy, d\zeta)\Big|\omega_0\Big]\Big] \\
&= \limsup_{N\to\infty} \int_{\Omega_0\times[0,1]} E\Big[\int_{\mathcal{Z}} \tilde{c}\left(\tilde{\omega}_0, \zeta, u, \int_{\mathbf{U}} uQ_N(du \times \mathbf{Y} \times \mathbf{S}), \int_{\mathbf{U}\times\mathbf{S}} \Lambda Q_N(du \times \mathbf{Y} \times d\zeta)\right) \\
&\qquad \times Q_N(du, dy, d\zeta)\Big|\tilde{\omega}_0\Big] P(d\tilde{\omega}_0),
\end{aligned} \tag{3.50}$$

where (3.50) follows from the fact that for all $N \in \mathbb{N}$, and for every $A^i \in \mathcal{B}(\mathbf{S})$

$$\int_{\Omega_0} P(\underline{\zeta}^1 \in A^1, \dots, \underline{\zeta}^N \in A^N | \omega_0) P(d\omega_0) = \int_{\Omega_0 \times [0,1]} \prod_{i=1}^{N} \eta(\underline{\zeta}^i \in A^i | z, \omega_0) P(dz, d\omega_0),$$

and with slightly abuse of notations we use the same notation, $\tilde{c}$, for the cost function after transformation in (3.50). The rest of the proof is identical to that of Theorem 3.3.1.

3.5.4 Proof of Theorem 3.3.3

We use individually randomized policies and we show that for every sequence of N-DM optimal policies, there exists a subsequence which converges to an optimal independently randomized policy for the mean-field limit under an appropriate topology defined by the product topology where each coordinate is endowed with the weak convergence topology. In (Step 1), we show that for each finite N-DM team problem, optimal policies are deterministic and symmetric and we consider the independently randomized policies induced by such policies $\{P_N\}_N$ (where $P_N \in \mathcal{P}(\mathbf{Y} \times \mathbf{U})$ for each DM satisfying (3.7)), as our sequence to be studied. We also define the sequence of empirical measures induced by these policies, Q_N, as (3.27).

In (Step 2), we show that for every sequence of policies satisfying a moment condition there exists a subsequence such that policies $\{P_n\}_n$ for each DM and a subsequence of empirical measures $\{Q_n\}_n$ induced by these policies (where $n \in \mathbb{I}$ is the index set of a convergent subsequence) converge weakly to a limit P-almost surely, that is, for a set of P-measure one, for every bounded function g which is continuous in actions and observations and measurable in uncertainties,

$$P\left(\left\{ \omega \in \Omega \middle| \lim_{n \to \infty} \left(\int g dQ_n - \int g dQ \right) = 0 \right\} \middle| \omega_0 \right) = 1.$$

To this end, we first show that for each DM a sequence $\{P_n\}_n$ is tight, then we show that the sequence of empirical measures $\{Q_n\}_n$ induced by these policies converges weakly to a limit P-almost surely.

In (Step 3), we show that the set of policies for each DM is closed under the weak convergence topology, hence, the limit policy satisfies the required measurability/conditional independence constraints (that is, the limit policy satisfies (3.7)). In (Step 4), we use the lower semi-continuity argument to show that the expected cost function under the induced limit policy is less than or equal to the expected cost achieved by the sequence of N-DM optimal policies.

(Step 1): Under Assumption 3.3.1, Assumption 3.3.1(c) and Assumption 3.3.2, and by condition (i) using [138, Theorem 5.2], or condition (ii) using [138, Theorem 5.6], there exists a deterministic optimal policy for each finite N-DM team problem. Action spaces are convex and the team problem is convex in policies, hence, using Lemma 3.3.1, one can consider a sequence of N-DM teams which are symmetrically optimal that defines ($\mathcal{P}_T^{N,MF}$) and whose limit is identified with ($\mathcal{P}_T^{\infty,MF}$). Hence, for each N-DM team problem, we consider symmetric randomized optimal policies.

(Step 2): In the following, we first show that the set of policies $P_N \in \mathcal{P}(\mathbf{Y} \times \mathbf{U})$ for each DM satisfying (3.7) and the moment condition is tight, then, by symmetry, we show that $Q_{N\,N}$ induced by this set of policies is tight. We use the fact that conditioned on ω_0, $(\underline{\gamma}_T^{*,N}(\underline{y}^i), \underline{y}^i, \underline{\zeta}^i)$ are i.i.d. random vectors (this follows from symmetry of the information structure and Lemma 3.3.1 since every DM applies the identical policy $\underline{\gamma}_T^{*,N}$) and also since the space of control policies is tight under the weak convergence for each DM (see e.g., [138, proof of Theorem 4.7]).

Since actions of DMs do not affect the observations of others, the policy spaces are decoupled from the actions of other decision makers. Since we can restrict the search for optimal policies over those satisfying the moment condition, the fact that $\nu \to \int \nu(dx)g(x)$ is lower semi-continuous for a continuous function g [138, proof of Theorem 4.7] implies that the marginals on $\mathbf{U}$ satisfying the moment condition are tight under the weak convergence topology. Hence, the collection of all probability measures with these tight marginals is also tight (see e.g., [137, Proof of Theorem 2.4]). This implies that the sequence of randomized policies satisfying the moment condition is tight.

Since every DM applies an identical policy and since observations are conditionally i.i.d., a countably infinite product of space of policies of DMs is tight (where each coordinate is tight in the weak convergence topology). Hence, there exists a subsequence of policies $\tilde{P}_n \in \mathcal{P}(\prod_i(\mathbf{Y} \times \mathbf{U}))$ (as a product of policies of DMs) converges weakly to a limit $\tilde{P}$ (each coordinate converges weakly) P-almost surely. Furthermore, since every DM applies an identical policy, conditioned on ω_0, actions induced by an identical randomized policies, observations and disturbances are i.i.d. through DMs. Hence, following from a similar argument as (Step 1) of the proof of Theorem 3.3.1, a subsequence of empirical measures $\{Q_n\}_{n\in\mathbb{I}}$ converges P-almost surely to Q in w-s topology. We note that the convergence is under the weak convergence topology, but since $\underline{\zeta}^i$s are exogenous with a fixed marginal, the convergence is also in the w-s topology.

(Step 3): In this step, we show that each coordinate of the space of policies (space of policies for each DM) is closed under the weak convergence topology. This in particular implies that the space of policies is closed under the product topology and using (Step 1), we can conclude that the space of control policies is compact under the product topology where each coordinate is weakly compact.

Assume P_n is a policy for DM^i induced by a randomized policy converging weakly to P_∞. In fact, conditions (i) or (ii) leads to the closedness of the set of policies (see (3.7)) induced by P_n. If Assumption 3.3.4 holds, then by the discussion in the proof of [138, Theorem 5.2], each coordinate of policy spaces corresponding to DM^i acting through time is closed under the weak convergence topology. Also, if condition (ii) holds, then [138, Theorem 5.6] leads to the same conclusion. Hence, each coordinate of space of policies (corresponding to DM^i) is closed under the weak convergence topology (since each coordinate of the space of policies is a finite product of space of policies for each DM at time instances $t = 0, \ldots, T-1$). Hence, following from (Step 2), there exists a subsequence $\{Q_n\}_{n\in\mathbb{I}}$ converges weakly to Q P-almost surely where Q is induced by a randomized policy in the set of policies satisfying (3.7) and the limit policy is admissible and satisfies the required measurability/conditional independence constraints.

Let for every $t = 0, \dots, T-1$, P_n^{*,ω_0} be a probability measure on actions, observations and uncertainties induced by optimal randomized policies for each DM (which is identical because of symmetry) for N-DM teams conditioned on ω_0, i.e., a probability measure that satisfies

$$\begin{aligned}
&\int g(\omega_0, x_0^i, \zeta_{0:t-1}^i, y_{0:t}^i, u_{n,0:t}^{i,*}) P_n^{*,\omega_0}(dx_0^i, d\zeta_{0:t-1}^i, dy_{0:t}^i, du_{n,0:t}^{i,*}) \\
&= \int g(\omega_0, x_0^i, \zeta_{0:t-1}^i, y_{0:t}^i, u_{n,0:t}^{i,*}) \mu^i(dx_0^i, d\zeta_{0:t-1}^i | \omega_0) \\
&\quad \times \prod_{k=0}^{t} \Pi_k^{*,n}(du_{n,k}^{*,i} | y_k^i) p_k(dy_k^i | \omega_0, x_0^i, \zeta_{0:k-1}^i, y_{0:k-1}^i, u_{n,0:k-1}^{i,*}),
\end{aligned} \tag{3.51}$$

for all bounded functions g which is continuous in actions and observations and measurable in other arguments. We denote $\underline{u}_n^{i,*} := (u_{n,0}^{i,*}, \dots, u_{n,T-1}^{i,*})$ as the action of DM^i through time induced by $\Pi_t^{*,n}$. Similarly, we denote P^{*,ω_0} as a probability measure induced by the limit policy, i.e., a probability measure satisfying (3.51) induced by $\Pi_k^{*,\infty}$ where $\underline{u}_\infty^{i,*} := (u_{\infty,0}^{i,*}, \dots, u_{\infty,T-1}^{i,*})$ is the action of DM^i through time induced by $\Pi_k^{*,\infty}$.

(Step 4): Now, we show that the expected cost function under the limit randomized policy is less than or equal to the expected cost achieved by $\limsup_{n\to\infty} J_T^n(\tilde{\underline{\gamma}}_T^{*,n})$. Since the cost function is continuous in states and actions, under the reduction (conditions (i) or (ii)), we have P-almost surely

$$\begin{aligned}
&\frac{1}{N}\sum_{i=1}^{N}\sum_{t=0}^{T-1} c\Big(\omega_0, x_t^i, u_t^i, \frac{1}{N}\sum_{p=1}^{N} u_t^p, \frac{1}{N}\sum_{p=1}^{N} x_t^p\Big) \prod_{i=1}^{N}\prod_{t=0}^{T-1} \psi_t\Big(y_t^i, \omega_0, x_0^i, \zeta_{0:t-1}^i, y_{0:t-1}^i, u_{0:t-1}^i\Big) \\
&\qquad = \frac{1}{N}\sum_{i=1}^{N} \bar{c}\Big(\omega_0, \underline{\zeta}^i, \underline{u}^i, \frac{1}{N}\sum_{p=1}^{N}\underline{u}^p, \frac{1}{N}\sum_{p=1}^{N}\Lambda(\underline{u}^p, \underline{\zeta}^p)\Big) \prod_{i=1}^{N} \underline{\psi}\Big(\underline{y}^i, \omega_0, \underline{\zeta}^i, \underline{u}^i\Big),
\end{aligned} \tag{3.52}$$

where (3.52) is true following from (3.2) and Assumption 3.3.1 for some functions $\bar{c} : \Omega_0 \times \mathbf{S} \times \mathbf{U} \times \mathbf{U} \times \mathbf{X} \to \mathbb{R}_+$ continuous in states and actions and a function $\Lambda : \mathbf{U} \times \mathbf{S} \to \mathbf{X}$ continuous in

actions and

$$\prod_{i=1}^{N}\underline{\psi}\left(\underline{y}^i,\omega_0,\underline{\zeta}^i,\underline{u}^i\right):=\prod_{i=1}^{N}\prod_{t=0}^{T-1}\psi_t\left(y_t^i,\omega_0,x_0^i,\zeta_{0:t-1}^i,y_{0:t-1}^i,u_{0:t-1}^i\right).$$

We have

$$\begin{aligned}
&\limsup_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}\sum_{t=0}^{T-1}E^{\underline{\gamma}_T^{*,1:N}}\left[c\left(\omega_0,x_t^i,u_t^i,\frac{1}{N}\sum_{p=1}^{N}u_t^p,\frac{1}{N}\sum_{p=1}^{N}x_t^p\right)\right]\\
&\geq\lim_{M\to\infty}\limsup_{N\to\infty}\int\int_{\mathcal{Z}}\min\left\{M,\tilde{c}\left(\omega_0,\zeta,u,\int_{\mathbf{U}}uQ_N(du\times\mathbf{Y}\times\mathbf{S}),\int_{\mathbf{U}\times\mathbf{S}}\Lambda Q_N(du\times\mathbf{Y}\times d\zeta)\right)\right\}\\
&\qquad\times Q_N(du,dy,d\zeta)\prod_{i=1}^{\infty}P_N^{*,\omega_0}(d\underline{u}_N^{i,*},d\underline{y}^i,d\underline{\zeta}^i)\prod_{i=1}^{\infty}\underline{\psi}(\underline{y}^i,\omega_0,\underline{\zeta}^i,\underline{u}_N^{i,*})P(d\omega_0) \qquad (3.53)\\
&\geq\lim_{M\to\infty}\lim_{n\to\infty}\int\int_{\mathcal{Z}}\min\left\{M,\tilde{c}\left(\omega_0,\zeta,u,\int_{\mathbf{U}}uQ_n(du\times\mathbf{Y}\times\mathbf{S}),\int_{\mathbf{U}\times\mathbf{S}}\Lambda Q_n(du\times\mathbf{Y}\times d\zeta)\right)\right\}\\
&\qquad\times Q_n(du,dy,d\zeta)\prod_{i=1}^{\infty}P_n^{*,\omega_0}(d\underline{u}_n^{i,*},d\underline{y}^i,d\underline{\zeta}^i)\prod_{i=1}^{\infty}\underline{\psi}(\underline{y}^i,\omega_0,\underline{\zeta}^i,\underline{u}_n^{i,*})P(d\omega_0) \qquad (3.54)\\
&=\lim_{M\to\infty}\int\lim_{n\to\infty}\int_{\mathcal{Z}}\min\left\{M,\tilde{c}\left(\omega_0,\zeta,u,\int_{\mathbf{U}}uQ_n(du\times\mathbf{Y}\times\mathbf{S}),\int_{\mathbf{U}\times\mathbf{S}}\Lambda Q_n(du\times\mathbf{Y}\times d\zeta)\right)\right\}\\
&\qquad\times Q_n(du,dy,d\zeta)\prod_{i=1}^{\infty}P_n^{*,\omega_0}(d\underline{u}_n^{i,*},d\underline{y}^i,d\underline{\zeta}^i)\prod_{i=1}^{\infty}\underline{\psi}(\underline{y}^i,\omega_0,\underline{\zeta}^i,\underline{u}_n^{i,*})P(d\omega_0) \qquad (3.55)\\
&=\lim_{M\to\infty}\int\int_{\mathcal{Z}}\min\left\{M,\tilde{c}\left(\omega_0,\zeta,u,\int_{\mathbf{U}}uQ(du\times\mathbf{Y}\times\mathbf{S}),\int_{\mathbf{U}\times\mathbf{S}}\Lambda Q(du\times\mathbf{Y}\times d\zeta)\right)\right\}\\
&\qquad\times Q(du,dy,d\zeta)\prod_{i=1}^{\infty}P^{*,\omega_0}(d\underline{u}_\infty^{i,*},d\underline{y}^i,d\underline{\zeta}^i)\prod_{i=1}^{\infty}\underline{\psi}(\underline{y}^i,\omega_0,\underline{\zeta}^i,\underline{u}_\infty^{i,*})P(d\omega_0) \qquad (3.56)\\
&=\int\int_{\mathcal{Z}}\tilde{c}\left(\omega_0,\zeta,u,\int_{\mathbf{U}}uQ(du\times\mathbf{Y}\times\mathbf{S}),\int_{\mathbf{U}\times\mathbf{S}}\Lambda Q(du\times\mathbf{Y}\times d\zeta)\right)\\
&\qquad\times Q(du,dy,d\zeta)\prod_{i=1}^{\infty}P^{*,\omega_0}(d\underline{u}_\infty^{i,*},d\underline{y}^i,d\underline{\zeta}^i)\prod_{i=1}^{\infty}\underline{\psi}(\underline{y}^i,\omega_0,\underline{\zeta}^i,\underline{u}_\infty^{i,*})P(d\omega_0) \qquad (3.57)
\end{aligned}$$

where (3.53) follows from the definition of empirical measures and by integrating over the set $(\prod_{i=n_l+1}^{\infty}\mathbf{Y}\times\mathbf{S})$ and since P-almost surely

$$\min\left\{M,\tilde{c}\left(\omega_0,\zeta,u,\int_{\mathbf{U}}uQ_N(du\times\mathbf{Y}\times\mathbf{S}),\int_{\mathbf{U}\times\mathbf{S}}\Lambda Q_N(du\times\mathbf{Y}\times d\zeta)\right)\right\}$$

$$\leq \tilde{c}\left(\omega_0, \zeta, u, \int_{\mathbf{U}} uQ_N(du \times \mathbf{Y} \times \mathbf{S}), \int_{\mathbf{U}\times\mathbf{S}} \Lambda Q_N(du \times \mathbf{Y} \times d\zeta)\right).$$

Inequality (3.54) is true since limsup is the greatest convergent subsequential limit for a bounded sequence and (3.55) follows from the dominated convergence theorem. We note that $\{Q_n\}_n$ is induced by $\underline{u}_n^{i,*}$ for each DM. Since $\{Q_n\}_{n\in I}$ converges weakly to Q P-almost surely, by the moment condition and Remark 3.3.4, a similar argument as (Step 3) of the proof of Theorem 3.3.1 implies that P-almost surely

$$\int_{\mathbf{U}} uQ_n(du \times \mathbf{Y} \times \mathbf{S}) \xrightarrow{n\to\infty} \int_{\mathbf{U}} uQ(du \times \mathbf{Y} \times \mathbf{S})$$
$$\int_{\mathbf{U}\times\mathbf{S}} \Lambda Q_n(du \times \mathbf{Y} \times d\zeta) \xrightarrow{n\to\infty} \int_{\mathbf{U}\times\mathbf{S}} \Lambda Q(du \times \mathbf{Y} \times d\zeta)$$

Hence, (3.56) follows from [123, Theorem 3.5] since

$$\min\left\{M, \tilde{c}\left(\omega_0, \zeta, u, \int_{\mathbf{U}} uQ_n(du \times \mathbf{Y} \times \mathbf{S}), \int_{\mathbf{U}\times\mathbf{S}} \Lambda Q_n(du \times \mathbf{Y} \times d\zeta)\right)\right\}$$

is bounded and non-negative, and continuously converges in u P-almost surely (see (3.45)). That is because, conditioned on ω_0, $\underline{y}^i$ are i.i.d. random vectors (thanks to the symmetry), the space of policies is compact under the product topology (with the weak convergence topology for each coordinate (for each DM)), $\prod_{i=1}^{\infty} \underline{\psi}(\underline{y}^i, \omega_0, \underline{\zeta}^i, \underline{u}_n^{i,*})$ converges in the product topology, and the cost function and $\underline{\psi}$ are continuous. Finally, (3.57) follows from the monotone convergence theorem. Hence, the proof is completed.

3.5.5 Proof of Theorem 3.4.3

We show $\limsup_{T\to\infty} J_T(\underline{\gamma}_T^*) = J(\underline{\gamma}_\infty^*)$ and invoke [120, Theorem 5] or [90, Theorem 1] to complete the proof. From (3.16), we have

$$\lim_{T\to\infty} |J_T(\underline{\gamma}_\infty^*) - J_T(\underline{\gamma}_T^*)|$$

$$= \lim_{T\to\infty} \frac{2}{T}\Bigg| E\Bigg[(x_0^1)^T(P_0^{(T)} - P)x_0^1\Bigg] + \frac{2}{T}\sum_{t=0}^{T-1} E\Bigg[(w_t^1)^T(P_t^{(T)} - P)w_t^1\Bigg] \tag{3.58}$$

$$+\frac{2}{T}\sum_{t=0}^{T-1} E\Bigg[\Bigg(E[x_0^2|x_0^1]\Bigg)^T\Bigg((L_t^{(T)})^T B^T P_{t+1}^{(T)} B L_t^{(T)} - (L_t^{(\infty)})^T B^T P B L_t^{(\infty)}\Bigg)E[x_0^2|x_0^1]\Bigg] \tag{3.59}$$

$$+\frac{2}{T}\sum_{t=0}^{T-1} E\Bigg[(x_0^1)^T(A^T)^{t-1}\Bigg(P_{t+1}^{(T)} B L_t^{(T)} - P B L_t^{(\infty)}\Bigg)E[x_0^2|x_0^1]\Bigg]\Bigg| \tag{3.60}$$

$$= 0, \tag{3.61}$$

where (3.58) is zero since $P_0^{(T)}$ converges to P using Lemma 3.4.1 since $P_{t+1}^{(T+1)} = P_t^{(T)}$. Expression (3.60) converges to zero since $L_t^{(T)}$ in (3.15) converges pointwise to $L_t^{(\infty)}$ as $T \to \infty$, we have $\sum_{s=t+1}^{\infty} B^T(A^T)^{s-t} P B L_s^{(\infty)} < \infty$, and this implies that $\lim_{s\to\infty} L_s^{(\infty)} = 0$. Hence, we have for every $\epsilon > 0$, there exists $\hat{N} > T$ such that for every $t > \hat{N}$, $|Tr[L_t^{(\infty)}(L_t^{(\infty)})^T]| < \epsilon$. We define $L_t^{(T)} = 0$ for $t > T$. Expression (3.59) is equal to zero following from Lemma 3.4.1 and the fact that $|Tr[L_t^{(\infty)}(L_t^{(\infty)})^T]| < \epsilon$ for every $t > \hat{N}$. Hence, equality (3.61) is true and global optimality follows from [120, Theorem 5]. The closed loop system is stable since $\limsup_{t\to\infty} E(||x_t^1||^2) < \infty$ following from $||A + BK|| < 1$ (all the eigenvalues of $A + BK$ are inside of the unit circle), and since $||L_t^{(\infty)}||$ is uniformly bounded.

3.5.6 Proof of Theorem 3.4.5

As $T \to \infty$, we have $X^{r,(\infty)} = \lim_{T\to\infty} X_t^{r,(T)}$ for (3.22) since the the recursion for $s \to s$ corresponds to the classical Riccati equation and $X_t^{r,(T)}$ is a continuous function of $X_t^{s,(T)}$ [84, Corollary 7]. We have,

$$\begin{aligned}
\lim_{T\to\infty} |J_T(\underline{\gamma}_{-\infty}^*) - J_T(\underline{\gamma}_T^*)| &= \lim_{T\to\infty} \frac{1}{T}| \sum_{i=1,\omega^i\to r}^{N} Tr([(X_0^{r,(T)})^{\{i\},\{i\}} - (X^{r,(\infty)})^{\{i\},\{i\}}] \times \mathbb{E}(x_0^i(x_0^i)^T)) \\
&\quad + \sum_{t=0}^{T-1}\sum_{i=1,\omega^i\to r}^{N} Tr([(X_{t+1}^{r,(T)})^{\{i\},\{i\}} - (X^{r,(\infty)})^{\{i\},\{i\}}] \times \mathbb{E}(\omega_t^i(\omega_t^i)^T))| \\
&= \lim_{T\to\infty} \frac{1}{T}\sum_{t=0}^{T-1}| \sum_{i=1,\omega^i\to r}^{N} Tr([(X_{t+1}^{r,(T)})^{\{i\},\{i\}}\Sigma_\omega^i - (X^{r,(\infty)})^{\{i\},\{i\}}\Sigma_\omega^i]| = 0,
\end{aligned}$$

where (3.23) and Lemma 3.4.1 implies the last equality. This is because (3.22) implies that $X_{t+1}^{r,(T+1)} = X_t^{r,(T)}$; hence, invoking [120, Theorem 5] implies global optimality of $\underline{\gamma}_{\infty}^{*}$. The stability argument follows from [84, Corollary 7] and [141].

Chapter 4

Large-Scale Exchangeable Stochastic Teams and Their Mean-Field Limit: Relaxing Convexity

In this chapter, we study stochastic (static and dynamic) team problems with finite but large as well as a countably infinite number of decision makers (DMs). We focus on existence and structural results for exchangeable stochastic team problems, where now the convexity condition imposed in the preceding chapters is relaxed. In addition, we establish approximation results for N-DM static and dynamic teams.

We now restate the setup for stochastic team problems.

4.1 Static Teams.

First, we consider static problems, and then, we introduce the dynamic ones.

Problem ($\mathcal{P}_N$): Let $\mathcal{N} = \{1, \dots, N\}$. Let $\underline{\gamma}_N = (\gamma^1, \cdots, \gamma^N)$ where $\Gamma_N = \prod_{i=1}^N \Gamma^i$. Given a (decentralized) static information structure, a static stochastic team problem is described with an expected cost function (to be minimized) of $\underline{\gamma}_N$ as

$$J_N(\underline{\gamma}_N) = E^{\underline{\gamma}_N}[c(\omega_0, \underline{u}_N)] := E[c(\omega_0, \gamma^1(y^1), \cdots, \gamma^N(y^N))], \tag{4.1}$$

for some Borel measurable cost function $c : \Omega_0 \times \prod_{k=1}^N \mathbb{U}^k \to \mathbb{R}_+$.

As we consider exchangeable team problems, we let action and observation spaces be identical through DMs $\mathbb{U}^i = \mathbb{U} \subseteq \mathbb{R}^n$ and $\mathbb{Y}^i = \mathbb{Y} \subseteq \mathbb{R}^m$ for all $i \in \mathcal{N}$, where n and m are positive integers. In this chapter, we also focus is on a class of exchangeable team problems satisfying an exchangeability assumption on the cost function.

Assumption 4.1.1. *The cost function is exchangeable with respect to actions for all ω_0, i.e., for any permutation σ of $\{1, \dots, N\}$,* (1.11) *for all ω_0.*

In particular, again for our main results, we focus on team problems with the following expected cost function instead of (4.1):

$$\frac{1}{N} \sum_{i=1}^{N} E^{\underline{\gamma}_N} \left[c\Big(\omega_0, u^i, \frac{1}{N} \sum_{p=1}^{N} u^p \Big) \right]. \tag{4.2}$$

Clearly, $\frac{1}{N} \sum_{i=1}^{N} c(\omega_0, u^i, \frac{1}{N} \sum_{p=1}^{N} u^p)$ satisfies Assumption 4.1.1. Now, we recall a stochastic team problem with countably infinite number of decision makers.

Problem ($\mathcal{P}_\infty$): Consider a stochastic team with a countably infinite number of decision makers, that is, $\mathcal{N} = \mathbb{N}$. Let $\Gamma = \prod_{i \in \mathbb{N}} \Gamma^i$ and $\underline{\gamma} = (\gamma^1, \gamma^2, \dots)$. Given a (decentralized) static information structure, a static stochastic team problem is described with an expected cost function (to be minimized) of a policy $\underline{\gamma}$ as

$$J(\underline{\gamma}) = \limsup_{N \to \infty} \frac{1}{N} E^{\underline{\gamma}} \left[\sum_{i=1}^{N} c\Big(\omega_0, u^i, \frac{1}{N} \sum_{p=1}^{N} u^p \Big) \right], \tag{4.3}$$

for some Borel measurable cost function $c : \Omega_0 \times \mathbb{U} \times \mathbb{U} \to \mathbb{R}_+$.

Our theorems require the following absolute continuity condition (policy-independent static reduction) under which we can equivalently view the observations of each DM as independent and also independent of ω_0 via change of measure argument.

Assumption 4.1.2. *Assume that for every $N \in \mathbb{N} \cup \{\infty\}$, there exists a probability measure Q^i on*

$\mathbb{Y}$ and a function f^i for all $i \in \mathcal{N}$ such that for all Borel set B^i in $\mathbb{Y}$ (with $B := B^1 \times \cdots \times B^N$)

$$\tilde{\mu}^N(B|\omega_0) = \prod_{i=1}^{N} \int_{B^i} f^i(y^i, \omega_0, y^1, \ldots, y^{i-1}) Q^i(dy^i), \tag{4.4}$$

where $\tilde{\mu}^N$ is the conditional distribution of observations $(y^1, \ldots, y^N)$ given ω_0.

The above allows us to introduce a suitable topology under which the space of randomized policies is Borel (see Section 4.4.1). In addition, our main Theorem 4.5.1 imposes the following assumptions on the observations and action space.

Assumption 4.1.3. *Assume that*

(i) Observations $\{y^i\}_{i\in\mathcal{N}}$ are i.i.d. conditioned on ω_0;

(ii) Action space $\mathbb{U}$ is compact.

We note that under Assumption 4.1.2 and Assumption 4.1.3(i), there exists an identical reference probability measure Q and function f such that the absolute continuity condition

$$\begin{aligned}\tilde{\mu}^N(B|\omega_0) &= \prod_{i=1}^{N} \hat{\mu}(B^i|\omega_0) \\ &= \prod_{i=1}^{N} \int_{B^i} f(y^i, \omega_0) Q(dy^i),\end{aligned}$$

where $\hat{\mu}$ is the conditional distribution of each observation y^i given ω_0. We note that the function f and the measure Q are identical through DMs since observations are identically distributed conditioned on ω_0. Furthermore, our main Theorem 4.5.1 imposes the following continuity assumption on the cost function.

Assumption 4.1.4. *The cost function in* (4.2)*, $c : \Omega_0 \times \mathbb{U} \times \mathbb{U} \to \mathbb{R}_+$, is continuous in its second and third arguments for all ω_0.*

For our results in Section 4.5, we impose Assumption 4.1.1 and Assumption 4.1.2, but we only impose Assumption 4.1.3 and Assumption 4.1.4 when they are needed.

4.1.1 Dynamic Teams.

Again, we consider exchangeable teams, and hence, we let action, observation, and state spaces, respectively, be identical through DMs $i \in \mathcal{N}$, and for simplicity, also through time $t = 0, \dots, T-1$, $\mathbb{U}_t^i = \mathbb{U} \subseteq \mathbb{R}^n$, $\mathbb{Y}_t^i = \mathbb{Y} \subseteq \mathbb{R}^{n'}$, $\mathbb{X}_t^i = \mathbb{X} \subseteq \mathbb{R}^{n''}$ for all $i \in \mathcal{N}$ and $t = 0, \dots, T-1$, where n, n' and n'' are positive integers. Define state dynamics and observation dynamics of DMs as follows:

$$x_{t+1}^i = f_t\bigg(x_t^i, u_t^i, \frac{1}{N}\sum_{p=1}^{N} x_t^p, \frac{1}{N}\sum_{p=1}^{N} u_t^p, w_t^i\bigg), \tag{4.5}$$

$$y_t^i = h_t\bigg(x_{0:t}^i, u_{0:t-1}^i, v_{0:t}^i\bigg), \tag{4.6}$$

where functions f_t and h_t are measurable functions and v_t^i and w_t^i are random vectors representing uncertainties in state dynamics and observations. We denote $x_{0:t}^i := (x_0^i, \dots, x_t^i)$, $u_{0:t-1}^i := (u_0^i, \dots, u_{t-1}^i)$, and $v_{0:t}^i := (v_0^i, \dots, v_t^i)$. Let the admissible policies $(\gamma_{0:T-1}^i)_{i\in\mathcal{N}}$ (with $\gamma_{0:T-1}^i := (\gamma_0^i, \dots, \gamma_{T-1}^i)$) be measurable control functions so that $u_t^i = \gamma_t^i(y_t^i)$ for all $i \in \mathcal{N}$ and $t = 0, \dots, T-1$.

Problem ($\mathcal{P}_T^N$): Consider N-DM mean-field dynamic teams with the expected cost function of $\underline{\gamma}^{1:N}$ as

$$J_T^N(\underline{\gamma}^{1:N}) = E^{\underline{\gamma}^{1:N}}\bigg[\frac{1}{N}\sum_{t=0}^{T-1}\sum_{i=1}^{N} c\bigg(\omega_0, x_t^i, u_t^i, \frac{1}{N}\sum_{p=1}^{N} u_t^p, \frac{1}{N}\sum_{p=1}^{N} x_t^p\bigg)\bigg], \tag{4.7}$$

where $\underline{\gamma}^{1:N} := (\gamma_{0:T-1}^1, \dots, \gamma_{0:T-1}^N)$ and $\gamma_{0:T-1}^i := (\gamma_0^i, \dots, \gamma_{T-1}^i)$.

Problem ($\mathcal{P}_T^\infty$): Consider mean-field dynamic teams with the expected cost function of $\underline{\gamma}$ as

$$J_T^\infty(\underline{\gamma}) = \limsup_{N\to\infty} J_T^N(\underline{\gamma}^{1:N}), \tag{4.8}$$

where $\underline{\gamma} := (\gamma_{0:T-1}^1, \gamma_{0:T-1}^2, \dots)$ and $\underline{\gamma}^{1:N} := (\gamma_{0:T-1}^1, \dots, \gamma_{0:T-1}^N)$.

Our solution technique for dynamic problems is similar to the static one, which requires more

technical arguments and additional assumptions. Our theorems for dynamic case impose an absolute continuity condition (see Assumption 4.6.1) to allow us to introduce a suitable topology on control policies and to facilitate our analysis (our main Theorem 4.7.1 requires an additional technical Assumption 4.7.3). Furthermore, our main Theorem 4.7.1 imposes the following:

Assumption 4.1.5.

(i) For $t = 0, \dots, T-1$, functions f_t and h_t in (4.5) *and* (4.6) *are continuous in the states and actions and f_ts are bounded;*

(ii) The cost function in (4.7)*, $c : \Omega_0 \times \mathbb{X} \times \mathbb{U} \times \mathbb{U} \times \mathbb{X} \to \mathbb{R}_+$, is continuous in the second, third, fourth, and fifth arguments.*

Assumption 4.1.6.

(i) $(x_0^i)_{i\in\mathcal{N}}$ are i.i.d. random vectors conditioned on ω_0;

(ii) For $t = 0, \dots, T-1$, $(w_t^i)_{i\in\mathcal{N}}$ are i.i.d. random vectors, and for $i \in \mathcal{N}$, $(w_t^i)_{t=0}^{T-1}$ are mutually independent, and independent of ω_0 and $(x_0^i)_{i\in\mathcal{N}}$. For $t = 0, \dots, T-1$, $(v_t^i)_{i\in\mathcal{N}}$ are i.i.d. random vectors, and for $i \in \mathcal{N}$, $(v_t^i)_{t=0}^{T-1}$ are mutually independent, and independent of ω_0, $(x_0^i)_{i\in\mathcal{N}}$, and w_t^is for $i \in \mathcal{N}$ and $t = 0, \dots, T-1$.

(iii) $\mathbb{U}$ is compact.

In view of Assumption 4.1.6(i), we note that ω_0 also introduces a correlation between initial states. For our results in Section 4.6, we impose Assumption 4.6.1, but we impose Assumption 4.7.3, Assumption 4.1.5, and Assumption 4.1.6 when they are needed.

4.1.2 Approximations.

Finally, we address the following problem in Section 4.8. If P_π^* is a (randomized) symmetric optimal policy for $(\mathcal{P}_\infty)$ $((\mathcal{P}_T^\infty))$ then there exist $\epsilon_N \geq 0$, with $\epsilon_N \to 0$ as $N \to \infty$, such that $P_\pi^*|_N$

is ϵ_N-optimal for $(\mathcal{P}_N)$ $((\mathcal{P}_T^N))$ where $P_\pi^*|_N$ is the restriction of P_π^* to the first N decision makers. We use our symmetry results and analysis for $(\mathcal{P}_\infty)$ $((\mathcal{P}_T^\infty))$.

The main goal of this chapter is to extend the existence and structural results in Chapters 2 and 3 for stochastic static and dynamic teams to exchangeable stochastic static and dynamic teams, where the convexity of the cost functions (in actions) and convexity of the action spaces are relaxed. Hence, our main objective in this chapter is to address the following questions:

Problem 1: Existence and structural properties of optimal policies of $(\mathcal{P}_N)$ and $(\mathcal{P}_\infty)$ $((\mathcal{P}_T^N)$ and $(\mathcal{P}_T^\infty))$. For mean-field team problems $(\mathcal{P}_\infty)$ $((\mathcal{P}_T^\infty))$, we are interested in the existence and structural properties of globally optimal policies. In particular, if there is a globally optimal policy and which is symmetric (by a symmetric policy we mean that a policy is identical through DMs) for these type of problems. The purpose of this chapter is to address this question for mean-field team problems where the problem can be non-convex. The non-convexity of the problem can arise as a result of non-convexity of the action space and/or non-convexity of the cost function in actions. Also, even if the action space is convex and the cost function is convex in actions, the information structure of the problem may lead to non-convexity of the problem in policies (see for example [140, Section 3.3]). A celebrated example is the counterexample of Witsenhausen [130].

Finally, we address the following problem in Section 4.8:

Problem 2: Approximations of (globally) optimal polices for $(\mathcal{P}_N)$ and $(\mathcal{P}_\infty)$ $(\mathcal{P}_T^N)$. One may be also interested in the relations between optimal policies for mean-field teams and approximation of the optimal policies for the pre-limit N-DM teams when N is large. Suppose P_π^* is a (randomized) symmetric optimal policy for $(\mathcal{P}^\infty)$ $((\mathcal{P}_T^\infty))$. Let $P_\pi^*|_N$ be the restriction of P_π^* to the first N decision makers to apply that. Is there exist $\epsilon_N \geq 0$ with $\epsilon_N \to 0$ as $N \to \infty$ such that $P_\pi^*|_N$ is ϵ_N-optimal for $(\mathcal{P}^N)$ $((\mathcal{P}_T^N))$?

In the view of the above discussion, this chapter addresses three main problems: (i) existence and structural results for static teams with a countably infinite number of DMs (Section 4.5) (ii) existence and structural results for dynamic teams with a countably infinite number of DMs (Section

4.6) (iii) approximation results for N-DM static and dynamic teams (Section 4.8).

4.2 Literature Review and Contributions

The main focus of this chapter is on teams with infinitely many decision makers. In this direction, we note that in the preceding chapters, we have studied static and dynamic teams, where under convexity and symmetry conditions, global optimality of the limit points of a sequence of N decision maker optimal policies has been established. We also established existence and structural results for convex static and dynamic teams with infinitely many decision makers. In this chapter, convexity is not imposed. We emphasize that mean-field teams under decentralized information structures generally correspond to dynamic team problems with non-classical information structures, hence, mean-field team problems may be non-convex even under the convexity of the cost function due to the non-classical information structures.

Related results to this chapter are those on the existence and convergence of equilibria from the mean-field games literature [81, 56, 82, 83, 80, 40], which have been discussed in the preceding chapters. In this chapter, we will adopt a different and novel approach. First, under symmetry of information structures and cost functions, we show that optimal policies are of an exchangeable type for both teams with a finite and countably infinite number of decision makers. Then, we will develop a de Finetti type representation theorem that characterizes the set of optimal policies as the extreme points of a convex set.

We also note that compared to the results on the existence of a globally optimal policy in team problems where (finite) N-decision maker team problems has been considered [138, 62, 140, 109], we study static team problems with a countably infinite number of decision makers. In our approach, we use randomized policies for our analysis and we define a topology on control policies for decentralized stochastic control. A consequence of our analysis is that, in the limit of countably infinitely many decision makers, one can characterize the set of optimal policies as the extreme points of a convex set of policies, which is, in turn, a subset of decentralized, independently randomized and identical policies. Such a result is not applicable to teams with finitely many decision

makers. This geometric representation of the set of policies is related to the celebrated de Finetti's theorem. De Finetti's theorem implies that infinitely exchangeable joint probability measures can be represented as mixtures (convex combination) of identical and independent probability measures [1, 65, 77].

There has been related work in the quantum information/mechanics literature. Let us first note, however, that in [49], it has been shown that a finite number of exchangeable probability measures can be approximated by a mixture of identical and independent probability measures, and this approximation asymptotically becomes more accurate when the number of exchangeable random variables increases. The de Finetti representation type results have been extended for quantum systems where conditional probability measures have been considered [32, 103, 45, 18, 42]. In fact, for permutation-symmetric conditional probability measures, approximation results have been obtained, provided that the non-signaling property holds (a conditional independence propety between local actions and other measurements given local measurement) [32, 103, 45, 18, 42]. We refer readers to [33, 101], for a review on the connection between the non-signaling conditional probability measures and the conditional probability measures with private and common randomness.

We note that, de Finetti type results developed for conditional probability measures in quantum information literature give us a geometric interpretation we require for strategic measures (a geometric connection between non-signaling infinitely exchangeable conditional probability measures and conditional probability measures induced by common and private randomness). However, in the team problem setup, in addition to show this geometric connection, one is require to show that the common randomness is independent of the observations. We address this issue by establishing a de Finetti type representation theorem on space of policies, properly defined and metrized.

4.3 Discussion on Main Results

In Lemma 4.5.1 for static N-DM teams and in Lemma 4.7.1 for dynamic N-DM teams, we first show that the optimal policies are of N-exchangeable type. Then, in Lemma 4.5.2 for static mean-field teams and in Lemma 4.7.2 for dynamic mean-field teams, we show the global optimality of infinitely-exchangeable optimal policies and we use de Finetti representation to establish symmetry of optimal policies. We establish the existence of a symmetric randomized optimal policy for static and dynamic mean-field teams in Theorem 4.5.1 and Theorem 4.7.1, respectively. In Section 4.8, based on our analysis for the existence and symmetry of optimal policy for mean-field teams, we establish approximation results for N-DM weakly coupled teams.

One of the main difficulties in studying non-convex mean-field team problems is to show that globally optimal policies for mean-field team problems are symmetric (identical for each DM). This difficulty stems from the observation that, in general, globally optimal policies are not symmetric for non-convex pre-limit N-DM team problems (which can be seen in Example 4.5.1). This is in contrast to the convex mean-field teams where symmetry can be established for both pre-limit N-DM and mean-field team problems (Chapters 2 and 3). In our approach:

(i) We introduce a topology on control polices which is used to establish a de Finetti representation result for probability measures on policies identified as randomized policies. In Theorem 4.4.1, we show that any infinitely-exchangeable randomized policies can be represented by elements of the set of randomized policies with common and private independent randomness where conditioned on common randomness, randomization of the policies are independent and identical through DMs.

(ii) In Section 4.5 for static and Section 4.6 for dynamic N-DM stochastic teams (see Lemma 4.5.1 and Lemma 4.7.1), we show that by exchangeablity of the cost function and considering symmetric information structures (under a causality condition for the dynamic case), one can establish N-exchangeability of optimal policies.

(iii) In Section 4.5 for static and Section 4.6 for dynamic mean-field teams (see Lemma 4.5.2 and Lemma 4.7.2) under regularity conditions on the cost function and dynamics, by constructing infinitely exchangeable policies by relabeling N-exchangeable optimal policies, as N goes to infinity, we show the asymptotic optimality of infinitely exchangeable optimal policies. Hence, this, following from our de Finetti type theorem (see Theorem 4.4.1), establishes asymptotic global optimality of symmetric and conditionally independent policies.

(iv) Using extreme point and lower semi-continuity arguments, we establish the existence of a symmetric optimal policy (which is privately randomized) for static and dynamic mean-field teams (see Theorem 4.5.1 and Theorem 4.7.1).

(v) In Section 4.8, using our analysis for mean-field problems, as N goes to infinity, we show that symmetric optimal policies of mean-field teams are asymptotically optimal for N-DM weakly coupled teams, hence, it establishes approximation results for this class of problems.

4.4 Topology on Control Policies and a de Finetti Representation Result

4.4.1 Topology on Control Policies

In this section, we introduce a topology using which, we can introduce Borel probability measures on policies. We first consider N-DM static team problems. Following from [138, 135], Assumption 4.1.2 allows us to reduce the problem as a static team problem where now the observation of each DM is independent of observations of other DMs and also independent of ω_0 (since under the measure transformation (4.4), a probability measure on the observation of each DM is Q^i, which is independent of observations of other DMs and ω_0). Hence, under Assumption 4.1.2, we can focus on each DM^i separately. Let us define

$$\Theta^i := \left\{ P \in \mathcal{P}(\mathbb{U} \times \mathbb{Y}) \middle| P(B) = \int_B 1_{\{g^i(y^i) \in du^i\}} Q^i(dy^i),\ g : \mathbb{Y} \to \mathbb{U},\ B \in \mathcal{B}(\mathbb{U} \times \mathbb{Y}) \right\}, \quad (4.9)$$

where $\mathcal{P}(\cdot)$ denotes the space of probability measures, and $1_{\{\cdot \in A\}}$ denotes the indicator function of the set A. The above set is the set of extreme points of the set of probability measures on $(\mathbb{U}\times\mathbb{Y})$ with fixed marginals Q^i on $\mathbb{Y}$, that is,

$$\mathcal{R}^i := \left\{ P \in \mathcal{P}(\mathbb{U}\times\mathbb{Y}) \middle| P(B) = \int_B \Pi^i(du^i|y^i)Q^i(dy^i),\ B \in \mathcal{B}(\mathbb{U}\times\mathbb{Y}) \right\}, \tag{4.10}$$

where Π^i is a stochastic kernel from $\mathbb{Y}$ to $\mathbb{U}$. Hence, it inherits Borel measurability and topological properties of that Borel measurable set [29]. We note that this set corresponds to Young measures [136] and this representation result is due to Borkar [29]. Now, we identify the set of relaxed policies Γ^i by $\mathcal{R}^i$ and we define convergence on policies as $\gamma^i_n \to \gamma^i$ if and only if $\gamma^i_n(du^i|y^i)Q^i(dy^i) \to \gamma^i(du^i|y^i)Q^i(dy^i)$ (in the weak convergence topology) as $n \to \infty$.

In view of the above standard Borel space formulation for Γ^i for each $i \in \mathcal{N}$, we can define the set of Borel probability measures on admissible policies Γ_N (which is referred to as a set of randomized policies) as $L^N := \mathcal{P}(\Gamma_N)$, where Borel σ-field $\mathcal{B}(\Gamma^i)$ is induced by the topology defined above. Define the set of randomized policies induced by a common and individual randomness as:

$$L^N_{\text{CO}} := \left\{ P_\pi \in L^N \middle| \text{for all } A_i \in \mathcal{B}(\Gamma^i) : \right.$$
$$\left. P_\pi(\gamma^1 \in A_1, \ldots, \gamma^N \in A_N) = \int_{z\in[0,1]} \prod_{i=1}^N P^i_\pi(\gamma^i \in A_i|z)\eta(dz),\ \eta \in \mathcal{P}([0,1]) \right\},$$

where η is the distribution of common, but independent (from intrinsic exogenous system variables), randomness, and for every fixed z, $P^i_\pi \in \mathcal{P}(\Gamma^i)$ indicates an identical independent randomized policy of each DM^i ($i = 1, \ldots, N$). Note that conditioned on a $[0,1]$-valued random variable Z, policies are independent. It can be shown that L^N_{CO} and L^N are identical (see Theorem 4.9.1 in the Appendix), and hence, the set of randomized policies L^N corresponds to randomized policies induced by an individual and a common randomness. Since individual and a common randomness do not improve the optimal expected cost, the relaxation of the problem to sets of randomized policies L^N is a legitimate relaxation for the team problems with N-DMs.

Now, we define the set of exchangeable randomized policies as:

$$L^N_{\text{EX}} := \Big\{ P_\pi \in L^N \Big| \text{for all } A_i \in \mathcal{B}(\Gamma^i) \text{ and for all } \sigma \in S_N : \\ P_\pi(\gamma^1 \in A_1, \dots, \gamma^N \in A_N) = P_\pi(\gamma^{\sigma(1)} \in A_1, \dots, \gamma^{\sigma(N)} \in A_N) \Big\}, \tag{4.11}$$

where S_N is the set of permutations of $\{1, \dots, N\}$. We note that L^N_{EX} is a convex subset of L^N. We also define the set $L^N_{\text{CO,SYM}}$ as the set of identical randomized policies induced by a common and individual randomness:

$$L^N_{\text{CO,SYM}} := \Big\{ P_\pi \in L^N \Big| \text{for all } A_i \in \mathcal{B}(\Gamma^i) : \\ P_\pi(\gamma^1 \in A_1, \dots, \gamma^N \in A_N) = \int_{z\in[0,1]} \prod_{i=1}^N \tilde{P}_\pi(\gamma^i \in A_i|z)\eta(dz),\ \eta \in \mathcal{P}([0,1]) \Big\},$$

where for all $i \in \mathcal{N}$, and fixed z, $\tilde{P}_\pi \in \mathcal{P}(\Gamma^i)$ indicates an identical independent randomized policy of each DM^i ($i = 1, \dots, N$). Also, define the set of randomized policies with only private independent randomness as:

$$L^N_{\text{PR}} := \Big\{ P_\pi \in L^N \Big| \text{for all } A_i \in \mathcal{B}(\Gamma^i) : \\ P_\pi(\gamma^1 \in A_1, \dots, \gamma^N \in A_N) = \prod_{i=1}^N P^i_\pi(\gamma^i \in A_i),\ \text{for } P^i_\pi \in \mathcal{P}(\Gamma^i) \Big\}. \tag{4.12}$$

Finally, define the set of randomized policies with identical and independent randomness:

$$L^N_{\text{PR,SYM}} := \Big\{ P_\pi \in L^N \Big| \text{for all } A_i \in \mathcal{B}(\Gamma^i) : \\ P_\pi(\gamma^1 \in A_1, \dots, \gamma^N \in A_N) = \prod_{i=1}^N \tilde{P}_\pi(\gamma^i \in A_i),\ \text{for } \tilde{P}_\pi \in \mathcal{P}(\Gamma^i) \Big\}. \tag{4.13}$$

For a team with a countably infinite number of decision makers, we define sets of randomized policies $L, L_{\text{EX}}, L_{\text{CO}}, L_{\text{CO,SYM}}, L_{\text{PR}}, L_{\text{PR,SYM}}$ similarly using Ionescu Tulcea extension theorem through the sequential formulation, by iteratively adding new coordinates for our probability measure (see e.g., [2, 64]). We define the set of randomized policies L on the infinite product Borel spaces $\Gamma = \prod_{i\in\mathbb{N}} \Gamma^i$ as $L := \mathcal{P}(\Gamma)$. Now, we define the set of infinitely exchangeable randomized policies as:

$$L_{\text{EX}} := \Bigg\{ P_\pi \in L \Bigg| \text{for all } A_i \in \mathcal{B}(\Gamma^i) \text{ and for all } N \in \mathbb{N}, \text{ and for all } \sigma \in S_N\text{:} \\ P_\pi(\gamma^1 \in A_1, \dots, \gamma^N \in A_N) = P_\pi(\gamma^{\sigma(1)} \in A_1, \dots, \gamma^{\sigma(N)} \in A_N) \Bigg\},$$

and we define

$$L_{\text{CO}} := \Bigg\{ P_\pi \in L \Bigg| \text{for all } A_i \in \mathcal{B}(\Gamma^i) : \\ P_\pi(\gamma^1 \in A_1, \gamma^2 \in A_2, \dots) = \int_{z\in[0,1]} \prod_{i\in\mathbb{N}} P_\pi^i(\gamma^i \in A_i|z)\eta(dz),\ \eta \in \mathcal{P}([0,1]) \Bigg\}.$$

Note that L_{CO} is a convex subset of L and its extreme points are in the set of randomized policies with private independent randomness:

$$L_{\text{PR}} := \Bigg\{ P_\pi \in L \Bigg| \text{for all } A_i \in \mathcal{B}(\Gamma^i) : \\ P_\pi(\gamma^1 \in A_1, \gamma^2 \in A_2, \dots) = \prod_{i\in\mathbb{N}} P_\pi^i(\gamma^i \in A_i),\ \text{for } P_\pi^i \in \mathcal{P}(\Gamma^i) \Bigg\}. \tag{4.14}$$

Also, we define

$$L_{\text{CO,SYM}} := \Bigg\{ P_\pi \in L \Bigg| \text{for all } A_i \in \mathcal{B}(\Gamma^i) : \\ P_\pi(\gamma^1 \in A_1, \gamma^2 \in A_2, \dots) = \int_{z\in[0,1]} \prod_{i\in\mathbb{N}} \tilde{P}_\pi(\gamma^i \in A_i|z)\eta(dz),\ \eta \in \mathcal{P}([0,1]) \Bigg\},$$

and

$$L_{\text{PR,SYM}} := \Big\{ P_\pi \in L \Big| \text{for all } A_i \in \mathcal{B}(\Gamma^i) : \\ P_\pi(\gamma^1 \in A_1, \gamma^2 \in A_2, \dots) = \prod_{i\in\mathbb{N}} \tilde{P}_\pi(\gamma^i \in A_i), \text{ for } \tilde{P}_\pi \in \mathcal{P}(\Gamma^i) \Big\}.$$

4.4.2 A de Finetti theorem for admissible team policies.

In view of the introduced topology and sets of Borel probability measures on policies (sets of randomized polices), we now establish a connection between L_{EX} and $L_{\text{CO,SYM}}$ using the classical de Finetti's theorem; that is, infinitely-exchangeable randomized policies are a mixture of i.i.d. randomized policies.

Theorem 4.4.1. *Any infinitely-exchangeable randomized policy $P_\pi \in L_{EX}$ is in the set of randomized policies $L_{CO,SYM}$ ($P_\pi \in L_{CO,SYM}$), i.e., for any $P_\pi \in L_{EX}$, there exists a $[0,1]$-valued random variable Z such that for any $A_i \in \mathcal{B}(\Gamma^i)$*

$$P_\pi(\gamma^1 \in A_1, \gamma^2 \in A_2, \dots) = \int_{z\in[0,1]} \prod_{i\in\mathbb{N}} \tilde{P}_\pi(\gamma^i \in A_i|z)\eta(dz),\ \eta \in \mathcal{P}([0,1]), \tag{4.15}$$

where for every fixed z, $\tilde{P}_\pi \in \mathcal{P}(\Gamma^i)$.

Proof. In view of the introduced weak convergence topology on Γ^i (using Borel measurable sets (4.10) and (4.9)), we have Γ^i is a closed subset of the Borel space $\mathcal{P}(\mathbb{U}\times\mathbb{Y})$, and hence, Γ^i is Borel for $i \in \mathbb{N}$. The proof follows from [76, Theorem 1.1] since $\Gamma = \prod_{i=1}^\infty \Gamma^i$ is Borel. We note that the de Finetti representation in [76, Theorem 1.1] is of the form $P_\pi(\gamma^1 \in A_1, \gamma^2 \in A_2, \dots) = \int_{\mathcal{P}(\Gamma^i)} \prod_{i=1}^\infty m(A^i)\hat{\eta}(dm)$ for $\hat{\eta} \in \mathcal{P}(\mathcal{P}(\Gamma^i))$ which can be written as in (4.15). That is because, $\mathcal{P}(\Gamma^i)$ is an (uncountable) Borel space [25, Corollary 7.25.1], and hence, by Borel-isomorphism Theorem (see for example, [25, Proposition 7.16]), it is Borel isomorphic to $[0,1]$. □

4.5 Existence and Structure of Optimal Policies for Symmetric Static Team Problems with Infinitely Many Decision Makers.

In this section, we consider static stochastic team problems where we impose Assumption 4.1.1 and Assumption 4.1.2. We again note that for our results in this section, we impose Assumption 4.1.1 and Assumption 4.1.2. Based on the definitions of randomized policies, we redefine the expected cost in $(\mathcal{P}_N)$ of a randomized policy $P_\pi \in L^N$ as:

$$\begin{aligned} J_N^\pi(\underline{\gamma}_N) &:= \int P_\pi(d\underline{\gamma})\mu^N(d\omega_0, d\underline{y})c^N(\underline{\gamma}, \underline{y}, \omega_0) \\ &:= \int \Big(\int c(\omega_0, u^1, \ldots, u^N) \prod_{k=1}^N \gamma^k(du^k|y^k) \Big) P_\pi(d\gamma^1, \ldots, d\gamma^N)\mu^N(d\omega_0, dy^1, \ldots, dy^N), \end{aligned} \tag{4.16}$$

where $c^N(\underline{\gamma}, \underline{y}, \omega_0) := \int c(\omega_0, u^1, \ldots, u^N) \prod_{k=1}^N \gamma^k(du^k|y^k)$, and μ^N is the joint probability measure on measurements $(y^1, \ldots, y^N)$ and ω_0. In the following, we characterize team problems in which the search for a randomized optimal policy can be restricted to policies in L^N_{EX} without losing global optimality.

Assumption 4.5.1. *Let observations of DMs, $(y^1, \cdots, y^N)$, be exchangeable conditioned on ω_0.*

Note that Assumption 4.5.1 is weaker than Assumption 4.1.3(i).

Lemma 4.5.1. *For a fixed N, consider an N-DM static team. Assume $\bar{L}^N$ is an arbitrary convex subset of L^N. If Assumption 4.5.1 holds, then*

$$\inf_{P_\pi \in \bar{L}^N} \int P_\pi(d\underline{\gamma})\mu^N(d\omega_0, d\underline{y})c^N(\underline{\gamma}, \underline{y}, \omega_0) = \inf_{P_\pi \in \bar{L}^N \cap L^N_{EX}} \int P_\pi(d\underline{\gamma})\mu^N(d\omega_0, d\underline{y})c^N(\underline{\gamma}, \underline{y}, \omega_0). \tag{4.17}$$

In the following, we present an existence result on globally optimal policies for static mean-field teams with infinitely many decision makers. First, we re-state the infinite decision maker mean-field team problem and its pre-limit.

Problem ($\mathcal{P}_N$): Consider an N-DM static team with the expected cost of a randomized policy $P_\pi^N \in L^N$ as:

$$\int P_\pi^N(d\underline{\gamma})\mu^N(d\omega_0, d\underline{y})c^N(\underline{\gamma}, \underline{y}, \omega_0) := \int \Big(\int \frac{1}{N}\sum_{i=1}^{N} c(\omega_0, u^i, \frac{1}{N}\sum_{p=1}^{N} u^p) \prod_{k=1}^{N} \gamma^k(du^k|y^k) \Big) \times P_\pi^N(d\gamma^1, \dots, d\gamma^N)\mu^N(d\omega_0, dy^1, \dots, dy^N). \quad (4.18)$$

The above problem is considered as a pre-limit problem for our infinite-decision maker team problem. This problem is a special case of $(\mathcal{P}_N)$ defined in the previous section since we have a special structure for the cost function c^N which satisfies Assumption 4.1.1.

Problem ($\mathcal{P}_\infty$): Consider infinite-DM static team with the following expected cost of a randomized policy $P_\pi \in L$ as

$$\limsup_{N\to\infty} \int P_{\pi,N}(d\underline{\gamma})\mu^N(d\omega_0, d\underline{y})c^N(\underline{\gamma}, \underline{y}, \omega_0), \quad (4.19)$$

where $P_{\pi,N}$ is the marginal of the $P_\pi \in L$ to the first N components and μ^N is the marginal of the fixed probability measure on $(\omega_0, y^1, y^2, \dots)$ to the first $N+1$ components.

In the following, we present a key result required for our main theorem. Under mild conditions, we show that the optimal expected cost function induced by L_{EX}^N and L_{EX} are equal as N goes to infinity. Hence, by Lemma 4.5.1, under symmetry, this allows us to show that without loss of global optimality, optimal policies of static mean-field teams with countably infinite number of DMs can be considered to be an infinitely-exchangeable type.

Lemma 4.5.2. *Suppose that Assumption 4.1.3 and Assumption 4.1.4 hold. Assume further that the cost function is bounded. Then*

$$\begin{aligned} &\limsup_{N\to\infty} \inf_{P_\pi^N \in L_{EX}^N} \int P_\pi^N(d\underline{\gamma})\mu^N(d\omega_0, d\underline{y})c^N(\underline{\gamma}, \underline{y}, \omega_0) \\ &= \limsup_{N\to\infty} \inf_{P_\pi \in L_{EX}} \int P_{\pi,N}(d\underline{\gamma})\mu^N(d\omega_0, d\underline{y})c^N(\underline{\gamma}, \underline{y}, \omega_0), \end{aligned} \quad (4.20)$$

where $P_{\pi,N}$ is the marginal of the $P_\pi \in L_{EX}$ to the first N components.

In the following, we establish an existence of a randomized optimal policy for ($\mathcal{P}_\infty$).

Theorem 4.5.1. *Consider a static team problem ($\mathcal{P}_\infty$) where Assumption 4.1.3 and Assumption 4.1.4 hold. Then, there exists a randomized optimal policy P^*_π for ($\mathcal{P}_\infty$) which is in $L_{PR,SYM}$:*

$$\inf_{P_\pi \in L_{PR,SYM}} \limsup_{N\to\infty} \int P_{\pi,N}(d\underline{\gamma})\mu^N(d\omega_0, d\underline{y})c^N(\underline{\gamma},\underline{y},\omega_0) := \limsup_{N\to\infty} \int P^*_{\pi,N}(d\underline{\gamma})\mu^N(d\omega_0, d\underline{y})c^N(\underline{\gamma},\underline{y},\omega_0)$$
$$= \inf_{P_\pi \in L_{PR}} \limsup_{N\to\infty} \int P_{\pi,N}(d\underline{\gamma})\mu^N(d\omega_0, d\underline{y})c^N(\underline{\gamma},\underline{y},\omega_0).$$

Here, we present an example where Theorem 4.5.1 can be applied but the existence result of Theorem the:ch2existence in Chapter 2 cannot be applied because the assumption that $\mathbb{U}^i$ for each DM is convex in Theorem 2.3.6 is violated.

Example 4.5.1. *Consider a team problem with the following expected cost function*

$$J(\underline{\gamma}) = \limsup_{N\to\infty} E^{\underline{\gamma}}\left[\left((\frac{1}{N}\sum_{i=1}^{N} u^i) - \frac{1}{2}\right)^2\right],$$

where σ-field $\sigma(y^i) = \{\emptyset, \Omega\}$ (this corresponds to a team setup where DMs have no measurement, hence measurable functions (policies) are constant functions), and we consider $u^i \in \{0,1\}$ for each DM. Clearly, an optimal policy that achieves zero is the one where a matching partition (such as even numbers vs. odd numbers) among DMs picking $u^i = 0$ and $u^i = 1$, that is because the cost function is non-negative. One can see that there is an optimal policy in $L_{PR,SYM}$ since each DM can choose independently an action zero or one with probability half and this achieves the expected cost of zero; however, there is no identically deterministic policy that achieves zero expected cost. We note also that the problem is not a convex problem, therefore the results in Theorem 2.3.6 and Proposition 2.3.1 are not applicable to show the existence of a symmetric randomized optimal policy, in particular, the action sets are not convex.

4.6 Finite Horizon Dynamic Team Problems with a Symmetric Information Structure

In this section, we study dynamic stochastic team problems. Similar to the static case, we first introduce the intrinsic model for general dynamic team problems, and then, we define a topology on control policies, and finally we establish our main results for dynamic problems.

4.6.1 A Revised intrinsic model for dynamic team problems

In this subsection, we recall the general (multi-stage) dynamic problems using the intrinsic model under deterministic policies. In the next subsections, we allow randomization equipped with a suitable topology.

According to the discussion above, by considering a collection of DMs as a single DM ($i = 1, \dots, N$) acting at different time instances ($t = 0, \dots, T-1$), we revise the intrinsic mode for (multi-stage) dynamic team problems with (NT)-DMs as a team with N-DMs (for $N \in \mathbb{N} \cup \{\infty\}$):

(i) Let the observation and action spaces be standard Borel spaces and be identical for each DM ($i = 1, \dots, N$) with $\mathbb{Y}_i := \mathbf{Y} = \prod_{t=0}^{T-1} \mathbb{Y}^t$, $\mathbb{U}_i := \mathbf{U} = \prod_{t=0}^{T-1} \mathbb{U}^t$, respectively (later on, for simplicity of our notation and analysis, we assume that action and observation spaces are also identical through time). For each DM^i, the set of all admissible policies are denoted by $\Gamma_i := \prod_{t=0}^{T-1} \Gamma^t$. Later on, these policies will be allowed to be randomized and accordingly the image will be $\mathcal{P}(\mathbf{U})$.

(ii) For $i = 1, \dots, N$, $y_t^i := h_t^i(x_0^{1:N}, \zeta_{0:t}^{1:N}, u_{0:t-1}^{1:N})$ represents the observation of DM^i at time t (h_t^is are Borel measurable functions). Let ν_t^N be a stochastic kernel characterizing the joint distribution of observations $y_t^{1:N} := (y_t^1, \dots, y_t^N)$ at time t induced by h_t^is given the available information, and let $(\underline{\zeta}^{1:N}) := (\underline{\zeta}^1, \dots, \underline{\zeta}^N)$ where $\underline{\zeta}^i := (x_0^i, \zeta_{0:T-1}^i)$ denotes all the uncertainty associated with DM^i including his/her initial states. We assume that $(\underline{\zeta}^i)$ takes values in Ω_ζ (where at each time instances t, it takes value in Ω_{ζ_t}). Let μ^N denote the law of $\underline{\zeta}^{1:N}$. To be consistent with our notations in our analysis of the static case, we used the same notation μ^N as the fixed probability measures on observations and ω_0 for the static

case; however, we note that in the dynamic case, the probability measures on uncertainties $(\underline{\zeta}^{1:N})$ is fixed and not probability measures on observations.

4.6.2 Topology on dynamic control policies.

Similar to Section 4.4.1, here, we allow randomization in policies, but first we introduce two reduction conditions (independent and nested reduction) that enable us to define sets of Borel probability measures on randomized policies for dynamic teams with different information structures by considering a policy of a single DM ($i = 1, \ldots, N$) acting at different time instances ($t = 0, \ldots, T-1$).

Assumption 4.6.1. *One of the following conditions holds:*

(i) (Independent reduction): for every $N \in \mathbb{N} \cup \{\infty\}$ and for $i = 1, \ldots, N$ and $t = 0, \ldots, T-1$, there exists a probability measure τ_t^i on $\mathbb{Y}^t$ and a function $\psi_t^i : \mathbb{Y}^t \times \Omega_0 \times \prod_{p=1}^N (\prod_{k=0}^{t-1} \Omega_{\zeta_k} \times \prod_{k=0}^{t-1} (\mathbb{U}^k \times \mathbb{Y}^k)) \to \mathbb{R}_+$ such that for all Borel sets A^i on $\mathbb{Y}^t$ (with $A = A^1 \times \ldots \times A^N$)

$$\nu_t^N(A|\omega_0, x_0^{1:N}, \zeta_{0:t-1}^{1:N}, y_{0:t-1}^{1:N}, u_{0:t-1}^{1:N}) = \prod_{i=1}^N \int_{A^i} \psi_t^i(y_t^i, \omega_0, x_0^{1:N}, \zeta_{0:t-1}^{1:N}, y_{0:t-1}^{1:N}, u_{0:t-1}^{1:N}) \tau_t^i(dy_t^i).$$

(ii) (Nested reduction): for every $N \in \mathbb{N} \cup \{\infty\}$ and for $i = 1, \ldots, N$ and $t = 0, \ldots, T-1$, there exists a probability measure η_t^i on $\mathbb{Y}^t$ and a function ϕ_t^i such that for all Borel sets A^i on $\mathbb{Y}^t$ (with $A = A^1 \times \ldots \times A^N$)

$$\begin{aligned}&\nu_t^N(A|\omega_0, x_0^{1:N}, \zeta_{0:t-1}^{1:N}, y_{0:t-1}^{1:N}, u_{0:t-1}^{1:N}) \\ &= \prod_{i=1}^N \int_{A^i} \phi_t^i(y_t^i, \omega_0, x_0^{-i}, \zeta_{0:t-1}^{-i}, y_{0:t-1}^{-i}, u_{0:t-1}^{-i}) \eta_t^i(dy_t^i | x_0^i, \zeta_{0:t-1}^i, y_{0:t-1}^i, u_{0:t-1}^i),\end{aligned}$$

and for each DM^i through time ($t = 0, \ldots, T-1$), there exists a static reduction with the classical information structure (i.e., under the reduction, the information structure of each DM through time is expanding such that $\sigma(y_t^i) \subset \sigma(y_{t+1}^i)$ for $t = 0, \ldots, T-1$).

We note that Assumption 4.6.1(i) allows us to obtain an independent measurements reduction

both through DMs and through time, $t = 0, \dots, T-1$ (see Appendix 4.9.5). Assumption 4.6.1(ii) holds if an independent static reduction exists through DMs and there exists a nested static reduction for each DM through time, i.e., under the reduction, the information is expanding for each DM through time (see Appendix 4.9.5). In view of the above reduction conditions, we introduce a suitable topology for randomized policies. Similar to Section 4.4.1, under Assumption 4.6.1(i), we define convergence on policies as:

$$\underline{\gamma}^i_n \xrightarrow{n\to\infty} \underline{\gamma}^i \text{ if and only if } \gamma^i_{t,n}(du^i_t|y^i_t)\tau^i_t(dy^i_t) \xrightarrow[\text{weakly}]{n\to\infty} \gamma^i_t(du^i_t|y^i_t)\tau^i_t(dy^i_t)\ \forall\, t = 0, \dots, T-1.$$

Under Assumption 4.6.1(ii), we define convergence on policies as:

$$\underline{\gamma}^i_n \xrightarrow{n\to\infty} \underline{\gamma}^i \text{ if and only if } \gamma^i_{t,n}(du^i_t|y^i_{0:t})\eta^i_t(dy^i_{0:t}) \xrightarrow[\text{weakly}]{n\to\infty} \gamma^i_t(du^i_t|y^i_{0:t})\eta^i_t(dy^i_{0:t})\ \forall\, t = 0, \dots, T-1.$$

Hence, under Assumption 4.6.1, we define all the sets of randomized policies defined in Section 4.4.1 for the dynamic teams by considering $\underline{\gamma}^i$.

Now, we provide examples under which either one of the conditions in Assumption 4.6.1 holds.

Example 4.6.1. *For each $i = 1, \dots, N$ and $t = 0, \dots, T-1$, let $x^i_{t+1} = f^i_t(x^{1:N}_{0:t}, u^{1:N}_{0:t}, w^i_t)$ and $y^i_t = h^i_t(\omega_0, x^{1:N}_{0:t}, \zeta^{1:N}_{0:t-1}, u^{1:N}_{0:t-1}) + v^i_t$, where $\zeta^i_t := (w^i_t, v^i_t)$, and v^i_t admits zero-mean Gaussian density function θ^i_t with positive-definite covariance.*

(i) If the information structure for each DM at time t is described as $I^i_t := \{y^i_t\}$ for all $i = 1, \dots, N$ and $t = 0, \dots, T-1$, then Assumption 4.6.1(i) holds.

(ii) If $I^i_t := \{y^i_{0:t}, u^i_{0:t-1}\}$ for all $i = 1, \dots, N$ and $t = 0, \dots, T-1$ (or equivalently, $I^i_t := \{\tilde{y}^i_t\}$ with $\tilde{y}^i_t := \tilde{h}^i_t(\omega_0, x^{1:N}_{0:t}, \zeta^{1:N}_{0:t-1}, u^{1:N}_{0:t-1}, v^i_{0:t})$ for some function $\tilde{h}^i_t$ and $\sigma(\tilde{y}^i_t) \subset \sigma(\tilde{y}^i_{t+1})$ and $\sigma(u^i_t) \subset \sigma(\tilde{y}^i_{t+1})$ for some function $\bar{h}_t$), then Assumption 4.6.1(ii) holds.

Part(i) is true since for all $t = 0, \dots, T-1$ and $i = 1, \dots, N$, we have

$$y^i_t = h^i_t(\omega_0, x^{1:N}_{0:t}, \zeta^{1:N}_{0:t-1}, u^{1:N}_{0:t-1}) + v^i_t = \kappa^i_t(\omega_0, x^{1:N}_0, \zeta^{1:N}_{0:t-1}, u^{1:N}_{0:t-1}) + v^i_t,$$

for some functions κ_t^i, and hence, we can define

$$\psi_t^i(y_t^i, \omega_0, x_0^{1:N}, \zeta_{0:t-1}^{1:N}, y_{0:t-1}^{1:N}, u_{0:t-1}^{1:N}) := \frac{\theta_t^i(y_t^i - \kappa_t^i(\omega_0, x_0^{1:N}, \zeta_{0:t-1}^{1:N}, u_{0:t-1}^{1:N}))}{\theta_t^i(y_t^i)},\ \tau_t^i(dy_t^i) := \theta_t^i(y_t^i)dy_t^i.$$

Part(ii) can be shown similarly by first applying the independent reduction as above, and then, considering the nested information structure through time for each DM.

Example 4.6.2. *Consider the following two information structures:*

(i) (Open-loop information structure): For each $i = 1, \dots, N$ and $t = 0, \dots, T-1$, let $x_{t+1}^i = f_t^i(x_{0:t}^{1:N}, u_{0:t}^{1:N}, w_t^i)$ and $y_t^i = h_t^i(\zeta_{0:t-1}^i, v_t^i)$ such that $\sigma(y_t^i) \subset \sigma(y_{t+1}^i)$, where $(\zeta_t^i)_t := (w_t^i, v_t^i)_t$ denotes the disturbances of DM^i (which is independent of disturbances of other DMs and independent of ω_0). If $I_t^i := \{y_t^i\}$ for all $i = 1, \dots, N$ and $t = 0, \dots, T-1$, then Assumption 4.6.1(ii) holds.

(ii) For each $i = 1, \dots, N$ and $t = 0, \dots, T-1$, let $x_{t+1}^i = f_t^i(\omega_0, x_{0:t}^{1:N}, u_{0:t}^{1:N}) + w_t^i$, where w_t^i admits zero-mean Gaussian density function θ_t^i with positive-definite covariance, and let $y_t^i = h_t^i(x_{0:t}^i, y_{0:t-1}^i, v_{0:t}^i)$ such that $\sigma(y_t^i) \subset \sigma(y_{t+1}^i)$, where $(v_t^i)_t$ are independent of disturbances of other DMs and independent of ω_0. If $I_t^i := \{y_t^i\}$ for all $i = 1, \dots, N$ and $t = 0, \dots, T-1$, then Assumption 4.6.1(ii) holds.

Part(i) follows from the fact that the information structure is open-loop and nested for each DM, and hence, under this information structure the problem is static with the classical information structure through time for each DM. Part(ii) is true since for all $t = 0, \dots, T-1$ and $i = 1, \dots, N$,

$$\hat{\phi}_t^i(x_t^i, \omega_0, x_{0:t-1}^{1:N}, u_{0:t-1}^{1:N}) := \frac{\theta_t^i(x_t^i - f_t^i(\omega_0, x_{0:t-1}^{1:N}, u_{0:t-1}^{1:N}))}{\theta_t^i(x_t^i)},\ \hat{\eta}_t^i(dx_t^i) := \theta_t^i(x_t^i)dx_t^i,$$

and since the information structure is nested through time for each DM.

4.7 Existence and Structure of Optimal Policies for Symmetric Dynamic Team Problems with Infinitely Many Decision Makers.

In the following, we study the existence and structure of globally optimal policies for dynamic team problems with a symmetric information structure (that are not necessarily partially nested) and with a finite but large and also infinitely many decision makers. In Chapter 3, convex mean-field team problems have been considered under the assumption that the action space is convex for each DM and the cost function is convex in policies. We note that even if the cost function is convex in actions when there is a mean-field coupling in dynamics, convexity rarely holds since the information structure under decentralized setup is non-classical, and that may lead to the non-convexity of the team problem in policies (see for example [140, Section 3.3]). In the following, convexity is not imposed. Again, for our results in this subsection, we impose Assumption 4.6.1.

4.7.1 Exchangeability of optimal policies for symmetric dynamic team problems with a finite but large number of decision makers.

We first focus on symmetric dynamic team problems with N-DMs, and we establish a structural result for optimal policies of this class of problems (which is more general than the pre-limit mean-field model ($\mathcal{P}_T^N$)). In the next subsection, we use this result to establish existence and structural properties of globally randomized optimal policies for mean-field dynamic team problems.

Now, we focus on dynamic teams with a symmetric information structure (see Definition 3.2). We note that pre-limit mean-field and mean-field dynamic team problems ($\mathcal{P}_T^N$) and ($\mathcal{P}_T^\infty$) introduced in Section 5.2 have a symmetric information structure. Before, we present the result for dynamic mean-field teams, we characterize team problems with symmetric information structures in which the search for an optimal policy can be restricted to policies in L_{EX}^N without losing global optimality. To this end, we focus on a more general setup of team problems within randomized

policies $P_\pi \in L^N$ as

$$\begin{aligned} J_N^\pi(\underline{\gamma}^{1:N}) &:= \int P_\pi(d\underline{\gamma})\mu^N(d\omega_0, d\underline{\zeta})c^N(\underline{\zeta}, \underline{\gamma}, \underline{y}, \omega_0)\nu^N(d\underline{y}|\underline{\zeta}, \underline{\gamma}, \omega_0) \\ &:= \int \left(\int c(\omega_0, \underline{\zeta}^{1:N}, \underline{u}^1, \dots, \underline{u}^N) \prod_{i=1}^N \underline{\gamma}^i(d\underline{u}^i|\underline{y}^i) \right) P_\pi(d\underline{\gamma}^1, \dots, d\underline{\gamma}^N)\mu^N(d\omega_0, d\underline{\zeta}^{1:N}) \\ &\quad \times \prod_{t=0}^{T-1} \nu_t^N \left(dy_t^{1:N} \middle| \omega_0, x_0^{1:N}, \zeta_{0:t-1}^{1:N}, y_{0:t-1}^{1:N}, u_{0:t-1}^{1:N} \right), \end{aligned} \tag{4.21}$$

where $c^N(\underline{\zeta}, \underline{\gamma}, \underline{y}, \omega_0) := \int c(\omega_0, \underline{\zeta}^{1:N}, \underline{u}^1, \dots, \underline{u}^N) \prod_{i=1}^N \underline{\gamma}^i(d\underline{u}^i|\underline{y}^i)$ and the following assumptions hold.

Assumption 4.7.1. *For any permutation σ of the set $\{1, \dots, N\}$, we have for all ω_0,*

$$c(\omega_0, (\underline{\zeta}^\sigma)^{1:N}, (\underline{u}^\sigma)^{1:N}) = c(\omega_0, \underline{\zeta}^{1:N}, \underline{u}^{1:N}), \tag{4.22}$$

where $(\underline{\zeta}^\sigma)^{1:N} := (\underline{\zeta}^{\sigma(1)}, \dots, \underline{\zeta}^{\sigma(N)})$ and $(\underline{u}^\sigma)^{1:N} := (\underline{u}^{\sigma(1)}, \dots, \underline{u}^{\sigma(N)})$.

Assumption 4.7.2.

(a) $(\underline{\zeta}^1, \dots, \underline{\zeta}^N)$ are exchangeable conditioned on ω_0;

(b) For all $t = 0, \dots, T-1$, and all Borel sets A^i on $\mathbb{Y}^t$ (with $A = A^1 \times \dots \times A^N$)

$$\nu_t^N \left(A \middle| \omega_0, x_0^{1:N}, \zeta_{0:t-1}^{1:N}, y_{0:t-1}^{1:N}, u_{0:t-1}^{1:N} \right) = \prod_{i=1}^N \nu_t^i \left(A^i \middle| \omega_0, x_0^i, \zeta_{0:t-1}^i, y_{0:t-1}^{1:N}, u_{0:t-1}^{1:N} \right),$$

where ν_t^i is a stochastic kernel of the observation DM^i at time t, y_t^i, induced by h_t (which is identical for each DM).

We note that dynamic mean-field team problems introduced in Section 4.1.1 with the cost function (4.7), dynamic (4.5), and observations (4.6), under Assumption 4.1.6 satisfy Assumption 4.7.1 and Assumption 4.7.2.

Lemma 4.7.1. *Consider a dynamic team problem with a symmetric information structure. Let Assumption 4.7.1 and Assumption 4.7.2 hold. Assume $\bar{L}^N$ is an arbitrary convex subset of L^N. Then*

$$\inf_{P_\pi \in \bar{L}^N} \int P_\pi(d\underline{\gamma})\mu^N(d\omega_0, d\underline{\zeta})c^N(\underline{\zeta}, \underline{\gamma}, \underline{y}, \omega_0)\nu^N(d\underline{y}|\underline{\zeta}, \underline{\gamma}, \omega_0)$$
$$= \inf_{P_\pi \in \bar{L}^N \cap L^N_{EX}} \int P_\pi(d\underline{\gamma})\mu^N(d\omega_0, d\underline{\zeta})c^N(\underline{\zeta}, \underline{\gamma}, \underline{y}, \omega_0)\nu^N(d\underline{y}|\underline{\zeta}, \underline{\gamma}, \omega_0).$$

4.7.2 Existence and structure of optimal policies for mean-field dynamic team problems

In the following, we establish the existence of a globally randomized optimal policy for dynamic mean-field teams with infinitely many decision makers. Define state dynamics and observations as (4.5) and (4.6). The information structure of DM^i at time t is $I^i_t = \{y^i_t\}$, and $\zeta^i_t := (w^i_t, v^i_t)$ (with $\zeta^i_0 := (x^i_0, w^i_0, v^i_0)$) denotes the uncertainty corresponding to dynamics and observations at time t for DM^i which are exogenous random vectors in standard Borel spaces. First, we re-state the infinite decision maker mean-field team problem and its pre-limit within randomized policies.

Problem ($\mathcal{P}^N_T$): Consider an N-DM dynamic team with the expected cost of a randomized policy $P^N_\pi \in L^N$ as:

$$\int P^N_\pi(d\underline{\gamma})\mu^N(d\omega_0, d\underline{\zeta})c^N(\underline{\zeta}, \underline{\gamma}, \underline{y}, \omega_0)\nu^N(d\underline{y}|\underline{\zeta}, \underline{\gamma}, \omega_0)$$
$$:= \int \Bigg(\int \frac{1}{N} \sum_{t=0}^{T-1} \sum_{i=1}^{N} c\Bigg(\omega_0, x^i_t, u^i_t, \frac{1}{N}\sum_{p=1}^{N} u^p_t(y^p_t), \frac{1}{N}\sum_{p=1}^{N} x^p_t\Bigg) \prod_{k=1}^{N} \underline{\gamma}^k(d\underline{u}^k|\underline{y}^k)\Bigg)$$
$$\times P^N_\pi(d\underline{\gamma}^1, \ldots, d\underline{\gamma}^N)\mu^N(d\omega_0, d\underline{\zeta}^{1:N}) \prod_{t=0}^{T-1} \nu^N_t\left(dy^{1:N}_t \middle| \omega_0, x^{1:N}_0, \zeta^{1:N}_{0:t-1}, y^{1:N}_{0:t-1}, u^{1:N}_{0:t-1}\right), \tag{4.23}$$

where

$$c^N(\underline{\zeta}, \underline{\gamma}, \underline{y}, \omega_0) := \int \frac{1}{N} \sum_{t=0}^{T-1} \sum_{i=1}^{N} c\Bigg(\omega_0, x^i_t, u^i_t, \frac{1}{N}\sum_{p=1}^{N} u^p_t(y^p_t), \frac{1}{N}\sum_{p=1}^{N} x^p_t\Bigg) \prod_{k=1}^{N} \underline{\gamma}^k(d\underline{u}^k|\underline{y}^k).$$

The above problem is considered as a pre-limit problem for our infinite-decision maker team

problem. We note that N-DM teams of $(\mathcal{P}_T^N)$ is a special case of (4.21) since we have a special structure for the cost function c^N and observations which satisfy Assumption 4.7.1 and symmetry, respectively.

Remark 4.7.1. *Our analysis below also allows a more general observations for each DM where the observations of each DM at time t can be explicitly functions of average of previous states and actions as*

$$y_t^i = h_t\bigg(x_{0:t}^i, u_{0:t-1}^i, \frac{1}{N}\sum_{p=1}^N x_{0:t-1}^p, \frac{1}{N}\sum_{p=1}^N u_{0:t-1}^p, v_{0:t}^i\bigg).$$

However, to simplify the presentations of theorems and proofs and emphasize in the decentralization of optimal policy, for the rest of this chapter, we consider (4.6)*.*

Problem $(\mathcal{P}_T^\infty)$: Consider infinite-DM static team with the following expected cost of a randomized policy $P_\pi \in L$ as:

$$\limsup_{N\to\infty} \int P_{\pi,N}(d\underline{\gamma})\mu^N(d\omega_0, d\underline{\zeta})c^N(\underline{\zeta}, \underline{\gamma}, \underline{y}, \omega_0)\nu^N(d\underline{y}|\underline{\zeta}, \underline{\gamma}, \omega_0) \tag{4.24}$$

where $P_{\pi,N}$ is the restriction of $P_\pi \in L$ to its first N components and μ^N is the marginal of the fixed probability measure on $(\omega_0, \underline{\zeta}^1, \underline{\zeta}^2, \dots)$ to the first $N+1$ components.

Assumption 4.7.3. *Assume Assumption 4.6.1 holds with functions ψ_t^i and ϕ_t^i are of the following forms for every $i \in \mathcal{N}$ and $t = 0, \dots, T-1$:*

$$\psi_t^i\bigg(y_t^i, \omega_0, x_0^i, \zeta_{0:t-1}^i, y_{0:t-1}^i, u_{0:t-1}^i, \frac{1}{N}\sum_{p=1}^N u_{0:t-1}^p, \frac{1}{N}\sum_{p=1}^N x_{0:t}^p\bigg),$$
$$\phi_t^i\bigg(y_t^i, \omega_0, \frac{1}{N}\sum_{p=1}^N u_{0:t-1}^p, \frac{1}{N}\sum_{p=1}^N x_{0:t}^p\bigg),$$

where ψ_t^i is continuous in the last three arguments (actions and the empirical mean of actions and states) and ϕ_t^i is continuous in the last two arguments (the empirical mean of actions and states).

Before presenting our main result for dynamic mean-field teams, we present sufficient conditions under which the expected cost function induced by randomized optimal policies in L_{EX}^N and L_{EX} are equal as N goes to infinity, and hence, following from Lemma 4.7.1, under symmetry, this shows that without loss of global optimality, optimal policies of dynamic mean-field teams can be considered to be an infinitely-exchangeable type.

Lemma 4.7.2. *Consider the team problem $(\mathcal{P}_T^N)$ where Assumption 4.7.3, Assumption 4.1.5, and Assumption 4.1.6 hold. Assume further that the cost function bounded, then*

$$\begin{aligned}&\limsup_{N\to\infty}\inf_{P_\pi^N\in L_{EX}^N}\int P_\pi^N(d\underline{\gamma})\mu^N(d\omega_0,d\underline{\zeta})c^N(\underline{\zeta},\underline{\gamma},\underline{y},\omega_0)\nu^N(d\underline{y}|\underline{\zeta},\underline{\gamma},\omega_0)\\&=\limsup_{N\to\infty}\inf_{P_\pi\in L_{EX}}\int P_{\pi,N}(d\underline{\gamma})\mu^N(d\omega_0,d\underline{\zeta})c^N(\underline{\zeta},\underline{\gamma},\underline{y},\omega_0)\nu^N(d\underline{y}|\underline{\zeta},\underline{\gamma},\omega_0),\end{aligned}\tag{4.25}$$

where $P_{\pi,N}$ is the restriction of $P_\pi\in L_{EX}$ to its first N components and μ^N is the marginal of the fixed probability measure on $(\omega_0,\underline{\zeta}^1,\underline{\zeta}^2,\dots)$ to the first $N+1$ components.

In the following, we establish an existence and structural result for a randomized optimal policy of $(\mathcal{P}_T^\infty)$.

Theorem 4.7.1. *Consider a mean-field team problem $(\mathcal{P}_T^\infty)$ with $(\mathcal{P}_T^N)$ having a symmetric information structure for every N. Let Assumption 4.7.3, Assumption 4.1.5, and Assumption 4.1.6 hold. Then, there exists a randomized optimal policy P_π^* for (P_T^∞) which is in $L_{PR,SYM}$,*

$$\begin{aligned}&\inf_{P_\pi\in L_{PR,SYM}}\limsup_{N\to\infty}\int P_{\pi,N}(d\underline{\gamma})\mu^N(d\omega_0,d\underline{\zeta})c^N(\underline{\zeta},\underline{\gamma},\underline{y},\omega_0)\nu^N(d\underline{y}|\underline{\zeta},\underline{\gamma},\omega_0)\\&:=\limsup_{N\to\infty}\int P^*_{\pi,N}(d\underline{\gamma})\mu^N(d\omega_0,d\underline{\zeta})c^N(\underline{\zeta},\underline{\gamma},\underline{y},\omega_0)\nu^N(d\underline{y}|\underline{\zeta},\underline{\gamma},\omega_0)\\&=\inf_{P_\pi\in L}\limsup_{N\to\infty}\int P_{\pi,N}(d\underline{\gamma})\mu^N(d\omega_0,d\underline{\zeta})c^N(\underline{\zeta},\underline{\gamma},\underline{y},\omega_0)\nu^N(d\underline{y}|\underline{\zeta},\underline{\gamma},\omega_0).\end{aligned}$$

4.8 Approximations of Optimal Policies for Symmetric N-DM Stochastic Team Problems

In this section, we present approximation results of optimal policies for N-DM team problems. We show that for large N, symmetric policies are nearly optimal and the restriction of the optimal infinite solution to the finite team problem is nearly optimal for large N. We first consider the static case. To present ideas more effectively, we first introduce the following set of probability measures on policies as:

$$L_{\mathrm{D}}^N := \left\{ P_\pi \in L^N \middle| \text{for all } A_i \in \mathcal{B}(\Gamma^i) : P_\pi(\gamma^1 \in A_1, \ldots, \gamma^N \in A_N) = \prod_{i=1}^N 1_{\{\tilde{\gamma}^i \in A_i\}}, \text{ for } \tilde{\gamma}^i \in \Gamma^i \right\},$$

where the above set corresponds to Dirac-delta measures in L_{PR}^N.

Theorem 4.8.1. *Consider a static team problem ($\mathcal{P}_N$) (see* (4.18)*) where Assumption 4.1.4 and Assumption 4.1.3 hold. Assume further the cost function is bounded. Then,*

(i)

$$\inf_{P_\pi^N \in L_{PR,SYM}^N} \int P_\pi^N(d\underline{\gamma}) \mu^N(d\omega_0, d\underline{y}) c^N(\underline{\gamma}, \underline{y}, \omega_0) \leq \inf_{P_\pi^N \in L_{CO}^N} \int P_\pi^N(d\underline{\gamma}) \mu^N(d\omega_0, d\underline{y}) c^N(\underline{\gamma}, \underline{y}, \omega_0) + \epsilon_N, \tag{4.26}$$

and

$$\inf_{P_\pi^N \in L_{PR,SYM}^N} \int P_\pi^N(d\underline{\gamma}) \mu^N(d\omega_0, d\underline{y}) c^N(\underline{\gamma}, \underline{y}, \omega_0) \leq \inf_{P_\pi^N \in L_{D}^N} \int P_\pi^N(d\underline{\gamma}) \mu^N(d\omega_0, d\underline{y}) c^N(\underline{\gamma}, \underline{y}, \omega_0) + \epsilon_N, \tag{4.27}$$

where $\epsilon_N \to 0$ as N goes to infinity.

(ii) If $P_\pi^* \in L_{PR,SYM}$ is a randomized optimal policy of ($\mathcal{P}_\infty$), then there exist $\bar{\epsilon}_N \geq 0$ where

$\bar{\epsilon}_N \to 0$ *as* N *goes to infinity and*

$$\int P^*_{\pi,N}(d\underline{\gamma})\mu^N(d\omega_0, d\underline{y})c^N(\underline{\gamma}, \underline{y}, \omega_0) \leq \inf_{P^N_\pi \in L^N_D} \int P^N_\pi(d\underline{\gamma})\mu^N(d\omega_0, d\underline{y})c^N(\underline{\gamma}, \underline{y}, \omega_0) + \epsilon_N + \bar{\epsilon}_N, \tag{4.28}$$

where $P^*_{\pi,N}$ *is the restriction of* P^*_π *to the first* N *components.*

The main idea for establishing Part(i) is to use Lemma 4.5.1 and Lemma 4.5.2 to provide an approximation of optimal expected cost by restricting the search for randomized policies to those that are restrictions of randomized policies in L_{EX} to the N first components. We note that since the set of policies L^N_{D} is not a convex subset of the set of randomized policies L^N, (4.26) does not immediately imply (4.27) using Lemma 4.5.1 but the result can be established using an extreme point argument and since policies in L^N_{D} are optimal among all randomized policies L^N_{PR} for N-DM teams thanks to Blackwell's irrelevant information theorem [28]. Part(ii) follows from Part(i) and Theorem 4.5.1, using the fact that a randomized optimal policy $P^*_\pi \in L_{\text{PR,SYM}}$ of $(\mathcal{P}_\infty)$ provides an approximation for the optimal expected cost when the search for randomized optimal policy for N-DM teams is restricted to those in $L^N_{\text{PR,SYM}}$.

Similarly, we present approximation results of optimal policies for symmetric dynamic N-DM team problems.

Theorem 4.8.2. *Consider a dynamic team problem* $(\mathcal{P}^N_T)$ *(see* (4.7)*). Let Assumption 4.7.3, Assumption 4.1.5, and Assumption 4.1.6 hold. If the cost function is bounded, then*

(i)

$$\begin{aligned}\inf_{P^N_\pi \in L^N_{PR,\ SYM}} &\int P^N_\pi(d\underline{\gamma})\mu^N(d\omega_0, d\underline{\zeta})c^N(\underline{\zeta}, \underline{\gamma}, \underline{y}, \omega_0)\nu^N(d\underline{y}|\underline{\zeta}, \underline{\gamma}, \omega_0) \\ &\leq \inf_{P^N_\pi \in L^N_{CO}} \int P^N_\pi(d\underline{\gamma})\mu^N(d\omega_0, d\underline{\zeta})c^N(\underline{\zeta}, \underline{\gamma}, \underline{y}, \omega_0)\nu^N(d\underline{y}|\underline{\zeta}, \underline{\gamma}, \omega_0) + \epsilon_N, \end{aligned} \tag{4.29}$$

and

$$\inf_{P_\pi^N \in L^N_{PR,\ SYM}} \int P_\pi^N(d\underline{\gamma})\mu^N(d\omega_0, d\underline{\zeta})c^N(\underline{\zeta},\underline{\gamma},\underline{y},\omega_0)\nu^N(d\underline{y}|\underline{\zeta},\underline{\gamma},\omega_0)$$
$$\leq \inf_{P_\pi^N \in L_D^N} \int P_\pi^N(d\underline{\gamma})\mu^N(d\omega_0, d\underline{\zeta})c^N(\underline{\zeta},\underline{\gamma},\underline{y},\omega_0)\nu^N(d\underline{y}|\underline{\zeta},\underline{\gamma},\omega_0) + \epsilon_N, \quad (4.30)$$

where $\epsilon_N \to 0$ as N goes to infinity.

(ii) If $P_\pi^* \in L_{PR,SYM}$ is a randomized optimal policy of ($\mathcal{P}_T^\infty$), then there exist $\bar{\epsilon}_N \geq 0$ where $\bar{\epsilon}_N \to 0$ as N goes to infinity and

$$\int P^*_{\pi,N}(d\underline{\gamma})\mu^N(d\omega_0, d\underline{\zeta})c^N(\underline{\zeta},\underline{\gamma},\underline{y},\omega_0)\nu^N(d\underline{y}|\underline{\zeta},\underline{\gamma},\omega_0)$$
$$\leq \inf_{P_\pi^N \in L_D^N} \int P_\pi^N(d\underline{\gamma})\mu^N(d\omega_0, d\underline{\zeta})c^N(\underline{\zeta},\underline{\gamma},\underline{y},\omega_0)\nu^N(d\underline{y}|\underline{\zeta},\underline{\gamma},\omega_0) + \epsilon_N + \bar{\epsilon}_N,$$

where $P^*_{\pi,N}$ is the restriction of P_π^* to the first N components.

Proof. Proof follows from a similar steps as the proof of Theorem 4.8.1 using the results of Lemma 4.7.2 and Theorem 4.7.1. □

4.9 Proofs

4.9.1 Connection between L^N_{CO} and L^N in Section 4.4.1.

In following theorem, we show that sets of randomized policies L^N_{CO} and L^N are identical.

Theorem 4.9.1. *The set of randomized policies L^N is identical to the set of randomized policies L^N_{CO}.*

Proof. Clearly, we have $L^N_{\text{CO}} \subseteq L^N$. In the following, we show $L^N \subseteq L^N_{\text{CO}}$. Following from [29], for each $i = 1, \ldots, N$, the set of marginals of randomized policies in L^N to each coordinate Γ^i is a convex combination of its extreme points which is in the set of Delta-dirac measures of elements

in Γ^i. Hence, for the set of extreme points of the convex set L^N (denoted by Extreme(L^N)), we have

$$\text{Extreme}(L^N) \subseteq \Bigg\{ P_\pi \in L^N \Bigg| \text{for all } A_i \in \mathcal{B}(\Gamma^i) : \\ P_\pi(\gamma^1 \in A_1, \ldots, \gamma^N \in A_N) = \prod_{i=1}^{N} 1_{\{\tilde{\gamma}^i \in A_i\}}, \text{ for } \tilde{\gamma}^i \in \Gamma^i \Bigg\}$$

Hence, Extreme$(L^N) \subset L^N_{\text{CO}}$ and since both sets L^N and L^N_{CO} are convex, we have $L^N \subseteq L^N_{\text{CO}}$, and this completes the proof. □

4.9.2 Proof of Lemma 4.5.1.

For any permutation $\sigma \in S_N$, we define a randomized policy $P^\sigma_\pi \in \bar{L}^N$ as a permutation, σ, of arguments of a randomized policy $P_\pi \in \bar{L}^N$, i.e., for $A^i \in \mathcal{B}(\Gamma^i)$

$$P^\sigma_\pi(\gamma^1 \in A^1, \ldots, \gamma^2 \in A^N) := P_\pi(\gamma^{\sigma(1)} \in A^1, \ldots, \gamma^{\sigma(N)} \in A^N).$$

We have

$$\begin{aligned}
&\int P^\sigma_\pi(d\underline{\gamma}) \mu^N(d\omega_0, d\underline{y}) c^N(\underline{\gamma}, \underline{y}, \omega_0) \\
&= \int c(\omega_0, u^1, \ldots, u^N) \prod_{k=1}^{N} \gamma^k(du^k|y^k) \tilde{\mu}^N(dy^1, \ldots, dy^N|\omega_0) P^\sigma_\pi(d\gamma^1, \ldots, d\gamma^N) \mathbb{P}_0(d\omega_0) \\
&= \int c(\omega_0, u^1, \ldots, u^N)) \prod_{k=1}^{N} \gamma^k(du^k|y^k) \tilde{\mu}^N(dy^1, \ldots, dy^N|\omega_0) P_\pi(d\gamma^{\sigma(1)}, \ldots, d\gamma^{\sigma(N)}) \mathbb{P}_0(d\omega_0) \qquad (4.31) \\
&= \int c(\omega_0, u^{\sigma(1)}, \ldots, u^{\sigma(N)})) \prod_{k=1}^{N} \gamma^{\sigma(k)}(du^{\sigma(k)}|y^{\sigma(k)}) \tilde{\mu}^N(dy^{\sigma(1)}, \ldots, dy^{\sigma(N)}|\omega_0) \\
&\times \quad P_\pi(d\gamma^1, \ldots, d\gamma^N) \mathbb{P}_0(d\omega_0) \qquad (4.32) \\
&= \int c(\omega_0, u^1, \ldots, u^N) \prod_{k=1}^{N} \gamma^k(du^k|y^k) \tilde{\mu}^N(dy^1, \ldots, dy^N|\omega_0) P_\pi(d\gamma^1, \ldots, d\gamma^N) \mathbb{P}_0(d\omega_0) \qquad (4.33)
\end{aligned}$$

$$= \int P_\pi(d\underline{\gamma})\mu^N(d\omega_0, d\underline{y})c^N(\underline{\gamma}, \underline{y}, \omega_0),$$

where $\tilde{\mu}^N$ is the joint conditional distribution of observations $(y^1, \ldots, y^N)$ given ω_0, and (4.31) follows from the definition of P_π^σ and (4.32) follows from relabeling $u^{\sigma(i)}, y^{\sigma(i)}$ with u^i, y^i for all $i = 1, \ldots, N$. Equality (4.33) follows from Assumption 4.1.1 and Assumption 4.5.1.

Let $\epsilon \geq 0$, and consider a randomized policy $P^*_{\pi,\epsilon} \in \bar{L}^N$ such that

$$\int P^*_{\pi,\epsilon}(d\underline{\gamma})\mu^N(d\omega_0, d\underline{y})c^N(\underline{\gamma}, \underline{y}, \omega_0) \leq \inf_{P_\pi \in \bar{L}^N} \int P_\pi(d\underline{\gamma})\mu^N(d\omega_0, d\underline{y})c^N(\underline{\gamma}, \underline{y}, \omega_0) + \epsilon.$$

Consider $\tilde{P}_{\pi,\epsilon}$ as a convex combination of all possible permutations of $P^*_{\pi,\epsilon}$ by averaging them. Since $\bar{L}^N$ is convex, we have $\tilde{P}_{\pi,\epsilon} \in \bar{L}^N$. Also, we have $\tilde{P}_{\pi,\epsilon} \in L^N_{\text{EX}}$, and for any permutation $\sigma \in S_N$, we have

$$\begin{aligned}\tilde{P}_{\pi,\epsilon}(d\gamma^1, \ldots, d\gamma^N) &:= \sum_{\sigma \in S_N} \frac{1}{|S_N|} P^{*,\sigma}_{\pi,\epsilon}(d\gamma^1, \ldots, d\gamma^N) \\ &= \tilde{P}^\sigma_{\pi,\epsilon}(d\gamma^1, \ldots, d\gamma^N),\end{aligned}$$

where $|S_N|$ denotes the cardinality of the set S_N, and the second equality follows from the fact that the sum is over all permutation σ by taking average of them. Therefore, a randomized policy $\tilde{P}_{\pi,\epsilon}$ is in $\bar{L}^N \cap L^N_{\text{EX}}$. We have,

$$\begin{aligned}\int \tilde{P}_{\pi,\epsilon}(d\underline{\gamma})\mu^N(d\omega_0, d\underline{y})c^N(\underline{\gamma}, \underline{y}, \omega_0) &:= \int (\sum_{\sigma \in S_N} \alpha_\sigma P^{*,\sigma}_{\pi,\epsilon})(d\underline{\gamma})\mu^N(d\omega_0, d\underline{y})c^N(\underline{\gamma}, \underline{y}, \omega_0) \\ &= \sum_{\sigma \in S_N} \alpha_\sigma \int P^{*,\sigma}_{\pi,\epsilon}(d\underline{\gamma})\mu^N(d\omega_0, d\underline{y})c^N(\underline{\gamma}, \underline{y}, \omega_0) \\ &= \sum_{\sigma \in S_N} \alpha_\sigma \int P^*_{\pi,\epsilon}(d\underline{\gamma})\mu^N(d\omega_0, d\underline{y})c^N(\underline{\gamma}, \underline{y}, \omega_0) \\ &\leq \inf_{P_\pi \in \bar{L}^N} \int P_\pi(d\underline{\gamma})\mu^N(d\omega_0, d\underline{y})c^N(\underline{\gamma}, \underline{y}, \omega_0) + \epsilon,\end{aligned}$$

where the second equality is true since the map $P_\pi \to \int P_\pi(d\underline{\gamma})\mu^N(d\omega_0, d\underline{y})c^N(\underline{\gamma}, \underline{y}, \omega_0)$ is linear

and the third equality follows from (4.33). Since $\tilde{P}_{N,\epsilon} \in \bar{L}^N \cap L^N_{\text{EX}}$, we have

$$\int \tilde{P}_{\pi,\epsilon}(d\underline{\gamma})\mu^N(d\omega_0, d\underline{y})c^N(\underline{\gamma}, \underline{y}, \omega_0) \geq \inf_{P_\pi \in \bar{L}^N \cap L^N_{\text{EX}}} \int P_\pi(d\underline{\gamma})\mu^N(d\omega_0, d\underline{y})c^N(\underline{\gamma}, \underline{y}, \omega_0).$$

Hence, for any $\epsilon \geq 0$, we have

$$\inf_{P_\pi \in \bar{L}^N \cap L^N_{\text{EX}}} \int P_\pi(d\underline{\gamma})\mu^N(d\omega_0, d\underline{y})c^N(\underline{\gamma}, \underline{y}, \omega_0) \leq \inf_{P_\pi \in \bar{L}^N} \int P_\pi(d\underline{\gamma})\mu^N(d\omega_0, d\underline{y})c^N(\underline{\gamma}, \underline{y}, \omega_0) + \epsilon.$$

Since ϵ is arbitrary, this completes the proof.

4.9.3 Proof of Lemma 4.5.2

To prove Lemma 4.5.2, we use two following results by Diaconis and Friedman [49, Theorem 13] and Aldous [1, Proposition 7.20] (see also [75] for more general results) which we recall for reader's convenience:

Theorem 4.9.2. *[49, Theorem 13] Let $Y = (Y_1, \dots, Y_n)$ be an n-exchangeable and $Z = (Z_1, Z_2, \dots)$ be an infinitely-exchangeable sequence of random variables with $\mathcal{L}(Z_1, \dots, Z_k) = \mathcal{L}(Y_{I_1}, \dots, Y_{I_k})$ for all $k \geq 1$ where the indices $(I_1, I_2, \dots)$ are i.i.d. random variables with the uniform distribution on the set $\{1, \dots, n\}$. Then, for all $m = 1, \dots, n$,*

$$\Big|\Big|\mathcal{L}(Y_1, \dots, Y_m) - \mathcal{L}(Z_1, \dots, Z_m)\Big|\Big|_{TV} \leq \frac{m(m-1)}{2n}, \tag{4.34}$$

where $\mathcal{L}(\cdot)$ denotes the law of random variables and $||\cdot||_{TV}$ is the total variation norm.

Theorem 4.9.3. *[1, Proposition 7.20] Let $X := (X_1, X_2, \dots)$ be an infinitely-exchangeable sequence of random variables taking values in a Polish space $\mathbb{X}$ and directed by a random measure α (i.e., α is a $\mathcal{P}(\mathbb{X})$-valued random variable and $Pr(X \in A) = \int_{\mathcal{P}(\mathbb{X})} \prod_{i=1}^{\infty} \xi(A^i) Pr(\alpha \in d\xi)$ where $A^i \in \mathcal{B}(\mathbb{X})$ and $(A = A^1 \times A^2 \times \dots)$, see [1, Definition 2.6]). Suppose that either for each*

n

(1) $X^{(n)} = (X_1^{(n)}, X_2^{(n)}, \dots)$ *is infinitely-exchangeable directed by* α_n*, or*

(2) $X^{(n)} = (X_1^{(n)}, \dots, X_n^{(n)})$ *is* n*-exchangeable with empirical measure* α_n*.*

Then, $X^{(n)}$ *converges in distribution to* X ($X^{(n)} \xrightarrow[n\to\infty]{d} X$) *if and only if* $\alpha_n \xrightarrow[n\to\infty]{d} \alpha$*.*

We note that by convergence in distribution to an infinite exchangeable sequence, we mean the following: $X^{(n)} \xrightarrow[n\to\infty]{\text{d}} X$ if and only if $(X_1^{(n)}, \dots, X_m^{(n)}) \xrightarrow[n\to\infty]{\text{d}} (X_1, \dots, X_m)$ for each $m \geq 1$ [1, page 55].

Using the above theorems, we now complete the Proof of Lemma 4.5.2. Since the action space $\mathbb{U}$ is compact and observations are i.i.d. with a fixed marginal (under Assumption 4.1.2, via a change of measure argument observations can be viewed to be independent of ω_0), the set of probability measures L^N is tight. Furthermore, by [138, Theorem 5.1], L^N is closed under the topology of weak convergence and hence L^N is compact. Using the argument in [138, Theorem 5.1] under Assumption 4.1.4, the expected cost function is lower semicontinuous in policies $P_N \in L^N$. Hence, there exists an optimal policy for ($\mathcal{P}_N$), and by Lemma 4.5.1, this optimal policy can be assumed to be in L^N_{EX}. Consider a sequence of N-exchangeable randomized policies $\{P_\pi^{*,N}\}_N$, where for every $N \geq 1$, $P_\pi^{*,N} \in L^N_{\text{EX}}$ and

$$\int P_\pi^{*,N}(d\underline{\gamma})\mu^N(d\omega_0, d\underline{y})c^N(\underline{\gamma}, \underline{y}, \omega_0) = \inf_{P_\pi^N \in L^N_{\text{EX}}} \int P_\pi^N(d\underline{\gamma})\mu^N(d\omega_0, d\underline{y})c^N(\underline{\gamma}, \underline{y}, \omega_0). \tag{4.35}$$

In the following, we show (4.20) in two steps. In the first step, for every N, we use the construction in Theorem 4.9.2 to construct an infinitely-exchangeable randomized policy $P_{\pi,N}^{*,\infty} \in L_{\text{EX}}$ using the N-exchangeable randomized policy $P_\pi^{*,N} \in L^N_{\text{EX}}$ by considering the indices as a sequence of i.i.d. random variables with uniform distribution on the set $\{1, \dots, N\}$, and then, we show that there exists a weakly convergent subsequence of joint measures on the first coordinate, observations, and the average of induced actions of randomized policies $P_{\pi,N}^{*,\infty} \in L_{\text{EX}}$. Then, we show that the expected cost function induced by the N-exchangeable randomized policy $P_\pi^{*,N} \in L^N_{\text{EX}}$ converges through a subsequence to a limit induced by an infinitely-exchangeable randomized policy $P_{\pi,N}^{*,\infty}$.

(Step 1): Let $(I_1, I_2, \dots)$ be i.i.d. random variables with uniform distribution on the set $\{1, \dots, N\}$. For a fixed N and for any N-exchangeable randomized policy $P_\pi^{*,N} \in L_{\text{EX}}^N$, we construct an infinitely-exchangeable randomized policy $P_{\pi,N}^{*,\infty} \in L_{\text{EX}}$ as follows: for every N and m and for all $A^i \in \mathcal{B}(\Gamma^i)$

$$P_{\pi,N}^{*,\infty}(\gamma^1 \in A^1, \dots, \gamma^m \in A^m) := P_\pi^{*,N}(\gamma^{I_1} \in A^1, \dots, \gamma^{I_m} \in A^m).$$

where $P_{\pi,N}^{*,\infty}$ is the restriction of $P_{\pi,P_\pi^N}^{*,\infty} \in L_{\text{EX}}$ to the first N components. We note that $P_{\pi,N}^{*,\infty} \in L_{\text{EX}}$ because we use i.i.d. sequence $(I_1, I_2, \dots)$ for indexing probability measures on the space of policies, hence, for every fixed N and N-exchangeable randomized policy $P_\pi^{*,N}$, a randomized policy $P_{\pi,N}^{*,\infty}$ is i.i.d through DMs and hence it is infinitely-exchangeable.

Let $u_N^{*,i}$ be the control action induced by γ_N^i where random variables $(\gamma_N^1, \dots, \gamma_N^N)$ are determined by N-exchangeable randomized policy $P_\pi^{*,N} \in L_{\text{EX}}^N$. Let $u_{\infty,N}^{*,i}$ be the control action induced by $\gamma_{N,\infty}^i$ where random variables $(\gamma_{N,\infty}^1, \dots, \gamma_{N,\infty}^N)$ are determined by infinitely-exchangeable randomized policy $P_{\pi,N}^{*,\infty} \in L_{\text{EX}}$. Since under the reduction (Assumption 4.1.2), observations are i.i.d. and also independent of ω_0, following from Theorem 4.9.2, we have for every $m \geq 1$

$$\begin{aligned}
&\left\| \mathcal{L}(\gamma_N^1, \dots, \gamma_N^m, y^1, \dots, y^m) - \mathcal{L}(\gamma_{N,\infty}^1, \dots, \gamma_{N,\infty}^m, y^1, \dots, y^m) \right\| \\
&= \left\| \mathcal{L}(\gamma_N^1, \dots, \gamma_N^m) \prod_{i=1}^m \mathcal{L}(y^i) - \mathcal{L}(\gamma_{N,\infty}^1, \dots, \gamma_{N,\infty}^m) \prod_{i=1}^m \mathcal{L}(y^i) \right\|_{TV} \xrightarrow[N\to\infty]{} 0, \qquad (4.36)
\end{aligned}$$

where (4.36) follows from the fact that $(\gamma_N^1, \dots, \gamma_N^N)$ and $(\gamma_{N,\infty}^1, \dots, \gamma_{N,\infty}^N)$ are random variables with joint probability measures $P_\pi^{*,N} \in L_{\text{EX}}^N$ and $P_{\pi,N}^{*,\infty} \in L_{\text{EX}}|_N$, respectively.

Since $\mathbb{U}$ is compact, the marginal of probability measures on $\mathbb{U}$ is tight. Since the probability measure on $\mathbb{Y}$ is fixed, the marginal on $\mathbb{Y}$ is also tight. Since marginals are tight, then the collection of all measures on $(\mathbb{U} \times \mathbb{Y})$ with these tight marginals is also tight (see e.g., [137, Proof of Theorem 2.4]), and hence, the set Γ^i is tight for each $i \in \mathbb{N}$. Hence, $\{\mathcal{L}(\gamma_{\infty,N}^i)\}_N$ is tight for each DM and by exchangeablity $\mathcal{L}(\gamma_{\infty,N}^i) = \mathcal{L}(\gamma_{\infty,N}^1)$. Hence, we can find a subsequence such that $\mathcal{L}(\gamma_{\infty,l}^i) \xrightarrow[l\to\infty]{}$

$\mathcal{L}(\gamma_\infty^i)$ for all $i \in \mathbb{N}$. Since marginals of $\{\mathcal{L}(\gamma_{\infty,l}^1, \dots, \gamma_{\infty,l}^m)\}_l$ are tight, for each $m \geq 1$, there exists a further subsequence

$$\mathcal{L}(\gamma_{\infty,n}^1, \dots, \gamma_{\infty,n}^m) \xrightarrow[n\to\infty]{} \mathcal{L}(\gamma_\infty^1, \dots, \gamma_\infty^m),$$

where $(\gamma_\infty^1, \gamma_\infty^2, \dots)$ is infinitely-exchangeable and induced by an infinitely-exchangeable randomized policy $P_\pi^{*,\infty} \in L_{\text{EX}}$ since the set of infinitely-exchangeable randomized policies is closed under the weak-convergence topology, where by weak convergence for an infinite sequence, we mean weak convergence of finite restrictions. That is because, if $P_\pi^{\sigma,*,\infty}$ is the limit in the weak convergence topology of the sequence randomized policies $\{P_{\pi,n}^{\sigma,*,\infty}\}_n$ as $n \to \infty$, where for $A^i \in \mathcal{B}(\Gamma^i)$ and for all $N \in \mathbb{N}$ and all finite permutations $\sigma \in S_N$

$$P_{\pi,n}^{\sigma,*,\infty}(\gamma^1 \in A^1, \gamma^2 \in A^2, \dots) := P_{\pi,n}^{*,\infty}(\gamma^{\sigma(1)} \in A^1, \gamma^{\sigma(2)} \in A^2, \dots).$$

Then, following from exchangeability, since sequences $\{P_{\pi,n}^{*,\infty}\}_n$ and $\{P_{\pi,n}^{\sigma,*,\infty}\}_n$ are identical, the limit in the weak convergence topology of both randomized policies $P_\pi^{*,\infty}$ and $P_\pi^{\sigma,*,\infty}$ are also identical, and hence, the limit $P_\pi^{*,\infty}$ is infinitely-exchangeable, $P_\pi^{*,\infty} \in L_{\text{EX}}$. Hence, following from (4.36), for each $m \geq 1$

$$\mathcal{L}(\gamma_n^1, \dots, \gamma_n^m) \xrightarrow[n\to\infty]{} \mathcal{L}(\gamma_\infty^1, \dots, \gamma_\infty^m).$$

By construction of random variables $u_n^{*,i}$ and $u_\infty^{*,i}$ and since random variables γ_n^is are independent of y^is, we have for each $m \geq 1$

$$(u_n^{*,1}, \dots, u_n^{*,m}) \xrightarrow[n\to\infty]{\text{d}} (u_\infty^1, \dots, u_\infty^m),$$

where $(u_\infty^1, u_\infty^2, \dots)$ is induced by an infinitely-exchangeable policy $P_\pi^{*,\infty} \in L_{\text{EX}}$. Following from

Theorem 4.9.3, we have for all $A \in \mathcal{U}$ and $\mathbb{P}$-almost surely

$$F_n(A) := F_n^{\omega}(A) := \frac{1}{n}\sum_{i=1}^{n} \delta_{u_n^{*,i}(\omega)}(A) \xrightarrow[n\to\infty]{\text{d}} \alpha^{\omega}(A), \tag{4.37}$$

where ω denotes the sample path dependency and α is the directing measure of an infinitely-exchangeable random variables $(u_\infty^1, u_\infty^2, \dots)$ (that is $\alpha(\omega, A) = Pr(u_\infty^{*,i} \in A|H)$ almost surely for all $A \in \mathcal{U}$ where H is the σ-field generated by $\mathcal{P}(\mathbb{U})$-valued random variable α [1]). Following from (4.37), since the action space $\mathbb{U}$ is compact, we have $\mathbb{P}$-almost surely

$$\bar{F}_n := \bar{F}_n^{\omega} := \frac{1}{n}\sum_{i=1}^{n} u_n^{*,i}(\omega) = \int_{\mathbb{U}} u F_n(du) \xrightarrow[n\to\infty]{\text{d}} \bar{F} := \int_{\mathbb{U}} u\alpha^{\omega}(du). \tag{4.38}$$

Define $\tilde{P}^{*,n}$ as the joint probability measure of $(u_n^{*,1}, \bar{F}_n, \underline{y})$ where marginals on $\underline{y} := (y^1, y^2, \dots)$ are fixed to be $\prod_{i=1}^{\infty} Q(dy^i)$. Since marginals on $(u_n^{*,1}, \bar{F}_n)$ are tight and marginals on $\underline{y}$ are fixed, $\{\tilde{P}^{*,n}\}_n$ is tight. Hence, there exists a subsubsequence $\{\tilde{P}^{*,k}\}_k$ converges weakly to $\tilde{P}^*$ as k goes to infinity. This implies that marginals $\{\tilde{P}^{*,k}\}_k$ on $(u_k^{*,1}, \bar{F}_k)$ converges to the marginals of $\tilde{P}^*$ on $(u^{*,1}, \bar{F})$, hence, $\tilde{P}^*$ is induced by $(u_\infty^1, u_\infty^2, \dots)$ which is infinitely-exchangeable and is induced by an infinitely-exchangeable randomized policy in L_{EX}.

(Step 2): We have

$$\begin{aligned}
&\limsup_{N\to\infty} \int P_\pi^{*,N}(d\underline{\gamma})\mu^N(d\omega_0, d\underline{y})c^N(\underline{\gamma}, \underline{y}, \omega_0) \\
&= \limsup_{N\to\infty} \frac{1}{N}\sum_{i=1}^{N} \int c(\omega_0, u^i, \frac{1}{N}\sum_{p=1}^{N} u^p) \prod_{k=1}^{N} \gamma^k(du^k|y^k) P_\pi^{*,N}(d\gamma^1, \dots, d\gamma^N) \\
&\times \prod_{i=1}^{N} \hat{\mu}(dy^i|\omega_0)\mathbb{P}_0(d\omega_0) \\
&= \limsup_{N\to\infty} \frac{1}{N}\sum_{i=1}^{N} \int c(\omega_0, u^i, \frac{1}{N}\sum_{p=1}^{N} u^p) \prod_{k=1}^{N} \gamma^k(du^k|y^k) P_\pi^{*,N}(d\gamma^1, \dots, d\gamma^N)
\end{aligned} \tag{4.39}$$

$$\times \prod_{i=1}^{N} f(\omega_0, y^i) Q(dy^i) \mathbb{P}_0(d\omega_0)$$

$$= \limsup_{N\to\infty} \int \int_{\prod_{i=N+1}^{\infty} \mathbb{Y}} c(\omega_0, u^1, \bar{F}_N) \tilde{P}^{*,N}(du^1, d\bar{F}_N, d\underline{y}) \prod_{i=1}^{\infty} f(\omega_0, y^i) \mathbb{P}_0(d\omega_0) \tag{4.40}$$

$$\geq \lim_{k\to\infty} \int \int_{\prod_{i=k+1}^{\infty} \mathbb{Y}} c(\omega_0, u^1, \bar{F}_k) \tilde{P}^{*,k}(du^1, d\bar{F}_k, d\underline{y}) \prod_{i=1}^{\infty} f(\omega_0, y^i) \mathbb{P}_0(d\omega_0) \tag{4.41}$$

$$= \int c(\omega_0, u^1, \bar{F}) \tilde{P}^{*}(du^1, d\bar{F}, , d\underline{y}) \prod_{i=1}^{\infty} f(\omega_0, y^i) \mathbb{P}_0(d\omega_0) \tag{4.42}$$

$$\geq \limsup_{N\to\infty} \inf_{P_\pi \in L_{\text{EX}}} \int P_{\pi,N}(d\underline{\gamma}) \mu^N(d\omega_0, d\underline{y}) c^N(\underline{\gamma}, \underline{y}, \omega_0). \tag{4.43}$$

where $\hat{\mu}$ is the conditional distribution of each observation y^i given ω_0, and (4.39) follows from Assumption 4.1.3(i), hence, under Assumption 4.1.2, in the new (equivalent) expected cost function, observations are i.i.d. and independent of ω_0. (4.40) follows from integrating over the set $\prod_{i=N+1}^{\infty} \mathbb{Y}$ and since $(u_N^{*,1}, \dots, u_N^{*,N})$ is N-exchangeable. Inequality (4.41) follows from the assumption that the cost function is bounded and limsup is the greatest subsequence limit of a bounded sequence where k is the index of the subsequence considered in (Step 1). Equality (4.42) follows from the dominated convergence theorem and following from Assumption 4.1.4 and since by (Step 1) $\{\tilde{P}^{*,k}\}_k$ converges weakly to $\tilde{P}^*$ as k goes to infinity. Inequality (4.43) follows from the fact that $\tilde{P}^*$ is the joint measure with the first coordinate $(u_\infty^1, u_\infty^2, \dots)$ which is infinitely-exchangeable and it is induced by an infinitely-exchangeable randomized policy in L_{EX}. The above inequalities become equalities since the opposite direction is true as well (that is because $L_{\text{EX}}|_N \subset L_{\text{EX}}^N$) and this completes the proof.

4.9.4 Proof of Theorem 4.5.1

We complete the proof in four steps.

(Step 1): Similar to the proof of Lemma 4.5.2, using [138, Theorem 5.1], we can show that there exists a randomized optimal policy for $(\mathcal{P}_N)$ which belongs to the set L^N, and by Lemma

4.5.1, this randomized optimal policy can be assumed to be in the set of N-exchangeable randomized policies L^N_{EX}. Consider a sequence of N-exchangeable randomized policies $\{P^{*,N}_\pi\}_N$, where for every $N \geq 1$, $P^{*,N}_\pi \in L^N_{\text{EX}}$ and

$$\int P^{*,N}_\pi(d\underline{\gamma})\mu^N(d\omega_0, d\underline{y})c^N(\underline{\gamma}, \underline{y}, \omega_0) = \inf_{P^N_\pi \in L^N_{\text{EX}}} \int P^N_\pi(d\underline{\gamma})\mu^N(d\omega_0, d\underline{y})c^N(\underline{\gamma}, \underline{y}, \omega_0).$$

(Step 2): In this step, we show that to establish an existence result, it is sufficient to show the convergence of the expected cost induced by a randomized optimal policy in $L^N_{\text{PR,SYM}}$ of N-DM teams to the expected cost induced by a randomized policy $L_{\text{PR,SYM}}$ of mean-field teams through a subsequence as N goes to infinity. We first lift the space of randomized admissible policies, and we represent any admissible randomized policy as a probability measure in L (which is convex) and $L_{\text{EX}} \subset L$. We have

$$\begin{aligned}
&\inf_{P_\pi \in L} \limsup_{N\to\infty} \int P_{\pi,N}(d\underline{\gamma})\mu^N(d\omega_0, d\underline{y})c^N(\underline{\gamma}, \underline{y}, \omega_0) \\
&\geq \limsup_{N\to\infty} \inf_{P^N_\pi \in L^N} \int P^N_\pi(d\underline{\gamma})\mu^N(d\omega_0, d\underline{y})c^N(\underline{\gamma}, \underline{y}, \omega_0) && (4.44)\\
&= \limsup_{N\to\infty} \inf_{P^N_\pi \in L^N_{\text{EX}}} \int P^N_\pi(d\underline{\gamma})\mu^N(d\omega_0, d\underline{y})c^N(\underline{\gamma}, \underline{y}, \omega_0) && (4.45)\\
&\geq \lim_{M\to\infty} \limsup_{N\to\infty} \inf_{P^N_\pi \in L^N_{\text{EX}}} \int P^N_\pi(d\underline{\gamma})\mu^N(d\omega_0, d\underline{y}) \min\{M, c^N(\underline{\gamma}, \underline{y}, \omega_0)\} && (4.46)\\
&= \lim_{M\to\infty} \limsup_{N\to\infty} \inf_{P_\pi \in L_{\text{EX}}} \int P_{\pi,N}(d\underline{\gamma})\mu^N(d\omega_0, d\underline{y}) \min\{M, c^N(\underline{\gamma}, \underline{y}, \omega_0)\} && (4.47)\\
&= \lim_{M\to\infty} \limsup_{N\to\infty} \inf_{P^N_\pi \in L^N_{\text{CO,SYM}}} \int P^N_\pi(d\underline{\gamma})\mu^N(d\omega_0, d\underline{y}) \min\{M, c^N(\underline{\gamma}, \underline{y}, \omega_0)\} && (4.48)\\
&= \lim_{M\to\infty} \limsup_{N\to\infty} \inf_{P^N_\pi \in L^N_{\text{PR,SYM}}} \int P^N_\pi(d\underline{\gamma})\mu^N(d\omega_0, d\underline{y}) \min\{M, c^N(\underline{\gamma}, \underline{y}, \omega_0)\} && (4.49)\\
&\geq \inf_{P_\pi \in L_{\text{PR,SYM}}} \limsup_{N\to\infty} \int P_{\pi,N}(d\underline{\gamma})\mu^N(d\omega_0, d\underline{y})c^N(\underline{\gamma}, \underline{y}, \omega_0) && (4.50)\\
&\geq \inf_{P_\pi \in L_{\text{CO,SYM}}} \limsup_{N\to\infty} \int P_{\pi,N}(d\underline{\gamma})\mu^N(d\omega_0, d\underline{y})c^N(\underline{\gamma}, \underline{y}, \omega_0) && (4.51)\\
&\geq \inf_{P_\pi \in L} \limsup_{N\to\infty} \int P_{\pi,N}(d\underline{\gamma})\mu^N(d\omega_0, d\underline{y})c^N(\underline{\gamma}, \underline{y}, \omega_0), && (4.52)
\end{aligned}$$

where (4.44) follows from exchanging limsup with inf and the fact that the restriction to N-first coordinate $P_{\pi,N} \in L^N$ for any randomized policy $P_\pi \in L$, and (4.45) follows from Lemma 4.5.1. Inequality (4.46) follows from $\min\{M, c^N(\underline{\gamma}, \underline{y}, \omega_0)\} \leq c^N(\underline{\gamma}, \underline{y}, \omega_0)$. Equality (4.47) follows from Lemma 4.5.2 and (4.48) follows from Theorem 4.4.1. The set of extreme points of the convex set $L^N_{\text{CO,SYM}}$ is in $L^N_{\text{PR,SYM}}$ (that is because, $L^N_{\text{CO,SYM}}$ corresponds to the randomized policies with common and individual independent randomness where each DM choose an identical randomized policy), hence, (4.49) is true since $L^N_{\text{CO,SYM}}$ is convex, and the map $\int P^N_\pi(d\underline{\gamma})\mu^N(d\omega_0, d\underline{y})c^N(\underline{\gamma}, \underline{y}, \omega_0) : L^N_{\text{CO,SYM}} \to \mathbb{R}$ is linear. Inequalities (4.51) and (4.52) follow from the fact that $L_{\text{PR,SYM}} \subset L_{\text{CO,SYM}} \subset L$. Hence, by (4.52), this chain of inequalities must be chain of equalities.

In the next two steps, we justify (4.50) through showing that there exists a subsequence of policies induced by symmetric/identical private randomization whose weak-limit achieves (4.50). First, we establish compactness of the set of randomized policies $L^N_{\text{PR,SYM}}$, and then, we show a lower semicontinuity of the induced expected cost function justifying (4.50).

(Step 3): Consider the set of randomized policies $L^N_{\text{PR,SYM}}$. For each DM, we can equivalently represent any randomized policy as a probability measure on $(\mathbb{U} \times \mathbb{Y})$, where the marginal on observations is fixed. Since the team is static, this decouples the policy spaces from the policies of the previous decision makers. Following from symmetry, we can represent each DM's privately randomized policy space as $\{P \in \mathcal{P}(\mathbb{U} \times \mathbb{Y}) | P(B) = \int_B \Pi(du^i|y^i)Q(dy^i)\}$ where $B \in \mathcal{B}(\mathbb{U} \times \mathbb{Y})$ and Π is an identical randomized policy from the set of stochastic kernels from space of observations to space of actions for each DM. Since $\mathbb{U}$ is compact, the marginals on $\mathbb{U}$ are relatively compact. Since the marginals are relatively compact, the collection of all measures with these relatively compact marginals are also relatively compact (see e.g., [137, Proof of Theorem 2.4]), and hence, the randomized policy space is relatively compact. Following from symmetry, the set of individual randomized policies for each DM is closed under product topology where each coordinate converges in the weak convergence topology. Hence, (Step 3) implies that there exists a subsequence of (symmetric) individually randomized policies for each DM that converges weakly

to the limit which is identical for each DM. In (Step 4), we show that the limit randomized policy is optimal by showing a lower semicontinuity of the induced expected cost function.

(Step 4): Define the empirical measure on actions and observations as follows:

$$\Lambda_N(B) := \frac{1}{N}\sum_{i=1}^{N} \delta_{\beta_N^i}(B),$$

where for each N, $\beta_N^i := (u_N^{i,*}, y^i)$, $B \in \mathcal{Z} := (\mathbb{U}\times\mathbb{Y})$, $u_N^{i,*}$ is the action induced by the randomized policy Π_N^* in (Step 3).

Now, we have

$$\begin{aligned}
&\lim_{M\to\infty}\limsup_{N\to\infty}\inf_{P_\pi^N\in L^N_{\text{PR,SYM}}}\int P_\pi^N(d\underline{\gamma})\mu^N(d\omega_0,d\underline{y})\min\{M,c^N(\underline{\gamma},\underline{y},\omega_0)\}\\
&=\lim_{M\to\infty}\limsup_{N\to\infty}\int\left(\int\min\left\{M,c\left(\omega_0,u,\int_{\mathbb{U}}u\Lambda_N(du\times\mathbb{Y})\right)\right\}\Lambda_N(du,dy)\right) &(4.53)\\
&\times\prod_{i=1}^{\infty}P_N^{*,\omega_0}(du^i,dy^i)\mathbb{P}_0(d\omega_0)\\
&\geq\lim_{M\to\infty}\lim_{n\to\infty}\int\left(\int\min\left\{M,c\left(\omega_0,u,\int_{\mathbb{U}}u\Lambda_n(du\times\mathbb{Y})\right)\right\}\Lambda_n(du,dy)\right) &(4.54)\\
&\times\prod_{i=1}^{\infty}P_n^{*,\omega_0}(du^i,dy^i)\mathbb{P}_0(d\omega_0)\\
&=\lim_{M\to\infty}\int\lim_{n\to\infty}\int\left(\int\min\left\{M,c\left(\omega_0,u,\int_{\mathbb{U}}u\Lambda_n(du\times\mathbb{Y})\right)\right\}\Lambda_n(du,dy)\right) &(4.55)\\
&\times\prod_{i=1}^{\infty}P_n^{*,\omega_0}(du^i,dy^i)\mathbb{P}_0(d\omega_0)\\
&\geq\lim_{M\to\infty}\int\int\left(\int\min\left\{M,c\left(\omega_0,u,\int_{\mathbb{U}}u\Lambda(du\times\mathbb{Y})\right)\right\}\Lambda(du,dy)\right) &(4.56)\\
&\times\prod_{i=1}^{\infty}P^{*,\omega_0}(du^i,dy^i)\mathbb{P}_0(d\omega_0)\\
&=\int\left(\int c\left(\omega_0,u,\int_{\mathbb{U}}u\Lambda(du\times\mathbb{Y})\right)\Lambda(du,dy)\right)\prod_{i=1}^{\infty}P^{*,\omega_0}(du^i,dy^i)\mathbb{P}_0(d\omega_0) &(4.57)\\
&=\limsup_{N\to\infty}\int\frac{1}{N}\sum_{i=1}^{N}c\Big(\omega_0,u^i,\frac{1}{N}\sum_{p=1}^{N}u^p\Big)\prod_{i=1}^{N}P^{*,\omega_0}(du^i,dy^i)\mathbb{P}_0(d\omega_0) &(4.58)
\end{aligned}$$

$$\geq \inf_{P_\pi \in L_{\text{PR,SYM}}} \limsup_{N\to N} \int P_{\pi,N}(d\underline{\gamma})\mu^N(d\omega_0, d\underline{y})c^N(\underline{\gamma}, \underline{y}, \omega_0) \tag{4.59}$$

where $P_N^{*,\omega_0}(du^i, dy^i) := \Pi_N^*(du^i|dy^i)\hat{\mu}(dy^i|\omega_0) = \Pi_N^*(du^i|y^i)f(\omega_0, y^i)Q(dy^i)$. Equality (4.53) follows from the definition of the empirical measure, and Assumption 4.1.3(i), and follows from symmetry of optimal policies since every DM apply an identical policy, the set of policies can be extended to infinite product space and then we can consider the expected cost by integrating over $\prod_{i=N}^{\infty}(\mathbb{U}\times\mathbb{Y})$. Inequality (4.54) follows from the fact that limsup is the greatest convergent subsequence limit for a bounded sequence, where we denoted the convergent subsequence of coordinates of policies in (Step 3) with $n \in \mathbb{I} \subset \mathbb{N}$. Equality (4.55) follows from the law of total expectation, and the dominated convergence theorem.

Fix the convergent subsequence n, following from symmetry and Assumption 4.1.2 and Assumption 4.1.3(i), we have $\beta_n^i = (u_n^{*,i}, y^i)$ are i.i.d. Now, using a similar argument as the proof of Theorem 3.3.1, through choosing a suitable subsubsequence and using the strong law of large numbers, we can show that for a continuous bounded function $g \in C_b(\mathcal{Z})$

$$\mathbb{P}\left(\left\{\omega\in\Omega : \lim_{n\to\infty}\left|\frac{1}{n}\sum_{i=1}^{n} g(\beta_n^i) - \mathbb{E}(g(\beta_\infty^i))\right| = 0\right\}\right) = 1. \tag{4.60}$$

By considering a countable family of measure determining functions $\mathcal{T} \subset C_b(\mathcal{Z})$, (4.60) implies that the empirical measures $\{\Lambda_n\}_n$ converges weakly to $\Lambda = \mathcal{L}(\beta_\infty^i)$ $\mathbb{P}$-almost surely, and Λ is induced by the limit randomized policy P^{*,ω_0}. We define the w-s topology on the above set of probability measures on $(\Omega_0\times\mathbb{U}\times\mathbb{Y})$. That is, the coarsest topology on $\mathcal{P}(\Omega_0\times\mathbb{U}\times\mathbb{Y})$ under which $\int \hat{f}(\omega_0, u, y)\kappa(d\omega_0, du, dy) : \mathcal{P}(\Omega_0\times\mathbb{U}\times\mathbb{Y}) \to \mathbb{R}$ is continuous for every measurable and bounded $\hat{f}$ which is continuous in u but need not to be continuous in y and ω_0 (see e.g., [121] and [138, Theorem 5.6]). Following from Assumption 4.1.5 and Assumption 4.1.3(ii), and since actions induced by identical policies are i.i.d. (thanks to symmetry), we have $\mathbb{P}$-almost surely

$$f_n := \min\left\{M, c\left(\omega_0, \cdot, \int_{\mathbb{U}} u\Lambda_n(du\times\mathbb{Y})\right)\right\} \xrightarrow{\text{cont}} f := \min\left\{M, c\left(\omega_0, \cdot, \int_{\mathbb{U}} u\Lambda(du\times\mathbb{Y})\right)\right\},$$

where we recall that the sequence f_n converges continuously to f ($f_n \xrightarrow{\text{cont}} f$) if and only if $f_n(a_n) \to f(a)$ whenever $a_n \to a$ as $n \to \infty$. Now, (4.56) follows from the generalized dominated convergence theorem for varying measures in [123, Theorem 3.5]. Equality (4.57) follows from the monotone convergence theorem, and (4.58) follows from the fact that the limit randomized policy, P^{*,ω_0}, does not depend on N and symmetry, hence, (4.58) is true using a similar analysis as (4.56). Inequality (4.59) follows from the fact that the limit policy, $P^{*,\omega_0}(du^i, dy^i) := \Pi^*(du^i|y^i)f(\omega_0, y^i)Q(dy^i)$, achieving (4.58) belongs to $L_{\text{PR,SYM}}$. That is because, following from (Step 3), for each DM, the set of randomized policies is closed under the product topology where each coordinate converges weakly, and hence, the limit policy is also a randomized policy induced by a subsequence of N-DM optimal policies (which are symmetric through DMs). This implies (4.59) and completes the proof.

4.9.5 Independent measurement reduction under Assumption 4.6.1

Under Assumption 4.6.1(i), we can represent the expected cost as

$$\begin{aligned}
J_N(\underline{\gamma}^{1:N}) :=& \int c(\omega_0, u^1_{0:T-1}, \ldots, u^N_{0:T-1})\mu^N(d\omega_0, d\underline{\zeta}^{1:N}) \\
&\times \prod_{i=1}^{N}\prod_{t=0}^{T-1} 1_{\{\gamma^i_t(y^i_t)\in du^i_t\}}\nu^i_t\left(dy^i_t \middle| \omega_0, x^{1:N}_0, \zeta^{1:N}_{0:t-1}, y^{1:N}_{0:t-1}, u^{1:N}_{0:t-1}\right) \\
=& \int c(\omega_0, u^1_{0:T-1}, \ldots, u^N_{0:T-1})\mu^N(d\omega_0, d\underline{\zeta}^{1:N}) \\
&\times \prod_{i=1}^{N}\prod_{t=0}^{T-1} 1_{\{\gamma^i_t(y^i_t)\in du^i_t\}}\psi^i_t\left(y^i_t, \omega_0, x^{1:N}_0, \zeta^{1:N}_{0:t-1}, y^{1:N}_{0:t-1}, u^{1:N}_{0:t-1}\right)\tau^i_t(dy^i_t) \\
=& \int c_s(\omega_0, \underline{\zeta}^{1:N}, u^{1:N}_{0:T-1}, y^{1:N}_{0:T-1})\mu^N(d\omega_0, d\underline{\zeta}^{1:N}) \prod_{i=1}^{N}\prod_{t=0}^{T-1} 1_{\{\gamma^i_t(y^i_t)\in du^i_t\}}\tau^i_t(dy^i_t),
\end{aligned} \tag{4.61}$$

where the new (equivalent) cost function is defined as

$$c_s(\omega_0, \underline{\zeta}^{1:N}, u^{1:N}_{0:T-1}, y^{1:N}_{0:T-1}) := c(\omega_0, u^{1:N}_{0:T-1}) \prod_{i=1}^{N}\prod_{t=0}^{T-1} \psi^i_t\left(y^i_t, \omega_0, x^{1:N}_0, \zeta^{1:N}_{0:t-1}, y^{1:N}_{0:t-1}, u^{1:N}_{0:t-1}\right),$$

and (4.61) follows from Assumption 4.6.1(i). Similar derivation holds when randomized policies are considered. Similarly, we can define the new (equivalent) cost function under Assumption 4.6.1(ii). We note that in the above, we considered control actions induced by deterministic policies; however, the above analysis can be extended to randomized policies by just replacing $\prod_{i=1}^{N}\prod_{t=0}^{T-1} 1_{\{\gamma_t^i(y_t^i)\in du_t^i\}}$ with $\prod_{i=1}^{N}\prod_{t=0}^{T-1}\gamma_t^i(du_t^i|y_t^i)$.

4.9.6 Proof of Lemma 4.7.1

We follow the steps of the proof of Lemma 4.5.1. For any permutation $\sigma \in S_N$, we define a randomized policy $P_\pi^\sigma \in \bar{L}^N$ as a permutation σ of arguments of a randomized policy $P_\pi \in \bar{L}^N$, i.e., for $A^i \in \mathcal{B}(\Gamma^i)$

$$P_\pi^\sigma(\underline{\gamma}^1 \in A^1, \ldots, \underline{\gamma}^2 \in A^N) := P_\pi(\underline{\gamma}^{\sigma(1)} \in A^1, \ldots, \underline{\gamma}^{\sigma(N)} \in A^N).$$

We have

$$\begin{aligned}
&\int P_\pi^\sigma(d\underline{\gamma})\mu^N(d\omega_0, d\underline{\zeta})c^N(\underline{\zeta},\underline{\gamma},\underline{y},\omega_0)\nu^N(d\underline{y}|\underline{\zeta},\underline{\gamma},\omega_0)\\
&= \int c(\omega_0,\underline{u}^1,\ldots,\underline{u}^N)\prod_{k=1}^{N}\underline{\gamma}^k(d\underline{u}^k|\underline{y}^k)P_\pi(d\underline{\gamma}^{\sigma(1)},\ldots,d\underline{\gamma}^{\sigma(N)}) \qquad (4.62)\\
&\times\tilde{\mu}^N(d\underline{\zeta}^{1:N}|\omega_0)\mathbb{P}_0(d\omega_0)\prod_{t=0}^{T-1}\prod_{i=1}^{N}\nu_t^i\left(dy_t^i|\omega_0,x_0^i,\zeta_{0:t-1}^i,y_{0:t-1}^{1:N},u_{t-1}^{1:N}\right)\\
&= \int c(\omega_0,\underline{u}^{\sigma(1)},\ldots,\underline{u}^{\sigma(N)})\prod_{k=1}^{N}\underline{\gamma}^{\sigma(k)}(d\underline{u}^{\sigma(k)}|\underline{y}^{\sigma(k)})P_\pi(d\underline{\gamma}^1,\ldots,d\underline{\gamma}^N) \qquad (4.63)\\
&\times\tilde{\mu}^N(d(\underline{\zeta}^\sigma)^{1:N}|\omega_0)\mathbb{P}_0(d\omega_0)\prod_{t=0}^{T-1}\prod_{i=1}^{N}\nu_t^i\left(dy_t^{\sigma(i)}\Big|\omega_0,x_0^{\sigma(i)},\zeta_{0:t-1}^{\sigma(i)},(y_{0:t-1}^\sigma)^{1:N},(u_{0:t-1}^\sigma)^{1:N}\right)\\
&= \int c(\omega_0,\underline{u}^1,\ldots,\underline{u}^N)\prod_{k=1}^{N}\underline{\gamma}^k(d\underline{u}^k|\underline{y}^k)P_\pi(d\underline{\gamma}^1,\ldots,d\underline{\gamma}^N) \qquad (4.64)\\
&\times\tilde{\mu}^N(d\underline{\zeta}^{1:N}|\omega_0)\mathbb{P}_0(d\omega_0)\prod_{t=0}^{T-1}\prod_{i=1}^{N}\nu_t^i\left(dy_t^i|\omega_0,x_0^i,\zeta_{0:t-1}^i,y_{0:t-1}^{1:N},u_{0:t-1}^{1:N}\right)\\
&= \int P_\pi(d\underline{\gamma})\mu^N(d\omega_0,d\underline{\zeta})c^N(\underline{\zeta},\underline{\gamma},\underline{y},\omega_0)\nu^N(d\underline{y}|\underline{\zeta},\underline{\gamma},\omega_0)
\end{aligned}$$

where $\tilde{\mu}^N$ is the conditional distribution of uncertainties $\underline{\zeta}^{1:N}$ given ω_0, and (4.62) follows from Assumption 4.7.2(b) and the definition of randomized policy P^{σ}_{π} and (4.63) follows from relabeling $\underline{u}^{\sigma(i)}, \underline{y}^{\sigma(i)}, \underline{\zeta}^{\sigma(i)}$ with $\underline{u}^i, \underline{y}^i, \underline{\zeta}^i$ for all $i = 1, \dots, N$ and the fact that $y^i_t = h_t(x^i_0, x^{-i}_0, \zeta^i_{0:t}, \zeta^{-i}_{0:t}, u^i_{0:t-1}, u^{-i}_{0:t-1})$. Equality (4.64) follows from Assumption 4.7.2(a), Assumption 4.7.1 and the hypothesis that the information structure is symmetric. The rest of the proof follows from similar steps in that of Lemma 4.5.1.

4.9.7 Proof of Lemma 4.7.2

We follow steps of the proof of Lemma 4.5.2. Under Assumption 4.7.3 and Assumption 4.1.5, for every finite N, there exists an optimal policy in L^N_{EX}. Consider a sequence $\{P^{*,N}_{\pi}\}_N$, where for every $N \geq 1$, $P^{*,N}_{\pi} \in L^N_{\text{EX}}$ and

$$\begin{aligned}
&\int P^{*,N}_{\pi}(d\underline{\gamma})\mu^N(d\omega_0, d\underline{\zeta})c^N(\underline{\zeta}, \underline{\gamma}, \underline{y}, \omega_0)\nu^N(d\underline{y}|\underline{\zeta}, \underline{\gamma}, \omega_0) \\
&= \inf_{P^N_{\pi} \in L^N_{\text{EX}}} \int P^N_{\pi}(d\underline{\gamma})\mu^N(d\omega_0, d\underline{\zeta})c^N(\underline{\zeta}, \underline{\gamma}, \underline{y}, \omega_0)\nu^N(d\underline{y}|\underline{\zeta}, \underline{\gamma}, \omega_0). \qquad (4.65)
\end{aligned}$$

(Step 1): Let $(I_1, I_2, \dots)$ be i.i.d. random variables with the uniform distribution on the set $\{1, \dots, N\}$. For a fixed N and for any $P^{*,N}_{\pi} \in L^N_{\text{EX}}$, we construct $P^{*,\infty}_{\pi,N} \in L_{\text{EX}}$ as follows: for every fixed N and for all $A^i \in \mathcal{B}(\Gamma^i)$

$$P^{*,\infty}_{\pi,N}(\underline{\gamma}^1 \in A^1, \dots, \underline{\gamma}^2 \in A^N) := P^{*,N}_{\pi}(\underline{\gamma}^{I_1} \in A^1, \dots, \underline{\gamma}^{I_N} \in A^N),$$

where $P^{*,\infty}_{\pi,N}$ is the restriction of $P^{*,\infty}_{\pi,P^N_{\pi}} \in L_{\text{EX}}$ to the first N components.

Let $u^{*,i}_{t,N}$ be the control action induced by $\gamma^i_{N,t}$ where random variables $(\gamma^1_{N,t}, \dots, \gamma^N_{N,t})$ for all $t = 0, \dots, T-1$ are determined by $P^{*,N}_{\pi} \in L^N_{\text{EX}}$. Let $u^{*,i}_{t,\infty,N}$ be the control action induced by $\gamma^i_{\infty,t}$ where random variables $(\gamma^1_{t,\infty,N}, \dots, \gamma^N_{t,\infty,N})$ are determined by $P^{*,\infty}_{\pi,N} \in L_{\text{EX}}$. Let $\underline{\gamma}^i_N := (\gamma^i_{N,0}, \dots, \gamma^i_{N,T-1})$, $\underline{\gamma}^i_{N,\infty} := (\gamma^i_{0,\infty,N}, \dots, \gamma^i_{T-1,\infty,N})$, $\underline{u}^i_N := (u^i_{N,0}, \dots, u^i_{N,T-1})$ and

$\underline{u}^i_{N,\infty} := (u^i_{0,\infty,N}, \dots, u^i_{T-1,\infty,N})$ for each DM. Since under the reduction (Assumption 4.6.1), observations are i.i.d. through DMs and also independent of ω_0, following from Theorem 4.9.2, we have for every $m \geq 1$

$$\begin{aligned}
&\left|\left|\mathcal{L}(\underline{\gamma}^1_N, \dots, \underline{\gamma}^m_N, \underline{y}^1, \dots, \underline{y}^m) - \mathcal{L}(\underline{\gamma}^1_{N,\infty}, \dots, \underline{\gamma}^m_{N,\infty}, \underline{y}^1, \dots, \underline{y}^m)\right|\right| \\
&= \left|\left|\mathcal{L}(\underline{\gamma}^1_N, \dots, \underline{\gamma}^m_N)\prod_{i=1}^m \mathcal{L}(\underline{y}^i) - \mathcal{L}(\underline{\gamma}^1_{N,\infty}, \dots, \underline{\gamma}^m_{N,\infty})\prod_{i=1}^m \mathcal{L}(\underline{y}^i)\right|\right|_{TV} \xrightarrow[N\to\infty]{} 0. \qquad (4.66)
\end{aligned}$$

where (4.36) follows from the fact that $(\underline{\gamma}^1_N, \dots, \underline{\gamma}^N_N)$ and $(\underline{\gamma}^1_{N,\infty}, \dots, \underline{\gamma}^N_{N,\infty})$ are random variables with joint probability measures $P^{*,N}_\pi \in L^N_{\text{EX}}$ and $P^{*,\infty}_{\pi,N} \in L_{\text{EX}}|_N$, respectively. Since $\mathbb{U}$ is compact, and under the reduction the probability measure on observation is fixed, any joint probability measures on acttions and observations is tight, hence, $\{\mathcal{L}(\underline{\gamma}^i_{\infty,N})\}_N$ is tight for each DM and by exchangeablity $\mathcal{L}(\underline{\gamma}^i_{\infty,N}) = \mathcal{L}(\underline{\gamma}^1_{\infty,N})$. Hence, we can find a subsequence such that $\mathcal{L}(\underline{\gamma}^i_{\infty,l}) \xrightarrow[l\to\infty]{} \mathcal{L}(\underline{\gamma}^i_\infty)$ for all $i \in \mathbb{N}$. Since marginals of $\{\mathcal{L}(\underline{\gamma}^1_{\infty,l}, \dots, \underline{\gamma}^m_{\infty,l})\}_l$ are tight, for each $m \geq 1$, there exists a further subsequence

$$\mathcal{L}(\underline{\gamma}^1_{\infty,n}, \dots, \underline{\gamma}^m_{\infty,n}) \xrightarrow[n\to\infty]{} \mathcal{L}(\underline{\gamma}^1_\infty, \dots, \underline{\gamma}^m_\infty),$$

where $(\underline{\gamma}^1_\infty, \underline{\gamma}^2_\infty, \dots)$ is infinitely-exchangeable and induced by $P^{*,\infty}_\pi \in L_{\text{EX}}$ since the set of infinitely-exchangeable random variables is closed under the weak-convergence topology. Hence, following from (4.36), for each $m \geq 1$

$$\mathcal{L}(\underline{\gamma}^1_n, \dots, \underline{\gamma}^m_n) \xrightarrow[n\to\infty]{} \mathcal{L}(\underline{\gamma}^1_\infty, \dots, \underline{\gamma}^m_\infty).$$

By construction $\underline{u}^{*,i}_n$ and $\underline{u}^{*,i}_\infty$ and since random variables $\underline{\gamma}^i_n$s are independent of $\underline{y}^i$s, we have for each $m \geq 1$

$$(\underline{u}^{*,1}_n, \dots, \underline{u}^{*,m}_n) \xrightarrow[n\to\infty]{\text{d}} (\underline{u}^1_\infty, \dots, \underline{u}^m_\infty),$$

where $(\underline{u}^1_\infty, \underline{u}^2_\infty, \dots)$ is induced by an infinitely-exchangeable policies $P^{*,\infty}_\pi \in L_{\text{EX}}$. Following from Theorem 4.9.3, $\mathbb{P}$-almost surely

$$F_{n,t}(A) := F^\omega_{n,t}(A) := \frac{1}{n}\sum_{i=1}^n \delta_{u^{*,i}_{n,t}(\omega)}(A) \xrightarrow[n\to\infty]{\text{d}} \alpha^{u,\omega}_t(A), \tag{4.67}$$

where $A \in \mathcal{U}$ and ω denotes the sample path dependence and α^u_t is the directing random measure of an infinitely-exchangeable random variables $(\underline{u}^1_{\infty,t}, \underline{u}^2_{\infty,t}, \dots)$. By (4.67), since the action space is compact, for all $t = 0, \dots, T-1$, we have $\mathbb{P}$-almost surely

$$\mu^u_{n,t} := \mu^{u,\omega}_{n,t} := \frac{1}{n}\sum_{i=1}^n u^{*,i}_{n,t} = \int_{\mathbb{U}} u F_{n,t}(du) \xrightarrow[n\to\infty]{\text{d}} \mu^u_t := \int_{\mathbb{U}} u\alpha^{u,\omega}_t(du). \tag{4.68}$$

(Step 2): Let $x^{*,i}_{t,n}$ be the state of DM^i at time t under $u^{*,i}_{0:t-1,n} := (u^{*,i}_{0,n}, \dots, u^{*,i}_{t-1,n})$:

$$x^{*,i}_{t,n} = f_{t-1}\bigg(x^{*,i}_{t-1,n}, u^{*,i}_{n,t-1}, \frac{1}{n}\sum_{p=1}^n x^{*,p}_{t-1,n}, \frac{1}{n}\sum_{p=1}^n u^{*,p}_{n,t-1}, w^i_{t-1}\bigg). \tag{4.69}$$

Let $t = 1$. We have

$$x^{*,i}_{1,n} = f_0\bigg(x^i_0, u^{*,i}_{n,0}, \frac{1}{n}\sum_{p=1}^n x^p_0, \frac{1}{n}\sum_{p=1}^n u^{*,p}_{n,0}, w^i_0\bigg). \tag{4.70}$$

Since initial states are i.i.d. conditioned on ω_0, by continuity of the function f_0 in actions and states, we have $x^{*,i}_{1,n} \xrightarrow[n\to\infty]{\text{d}} x^{*,i}_{1,\infty}$ for all DMs. Hence, $\{\mathcal{L}(x^{*,1}_{1,n}, \dots, x^{*,n}_{1,n})\}_n$ is tight, and hence, for each $m \geq 1$, there exists a subsubsequence k such that $(x^{*,1}_{1,k}, \dots, x^{*,m}_{1,k}) \xrightarrow[k\to\infty]{\text{d}} (x^{*,1}_{1,\infty}, \dots, x^{*,m}_{1,\infty})$. Following from Theorem 4.9.3, since f_0 is bounded, we have $\mathbb{P}$-almost surely

$$\mu^x_{k,1} := \frac{1}{k}\sum_{i=1}^n x^{*,i}_{1,k} = \mu^{x,\omega}_{k,1} = \int_{\mathbb{X}} x \frac{1}{k}\sum_{i=1}^k \delta_{x^{*,i}_{1,k}}(dx) \xrightarrow[k\to\infty]{\text{d}} \mu^x_1 := \int_{\mathbb{X}} x\alpha^{x,\omega}_1(dx), \tag{4.71}$$

where α_1^x is the directing measure for $(x_{\infty,1}^{*,1}, x_{\infty,1}^{*,2}, \dots)$. Similarly, we can show that for $t = 2$,

$$x_{2,k}^{*,i} = f_1\Big(x_{1,k}^{*,i}, u_{k,1}^{*,i}, \mu_{k,1}^x, \mu_{k,1}^u, w_1^i\Big). \tag{4.72}$$

By continuity of the function f_1 and the analysis for $t = 1$, we have $x_{2,k}^{*,i} \xrightarrow[k\to\infty]{\mathrm{d}} x_{2,\infty}^{*,i}$ for all DMs. Hence, $\{\mathcal{L}(x_{2,k}^{*,1}, \dots, x_{2,k}^{*,k})\}_l$ is tight and for each $m \geq 1$, there exists a further subsubsequence k_l such that $(x_{2,k_l}^{*,1}, \dots, x_{2,k_l}^{*,m}) \xrightarrow[k_l\to\infty]{\mathrm{d}} (x_{2,\infty}^{*,1}, \dots, x_{2,\infty}^{*,m})$. Following from Theorem 4.9.3, since f_1 is bounded, , we have $\mathbb{P}$-almost surely

$$\mu_{k_l,2}^x := \mu_{k_l,2}^{x,\omega} = \int_{\mathbb{X}} x \frac{1}{k_l} \sum_{i=1}^{k_l} \delta_{x_{2,k_l}^{*,i}}(dx) \xrightarrow[k_l\to\infty]{\mathrm{d}} \mu_2^x := \int_{\mathbb{X}} x \alpha_2^{x,\omega}(dx), \tag{4.73}$$

where α_2^x is the directing measure for $(x_{\infty,2}^{*,1}, x_{\infty,2}^{*,2}, \dots)$. By induction, for each $m \geq 1$, there exists a further subsubsequence n (which we indicate by n to omit further sub-subscript) such that $(\underline{x}_n^{*,1}, \dots, \underline{x}_n^{*,m}) \xrightarrow[n_l\to\infty]{\mathrm{d}} (\underline{x}_\infty^{*,1}, \dots, \underline{x}_\infty^{*,m})$ and $\mu_{n_l,t}^x \xrightarrow[n\to\infty]{\mathrm{d}} \mu_t^x$ for all $t = 0, \dots, T-1$.

Now, we follow the steps of Lemma 4.5.2, however, in addition to actions and observations, we consider states and disturbances in our analysis and we use the result of (Step 2). Define $\tilde{P}^{*,n}$ as the joint probability measures of $(\underline{u}_n^{*,1}, \underline{x}_n^{*,1}, \mu_{n,0:T-1}^u, \mu_{n,0:T-1}^x, \underline{y}, \underline{\zeta})$. Since marginals on $(\underline{u}_n^{*,1}, \underline{x}_n^{*,1}, \mu_{n,0:T-1}^u, \mu_{n,0:T-1}^x)$ are tight and under the reduction marginals on $(\underline{y}, \underline{\zeta})$ are fixed, $\{\tilde{P}^{*,n}\}_n$ is tight. Hence, there exists a further subsubsequence $\{\tilde{P}^{*,n_k}\}_{n_k}$ converges weakly to $\tilde{P}^*$ as n_k goes to infinity. This implies that marginals $\{\tilde{P}^{*,n_k}\}_{n_k}$ converge to the marginals of $\tilde{P}^*$, hence, $\tilde{P}^*$ is induced by $(\underline{u}_\infty^1, \underline{u}_\infty^2, \dots)$ which is infinitely-exchangeable and is induced by a policy in L_{EX}.

(Step 3): Since the cost function is continuous in states and actions, under the reduction (Assumption 4.7.3), we have $\mathbb{P}$-almost surely

$$\begin{aligned}
&\frac{1}{N}\sum_{i=1}^{N}\sum_{t=0}^{T-1} c\Big(\omega_0, x_t^i, u_t^i, \frac{1}{N}\sum_{p=1}^{N} u_t^p, \frac{1}{N}\sum_{p=1}^{N} x_t^p\Big) \\
&\times \prod_{i=1}^{N}\prod_{t=0}^{T-1} \phi_t^i\Big(y_t^i, \omega_0, x_0^i, \zeta_{0:t-1}^i, y_{0:t-1}^i, u_{0:t-1}^i, \frac{1}{N}\sum_{p=1}^{N} u_{0:t-1}^p, \frac{1}{N}\sum_{p=1}^{N} x_{0:t}^p\Big)
\end{aligned}$$

$$= \frac{1}{N}\sum_{i=1}^{N} \bar{c}\Big(\omega_0, \underline{\zeta}^i, \underline{x}^i, \underline{u}^i, \frac{1}{N}\sum_{p=1}^{N}\underline{u}^p, \frac{1}{N}\sum_{p=1}^{N}\underline{x}^p\Big) \tag{4.74}$$
$$\prod_{i=1}^{N}\underline{\phi}^i\Big(\underline{y}^i, \omega_0, \underline{\zeta}^i, \underline{u}^i, \frac{1}{N}\sum_{p=1}^{N}\underline{u}^p, \frac{1}{N}\sum_{p=1}^{N}\underline{x}^p\Big),$$

where (4.74) is true following from (4.5) and Assumption 4.1.5 for some function $\hat{c} : \Omega_0 \times \mathbf{S} \times \mathbf{X} \times \mathbf{U} \times \mathbf{U} \times \mathbf{X} \to \mathbb{R}_+$ which is continuous in states and actions and

$$\underline{\phi}^i\Big(\underline{y}^i, \omega_0, \underline{\zeta}^i, \underline{u}^i, \frac{1}{N}\sum_{p=1}^{N}\underline{u}^p, \frac{1}{N}\sum_{p=1}^{N}\underline{x}^p\Big)$$
$$:= \prod_{t=0}^{T-1}\phi_t^i\Big(y_t^i, \omega_0, x_0^i, \zeta_{0:t-1}^i, y_{0:t-1}^i, u_{0:t-1}^i, \frac{1}{N}\sum_{p=1}^{N}u_{0:t-1}^p, \frac{1}{N}\sum_{p=1}^{N}x_{0:t}^p\Big)$$

We have,

$$\limsup_{N\to\infty}\inf_{P_\pi^N\in L_{\text{EX}}^N}\int P_\pi^N(d\underline{\gamma})\mu^N(d\omega_0, d\underline{\zeta})c^N(\underline{\zeta}, \underline{\gamma}, \underline{y}, \omega_0)\nu(d\underline{y}|\underline{\zeta}, \underline{\gamma}, \omega_0)$$
$$= \limsup_{N\to\infty}\int\int_{\prod_{i=n_l+1}^{\infty}\mathbf{Y}\times\mathbf{S}} \bar{c}\Big(\omega_0, \underline{\zeta}^i, \underline{x}^i, \underline{u}^i, \mu_{N,0:T-1}^u, \mu_{N,0:T-1}^x\Big) \tag{4.75}$$
$$\times\tilde{P}^{*,N}(d\underline{u}^{*,i}, d\underline{x}^{*,i}, d\mu_{N,0:T-1}^u, d\mu_{N,0:T-1}^x, \underline{y}, \underline{\zeta})$$
$$\times\prod_{i=1}^{\infty}\underline{\phi}^i\Big(\underline{y}^i, \omega_0, \underline{\zeta}^i, \underline{u}^{*,i}, \mu_{N,0:T-1}^u, \mu_{N,0:T-1}^x\Big)\mathbb{P}_0(d\omega_0)$$
$$\geq \lim_{n_k\to\infty}\int\int_{\prod_{i=n_k+1}^{\infty}\mathbf{Y}\times\mathbf{S}} \bar{c}\Big(\omega_0, \underline{\zeta}^1, \underline{x}^1, \underline{u}^1, \mu_{n_k,0:T-1}^u, \mu_{n_k,0:T-1}^x\Big) \tag{4.76}$$
$$\times\tilde{P}^{*,n_k}(d\underline{u}^{*,i}, d\underline{x}^{*,i}, d\mu_{n_k,0:T-1}^u, d\mu_{n_k,0:T-1}^x, \underline{y}, \underline{\zeta})$$
$$\times\prod_{i=1}^{\infty}\underline{\phi}^i\Big(\underline{y}^i, \omega_0, \underline{\zeta}^i, \underline{u}^{*,i}, \mu_{n_k,0:T-1}^u, \mu_{n_k,0:T-1}^x\Big)\mathbb{P}_0(d\omega_0)$$
$$= \int \bar{c}\Big(\omega_0, \underline{\zeta}^1, \underline{x}^1, \underline{u}^1, \mu_{0:T-1}^u, \mu_{0:T-1}^x\Big)\tilde{P}^*(d\underline{u}^{*,i}, d\underline{x}^{*,i}, d\mu_{0:T-1}^u, d\mu_{0:T-1}^x, \underline{y}, \underline{\zeta}) \tag{4.77}$$
$$\times\prod_{i=1}^{\infty}\underline{\phi}^i\Big(\underline{y}^i, \omega_0, \underline{\zeta}^i, \underline{u}^{*,i}, \mu_{0:T-1}^u, \mu_{0:T-1}^x\Big)\mathbb{P}_0(d\omega_0)$$
$$\geq \limsup_{N\to\infty}\inf_{P_\pi\in L_{\text{EX}}}\int P_{\pi,N}(d\underline{\gamma})\mu^N(d\omega_0, d\underline{\zeta})c^N(\underline{\zeta}, \underline{\gamma}, \underline{y}, \omega_0)\nu(d\underline{y}|\underline{\zeta}, \underline{\gamma}, \omega_0). \tag{4.78}$$

where (4.75) follows from integrating over the set $(\prod_{i=n_l+1}^{\infty} \mathbf{Y} \times \mathbf{S})$ and the fact that under the reduction, observations and disturbances, initial states are i.i.d. and $(u_N^{*,1}, \ldots, u_N^{*,N})$ is N-exchangeable. Inequality (4.76) follows from the assumption that the cost function is bounded and limsup is the greatest subsequence limit of a bounded sequence. Equality (4.77) follows from the dominated convergence theorem and following from Assumption 4.1.5 and Assumption 4.7.3 and since probability measures on observations disturbances are fixed and since by (Step 2) $\{\tilde{P}^{*,n_k}\}_{n_k}$ converges weakly to $\tilde{P}^*$ and $\prod_{i=1}^{\infty} \underline{\phi}^i(\underline{y}^i, \omega_0, \underline{\zeta}^i, \underline{u}^{*,i}, \mu^u_{n_k,0:T-1}, \mu^x_{n_k,0:T-1})$ converges weakly to the limit in the product topology as n_k goes to infinity. Inequality (4.78) follows from the fact that $\tilde{P}^*$ is the joint measure with the first coordinate $(u_\infty^1, u_\infty^2, \ldots)$ which is infinitely-exchangeable and is induced by a policy in L_{EX}. The above inequalities are equalities since the opposite direction is true (that is because $L_{\mathrm{EX}}|_N \subset L_{\mathrm{EX}}^N$) and this completes the proof.

4.9.8 Proof of Theorem 4.7.1

We complete the proof in five steps where the steps are similar to the proof of Theorem 4.5.1.

(Step 1): Under Assumption 4.1.5 and Assumption 4.7.3, by Lemma 4.7.1, for every finite N, there exists a randomized optimal policy in L_{EX}^N. Consider a sequence $\{P_\pi^{*,N}\}_N$, where for every $N \geq 1$, $P_\pi^{*,N} \in L_{\mathrm{EX}}^N$ and

$$\begin{aligned} &\int P_\pi^{*,N}(d\underline{\gamma})\mu^N(d\omega_0, d\underline{\zeta})c^N(\underline{\zeta}, \underline{\gamma}, \underline{y}, \omega_0)\nu^N(d\underline{y}|\underline{\zeta}, \underline{\gamma}, \omega_0) \\ &= \inf_{P_\pi^N \in L_{\mathrm{EX}}^N} \int P_\pi^N(d\underline{\gamma})\mu^N(d\omega_0, d\underline{\zeta})c^N(\underline{\zeta}, \underline{\gamma}, \underline{y}, \omega_0)\nu^N(d\underline{y}|\underline{\zeta}, \underline{\gamma}, \omega_0). \end{aligned} \tag{4.79}$$

(Step 2): Similar to (Step 2) of the proof of Theorem 4.5.1 using Lemma 4.7.2 and Theorem 4.4.1, we can show that to complete the proof, it is sufficient to show

$$\begin{aligned} &\lim_{M\to\infty} \limsup_{N\to\infty} \inf_{P_\pi^N \in L_{\mathrm{PR,SYM}}^N} \int P_\pi^N(d\underline{\gamma})\mu^N(d\omega_0, d\underline{\zeta}) \min\{M, c^N(\underline{\zeta}, \underline{\gamma}, \underline{y}, \omega_0)\}\nu^N(d\underline{y}|\underline{\zeta}, \underline{\gamma}, \omega_0) \\ &\geq \inf_{P_\pi \in L_{\mathrm{PR,SYM}}} \limsup_{N\to\infty} \int P_{\pi,N}(d\underline{\gamma})\mu^N(d\omega_0, d\underline{\zeta})c^N(\underline{\zeta}, \underline{\gamma}, \underline{y}, \omega_0)\nu^N(d\underline{y}|\underline{\zeta}, \underline{\gamma}, \omega_0). \end{aligned} \tag{4.80}$$

In the next two steps, we justify (4.80) through showing that there exists a subsequence of randomized policies induced by symmetric/identical private randomization whose weak subsequent limit achieves the right hand side of (4.80).

(Step 3): Consider the set of randomized policies $L^N_{\text{PR,SYM}}$. We note that under a symmetric information structure and since each DM applies an identical policy, under Assumption 4.6.1, $\underline{y}^i$ are i.i.d. through DMs and also independent of ω_0. Hence, following from the information structure, the randomized policy spaces of each DM is separated from the policies of the other decision makers. Hence, we can equivalently represent any privately randomized policy for each DM acting through time separately as a probability measures induced by symmetric (identical randomized policies), i.e., probability measures q on $(\mathbf{U} \times \mathbf{Y})$ where randomized policies of each DM for every $t = 0, \ldots, T-1$ satisfy

$$\int g(\omega_0, x^i_0, \zeta^i_{0:t-1}, y^i_{0:t}, u^i_{0:t}) q(dy^i_{0:t}, du^i_{0:t} | \omega_0, x^i_0, \zeta^i_{0:t-1})$$
$$= \int g(\omega_0, x^i_0, \zeta^i_{0:t-1}, y^i_{0:t}, u^i_{0:t}) \prod_{k=0}^{t} \Pi^N_k(du^i_k | y^i_k) \eta_k(dy^i_k | \omega_0, x^i_0, \zeta^i_{0:k-1}, y^i_{0:k-1}, u^i_{0:k-1}),$$

for all bounded functions g which is continuous in actions and observations and measurable in other arguments and for some stochastic kernel Π^N_k representing a randomized policy of DMs at time k (which is identical through DMs).

Since $\mathbb{U}$ is compact, the marginals on $\mathbf{U}$ is relatively compact under the weak convergence topology. Hence, the collection of all probability measures with these relatively compact marginals are also relatively compact (see e.g., [137, Proof of Theorem 2.4]). Since every DM applies an identical policy and since observations are i.i.d., the randomized policy space is relatively compact (where each coordinate is relatively compact in the weak convergence topology), and hence, there exists a subsequence of randomized policies $\tilde{q}_n \in \mathcal{P}(\prod_i (\mathbf{Y} \times \mathbf{U}))$ converges weakly (each coordinate converges weakly) to a limit $\tilde{q}$ (as an infinite product of policies of DMs), where n is the index of the subsequence and n goes to infinity. Now, we show that randomized policy space is closed

under the weak convergence topology. Assume $\hat{q}_n \in \mathcal{P}(\mathbf{Y} \times \mathbf{U})$ (induced by identical randomized policies Π_t^n for each DM at time $t = 0, \ldots, T-1$) converges weakly to $\hat{q}$. If Assumption 4.6.1(i) (under the structure Assumption 4.7.3) holds, then there exists an independent static reduction for each DM through time, and hence, following from the discussion in the proof of [138, Theorem 5.2], each coordinate of policy spaces corresponds to DM^i at time t is closed under the weak convergence topology. Also, if Assumption 4.6.1(ii) (under the structure 4.7.3) holds, then [138, Theorem 5.6] leads to the same conclusion. Hence, this implies that for $\tilde{q}_N^* \in \mathcal{P}(\prod_{i=1}^N (\mathbf{Y} \times \mathbf{U}))$ induced by optimal randomized policies $\Pi_t^{*,N}$ for each DM at time t, there exists a subsequence $\tilde{q}_n^* \in \mathcal{P}(\prod_{i=1}^\infty (\mathbf{Y} \times \mathbf{U}))$ (as an infinite product of policies of DMs $\Pi_t^{*,n}$) converges weakly (each coordinate converges weakly) to a limit $\tilde{q}^*$ which is in $L_{\text{PR,SYM}}$ and it is induced by a randomized policy $\Pi_t^{*,\infty}$ for each DM at time t.

(Step 4): Let $\{\hat{q}_N^*\}_N$ be a policy for each DM induced by optimal randomized policies $\Pi_t^{*,N}$ for N-DM team problems, and let $\underline{u}_N^{i,*} := (u_{N,0}^{i,*}, \ldots, u_{N,T-1}^{i,*})$ be the action of DM^i through time induced by $\Pi_t^{*,N}$. Following from (Step 3), there exists a weak subsequential limit $\hat{q}^*$ of $\{\hat{q}_n^*\}_n$ as $n \to \infty$ for each DM, which is induced by $\Pi_t^{*,\infty}$. Let $\underline{u}_\infty^{i,*} := (u_{\infty,0}^{i,*}, \ldots, u_{\infty,T-1}^{i,*})$ be the action of DM^i induced by the identically randomized policy $\Pi_t^{*,\infty}$. Define

$$\Upsilon_N(B) := \frac{1}{N} \sum_{i=1}^N \delta_{(\underline{x}_N^i, \tilde{\alpha}_N^i)}(B), \tag{4.81}$$

where $\tilde{\alpha}_N^i := (\underline{u}_N^{i,*}, \underline{y}^i, \underline{\zeta}^i)$, $B \in \mathbf{X} \times \mathcal{Z} := \mathbf{X} \times \mathbf{U} \times \mathbf{Y} \times \mathbf{S}$, $\mathbf{U} := (\prod_{t=0}^{T-1} \mathbb{U})$, $\mathbf{Y} := (\prod_{t=0}^{T-1} \mathbb{Y})$, $\mathbf{S} := (\prod_{t=0}^{T-1} \mathbb{S}) = \mathbb{X} \times (\prod_{t=0}^{T-1} \mathbb{W} \times \mathbb{V})$, $\mathbf{X} := (\prod_{t=0}^{T-1} \mathbb{X})$, $\underline{y}^i := (y_0^i, \ldots, y_{T-1}^i)$, $\underline{\zeta}^i := (\zeta_0^i, \ldots, \zeta_{T-1}^i)$, and $\underline{x}_N^i := (x_0^i, \ldots, x_{T-1}^i)$ with states are driven by a sequence of N-DM randomized optimal policies of $\Pi_t^{*,N}$. In the following, we show that, conditioned on ω_0, the subsequence of empirical measures $\{\Upsilon_n\}_n$ converges to $\Upsilon := \mathcal{L}((\underline{x}_\infty^1, \tilde{\alpha}_\infty^1)|\omega_0)$ in w-s topology, where $\tilde{\alpha}_\infty^i = (\underline{u}_\infty^{*,i}, \underline{y}^i, \underline{\zeta}^i)$ and $\underline{x}_\infty^i$ denotes that states of DM^i driven by $\underline{u}_\infty^{*,i}$ (we note that the convergence is weakly, but since $\underline{\zeta}^i$s are exogenous random variables with a fixed marginal, the convergence is also in the w-s

topology). Define

$$\bar{Q}_N(B) := \frac{1}{N}\sum_{i=1}^{N} \delta_{\tilde{\alpha}^i_N}(B), \tag{4.82}$$

where $B \in \mathcal{Z}$. Under the reduction in Assumption 4.6.1, observations of each DM are independent of actions and observations of other DMs through time, $t = 0, \ldots, T-1$, and hence, a similar argument, which is used to show (4.60), implies that the subsequence of empirical measures $\{\bar{Q}_n\}_{n\in\mathbb{I}}$ converges $\mathbb{P}$-almost surely to $\bar{Q} = \mathcal{L}(\tilde{\alpha}^i_\infty|\omega_0)$ in w-s topology. Define

$$\Upsilon^t_n(A) := \frac{1}{n}\sum_{i=1}^{n} \delta_{(\underline{x}^i_{t,n},\tilde{\alpha}^i_{t,n})}(A), \tag{4.83}$$

where $\tilde{\alpha}^i_{t,n} := (u^{i,*}_{n,t}, y^i_t, \zeta^i_t)$, $A \in \mathbb{X}\times\mathbb{U}\times\mathbb{Y}\times\mathbb{S}$. Since conditioned on ω_0, initial states are i.i.d, the empirical measure of initial states converges weakly to $\mathcal{L}(x^1_0|\omega_0)$ $\mathbb{P}$-almost surely. Since $\{\bar{Q}_n\}_n$ converges $\mathbb{P}$-almost surely to $\bar{Q}$ in w-s topology, we can conclude that Υ^0_N converges $\Upsilon^0 := \mathcal{L}((x^i_0, \tilde{\alpha}^i_{0,\infty})|\omega_0)$ in w-s topology $\mathbb{P}$-almost surely. Following from (4.5), for $t = 0$, we have for every continuous and bounded function $g \in C_b(\mathbb{X})$, conditioned on ω_0, $\mathbb{P}$-almost surely

$$\begin{aligned}
\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n} g(x^i_{1,n}) &= \lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n} g\bigg(f_0\bigg(x^i_0, u^{*,i}_{n,0}, \frac{1}{n}\sum_{p=1}^{n} x^p_0, \frac{1}{n}\sum_{p=1}^{n} u^{*,p}_{n,0}, w^i_0\bigg)\bigg) \\
&= \lim_{n\to\infty}\int g\bigg(f_0\bigg(x, u, \int x\Upsilon^0_n(dx\times\mathbb{U}\times\mathbb{Y}\times\mathbb{S}), \int u\Upsilon^0_n(\mathbb{X}\times du\times\mathbb{Y}\times\mathbb{S}), \zeta\bigg)\bigg) \qquad (4.84)\\
&\quad \times\Upsilon^0_n(dx, du, dy, d\zeta) \\
&= \int g\bigg(f_0\bigg(x, u, \int x\Upsilon^0(dx\times\mathbb{U}\times\mathbb{Y}\times\mathbb{S}), \int u\Upsilon^0(\mathbb{X}\times du\times\mathbb{Y}\times\mathbb{S}), \zeta\bigg)\bigg) \qquad (4.85)\\
&\quad \times\Upsilon^0(dx, du, dy, d\zeta)
\end{aligned}$$

where (4.84) follows from (4.83), and (4.85) follows from the generalized dominated convergence theorem for varying measures. That is because, function g is continuous and bounded, f_0 is a bounded function which is continuous in actions and observations and measurable in uncertainties, and the fact that under the reduction, conditioned on ω_0, Υ^0_N converges $\Upsilon^0 := \mathcal{L}((x^i_0, \tilde{\alpha}^i_{0,\infty})|\omega_0)$ in

w-s topology $\mathbb{P}$-almost surely. Hence, since $\{\bar{Q}_n\}_n$ converges $\mathbb{P}$-almost surely to $\bar{Q}$ in w-s topology conditioned on ω_0, Υ_N^1 converges $\Upsilon^1 := \mathcal{L}((x^i_{1,\infty}, \tilde{\alpha}^i_{1,\infty})|\omega_0)$ in w-s topology $\mathbb{P}$-almost surely. By induction, one can show that conditioned on ω_0, Υ_N^t converges $\Upsilon^t := \mathcal{L}((x^i_{t,\infty}, \tilde{\alpha}^i_{t,\infty})|\omega_0)$ in w-s topology $\mathbb{P}$-almost surely for $t = 0, \ldots, T-1$. Hence, conditioned on ω_0, $\{\Upsilon_n\}_{n\in\mathbb{I}}$ converges to $\Upsilon := \mathcal{L}((\underline{x}^i_\infty, \tilde{\alpha}^i_\infty)|\omega_0)$ in w-s topology.

(Step 5): By Assumption 4.1.5, similar to the proof of Theorem 4.7.2, we have (4.74). Under the reduction, we can consider policy spaces for each DM individually. Let for every $t = 0, \ldots, T-1$, P_n^{*,ω_0} be a probability measure on actions, observations and uncertainties induced by optimal randomized policies for each DM (which is identical because of symmetry) for N-DM teams conditioned on ω_0, i.e., a probability measure that satisfies

$$\begin{aligned}
&\int g(\omega_0, x^i_0, \zeta^i_{0:t-1}, y^i_{0:t}, u^{i,*}_{n,0:t}) P_n^{*,\omega_0}(dx^i_0, d\zeta^i_{0:t-1}, dy^i_{0:t}, du^{i,*}_{n,0:t}|\omega_0) \\
&= \int g(\omega_0, x^i_0, \zeta^i_{0:t-1}, y^i_{0:t}, u^{i,*}_{n,0:t}) \mu^i(dx^i_0, d\zeta^i_{0:t-1}|\omega_0) \\
&\quad \times \prod_{k=0}^{t} \Pi_k^{*,n}(du^{*,i}_{n,k}|y^i_k) \eta_k(dy^i_k|\omega_0, x^i_0, \zeta^i_{0:k-1}, y^i_{0:k-1}, u^{i,*}_{n,0:k-1}),
\end{aligned} \tag{4.86}$$

for all bounded functions g which is continuous in actions and observations and measurable in other arguments. Similarly, we denote P^{*,ω_0} as a probability measure induced by the limit policy, i.e., a probability measure satisfying (4.86) induced by $\Pi_k^{*,\infty}$. Hence, following from a similar argument as in the (Step 4) of the proof of Theorem 4.7.1, we have

$$\begin{aligned}
&\lim_{M\to\infty} \limsup_{N\to\infty} \inf_{P_\pi^N \in L^N_{\text{PR,SYM}}} \int P_\pi^N(d\underline{\gamma}) \mu^N(d\omega_0, d\underline{\zeta}) \nu(d\underline{y}|\underline{\zeta}, \underline{\gamma}, \omega_0) \min\{M, c^N(\underline{\zeta}, \underline{\gamma}, \underline{y}, \omega_0)\} \\
&\geq \lim_{M\to\infty} \lim_{n\to\infty} \int \int \min\left\{ M, \bar{c}\left(\omega_0, \underline{\zeta}, \underline{x}, \underline{u}, \int \underline{u} \Upsilon_n(\mathbf{X} \times d\underline{u} \times \mathbf{Y} \times \mathbf{S}), \int \underline{x} \Upsilon_n(d\underline{x} \times \mathbf{U} \times \mathbf{Y} \times \mathbf{S})\right)\right\}
\end{aligned} \tag{4.87}$$

$$\times \Upsilon_n(d\underline{x}, d\underline{u}, d\underline{y}, d\underline{\zeta}) \prod_{i=1}^{\infty} P_n^{*,\omega_0}(d\underline{u}^{i,*}_n, d\underline{y}^i, d\underline{\zeta}^i) \prod_{i=1}^{\infty} \underline{\phi}^i\left(\underline{y}^i, \omega_0, \underline{\zeta}^i, \underline{u}^{i,*}_n, \frac{1}{n}\sum_{p=1}^{n} \underline{u}^{p,*}_n, \frac{1}{n}\sum_{p=1}^{n} \underline{x}^p_n\right) \mathbb{P}_0(d\omega_0)$$

$$= \lim_{M\to\infty} \int \lim_{n\to\infty} \int \min\left\{M, \bar{c}\left(\omega_0, \underline{\zeta}, \underline{x}, \underline{u}, \int \underline{u}\Upsilon_n(\mathbf{X}\times d\underline{u}\times\mathbf{Y}\times\mathbf{S}), \int \underline{x}\Upsilon_n(d\underline{x}\times\mathbf{U}\times\mathbf{Y}\times\mathbf{S})\right)\right\} \tag{4.88}$$

$$\times\Upsilon_n(d\underline{x}, d\underline{u}, d\underline{y}, d\underline{\zeta}) \prod_{i=1}^{\infty} P_n^{*,\omega_0}(d\underline{u}_n^{i,*}, d\underline{y}^i, d\underline{\zeta}^i) \prod_{i=1}^{\infty} \underline{\phi}^i\left(\underline{y}^i, \omega_0, \underline{\zeta}^i, \underline{u}_n^{i,*}, \frac{1}{n}\sum_{p=1}^{n}\underline{u}_n^{p,*}, \frac{1}{n}\sum_{p=1}^{n}\underline{x}_n^p\right)\mathbb{P}_0(d\omega_0)$$

$$= \lim_{M\to\infty} \int\int \min\left\{M, \bar{c}\left(\omega_0, \underline{\zeta}, \underline{x}, \underline{u}, \int \underline{u}\Upsilon(\mathbf{X}\times d\underline{u}\times\mathbf{Y}\times\mathbf{S}), \int \underline{x}\Upsilon(d\underline{x}\times\mathbf{U}\times\mathbf{Y}\times\mathbf{S})\right)\right\} \tag{4.89}$$

$$\times\Upsilon(d\underline{x}, d\underline{u}, d\underline{y}, d\underline{\zeta}) \prod_{i=1}^{\infty} P^{*,\omega_0}(d\underline{u}_\infty^{i,*}, d\underline{y}^i, d\underline{\zeta}^i) \prod_{i=1}^{\infty} \underline{\phi}^i\left(\underline{y}^i, \omega_0, \underline{\zeta}^i, \underline{u}_\infty^{i,*}, E[\underline{u}_\infty^{1,*}|\omega_0], E[\underline{x}_\infty^1|\omega_0]\right)\mathbb{P}_0(d\omega_0)$$

$$= \int\int \bar{c}\left(\omega_0, \underline{\zeta}, \underline{x}, \underline{u}, \int \underline{u}\Upsilon(\mathbf{X}\times d\underline{u}\times\mathbf{Y}\times\mathbf{S}), \int \underline{x}\Upsilon(d\underline{x}\times\mathbf{U}\times\mathbf{Y}\times\mathbf{S})\right) \tag{4.90}$$

$$\times\Upsilon(d\underline{x}, d\underline{u}, d\underline{y}, d\underline{\zeta}) \prod_{i=1}^{\infty} P^{*,\omega_0}(d\underline{u}_\infty^{i,*}, d\underline{y}^i, d\underline{\zeta}^i) \prod_{i=1}^{\infty} \underline{\phi}^i\left(\underline{y}^i, \omega_0, \underline{\zeta}^i, \underline{u}_\infty^{i,*}, E[\underline{u}_\infty^{1,*}|\omega_0], E[\underline{x}_\infty^1|\omega_0]\right)\mathbb{P}_0(d\omega_0)$$

$$\geq \inf_{P_\pi\in L_{\text{PR,SYM}}} \limsup_{N\to N} \int P_{\pi,N}(d\underline{\gamma})\mu^N(d\omega_0, d\underline{\zeta})c^N(\underline{\zeta}, \underline{\gamma}, \underline{y}, \omega_0)\nu(d\underline{y}|\underline{\zeta}, \underline{\gamma}, \omega_0),$$

where (4.87) follows from (4.81), (4.74), and since limsup is the greatest convergent subsequence limit for a bounded sequence, and (4.88) follows from the dominated convergence theorem. Following from a similar argument as the analysis in (Step 4) of the proof of Theorem 4.5.1, since $\{\Upsilon_n\}_{n\in\mathbb{I}}$ converges weakly to Υ $\mathbb{P}$-almost surely, an argument based on the generalized dominated convergence theorem for varying measures in [123, Theorem 3.5] implies (4.89), and (4.90) follows from the monotone convergence theorem. Hence, (4.80) holds and this completes the proof.

4.9.9 Proof of Theorem 4.8.1

(i) We first show (4.26). We have

$$\inf_{P_\pi^N\in L_{\text{CO}}^N} \int P_\pi^N(d\underline{\gamma})\mu^N(d\omega_0, d\underline{y})c^N(\underline{\gamma}, \underline{y}, \omega_0)$$

$$\geq \inf_{P_\pi^N\in L_{\text{CO}}^N\cap L_{\text{EX}}\big|_N} \int P_\pi^N(d\underline{\gamma})\mu^N(d\omega_0, d\underline{y})c^N(\underline{\gamma}, \underline{y}, \omega_0) - \epsilon_N \tag{4.91}$$

$$= \inf_{P_\pi^N\in L_{\text{PR,SYM}}^N} \int P_\pi^N(d\underline{\gamma})\mu^N(d\omega_0, d\underline{y})c^N(\underline{\gamma}, \underline{y}, \omega_0) - \epsilon_N, \tag{4.92}$$

where $L_{\text{EX}}|_N$ denotes the set of N-DM randomized policies which are the restrictions of policies in L_{EX} to the N first components. By Lemma 4.5.1 since L^N_{CO} is convex, without losing global optimality, we can optimize over $L^N_{\text{CO}} \cap L^N_{\text{EX}}$. Let $\epsilon > 0$, and consider $P^{*,N}_{\pi,\epsilon} \in L^N_{\text{CO}} \cap L^N_{\text{EX}}$ such that

$$\inf_{P^N_\pi \in L^N_{\text{CO}} \cap L^N_{\text{EX}}} \int P^N_\pi(d\underline{\gamma})\mu^N(d\omega_0, d\underline{y})c^N(\underline{\gamma}, \underline{y}, \omega_0) \geq \int P^{*,N}_{\pi,\epsilon}(d\underline{\gamma})\mu^N(d\omega_0, d\underline{y})c^N(\underline{\gamma}, \underline{y}, \omega_0) - \epsilon. \tag{4.93}$$

Following from the proof of Lemma 4.5.2, using $P^{*,N}_{\pi,\epsilon} \in L^N_{\text{CO}} \cap L^N_{\text{EX}}$ and by considering the indexes as a sequence of i.i.d. random variables with uniform distribution on the set $\{1, \ldots, N\}$, we can construct an infinitely-exchangeable policy $P^{*,\infty}_{\pi,\epsilon}$ where the restriction of an infinitely-exchangeable policy to N first components $P^{*,\infty}_{\pi,N,\epsilon} \in L^N_{\text{CO}} \cap L_{\text{EX}}|_N$, satisfies

$$\int P^{*,\infty}_{\pi,N,\epsilon}(d\underline{\gamma})\mu^N(d\omega_0, d\underline{y})c^N(\underline{\gamma}, \underline{y}, \omega_0) \int P^{*,N}_{\pi,\epsilon}(d\underline{\gamma})\mu^N(d\omega_0, d\underline{y})c^N(\underline{\gamma}, \underline{y}, \omega_0) + \epsilon_N. \tag{4.94}$$

Hence, (4.93) and (4.94) imply that

$$\begin{aligned}\inf_{P^N_\pi \in L^N_{\text{CO}} \cap L^N_{\text{EX}}} &\int P^N_\pi(d\underline{\gamma})\mu^N(d\omega_0, d\underline{y})c^N(\underline{\gamma}, \underline{y}, \omega_0) \\ &\geq \inf_{P^N_\pi \in L^N_{\text{CO}} \cap L_{\text{EX}}|_N} \int P^N_\pi(d\underline{\gamma})\mu^N(d\omega_0, d\underline{y})c^N(\underline{\gamma}, \underline{y}, \omega_0) - \epsilon - \epsilon_N.\end{aligned}$$

Since ϵ is arbitrary, this implies (4.91). By Theorem 4.4.1, without losing optimality, we can optimize over $L^N_{\text{CO,SYM}}$. Equality (4.92) is true since $L^N_{\text{CO,SYM}}$ is convex with extreme points in $L^N_{\text{PR,SYM}}$, and the map $\int P^N_\pi(d\underline{\gamma})\mu^N(d\omega_0, d\underline{y})c^N(\underline{\gamma}, \underline{y}, \omega_0) : L^N_{\text{CO,SYM}} \to \mathbb{R}$ is linear.

Now, we show (4.27) holds. We have

$$\begin{aligned}\inf_{P^N_\pi \in L^N_{\text{D}}} \int P^N_\pi(d\underline{\gamma})\mu^N(d\omega_0, d\underline{y})c^N(\underline{\gamma}, \underline{y}, \omega_0) &= \inf_{P^N_\pi \in L^N_{\text{PR}}} \int P^N_\pi(d\underline{\gamma})\mu^N(d\omega_0, d\underline{y})c^N(\underline{\gamma}, \underline{y}, \omega_0) \qquad (4.95) \\ &\geq \inf_{P^N_\pi \in L^N_{\text{PR,SYM}}} \int P^N_\pi(d\underline{\gamma})\mu^N(d\omega_0, d\underline{y})c^N(\underline{\gamma}, \underline{y}, \omega_0) - \epsilon_N, \qquad (4.96)\end{aligned}$$

where (4.95) follows from Blackwell's irrelevant information theorem [28] and since L_{CO}^N is convex with extreme points in L_{PR}^N and the map $\int P_\pi^N(d\underline{\gamma})\mu^N(d\omega_0, d\underline{y})c^N(\underline{\gamma}, \underline{y}, \omega_0) : L_{\text{CO}}^N \to \mathbb{R}$ is linear, hence, without losing optimality, we can optimaize over L_{CO}^N. Inequality (4.96) follows from (4.26) and this completes the proof of (i).

(ii) Let $P_\pi^* \in L_{\text{PR,SYM}}$ be an optimal policy of $(\mathcal{P}_\infty)$ and $P_{\pi,N}^*$ is the restriction of P_π^* to the first N components. Define for all $N \in \mathbb{N}$

$$a_N := \int P_{\pi,N}^*(d\underline{\gamma})\mu^N(d\omega_0, d\underline{y})c^N(\underline{\gamma}, \underline{y}, \omega_0)$$
$$b_N := \inf_{P_\pi^N \in L_{\text{PR,SYM}}^N} \int P_\pi^N(d\underline{\gamma})\mu^N(d\omega_0, d\underline{y})c^N(\underline{\gamma}, \underline{y}, \omega_0).$$

Following from (Step 4) of the proof of Theorem 4.5.1, since the cost function is bounded,

$$\limsup_{N\to\infty} \int P_{\pi,N}^*(d\underline{\gamma})\mu^N(d\omega_0, d\underline{y})c^N(\underline{\gamma}, \underline{y}, \omega_0)$$
$$= \limsup_{N\to\infty} \inf_{P_\pi^N \in L_{\text{PR,SYM}}^N} \int P_\pi^N(d\underline{\gamma})\mu^N(d\omega_0, d\underline{y})c^N(\underline{\gamma}, \underline{y}, \omega_0). \tag{4.97}$$

Hence, $\limsup_{N\to\infty} a_N = \limsup_{N\to\infty} b_N$. Following from (Step 4) of the proof of Theorem 4.5.1, and symmetry, $\lim_{N\to\infty} a_N = a < \infty$ and also there exists a subsequence such that $\lim_{k\to\infty} b_{N_k} = a < \infty$. On the other hand, since $a_N \geq b_N$ for all $N \in \mathbb{N}$, we can find $\tilde{\epsilon}_N \geq 0$ such that $a_N = b_N + \tilde{\epsilon}_N$. Taking limit as k goes to infinity from both sides, we have $a = \lim_{k\to\infty}(b_{N_k} + \epsilon_{N_k}) = a + \lim_{k\to\infty} \epsilon_{N_k}$. Hence, $\lim_{k\to\infty} \epsilon_{N_k} = 0$ since $\tilde{\epsilon}_N \geq 0$. Hence, there exists $\bar{\epsilon}_N \geq 0$ where $\bar{\epsilon}_N \to 0$ as N goes to infinity such that

$$\int P_{\pi,N}^*(d\underline{\gamma})\mu^N(d\omega_0, d\underline{y})c^N(\underline{\gamma}, \underline{y}, \omega_0)$$
$$\leq \inf_{P_\pi^N \in L_{\text{D}}^N} \int P_\pi^N(d\underline{\gamma})\mu^N(d\omega_0, d\underline{y})c^N(\underline{\gamma}, \underline{y}, \omega_0) + \epsilon_N + \bar{\epsilon}_N \tag{4.98}$$

where (4.98) follows from (4.27), and this completes the proof of (ii).

Chapter 5

Policy-Independent and Policy-Dependent Static Reductions of Stochastic Dynamic Teams

We observed earlier that the static reduction of a dynamic team problem facilitates optimality analysis. In this chapter, we study static reductions for stochastic dynamic team problems with a finite number of decision makers. In view of the intrinsic model, and the brief introduction on three types of static reductions in Chapter 1, we introduce the main objective of this chapter, which is to address mainly the following problems:

Problem 1: Connections between three optimality concepts (global optimality, person-by-person optimality, and stationarity) for stochastic dynamic teams and their policy-independent static reductions. Suppose that a policy-independent static reduction holds for a dynamic team problem. Let $\underline{\gamma}^{D,*} := \{\gamma_i^{D*}\}_{i\in\mathcal{N}}$ be a person-by-person optimal (globally optimal) policy for a dynamic problem. Is it true that $\underline{\gamma}^{D,*}$ is also a person-by-person optimal (globally optimal) policy for a dynamic problem under policy-independent static reductions? Is the converse statement valid?

Problem 2: Connections between three optimality concepts (global optimality, person-by-person optimality, and stationarity) for stochastic dynamic teams and their policy-dependent static reductions. Suppose that a policy-dependent static reduction holds for a dynamic team problem with a partially nested information structure. Let $\underline{\gamma}^{D,*} := \{\gamma_i^{D*}\}_{i\in\mathcal{N}}$ be a person-by-person

optimal (stationary, globally optimal) policy for a dynamic problem. Construct an admissible policy $\underline{\gamma}^{S,*} := \{\gamma_i^{S*}\}_{i\in\mathcal{N}}$ for the policy-dependent static reduction of the dynamic problem, realizing almost surely an identical control action ($u^i = \gamma_i^{D*}(y_i^D) = \gamma_i^{S*}(y_i^S)$). Is it true that $\underline{\gamma}^{S,*}$ is a person-by-person optimal (stationary, globally optimal) policy for the dynamic problem under the policy-dependent static reduction? Is the converse statement valid?

5.1 Literature Review and Contributions

As we have discussed in the introductory chapter and through our analysis for large-scale teams, the static reduction method has been shown to be very effective in arriving at existence, structural and approximation results. For existence results building on this approach, we refer the reader to [62, 140, 109, 138] and for approximations with finite models to [112]. For game problems, for an application on existence of saddle point equilibria, see [68]. Also, it has also been shown that for any sequential team problem with arbitrary information structures, a dynamic programming formulation can be formulated, see [133] for countable spaces and [138] for general spaces which lead to general existence results.

In this chapter, we study the connections between stationary (pbp optimal, globally optimal) policies for both types of static reductions. An interpretation of the connections between policies of dynamic teams and their static reductions is *(path-dependent) feedback* policies versus *noise feedforward* policies [24]. A subtlety of these connections for stationary (pbp optimal) policies can stem from the following observation: for dynamic teams, deviating a policy of a DM and fixing policies of others, requires a multi-directional deviation analysis since observations of frozen DMs depend on the deviating DM's policy. However, under a static reduction since observations of frozen DMs do not depend on the actions of the deviating DMs, only considering a single-directional deviation analysis of control actions is sufficient. Therefore, in general, establishing such connections between optimal/stationary policies are not obvious at all and can fail to hold even under a partially nested IS. In this chapter, we present negative results and also sufficient conditions for positive results on the connections between optimality concepts of dynamic teams

and their policy-independent and policy-dependent static reductions.

As we discussed in the introductory chapter, in the game theory literature, an interpretation of policies of dynamic teams and their static reductions is *closed-loop* versus *open-loop* policies. Related to the subtlety of these connections between open-loop and closed-loop policies, we refer the reader to the results in [41, 132, 129, 11, 17, 21, 46] on zero-sum games, and [9, 17] on nonzero-sum games. We again note that as we have observed through our analysis in Chapters 3 and 4, these connections are particularly important since the compactness and convexity conditions often can be established under more relaxed assumptions for policies under static reductions (open-loop policies), and hence, existence of such connections can facilitate the analysis for the convergence and existence of optimal policies in the context of large-scale teams/games and their mean-field limit (see also [56, 81, 39, 72, 83] for games). In Section 5.2.4, we focus on multi-stage team problems, where we first introduce two types of static reductions suitable for these problems, and then we establish connections between optimal policies of dynamic teams and their static reductions. Our preliminary results in Section 5.2.4 have shown to be important for analysis of large-scale teams (Chapters 3 and 4), but the detail study on the equivalence relation, approximation and convergence of open-loop and closed-loop (player-wise) pbp/Nash equilibrium policies for (mean-field) weakly interacting teams/games is left for future research. Notable results in this direction are: [83] on mean-field games, where it has been shown that in a sense open-loop and closed-loop converge together to mean-field equilibria; [58] on large games in economic theory literature, where it has been shown that as the number of players becomes large (the influence of individual player becomes small), open-loop and closed-loop equilibria are approximately the same.

Contributions. In view of the discussion above, this chapter makes the following observations and contributions.

(i) For policy-dependent static reductions of stochastic teams with partially nested information structure, in the reduced form, the cost functions are unaltered; this is not the case for the policy-independent static reduction of stochastic teams. On the other hand, the probability measure on the exogenous random variables do not change under the policy-dependent

static reduction, but does so in the policy-independent static reduction. Furthermore, for the policy-dependent case, the static-reduced policy/map from exogenous variables to actions change depending on the policies that precedent decision makers choose, but in the policy-independent case, this map is unaltered (as the measurement variables are interpreted as exogenous variables). For policy-dependent static reductions, it is essential that policies are deterministic; however, policy-independent static reduction applies even when the policies are randomized.

(ii) We show in Theorem 5.3.1 that there is a bijection between pbp optimal (globally optimal) policies (and under a further condition between stationary policies) of dynamic teams and their policy-independent static reductions (these connections are depicted in Fig. 5.1). This equivalence relationship follows from the fact that this static reduction is policy-independent.

(iii) While for global optimality, policy-dependent static reductions have equivalence optimality properties with the dynamic information setup; for the policy-dependent case, when one considers person-by-person optimal or stationary policies, significant subtleties emerge: a policy which is person-by-person optimal in one form may not be so in the other form (see Proposition 5.4.1). Under sufficient convexity and minor regularity conditions on the cost function (see Assumption 5.4.1) and a further regularity condition on policies and observations (see Condition (C)), we show in Theorem 5.4.2 that there is a bijection between stationary (pbp optimal) policies of dynamic teams and their policy-dependent static reductions (these connections are depicted in Fig. 5.2).

(iv) We define the reduction of dynamic stochastic teams with partially nested with control-sharing IS to ones with static measurements with control-sharing IS as static measurements with control-sharing reduction. We show that this reduction is independent of policies (see Theorems 5.5.1 and 5.5.2). Under this reduction, the connections between optimality concepts can be established under relaxed conditions at the price of introducing informational nonuniqueness in the representations of policies (see Section 5.5 for details and see Fig. 5.3

for summary of the connections).

(v) The above static measurements with control-sharing reduction has implications on convexity properties: a partially nested dynamic team is convex (in policies) if and only if its static measurements with control-sharing reduction is convex (see [140, Section 3]). This suggests that for convexity the policy-dependent static reduction requires actions to be shared (see Section 5.5.2).

(vi) Several examples are presented to explain both positive and negative results. Due to uniqueness of pbp optimal policies for LQG team problems under the policy-dependent static reduction, we establish stronger results for LQG models with a partially nested IS (Corollary 5.6.1).

(vii) We also study multi-stage team problems as a special setup of dynamic teams, and the connections between DM-wise/player-wise pbp optimality under two classes of static reductions, (i) independent data reduction under which the policy-independent reduction holds through players and time, and (ii) nested reduction under which measurements are independent through players but nested through time. We show that there is a bijection between player-wise person-by-person optimal policies (globally optimal policies) under both classes of reductions, but there is no bijection between decision maker-wise person-by-person optimal policies in general under the nested reduction (Corollary 5.7.1). Furthermore, we discuss the impact of the independent-data and nested reductions on the variational analysis (Corollary 5.7.2).

The organization of the chapter is as follows: In Section 5.2, we present preliminaries and provide a description of team problems within policy-independent and policy-dependent static reductions. In Sections 5.3, 5.4, and 5.5, we present results for dynamic teams under policy-independent, policy-dependent static reductions, and static measurements with control-sharing reduction, respectively. In Section 5.6, partially nested LQG problems under policy-dependent static reductions are studied, and Multi-stage team problems are studied in Section 5.7.

5.2 Information Structures and Policy-Independent and Policy-Dependent Static Reductions of Dynamic Teams

5.2.1 Dynamic Teams under Policy-Independent Static Reductions

Let action and observation spaces be subsets of appropriate Euclidean spaces, i.e., $\mathbb{U}^i \subseteq \mathbb{R}^{n_i}$ and $\mathbb{Y}^i \subseteq \mathbb{R}^{m_i}$, for $i \in \mathcal{N}$, where n_i and m_i are positive integers. We again recall a stochastic dynamic team problem as follows:

Problem ($\mathcal{P}$): Consider a team problem within the intrinsic model with observations for each DM given by (1.2). Let the IS of DM^i be $I^i = \{y^i\}$ (or $I^i = \{y^k\}_{k\in K_i}$ for a subset $K_i \subseteq \{1, \ldots, i\}$). An expected cost function (to be minimized) under a policy $\underline{\gamma} = (\gamma^1, \cdots, \gamma^N) \in \Gamma$ as

$$J(\underline{\gamma}) = E^{\underline{\gamma}}[c(\omega_0, \underline{u})] := E[c(\omega_0, \gamma^1(y^1), \cdots, \gamma^N(y^N))], \tag{5.1}$$

for some Borel measurable cost function $c : \Omega_0 \times \prod_{i=1}^N \mathbb{U}^i \to \mathbb{R}_+$ and $\underline{u} := \{u^1, \cdots, u^N\}$. □

We first recall Witsenhausen's static reduction (see [135, 138]), and then provide a description of team problems under this static reduction. Now, we introduce an absolute continuity condition under which a policy-independent static reduction exists (see [138] for further discussions).

Assumption 5.2.1. *If for every $i \in \mathcal{N}$, there exists a probability measure Q^i on $\mathbb{Y}^i$ and a function f^i such that for any Borel set A^i,* (1.3) *holds.*

Denote the joint distribution on $(\omega_0, u^1, \ldots, u^N, y^1, \ldots, y^N)$ by $\mathbb{P}$, and the distribution of ω_0 by $\mathbb{P}^0$. If the preceding absolute continuity condition holds, then there exists a joint reference distribution $\mathbb{Q}$ on $(\omega_0, u^1, \ldots, u^N, y^1, \ldots, y^N)$ such that the distribution $\mathbb{P}$ is absolutely continuous with respect to $\mathbb{Q}$ ($\mathbb{P} \ll \mathbb{Q}$), where for every Borel set A on $(\Omega_0 \times \prod_{i=1}^N (\mathbb{U}^i \times \mathbb{Y}^i))$

$$\mathbb{P}(A) = \int_A \frac{d\mathbb{P}}{d\mathbb{Q}} \mathbb{Q}(d\omega_0, du^1, \ldots, du^N, dy^1, \ldots, dy^N), \tag{5.2}$$

where a joint reference distribution and Radon-Nikodym derivative (defined P-almost surely) are as follows:

$$\mathbb{Q}(d\omega_0, du^1, \ldots, du^N, dy^1, \ldots, dy^N) := \mathbb{P}^0(d\omega_0) \prod_{i=1}^{N} Q^i(dy^i) 1_{\{\gamma^i(y^i) \in du^i\}}, \tag{5.3}$$

$$\frac{d\mathbb{P}}{d\mathbb{Q}} := \frac{d\mathbb{P}}{d\mathbb{Q}}(\omega_0, u^1, \ldots, u^1, y^1 \ldots, y^N) := \prod_{i=1}^{N} f^i(y^i, \omega_0, u^1, \ldots, u^{i-1}, y^1, \ldots, y^{i-1}). \tag{5.4}$$

We note that, if $I^i = \{y^k\}_{k \in K_i}$ for a subset $K_i \subseteq \{1, \ldots, i\}$, in the above $1_{\{\gamma^i(y^i) \in du^i\}}$ needs to be replaced by $1_{\{\gamma^i(I^i) \in du^i\}}$. In view of the above derivations, we now formally introduce policy-independent static reductions as follows:

Definition (**Policy-Independent Static Reduction).** *For stochastic teams ($\mathcal{P}$) with a given IS under Assumption 5.2.1, a policy-independent static reduction is defined as a change of measure* (5.2) *under which measurements y^i for each DM (defined in* (1.2)*) have a distribution Q^i and the expected cost function can be written as follows:*

$$J(\underline{\gamma}) := E^{\underline{\gamma}}_{\mathbb{Q}}[\tilde{c}(\omega_0, u^1, \ldots, u^N, y^1, \ldots, y^N)], \tag{5.5}$$

where the new cost function under the reduction is

$$\tilde{c}(\omega_0, u^1, \ldots, u^N, y^1, \ldots, y^N) := c(\omega_0, u^1, \ldots, u^N) \frac{d\mathbb{P}}{d\mathbb{Q}}. \tag{5.6}$$

□

We now recall definitions of global optimality, person-by-person (pbp) optimality, and stationarity policies for team problems ($\mathcal{P}$).

Definition (**Optimality concepts for dynamic problem ($\mathcal{P}$)).** *For stochastic teams ($\mathcal{P}$) with a given IS:*

- *a policy* $\underline{\gamma}^* := (\gamma^{1*}, \ldots, \gamma^{N*}) \in \Gamma$ *is globally optimal if*

$$J(\underline{\gamma}^*) = \inf_{\underline{\gamma} \in \Gamma} J(\underline{\gamma}) := \inf_{\underline{\gamma} \in \Gamma} E^{\underline{\gamma}}_{\mathbb{P}}[c(\omega_0, u^1, \ldots, u^N)],$$

- *a policy* $\underline{\gamma}^* \in \Gamma$ *is pbp optimal if for all* $\beta \in \Gamma^i$ *and all* $i \in \mathcal{N}$*,*

$$J(\underline{\gamma}^*) \leq J(\underline{\gamma}^{-i*}, \beta) := E^{(\underline{\gamma}^{-i*}, \beta)}_{\mathbb{P}}[c(\omega_0, u^1, \ldots, u^N)],$$

where $(\underline{\gamma}^{-i*}, \beta) := (\gamma^{1*}, \ldots, \gamma^{(i-1)*}, \beta, \gamma^{(i+1)*}, \ldots, \gamma^{N*})$*,*

- *a policy* $\underline{\gamma}^* \in \Gamma$ *is stationary if, for all* $i \in \mathcal{N}$*,*

$$\nabla_{u^i} E_{\mathbb{P}}\Big[c\Big(\omega_0, \gamma^{1*}(y^{1*}), \ldots, \gamma^{(i-1)*}(y^{(i-1)*}), u^i, \gamma^{(i+1)*}(y^{i+1}_{u^i}), \ldots, \gamma^{N*}(y^N_{u^i})\Big)\Big| y^{i*}\Big]\Big|_{u^i = \gamma^{i*}(y^{i*})} = 0 \ P\text{-a.s.},$$

where for any $j \in \mathcal{N}$

$$y^{j*} = h^j\Big(\omega_0, \omega_j, \gamma^{1*}(y^{1*}), \ldots, \gamma^{(j-1)*}(y^{(j-1)*}), y^{1*}, \ldots, y^{(j-1)*}\Big), \tag{5.7}$$

$$y^{j*}_{u^i} = h^j\Big(\omega_0, \omega_j, \gamma^{1*}(y^{1*}), \ldots, \gamma^{(i-1)*}(y^{(i-1)*}), u^i, \gamma^{(i+1)*}(y^{(i+1)*}_{u^i}), \ldots, \gamma^{(j-1)*}(y^{(j-1)*}_{u^i}),$$
$$y^{1*}, \ldots, y^{i*}, y^{(i+1)*}_{u^i}, \ldots, y^{(j-1)*}_{u^i}\Big). \tag{5.8}$$

Next, we recall the definitions of stationary (pbp optimal, globally optimal) policies for dynamic team problems under the preceding policy-independent static reduction.

Definition (**Optimality concepts under policy-independent static reduction).** *For stochastic teams (*$\mathcal{P}$*) with a given IS under a policy-independent static reduction:*

- *a policy* $\underline{\gamma}^* \in \Gamma$ *is globally optimal if*

$$J(\underline{\gamma}^*) = \inf_{\underline{\gamma} \in \Gamma} J(\underline{\gamma}) := \inf_{\underline{\gamma} \in \Gamma} E^{\underline{\gamma}}_{\mathbb{Q}}[\tilde{c}(\omega_0, u^1, \ldots, u^N, y^1, \ldots, y^N)],$$

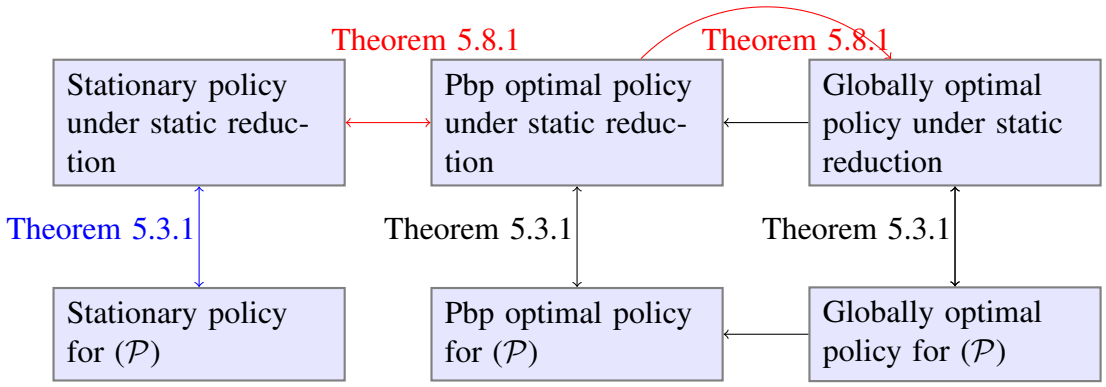

Figure 5.1: Diagram of the connections between three optimality concepts in dynamic teams and their policy-independent static reductions. The blue arrow connection is established under additional condition (5.22), and the red arrow connections are true under the convexity and regularity assumption on the cost function under policy-independent static reduction.

- *a policy* $\underline{\gamma}^* \in \Gamma$ *is pbp optimal if for all* $\beta^i \in \Gamma^i$ *and all* $i \in \mathcal{N}$,

$$J(\underline{\gamma}^*) \leq J(\underline{\gamma}^{-i*}, \beta) := E_{\mathbb{Q}}^{(\underline{\gamma}^{-i*}, \beta)}[\tilde{c}(\omega_0, u^1, \ldots, u^N, y^1, \ldots, y^N)],$$

- *a policy* $\underline{\gamma}^* \in \Gamma$ *is a stationary policy if* P*-almost surely,*

$$\nabla_{u^i} E_{\mathbb{Q}}\left[\tilde{c}\Big(\omega_0, \gamma^{1*}(y^1), \ldots, \gamma^{(i-1)*}(y^{i-1}), u^i, \gamma^{(i+1)*}(y^{i+1}) \ldots, \gamma^{N*}(y^N), \underline{y}\Big)\,\Big|\, y^i\right]\Big|_{u^i=\gamma^{i*}(y^i)} = 0$$

where $\underline{y} := (y^1, \ldots, y^N)$. □

One of our goals here is to study the connections between Definition 5.2.1 and Definition 5.2.1. In Section 5.3, we show the existence of a bijection between pbp optimal (globally optimal) policies of dynamic teams and their policy-independent static reductions, and between stationary policies of dynamic teams and their policy-independent static reductions under a further condition on the Randon-Nikodym derivative (see (5.22)). These connections are depicted in Fig 5.1.

5.2.2 Partially Nested Dynamic Teams under Policy-Dependent Static Reduction

In the following, we first briefly recall Ho and Chu's static reduction [66, 67] (the policy-dependent static reduction), and then we provide a description of dynamic team problems under the policy-dependent static reduction.

Consider dynamic teams with partially nested IS, where observations of DMs are defined as (1.6). Let $I_D^i = \{y_i^D\}$, and the set of admissible policies for dynamic teams be as

$$\Gamma^D := \left\{ \underline{\gamma}^D := (\gamma_1^D, \ldots, \gamma_N^D) \;\middle|\; u_i^D = \gamma_i^D(y_i^D) \text{ for all } i \in \mathcal{N} \right\}.$$

Under the above formulation, we introduce a class of dynamic team problems as follows:

Problem ($\mathcal{P}^D$): For stochastic teams with I_D^i (with measurements y_i^D defined in (1.6)) for all $i \in \mathcal{N}$, consider an expected cost function under the policy $\underline{\gamma}^D$ as (5.1). Derive a policy $\underline{\gamma}^{D,*} := (\gamma_1^{D,*}, \ldots, \gamma_N^{D,*}) \in \Gamma^D$ that is *globally optimal* for ($\mathcal{P}^D$), that is

$$J(\underline{\gamma}^{D,*}) = \inf_{\underline{\gamma}^D \in \Gamma^D} J(\underline{\gamma}^D).$$

Furthermore, obtain a policy $\underline{\gamma}^{D,*} \in \Gamma^D$ that is *pbp optimal* for ($\mathcal{P}^D$), that is

$$J(\underline{\gamma}^{D,*}) = \inf_{\gamma_i^D \in \Gamma_i^D} J(\gamma_i^D, \underline{\gamma}_{-i}^{D*}) \quad \text{for all} \quad i \in \mathcal{N}, \tag{5.9}$$

where $\underline{\gamma}_{-i}^{D*} = (\gamma_1^{D*}, \ldots, \gamma_{i-1}^{D*}, \gamma_{i+1}^{D*}, \ldots, \gamma_N^{D*})$. □

Now, we introduce an assumption under which the policy-dependent static reduction exists [66, 67].

Assumption 5.2.2. *For all $i \in \mathcal{N}$ and for every fixed $u_{\downarrow i}^D$, the function $g_i(\cdot, u_{\downarrow i}^D) : h_i(\zeta) \mapsto \hat{y}_i^D$ is invertible for all realizations of ζ.*

Following the policy-dependent static reduction introduced in [66, 67], and under Assumption

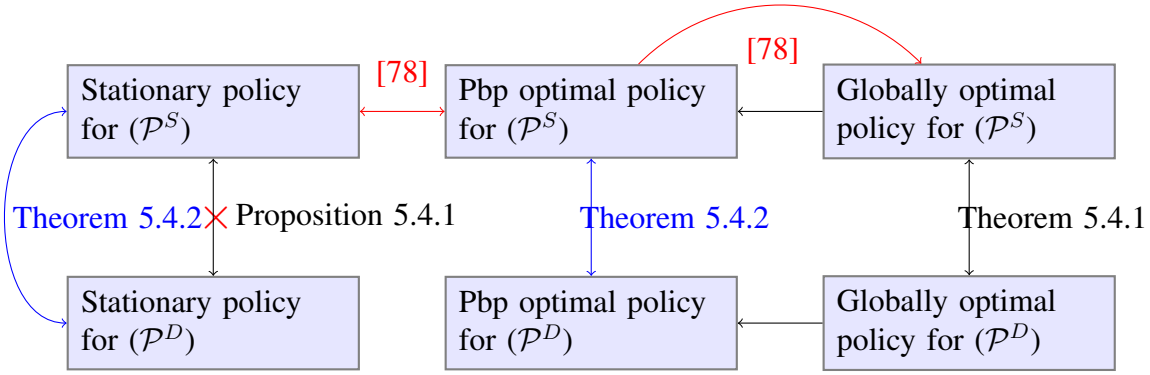

Figure 5.2: Diagram of the connections between three optimality concepts for dynamic teams with partially nested ISs and their policy-dependent reductions. The red arrow connections hold under Assumption 5.4.1. The blue arrow connections require in addition to Assumption 5.4.1 some further conditions (see assumptions in Theorem 5.4.2).

5.2.2, given a policy $\underline{\gamma}^D$, we can define the observations within the policy-dependent static reduction as (1.8). Let the IS of DM^i be $I_S^i = \{y_i^S\}$. Define the set of admissible policies for dynamic teams within the policy-dependent static reduction as follows:

$$\Gamma^S := \left\{ \underline{\gamma}^S = (\gamma_1^S, \ldots, \gamma_N^S) \;\middle|\; u_i^S = \gamma_i^S(y_i^S) \text{ for all } i \in \mathcal{N} \right\}.$$

To see various examples, where the above policy-dependent static reduction exists, we refer the reader to [66, 67]. We define team problems under the policy-dependent static reduction as follows:

Problem ($\mathcal{P}^S$): For stochastic teams with I_S^i (with measurements y_i^S defined in (1.8)) for all $i \in \mathcal{N}$, consider an expected cost function under policy $\underline{\gamma}^S$ as (5.1). Derive a policy $\underline{\gamma}^{S,*} := (\gamma_1^{S,*}, \ldots, \gamma_N^{S,*}) \in \Gamma^S$ that is *globally optimal* for ($\mathcal{P}^S$)

$$J(\underline{\gamma}^{S,*}) = \inf_{\underline{\gamma}^S \in \Gamma^S} J(\underline{\gamma}^S).$$

Furthermore, obtain a policy $\underline{\gamma}^{S,*} \in \Gamma^S$ that is *pbp optimal* for $(\mathcal{P}^S)$, that is

$$J(\underline{\gamma}^{S,*}) = \inf_{\gamma_i^S \in \Gamma_i^S} J(\gamma_i^S, \underline{\gamma}_{-i}^{S*}) \quad \text{for all} \quad i \in \mathcal{N}. \tag{5.10}$$

□

In view of the above derivations, we now formally introduce policy-dependent static reductions as follows:

Definition (Policy-Dependent Static Reduction). *Consider a partially nested stochastic dynamic team $(\mathcal{P}^D)$ with a given IS, I_D^i, where Assumption 5.2.2 holds. Policy-dependent static reduction is defined as the reduction of a stochastic dynamic team $(\mathcal{P}^D)$ to a static one $(\mathcal{P}^S)$ (which has an equivalent IS, I_S^i), where under the reduction, the cost function is unaltered and measurements are static), and for a given admissible policy $\underline{\gamma}^D \in \Gamma^D$, an admissible policy $\underline{\gamma}^S \in \Gamma^S$ can be constructed through a relation*

$$u^i = \gamma_i^S(y_i^S) = \gamma_i^D(y_i^D) \text{ for } P\text{-almost every } \zeta \tag{5.11}$$

for all $i \in \mathcal{N}$. □

Some of our results in this chapter address the following question: Given a stationary (pbp optimal, globally optimal) policy $\underline{\gamma}^{S,*} \in \Gamma^S$ for $(\mathcal{P}^S)$, is a policy $\underline{\gamma}^{D,*} \in \Gamma^D$, constructed through relation (5.11), stationary (pbp optimal, globally optimal) policy for $(\mathcal{P}^D)$? Is the converse statement also true?

Remark 5.2.1.

(i) In contrast to the policy-independent static reduction, which also holds for randomized policies, the above policy-dependent static reduction requires DMs to have access to the actions of precedent DMs according to a partially nested IS; hence, it requires the policies to be deterministic (that is pure). On the other hand, for team problems, globally optimal policies can be chosen among those that are deterministic without any loss of optimality [140, Theorems 2.3 and 2.5].

(ii) We also note that in the policy-dependent static reduction, in contrast to the policy-independent static reduction, the cost function will not change under the static reduction.

(iii) A relation (5.11) *in the construction of policies under the policy-dependent static reduction can be viewed as a composition of policies with a bijection* $F^i_{\underline{\gamma}} : y^D_i \mapsto y^S_i$ *with the inverse* $(F^i_{\underline{\gamma}})^{-1} : y^S_i \mapsto y^D_i$ *for all* $i \in \mathcal{N}$*, where the existence of this bijection follows from Assumption 5.2.2 (we note that the subscript* $\underline{\gamma}$ *in a bijection denotes the fact that a bijection depends on the precedent policies for each* $i \in \mathcal{N}$*). That is, for any given policy* $\underline{\gamma}^D \in \Gamma^D$*, an admissible policy* $\underline{\gamma}^S \in \Gamma^S$ *can be constructed as* $\gamma^S_i := \gamma^D_i \circ F^i_{\gamma^S_{\downarrow i}}$*, and for any given policy* $\underline{\gamma}^S$*, an admissible policy* $\underline{\gamma}^D$ *can be constructed as* $\gamma^D_i := \gamma^S_i \circ (F^i_{\gamma^D_{\downarrow i}})^{-1}$.

In Section 5.4, we first show that the answer to the aforementioned question is affirmative for globally optimal policies of dynamic teams and their policy-dependent static reduction. For pbp optimal and stationary policies, however, our results show that, under the policy-dependent static reduction, there might not exist a bijection between stationary (pbp optimal) policies of dynamic teams and their policy-dependent static reduction in general. These connections are depicted in Fig. 5.2. This is quite opposite of the result for the policy-independent static reduction (see Theorem 5.3.1), where there is a bijection between stationary (pbp optimal, globally optimal) policies of dynamic teams and their policy-independent static reduction. We also present sufficient conditions such that the answer to the above question is affirmative for pbp optimal and stationary policies. Several examples including LQG models are presented.

5.2.3 Partially Nested with Control-Sharing IS and Static Measurements with Control-Sharing Reduction

To establish connections between pbp optimality and convexity of dynamic teams and their policy-dependent static reductions, we introduce *(dynamic) partially nested with control-sharing* team problems, where we expand the information structure such that in addition to observations, actions are also shared (this expansion is consistent with partially nested IS), i.e., for each DM^i,

observations are given by (1.9) with $I_i^{D,CS} := \{y_i^{D,CS}\}$.

Problem ($\mathcal{P}^{D,CS}$): For stochastic teams with $I_i^{D,CS}$ with measurements (1.9) for all $i \in \mathcal{N}$, consider an expected cost function (to be minimized) under policy $\underline{\gamma}^{D,CS}$ as (5.1). □

Under the invertibility condition (Assumption 5.2.2), there is a bijection between dynamic observations y_i^D and static one y_i^S for each $i \in \mathcal{N}$, and hence, this allows us to reduce the dynamic problem to another dynamic problem where measurements are static. We refer to this reduction as *static measurements with control-sharing*. Observations within this reduction are given by (1.10) with $I_i^{CS} := \{y_i^{CS}\}$. □

Problem ($\mathcal{P}^{CS}$): For stochastic teams with I_i^{CS} with measurements (1.10) for all $i \in \mathcal{N}$, consider an expected cost function (to be minimized) under policy $\underline{\gamma}^{CS}$ as (5.1). □

We refer the above problems as *static measurements control-sharing* stochastic team problems. We now formally introduce *static measurements with control-sharing* reduction as follows:

Definition (Static Measurements with Control-Sharing Reduction). *Consider a dynamic partially nested control-sharing stochastic dynamic team ($\mathcal{P}^{D,CS}$) with a given IS, $I_i^{D,CS}$, where Assumption 5.2.2 holds. Static measurements with control-sharing reduction is defined as the reduction of a stochastic dynamic team ($\mathcal{P}^{D,CS}$) to a static measurement control-sharing problem ($\mathcal{P}^{CS}$) with IS, I_i^{CS}, where under the reduction, the cost is unaltered and the measurements are static (see* (1.10)*), and for a given admissible policy $\underline{\gamma}^{D,CS}$ for ($\mathcal{P}^{D,CS}$), an admissible policy $\underline{\gamma}^{CS}$ for ($\mathcal{P}^{CS}$) can be constructed for each $i \in \mathcal{N}$, through the relation*

$$\gamma_i^{D,CS}(y_i^{D,CS}) = \gamma_i^{CS}(y_i^{CS}) \text{ for } P\text{-almost every } \zeta \text{ and } u^{\downarrow i} \tag{5.12}$$

for all $i \in \mathcal{N}$. □

In Section 5.4, based on static measurements with control-sharing reduction, we establish the connections between pbp optimal (globally optimal) policies ($\mathcal{P}^D$), ($\mathcal{P}^S$), ($\mathcal{P}^{CS}$), and ($\mathcal{P}^{D,CS}$). We

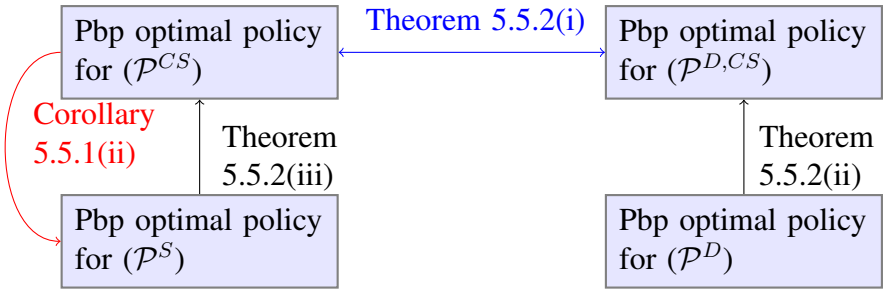

Figure 5.3: Diagram of the connections between three optimality concepts for dynamic teams with partially nested ISs and static measurements with control-sharing reduction. The blue arrow connection hold under Assumption 5.2.2 and the red arrow connection hold under further Assumption 5.4.1 and a regularity condition on policies.

also study convexity of team problems under the policy-dependent static reduction and under the static measurements with control-sharing reduction. The summary of the connections between pbp optimal policies ($\mathcal{P}^{D}$), ($\mathcal{P}^{S}$), ($\mathcal{P}^{CS}$), and ($\mathcal{P}^{D,CS}$) are depicted in Fig. 5.3.

5.2.4 Multi-Stage Team Problems and Static Reductions

In this subsection, we consider multi-stage stochastic dynamic team problems and introduce two static reductions in the connection to policy-independent and policy-dependent static reductions introduced in preceding subsections for single-stage team problems. We will introduce a new static reduction concept building on the one introduced by Witsenhausen (called independent-data reduction) [135, Section 2.4] and the other one in Chapter 3. The underlying idea is to view DMs acting in sequence with increasing information as a single player with a larger action space. This facilitates our optimality analysis. We note that this approach was utilized to establish structural and existence results in Chapter 3 ([119, Section 3.2]).

Problem ($\mathcal{P}^{\text{Multi}}$): Consider the following formulation of multi-stage stochastic teams:

(i) The state dynamics and observations for $t \in \mathcal{T} := \{0, \dots, T-1\}$ are given by

$$x_{t+1} = f_t(x_{0:t-1}, u_{0:t-1}^{1:N}, w_t), \tag{5.13}$$

$$y_t^i = h_t^i(x_{0:t-1}, u_{0:t-1}^{1:N}, v_t^i), \tag{5.14}$$

where f_t and h_t^i are measurable functions and $x_{0:t-1} := (x_0, \ldots, x_{t-1})$, and $w_t, v_t^1, \ldots, v_t^N$ for all $t \in \mathcal{T}$ are random variables taking values in standard Borel spaces, and also we denote $u_{0:t-1}^{1:N} := (u_0^1, \ldots, u_{t-1}^1, \ldots, u_0^N, \ldots, u_{t-1}^N)$. Introduce appropriate collections of DMs as players, with the i-th player (PL^i) for $i \in \mathcal{N}$ acting at different time instances $t \in \mathcal{T}$ and comprised of DM$_0^i$, ..., DM$_{T-1}^i$.

(ii) The observation, action, state, and disturbances spaces are standard Borel spaces with $\mathbf{Y}^i := \prod_{t=0}^{T-1} \mathbb{Y}_t^i$, $\mathbf{U}^i := \prod_{t=0}^{T-1} \mathbb{U}_t^i$, $\mathbf{X} := \prod_{t=0}^{T-1} \mathbb{X}_t$, $\mathbf{W}^i := \prod_{t=0}^{T-1} \mathbb{W}_t^i$, $\mathbf{V}^i := \prod_{t=0}^{T-1} \mathbb{V}_t^i$, respectively.

(iii) An admissible policy for PL^i is defined as $\boldsymbol{\gamma^i} \in \mathbf{\Gamma^i}$, where $\boldsymbol{\gamma^i} := (\gamma_0^i, \ldots, \gamma_{T-1}^i)$ and $\mathbf{\Gamma^i} = \prod_{t=0}^{T-1} \Gamma_t^i$. Given an IS $I_t^i \subseteq \{y_{0:t}^{1:N}, u_{0:t-1}^{1:N}\}$, each admissible policy γ_t^i is a measurable function with $u_t^i = \gamma_t^i(I_t^i)$.

(iv) A multi-stage expected cost function under a policy $\underline{\boldsymbol{\gamma}}$ is given as

$$J(\underline{\boldsymbol{\gamma}}) = E^{\underline{\boldsymbol{\gamma}}}\left[\sum_{t=0}^{T-1} c_t(\omega_0, x_t, u_t^1, \ldots, u_t^N)\right], \tag{5.15}$$

for some Borel measurable cost function $c : \Omega_0 \times \mathbb{X}_t \times \prod_{i=1}^N \mathbb{U}_t^i \to \mathbb{R}_+$, where $\underline{\boldsymbol{\gamma}} := (\boldsymbol{\gamma^1}, \boldsymbol{\gamma^2}, \ldots, \boldsymbol{\gamma^N})$, and again ω_0 is a Ω_0-valued cost function-relevant exogenous random variable, $\omega_0 : (\Omega, \mathcal{F}, P) \to (\Omega_0, \mathcal{F}_0)$, where Ω_0 is a Borel space with its Borel σ-field $\mathcal{F}_0$.

Determine (existence and characterization of) the policy $\underline{\boldsymbol{\gamma}}$ that minimizes (5.15) over $\prod_{i=1}^N \mathbf{\Gamma^i}$.

□

Definition. *For multi-stage stochastic teams, a policy $\underline{\boldsymbol{\gamma}}^*$ is player-wise pbp optimal if for all $i \in \mathcal{N}$ and for all $\boldsymbol{\beta} \in \mathbf{\Gamma^i}$,*

$$J(\underline{\boldsymbol{\gamma}}^*) \leq J(\boldsymbol{\gamma}^{-i,*}, \boldsymbol{\beta}), \tag{5.16}$$

where $\boldsymbol{\gamma}^{-i,*} := (\boldsymbol{\gamma}^{1*}, \ldots, \boldsymbol{\gamma}^{i-1*}, \boldsymbol{\gamma}^{i+1*}, \ldots, \boldsymbol{\gamma}^{N*})$. *A policy* $\underline{\boldsymbol{\gamma}}^*$ *is player-wise globally optimal if for all* $\underline{\boldsymbol{\gamma}} \in \prod_{i=1}^{N} \mathbf{\Gamma^i}$,

$$J(\underline{\boldsymbol{\gamma}}^*) \leq J(\underline{\boldsymbol{\gamma}}). \tag{5.17}$$

□

Comparing Definition 5.2.4 and Definition 5.2.1, we can see that concepts of player-wise and (DM-wise) global optimality are equivalent. Also, every player-wise pbp optimal policy is DM-wise pbp optimal; however, the converse statement is not true in general. That is because, in the definition of player-wise pbp optimality, in contrast to the definition of (DM-wise) pbp optimality, policies $(\gamma_0^i, \ldots, \gamma_{t-1}^i, \gamma_{t+1}^i, \ldots, \gamma_{T-1}^i)$ are not frozen. Later on, we provide sufficient conditions for the converse statement to hold based on our static reductions for multi-stage problems. However, we note that clearly if there is a unique DM-wise pbp optimal policy, then it is a unique player-wise pbp optimal policy. In the following, we first, introduce two assumptions which will be used for our static reductions, and then introduce a static reduction for multi-stage stochastic teams.

Assumption 5.2.3. *Assume that for every PL*^{i} *and for every* $t \in \mathcal{T}$, *there exists a probability measure* $\tilde{Q}_t^i$ *on* $\mathbb{Y}_t^i$ *and measurable function* ϕ_t^i *such that for all Borel sets* $A = A^1 \times \ldots \times A^N$ *with* A^i *in* $\mathbb{Y}_t^i$, *we have*

$$\begin{aligned} &P\Big((y_t^1, \ldots, y_t^N) \in A \;\Big|\; \omega_0, x_0, v_{0:t-1}^{1:N}, w_{0:t-1}^{1:N}, y_{0:t-1}^{1:N}, u_{0:t-1}^{1:N}\Big) \\ &\quad = \prod_{i=1}^{N} \int_{A^i} \phi_t^i(y_t^i, \omega_0, x_0, v_{0:t-1}^{1:N}, w_{0:t-1}^{1:N}, y_{0:t-1}^{1:N}, u_{0:t-1}^{1:N}) \tilde{Q}_t^i(dy_t^i). \end{aligned} \tag{5.18}$$

Let $\tilde{\mathbb{P}}$ be the joint distribution on $(\omega_0, x_0, \underline{\boldsymbol{w}}, \underline{\boldsymbol{v}}, \underline{\boldsymbol{u}}, \underline{\boldsymbol{y}})$, and μ be the fixed joint distribution on $(\omega_0, x_0, \underline{\boldsymbol{w}}, \underline{\boldsymbol{v}})$. Let $\underline{\boldsymbol{z}} := (\boldsymbol{z}^1, \ldots, \boldsymbol{z}^N)$ and $\boldsymbol{z}^i := (z_0^i, \ldots, z_{T-1}^i)$ for $z = u, y, w, v$ and $i \in \mathcal{N}$. Hence, under the preceding change of measure (5.18), there exists a joint reference distribution $\tilde{\mathbb{Q}}$ on $(\omega_0, x_0, \underline{\boldsymbol{w}}, \underline{\boldsymbol{v}}, \underline{\boldsymbol{u}}, \underline{\boldsymbol{y}})$ such that $\tilde{\mathbb{P}} \ll \tilde{\mathbb{Q}}$, where for every Borel set B on $(\Omega_0 \times \mathbb{X}_0 \times \prod_{i=1}^{N} (\mathbf{W}^i \times$

$\mathbf{V}^i \times \mathbf{U}^i \times \mathbf{Y}^i))$

$$\tilde{\mathbb{P}}(B) = \int_B \frac{d\tilde{\mathbb{P}}}{d\tilde{\mathbb{Q}}} \tilde{\mathbb{Q}}(d\omega_0, dx_0, d\underline{\boldsymbol{w}}, d\underline{\boldsymbol{v}}, d\underline{\boldsymbol{u}}, d\underline{\boldsymbol{y}}), \tag{5.19}$$

where

$$\tilde{\mathbb{Q}}(d\omega_0, dx_0, d\underline{\boldsymbol{w}}, d\underline{\boldsymbol{v}}, d\underline{\boldsymbol{u}}, d\underline{\boldsymbol{y}}) := \mu(d\omega_0, dx_0, d\underline{\boldsymbol{w}}, d\underline{\boldsymbol{v}}) \prod_{t=0}^{T-1} \prod_{i=1}^{N} \tilde{Q}_t^i(dy_t^i) 1_{\{\gamma_t^i(y_t^i) \in du_t^i\}},$$

$$\frac{d\tilde{\mathbb{P}}}{d\tilde{\mathbb{Q}}} := \prod_{t=0}^{T-1} \prod_{i=1}^{N} \phi_t^i(y_t^i, \omega_0, x_0, v_{0:t-1}^{1:N}, w_{0:t-1}^{1:N}, y_{0:t-1}^{1:N}, u_{0:t-1}^{1:N}).$$

Assumption 5.2.4. *Assume that the dynamics and observations are decoupled through players, that is, for $i \in \mathcal{N}$*

$$x_{t+1}^i = f_t^i(x_{0:t-1}^i, u_{0:t-1}^i, w_t^i),$$
$$y_t^i = h_t^i(x_{0:t-1}^i, u_{0:t-1}^i, v_t^i),$$

where f_t^i and h_t^i are measurable functions and $x_t := \{x_t^i\}_{i\in\mathcal{N}}$, $w_t := \{w_t^i\}_{i\in\mathcal{N}}$.

Definition (Independent-Data and Nested Reductions). *Consider multi-stage stochastic teams ($\mathcal{P}^{Multi}$) with a given IS. Define the following two static reductions for ($\mathcal{P}^{Multi}$):*

(i) (Independent-data reduction) Let Assumption 5.2.3 hold, then independent-data reduction is defined as change of measure (5.19) *under which the measurements driven by* (5.14) *have distributions $\tilde{Q}_t^i$, and the expected cost function can be written as follows:*

$$J(\underline{\boldsymbol{\gamma}}) := E_{\tilde{\mathbb{P}}}^{\underline{\boldsymbol{\gamma}}}\left[\sum_{t=0}^{T-1} c_t(\omega_0, x_t, u_t^1, \ldots, u_t^N)\right] = E_{\tilde{\mathbb{Q}}}^{\underline{\boldsymbol{\gamma}}}\left[\hat{c}(\omega_0, x_0, \underline{\boldsymbol{w}}, \underline{\boldsymbol{v}}, \underline{\boldsymbol{u}}, \underline{\boldsymbol{y}})\right], \tag{5.20}$$

where the new cost function is

$$\hat{c}(\omega_0, x_0, \underline{\boldsymbol{w}}, \underline{\boldsymbol{v}}, \underline{\boldsymbol{u}}, \underline{\boldsymbol{y}}) := \sum_{t=0}^{T-1} c_t(\omega_0, x_t, u_t^1, \ldots, u_t^N) \frac{d\tilde{\mathbb{P}}}{d\tilde{\mathbb{Q}}}. \tag{5.21}$$

The team problem under this static reduction can be viewed as the one that Witsenhausen referred to as a static problem with independent data [135];

(ii) (Nested reduction) Let Assumption 5.2.4 hold, then nested reduction is defined as a reduction under which for each PL^i through $t \in \mathcal{T}$, the information structure is nested, i.e., $\sigma(y_t^i) \subset \sigma(y_{t+1}^i)$. □

Remark 5.2.2.

(i) We note that it has been shown in Chapters 3 and 4, the above static reduction under mild assumptions on the action and observation spaces leads to a closedness of a set of policies for each player through times under an appropriate topology, which is desirable for establishing existence and/or convergence results.

(ii) We also note that the infinite horizon team problem under a nested reduction is more tractable compared to an independent-data static reduction. That is because nested reductions allow players to have nested information structures without requiring independent data (which can be viewed as the total recall property of the private history for players, where measurements may not necessarily be independent random variables under the reduction). Furthermore, using nested reductions leads to richness in the variational analysis since for multi-stage team problems joint perturbations through times of a given player are allowed (see Corollary 5.7.2 and Remark 5.7.1).

5.3 Optimal policies for dynamic teams under policy-independent static reduction

In this section, we establish connections between stationary (pbp optimal, globally optimal) policies for dynamic teams and their policy-independent static reductions.

Theorem 5.3.1. *Consider stochastic dynamic teams ($\mathcal{P}$) with a policy-independent static reduction* (1.3).

(i) A policy $\underline{\gamma}^*$ is pbp optimal (globally optimal) for dynamic teams ($\mathcal{P}$) if and only if $\underline{\gamma}^*$ is pbp optimal (globally optimal) for a policy-independent static reduction of dynamic teams ($\mathcal{P}$);

(ii) Let a policy $\underline{\gamma}^*$ satisfy P-almost surely

$$\nabla_{u^i} E_{\mathbb{Q}}^{\gamma^{-i*}}\left[\frac{d\mathbb{P}}{d\mathbb{Q}}\middle| y^i\right]\Bigg|_{u^i=\gamma^{i*}(y^i)} = 0 \ \forall i \in \mathcal{N}, \tag{5.22}$$

where $\frac{d\mathbb{P}}{d\mathbb{Q}}$ is defined in (5.4)*. Then, $\underline{\gamma}^*$ is stationary for dynamic teams ($\mathcal{P}$) if and only if $\underline{\gamma}^*$ is stationary for a policy-independent static reduction of dynamic teams ($\mathcal{P}$).*

We note that (5.22) implies that at a stationary point $\underline{\gamma}^*$, the control has no impact locally on the reduction (on the Radon-Nikodym derivative term $\frac{d\mathbb{P}}{d\mathbb{Q}}$ defined in (5.4)). Now, in view of the above connections and sufficient conditions for stationary policies provided in Theorem 5.3.1, we use [140, Theorem 3.3 and 3.4] (which is a generalization of [78, Theorem 2 and 3] by using an IS dependent nature of convexity under policy-independent static reductions) to introduce sufficient conditions for stationary policies of ($\mathcal{P}$) to be globally optimal for dynamic teams ($\mathcal{P}$) with a policy-independent static reduction. We note that if the team is static, then (5.22) holds since $\frac{d\mathbb{P}}{d\mathbb{Q}}$ is the identity map, and hence, the following Corollary provide a refinement for [140, Theorem 3.3 and 3.4] and [78, Theorem 2 and 3].

Corollary 5.3.1. *Consider stochastic dynamic teams ($\mathcal{P}$) with a policy-independent static reduction* (1.3)*. Assume that*

(i) The cost function, c, and the Radon-Nikodym derivative, $\frac{d\mathbb{P}}{d\mathbb{Q}}$, are continuously differentiable in $u^1, \ldots, u^N$,

(ii) The cost function under the policy-independent static reduction, $\tilde{c}$ (defined in (5.6)*), is convex in $u^1, \ldots, u^N$.*

Suppose that $\underline{\gamma}^*$ is a stationary policy for the dynamic team ($\mathcal{P}$) and satisfies (5.22)*. Let for all*

$\underline{\gamma} \in \Gamma$ with $E^{\underline{\gamma}}_{\mathbb{P}}[c(\cdot)] < \infty$, the following two conditions hold for all $i \in \mathcal{N}$:

$$E_{\mathbb{P}}\left[\nabla_{u^i} c(\omega_0, \underline{\gamma}^*(\underline{y}))\left(\gamma^i(y^{i*}) - \gamma^{i*}(y^{i*})\right)\right] < \infty, \tag{5.23}$$

$$E_{\mathbb{Q}}\left[\left(\nabla_{u^i} \frac{d\mathbb{P}}{d\mathbb{Q}}(\omega_0, \underline{y}, \underline{\gamma}^*(\underline{y}))\right) c(\omega_0, \underline{\gamma}^*(\underline{y}))\left(\gamma^i(y^i) - \gamma^{i*}(y^i)\right)\right] < \infty, \tag{5.24}$$

where

$$\nabla_{u^i} c(\omega_0, \underline{\gamma}^*(\underline{y})) := \nabla_{u^i} c\Big(\omega_0, \gamma^{1*}(y^{1*}), \ldots, \gamma^{(i-1)*}(y^{(i-1)*}), u^i, \gamma^{(i+1)*}(y^{i+1}_{u^i}), \ldots, \gamma^{N*}(y^N_{u^i})\Big)\Big|_{u^i = \gamma^{i*}(y^{i*})},$$

$$\nabla_{u^i} \frac{d\mathbb{P}}{d\mathbb{Q}}(\omega_0, \underline{y}, \underline{\gamma}^*(\underline{y})) := \nabla_{u^i} \frac{d\mathbb{P}}{d\mathbb{Q}}\Big(\omega_0, \underline{y}, \gamma^{1*}(y^1), \ldots, \gamma^{(i-1)*}(y^{i-1}), u^i, \gamma^{(i+1)*}(y^{i+1}), \ldots, \gamma^{N*}(y^N)\Big)\Big|_{u^i = \gamma^{i*}(y^i)}.$$

Then, $\underline{\gamma}^*$ is globally optimal for $(\mathcal{P})$. Moreover, if the cost function $\tilde{c}$ is strictly convex in $(u^1, \ldots, u^N)$, $\underline{\gamma}^*$ is the unique globally optimal policy for $(\mathcal{P})$.

Remark 5.3.1.

(i) We note that for dynamic teams with a given IS, if c is convex in $u^1, \ldots, u^N$ for all $\omega_0, \omega_1, \ldots, \omega_N$, then $\tilde{c}$ (see (5.6)*) is not necessarily convex in $u^1, \ldots, u^N$ for all $\omega_0, \omega_1, \ldots, \omega_N$. In particular, the celebrated Witsenhausen's counterexample [130] is an example of non-convexity becoming evident under a policy-independent static reduction (this has been precisely shown in [140, equation (3.5)]).*

(ii) We also note the result in [140, Theorem 3.6], where it has been shown that dynamic teams with a policy-independent static reduction is convex in policies (see, Definition 3.2) if and only if its policy-independent static reduction of the problem is.

(iii) In view of Corollary 5.3.1, even if the cost function $\tilde{c}$ in (5.6) *is assumed to be convex and continuously differentiable in actions under policy-independent static reductions,* (5.23) *might not be sufficient to establish global optimality of a stationary policy of $\underline{\gamma}^*$ for $(\mathcal{P})$, in general. That is because, the effect of the deviating controllers (or a deviating controller in the definition of the stationary policy u^i) on the probability measures of observations has not*

been taken into account. This observation suggests that in dynamic teams, variational analysis requires to take into account the effect of the deviating controllers on the probability measures of observations (which has been considered in (5.22) *and* (5.24)*).*

5.4 Optimal Policies for Dynamic Teams under Policy-Dependent Static Reductions

Here, we study the connections between stationary (pbp optimal, globally optimal) policies of dynamic teams and their policy-dependent static reductions. Consider the setup of Section 5.2.2.

Theorem 5.4.1. *Consider stochastic dynamic teams ($\mathcal{P}^D$) with partially nested IS. Let Assumption 5.2.2 hold. Then, $\underline{\gamma}^{D,*}$ is a globally optimal policy for ($\mathcal{P}^D$) if and only if $\underline{\gamma}^{S,*}$ is a globally optimal policy for ($\mathcal{P}^S$) under the policy-dependent static reduction (see* (5.11)*).*

Proof. Since the IS is partially nested and policies are deterministic, under Assumption 5.2.2, there is a bijection from the set of policies Γ^D to the set of policies Γ^S. Therefore, global optimality in one domain implies global optimality in the other domain. □

5.4.1 Stationary and Person-by-Person Optimal Policies for Dynamic Teams and their Policy-Dependent Static Reductions

Here, we provide three examples that serve to demonstrate the subtlety of the connections between stationary (pbp optimal) policies of ($\mathcal{P}^D$) and ($\mathcal{P}^S$). In fact, we present counterexamples to show that, in contrast to the case of globally optimal policies, the isomorphic connections between stationary (pbp optimal) policies of ($\mathcal{P}^D$) and ($\mathcal{P}^S$) are no longer true, in general (under Assumption 5.2.2). We first introduce a regularity and convexity condition on the cost function, needed for our results.

Assumption 5.4.1. *Assume, for every ω_0 that*

(a) the cost function c is continuously differentiable in $(u^1, \dots, u^N)$,

(b) the cost function c is (jointly) convex in $(u^1, \dots, u^N)$.

In the following example, we show that a policy $\underline{\gamma}^{S,*}$ is stationary (also pbp optimal) for $(\mathcal{P}^S)$, but $\underline{\gamma}^{D,*}$, satisfying the policy-dependent static reduction (constructed under the relation (5.11)), is not pbp optimal for $(\mathcal{P}^D)$.

Example 5.4.1. *Consider a 2-DM stochastic team $(\mathcal{P}^D)$ with $I^1 = \{y_1^D\}$ and $I^2 := \{y_2^D\} = \{y_1^D, \hat{y}_2^D\}$, where $\hat{y}_2^D = \omega_2 + u^1$, and $\omega_2 =: \hat{y}_2^S$ and $y_1^D =: y_1^S = \omega_1$ are primitive random variables. Let the expected cost function be given as*

$$E[c(\omega_2, u^1, u^2)] := E[(u^1 - u^2 + \omega_2)^2 - \alpha(u^1)^2], \tag{5.25}$$

for a given $\alpha \in (0, 1)$.

- *A policy $\underline{\gamma}^{S,*} = (\gamma_1^{S,*}, \gamma_2^{S,*}) = (0, (0, I))$ (where the policy $(0, (0, I))$ denotes $\gamma_1^{S,*} \equiv 0$, $\gamma_{2,1}^{S,*} \equiv 0$, and $\gamma_{2,2}^{S,*}$ is the identity map, I, that is, $u^{1*} = \gamma_1^{S,*}(y_1^S) = 0$ and $u^{2*} = \gamma_2^{S,*}(y_1^S, \hat{y}_2^S) = \hat{y}_2^S$) is pbp optimal for $(\mathcal{P}^S)$.*

- *However, a policy $\underline{\gamma}^{D,*} = (\gamma_1^{D,*}, \gamma_2^{D,*}) = (0, (-\gamma_1^{S*}, I))$ constructed under a relation (5.11) (where the policy $(0, (-\gamma_1^{S*}, I))$ denotes $\gamma_1^{D,*} \equiv 0$, $\gamma_{2,1}^{D,*} = -\gamma_1^{S*}$, and $\gamma_{2,2}^{D,*}$ is the identity map, that is, $u^{1*} = \gamma_1^{D,*}(y_1^D) = 0$ and $u^{2*} = \hat{y}_2^D - \gamma_1^{S,*}(y_1^D)$) is not pbp optimal for $(\mathcal{P}^D)$ since fixing a policy of DM^2 to $\gamma_2^{D,*}$ such that $u^{2*} = \hat{y}_2^D - \gamma_1^{S,*}(y_1^D)$, the expected cost function will be concave in u^1 ($c(u^1, u^{2*}) = -\alpha(u^1)^2$). We note, however, that $\underline{\gamma}^{D,*}$ is a stationary policy for $(\mathcal{P}^D)$. Here, Assumption 5.4.1(b) fails to hold.*

In the following example, we show that a policy $\underline{\gamma}^{D,*}$ is stationary (also pbp optimal) for $(\mathcal{P}^D)$, but $\underline{\gamma}^{S,*}$ under the policy-dependent static reduction, is not pbp optimal for $(\mathcal{P}^S)$.

Example 5.4.2. *Consider a 2-DM stochastic team $(\mathcal{P}^D)$ with $I^1 = \{y_1^D\}$ and $I^2 := \{y_2^D\} = \{y_1^D, \hat{y}_2^D\}$, where $\hat{y}_2^D = \omega_2 + u^1$, and $\omega_2 =: \hat{y}_2^S = \omega_1$ and $y_1^D =: y_1^S$ are primitive random variables. Let the expected cost function be given as*

$$E[c(\omega_2, u^1, u^2)] := E[\alpha(u^1)^2 + \beta(u^2 - \omega_2)^2 - (u^1 - u^2 + \omega_2)^2], \tag{5.26}$$

for a given $\alpha \in (0,1)$ *and* $\beta > 1$.

- *A policy* $\underline{\gamma}^{D,*} = (\gamma_1^{D,*}, \gamma_2^{D,*}) = (0,(0,I))$ *(where* $\gamma_1^{D,*} \equiv 0$, $\gamma_{2,1}^{D,*} \equiv 0$ *and* $\gamma_{2,2}^{D,*}$ *is the identity map, that is,* $u^{1*} = \gamma_1^{D,*}(y_1^D) = 0$ *and* $u^{2*} = \hat{y}_2^D$*) is pbp optimal for* $(\mathcal{P}^D)$ *since fixing a policy of DM*^{2} *to* $\gamma_2^{D,*}$*, the expected cost function will be convex in* u^1 *(*$c(u^1,u^2) = (\alpha+\beta)(u^1)^2$*), and fixing a policy of DM*^{1} *to* $\gamma_1^{D,*}$ *such that* $u^1 = \gamma_1^{D,*}(y_1^D) = 0$*, the expected cost function will be convex in* u^2 *(*$c(u^1,u^2) = (\beta-1)(u^2-\omega_2)^2$*).*

- *However, under the policy-dependent static reduction, a policy* $\underline{\gamma}^{S,*} = (\gamma_1^{S,*}, \gamma_2^{S,*}) = (0,(-\gamma_1^{D,*}, I))$ *constructed under a relation* (5.11)*, is not pbp optimal for* $(\mathcal{P}^S)$ *since fixing a policy of DM*^{2} *to* $\gamma_2^{S,*}$ *such that* $u^2 = \gamma_2^{S,*}(y_1^S, \hat{y}_2^S) = \hat{y}_2^S - \gamma_1^{D,*}(y_1^S)$*, the expected cost function will be concave in* u^1 *(*$c(u^1,u^2) = (\alpha-1)(u^1)^2$*).*

Next, we provide an example, in which a policy $\underline{\gamma}^{D,*}$ is stationary (pbp optimal) for $(\mathcal{P}^D)$, but a policy $\underline{\gamma}^{S,*}$ under the policy-dependent static reduction (see (5.11)), is not stationary for $(\mathcal{P}^S)$.

Example 5.4.3. *Consider a 2-DM stochastic team* $(\mathcal{P}^D)$ *with* $I^1 = \{y_1^D\}$ *and* $I^2 = \{y_2^D\} := \{y_1^D, \hat{y}_2^D\}$*, where* $\hat{y}_2^D = \omega_2 + \sqrt{u^1}$*, and* ω_2 *and* $y_1^D = y_1^S := \omega_1$ *are primitive random variables. Let the expected cost function be given as*

$$E[c(\omega_2, u^1, u^2)] := E[(\sqrt{u^1} - u^2 + \omega_2)^2], \tag{5.27}$$

where $\mathbb{U}^1 = \mathbb{R}_+$.

- *A policy* $\underline{\gamma}^{D,*} = (\gamma_1^{D,*}, \gamma_2^{D,*}) = (0,(0,I))$ *(where* $\gamma_1^{D,*} \equiv 0$, $\gamma_{2,1}^{D,*} \equiv 0$ *and* $\gamma_{2,2}^{D,*}$ *is the identity map, that is,* $u^{1*} = 0$ *and* $u^{2*} = \hat{y}_2^D$*) is stationary for* $(\mathcal{P}^D)$*.*

- *However, under the policy-dependent static reduction, a policy* $\underline{\gamma}^{S,*} = (\gamma_1^{S,*}, \gamma_2^{S,*}) = (0,(\sqrt{\gamma_1^{D*}}, I))$ *constructed under a relation* (5.11) *(where* $\gamma_1^{S,*} \equiv 0$, $\gamma_{2,1}^{S,*} = \sqrt{\gamma_1^{D*}}$*, and* $\gamma_{2,2}^{S,*}$ *is the identity map, that is,* $u^1 = 0$ *and* $u^2 = \omega_2 + \sqrt{\gamma_1^{D*}(y_1^S)}$*) is not stationary (although it is pbp optimal)*

for ($\mathcal{P}^S$). Since fixing a policy of DM^2 to $\gamma_2^{S,*}$ such that $u^2 = \omega_2$, the derivative of the expected cost function with respect to u^1 is always 1. Hence, the criterion for stationarity does not lead to any solution. Here, Assumption 5.4.1(a) fails to hold at $\underline{\gamma}^{S,*} = (0, (\sqrt{\gamma_1^{D*}}, I))$.

Hence, in view of the preceding examples, we have the following negative result.

Proposition 5.4.1. *Consider stochastic dynamic teams ($\mathcal{P}^D$) with partially nested IS. Let Assumption 5.2.2 hold. Then:*

(i) If $\underline{\gamma}^{D,*}$ is a stationary (pbp optimal) policy for ($\mathcal{P}^D$), then $\underline{\gamma}^{S,*}$ is not necessarily a stationary (pbp optimal) policy for ($\mathcal{P}^S$) under the policy-dependent static reduction (see (5.11)*);*

(ii) If $\underline{\gamma}^{S,*}$ is a stationary (pbp optimal) policy for ($\mathcal{P}^S$), then $\underline{\gamma}^{D,*}$, satisfying the policy-dependent static reduction relation (5.11)*, is not necessarily pbp optimal for ($\mathcal{P}^D$).*

Proof. This is a direct consequence of the examples above, where Examples 5.4.2 and 5.4.3 imply Part(i), and Example 5.4.1 implies Part(ii). □

Now, we introduce sufficient conditions to establish connections between stationary policies of dynamic teams and their policy-dependent static reductions.

Condition (C5). *A policy $\underline{\gamma}^D$ satisfies (C) if for all $i \in \mathcal{N}$, $\gamma_i^D(\{g_j(h_j(\zeta), u^{\downarrow j})\}_{j\in\downarrow i}, g_i(h_i(\zeta), u^{\downarrow i}))$ is differentiable P-almost surely in $u^{\downarrow i}$ (for realizations of ζ on a set of measure one).*

Theorem 5.4.2. *Consider stochastic dynamic teams ($\mathcal{P}^D$) with partially nested IS. Let Assumptions 5.2.2 and 5.4.1(a) hold. Then:*

(i) Under Assumption 5.4.1(b), a policy $\underline{\gamma}^{D,*}$ satisfying (C) is stationary for ($\mathcal{P}^D$) if and only if $\underline{\gamma}^{S,*}$ is pbp optimal (stationary) for ($\mathcal{P}^S$) under the policy-dependent static reduction (see (5.11)*).*

(ii) A policy $\underline{\gamma}^{D,*}$, where $\gamma_i^D(\{g_j(h_j(\zeta), u^{\downarrow j})\}_{j\in\downarrow i}, g_i(h_i(\zeta), u^{\downarrow i}))$ is linear (affine) in $u^{\downarrow i}$ for all $i \in \mathcal{N}$ and ζ, is stationary for ($\mathcal{P}^D$) if and only if $\underline{\gamma}^{S,*}$ is stationary for ($\mathcal{P}^S$) under the

policy-dependent static reduction (see (5.11)*). Furthermore, if Assumption 5.4.1(b) holds, then* $\underline{\gamma}^{D,*}$ *and* $\underline{\gamma}^{S,*}$ *are pbp optimal policies for* $(\mathcal{P}^D)$ *and* $(\mathcal{P}^S)$*, respectively.*

Remark 5.4.1. *We note that for a policy* $\underline{\gamma}^{D,*}$*, satisfying the linearity assumption in Part(ii), (C) holds. However, for the first claim of Part(ii), Assumption 5.4.1(b) is not imposed since in the proof, there is no need for exchanging the expectation and the limit (for example, this applies to the linear exponential Gaussian team (LEGT) problems with partially nested IS and a negative exponential coefficient (risk preferring case), where seeking for pbp optimal (globally optimal) policies apriori is restricted to affine policies; for static LEGT problems with a positive exponential coefficient (risk averse case), we refer the reader to [79] and [139, p. 68]).*

5.5 Optimality and Convexity under Static Measurements with Control-Sharing Reduction

In this section, we present our results for the static measurement with control-sharing reduction and its impact on optimality and convexity of dynamic team problems.

5.5.1 Optimal Policies under Static Measurements with Control-Sharing Reduction

In this subsection, we consider team problems with partially nested with control-sharing IS (see Section 5.2.3), and we establish connections between pbp optimal (globally optimal) policies of $(\mathcal{P}^D)$, $(\mathcal{P}^S)$, $(\mathcal{P}^{D,CS})$, and $(\mathcal{P}^{CS})$.

Now, we state the following observation as a theorem, since it will be consequential later on:

Theorem 5.5.1. *Consider stochastic dynamic teams with partially nested IS, where Assumption 5.2.2 hold. The static measurements with control-sharing reduction is policy-independent.*

Proof. Since Assumption 5.2.2 holds and DMs have access to the precedent DMs' actions, $u^{\downarrow i}$, the static measurements with control-sharing reduction to $(\mathcal{P}^{CS})$ for each DM is independent of precedent DMs' policies. That is because, given $\underline{\gamma}^{D,CS}$, a policy $\underline{\gamma}^{CS}$ can be constructed through (5.12), i.e., for every $i \in \mathcal{N}$, $u^i = \gamma_i^{D,CS}(y^D_{\downarrow i}, u^{\downarrow i}, g_i(h_i(\zeta), u^{\downarrow i})) = \gamma_i^{CS}(y^S_{\downarrow i}, u^{\downarrow i}, \hat{y}_i^S)$ for every $u^{\downarrow i}$

and P-almost every ζ. The converse of the above statement is also true. The fact that the expected cost function does not change under the above reduction, completes the proof. □

In view of similar examples as Examples 5.4.1, 5.4.2, and 5.4.3, the isomorphic connections between stationary (pbp optimal) policies of ($\mathcal{P}^{CS}$) (($\mathcal{P}^{D,CS}$)) and ($\mathcal{P}^{S}$) and/or ($\mathcal{P}^{D}$) fail to hold in general. Now, we provide some positive results using the static measurements with control-sharing reduction.

Theorem 5.5.2. *Consider stochastic dynamic teams with partially nested IS.*

(i) If Assumption 5.2.2 holds, then a policy $\underline{\gamma}^{D,CS}$ is pbp optimal (stationary, globally optimal) for ($\mathcal{P}^{D,CS}$) if and only if $\underline{\gamma}^{CS}$ is a pbp optimal (stationary, globally optimal) policy for ($\mathcal{P}^{CS}$) under the static measurements with control-sharing reduction (see (5.12)*).*

(ii) If a policy $\underline{\gamma}^{D,*}$ is pbp optimal (stationary) for ($\mathcal{P}^{D}$), then a policy $\underline{\gamma}^{D,CS,*}$ is pbp optimal (stationary) for ($\mathcal{P}^{D,CS}$), where for all $i \in \mathcal{N}$ $\gamma_i^{D,*}(y_i^D) = \gamma_i^{D,CS,*}(y_i^{D,CS})$ for all $u^{\downarrow i}$ and almost all ζ (a representation of $\underline{\gamma}^{D,CS,*}$ is $\underline{\gamma}^{D,*}$, where the extra information on the actions of precedent DMs is ignored).

(iii) If a policy $\underline{\gamma}^{S,*}$ is pbp optimal (stationary) for ($\mathcal{P}^{S}$), then there exists a pbp optimal (stationary) policy $\underline{\gamma}^{CS,*}$ for ($\mathcal{P}^{CS}$), where for every $i \in \mathcal{N}$, $\gamma_i^{S,*}(y_i^S) = \gamma_i^{CS,*}(y_i^{CS})$ for almost all ζ (a representation of $\underline{\gamma}^{CS,*}$ is $\underline{\gamma}^{S,*}$, where the extra information on the actions of precedent DMs is ignored).

Now, we provide a result as a corollary to Theorems 5.4.2, 5.5.1, and 5.5.2.

Corollary 5.5.1. *Consider stochastic dynamic teams ($\mathcal{P}^{D}$) with partially nested IS, where Assumption 5.2.2 holds.*

(i) A policy $\underline{\gamma}^{CS,*}$ is globally optimal for ($\mathcal{P}^{CS}$) if and only if policies $\underline{\gamma}^{D,*}$ and $\underline{\gamma}^{S,*}$ are globally optimal for ($\mathcal{P}^{D}$) and ($\mathcal{P}^{S}$), respectively with a relation $\gamma_i^{D,*}(y_i^D) = \gamma_i^{S,*}(y_i^S) = \gamma_i^{CS,*}(y_i^{CS})$ for all $i \in \mathcal{N}$ and almost all ζ (for any static representation of $\underline{\gamma}^{S,*}$, there may exist multiple representations for $\underline{\gamma}^{D,*}$ and $\underline{\gamma}^{CS,*}$).

(ii) Under Assumption 5.4.1(a), if a pbp optimal policy $\underline{\gamma}^{CS,*}$ for $(\mathcal{P}^{CS})$ is linear in actions, then $\underline{\gamma}^{S,*}$, satisfying a relation $\gamma_i^{S,*}(y_i^S) = \gamma_i^{CS,*}(y_i^{CS})$ for all $i \in \mathcal{N}$ and almost all ζ, is pbp optimal for $(\mathcal{P}^S)$.

Remark 5.5.1. *Define $\tilde{\Gamma}^{CS}$ as a space of admissible policies where observations* (1.10) *is replaced by $\tilde{y}_i^{CS} := \left\{ y_{\downarrow i}^S, u^{K_i}, \hat{y}_i^S := h^i(\zeta) \right\}$, $K_i \subseteq \downarrow i$, and $\tilde{I}^{CS} = \{\tilde{y}_i^{CS}\}$, that is, for each $i \in \mathcal{N}$, control action of DM^i is only shared to a subset $K_i \subseteq \downarrow i$ of DM^{j}s with $j \in \downarrow i$ (similarly, we can define $\tilde{\Gamma}^{D,CS}$ with revising observations* (1.9)*). Following from Theorem 5.5.2, for teams with partially nested IS, we have*

(i) any pbp optimal policy $\underline{\gamma}^{D,*} \in \Gamma^D$ for $(\mathcal{P}^D)$ constitutes a pbp optimal policy on the enlarged space $\Gamma^{D,CS}$;

(ii) any pbp optimal policy $\underline{\gamma}^{S,*} \in \Gamma^S$ for $(\mathcal{P}^S)$ constitutes a pbp optimal policy on the enlarged space Γ^{CS}.

(iii) The results of Theorem 5.5.2 (i) and (ii) remain valid if the enlarged space $\Gamma^{D,CS}$ (or the enlarged space Γ^{CS}) is replaced by $\tilde{\Gamma}^{D,CS}$ where $\Gamma^D \subseteq \tilde{\Gamma}^{D,CS} \subseteq \Gamma^{D,CS}$ ($\tilde{\Gamma}^{CS}$ where $\Gamma^S \subseteq \tilde{\Gamma}^{CS} \subseteq \Gamma^{CS}$).

Next, we present two results on the existence and uniqueness of optimal policies for $(\mathcal{P}^D)$, $(\mathcal{P}^S)$, $(\mathcal{P}^{D,CS})$, and $(\mathcal{P}^{CS})$, using Theorems 5.4.1, Corollary 5.5.1, and Examples 5.4.1 and 5.4.2.

Proposition 5.5.1. *Consider stochastic dynamic teams $(\mathcal{P}^D)$ with partially nested IS, where Assumption 5.2.2 holds. Then:*

(i) The existence of a pbp optimal policy for $(\mathcal{P}^{CS})$ ($(\mathcal{P}^{D,CS})$) does not imply the existence of a pbp optimal policy for $(\mathcal{P}^S)$ ($(\mathcal{P}^D)$).

(ii) The globally optimal policy is unique for $(\mathcal{P}^S)$ if and only if the globally optimal policy is unqiue (or essentially unique) for $(\mathcal{P}^D)$;

(iii) If the globally optimal policy for ($\mathcal{P}^S$) and/or ($\mathcal{P}^D$) is (essentially) unique, then there exists an essentially unique globally optimal policy for ($\mathcal{P}^{CS}$) (($\mathcal{P}^{D,CS}$)) (that is, there can exist multiple representations of $\underline{\gamma}^{CS,*}$ and $\underline{\tilde{\beta}}$ with $\gamma_i^{CS,*}(y_i^{CS}) = \tilde{\beta}^{i*}(y_i^{CS})$ for almost all ζ).

Proof. Examples 5.4.1 and 5.4.2 imply Part(i). Theorem 5.4.1 implies Part(ii), and Theorem 5.5.1(i) implies Part(iii). □

5.5.2 Convexity of Dynamic Team Problems and their Static Measurements with Control-Sharing Reduction

Next, we study convexity of team problems ($\mathcal{P}^D$), ($\mathcal{P}^S$), ($\mathcal{P}^{D,CS}$), and ($\mathcal{P}^{CS}$) in policies (see, [140, Definition 3.1]). We first present non-convex dynamic team problems (in policies) with a partially nested IS, where the cost function satisfies Assumption 5.4.1. Hence, we show inadequacy of Assumption 5.4.1(ii) (convexity of the cost function in actions) to imply convexity of the team problem in policies for dynamic teams even for those with a partially nested IS.

Example 5.5.1. *Consider a 2-DM stochastic dynamic team ($\mathcal{P}^D$) where the expected cost function is given by*

$$E[c(\omega_0, u^1, u^2)] := E[(u^1 + \omega_0)^2 + (u^2)^2],$$

where ω_0 is a primitive random variable.

- *Let $I^1 = \{y_1^D\}$ and $I^2 = \{y_2^D\} := \{y_1^D, \hat{y}_2^D\}$, where $\hat{y}_2^D = \hat{y}_2^S + u^1$, and y_1^D and $\hat{y}_2^S$ are primitive random variables. If $\gamma_2^D(y_2^D) = \sqrt[4]{y_2^D}$, then the above dynamic team problem ($\mathcal{P}^D$) is not convex in (γ_1^D, γ_1^D) since for arbitrary $u = \gamma_1^D(y_1^D)$ and $\hat{u} = \hat{\gamma}_1^D(y_1^D)$ when a convex combination of policies is applied by DM^1, we have $\gamma_2^D(y_2^D) = \sqrt[4]{\hat{y}_2^S + \alpha\gamma_1^D(y_1^D) + (1-\alpha)\hat{\gamma}_1^D(y_1^D)}$, and hence,*

$$\begin{aligned}
&E[c(\omega_0, \alpha\gamma_1^D(y_1^D) + (1-\alpha)\hat{\gamma}_1^D(y_1^D), \gamma_2^D(\hat{y}_2^D)] \\
&= E[(\omega_0 + \alpha\gamma_1^D(y_1^D) + (1-\alpha)\hat{\gamma}_1^D(y_1^D))^2 + \sqrt{\hat{y}_2^S + \alpha\gamma_1^D(y_1^D) + (1-\alpha)\hat{\gamma}_1^D(y_1^D)}]
\end{aligned}$$

which implies that the above dynamic team problem ($\mathcal{P}^D$) is not convex in policies.

• *If $I_2^{CS} := \{y_2^{CS}\} := \{y_1^D, u^1, \hat{y}_2^S\}$, then for any $u = \gamma_1^D(y_1^D)$ and $\hat{u} = \hat{\gamma}_1^D(y_1^D)$, policies β^2 and $\hat{\beta}^2$ in Γ^{CS} can be constructed satisfying $\beta^2(y_1^D, \hat{y}_2^S) = \gamma_2^{CS}(y_1^D, \hat{y}_2^S, \gamma_1^D(y_1^D))$ and $\hat{\beta}^2(y_1^D, \hat{y}_2^S) = \gamma_2^{CS}(y_1^D, \hat{y}_2^S, \hat{\gamma}_1^D(y_1^D))$ such that*

$$\begin{aligned}&E[c(\omega_0, \alpha u + (1-\alpha)\hat{u}, \gamma_2^{CS}(y_1^D, \hat{y}_2^S, \alpha u + (1-\alpha)\hat{u}))]\\ &= E[(\omega_0 + \alpha u + (1-\alpha)\hat{u})^2 + (\alpha\beta^2(y_1^S, \hat{y}_2^S) + (1-\alpha)\hat{\beta}^2(y_1^S, \hat{y}_2^S))^2],\end{aligned}$$

which implies that under the static measurements with control-sharing reduction, the team problem above (with static measurements with control-sharing IS) is convex in policies. We can show that the dynamic team problem above with $I_2^{D,CS} := \{y_2^{D,CS}\} := \{y_1^D, u^1, \hat{y}_2^D\}$ is convex in policies using the static measurements with control-sharing reduction since the reduction is policy-independent. □

In the following example, we further discuss the fragility of the convexity (in policies) of a team problem under different reductions and information structures. In particular, similar to Example ex:non-cvx, we show that a dynamic partially nested team problem, admitting a policy-dependent static reduction and satisfying the convexity condition in Assumption 5.4.1, can be non-convex (in policies); however, we further discuss that control-sharing (partial control-sharing) under static measurements with control-sharing reduction leads to the convexity of the team problem (in policies).

Example 5.5.2. *Consider a $N+1$-DM stochastic dynamic team ($\mathcal{P}^D$) where the expected cost function is given by*

$$\frac{1}{N}\sum_{i=1}^{N} E[c(\omega_0, u^i, u^{N+1}, \frac{1}{N}\sum_{p=1}^{N} u^p)] := \frac{1}{N}\sum_{i=1}^{N} E[(\omega_0 + R^i u^i + R^{N+1}u^{N+1} + \frac{\tilde{R}}{N}\sum_{p=1}^{N} u^p)^2],$$

where for all $i = 1, \ldots, N+1$, R^i and $\tilde{R}$ are positive real numbers. Let for every $i = 1, \ldots, N+1$,

y_i^S be exogenous random variables, and let $y_{N+1}^D = \frac{1}{N}\sum_{i=1}^N u^i + K y_{N+1}^S$ for $K \in \mathbb{R}$. We consider convexity under the following partially nested information structures:

(i) $I_{N+1}^S := \{y_1^S, \dots, y_{N+1}^S\}$ *and* $I_i^S := \{y_i^S\}$;

(ii) $I_{N+1}^{CS} := \{y_1^S, \dots, y_N^S, y_{N+1}^S, u^1, \dots, u^N\}$ *and* $I_i^S := \{y_i^S\}$;

(iii) $I_{N+1}^{D,CS} := \{y_1^S, \dots, y_N^S, y_{N+1}^D, u^1, \dots, u^N\}$ *and* $I_i^S := \{y_i^S\}$;

(iv) $I_{N+1}^D := \{y_1^S, \dots, y_N^S, y_{N+1}^D\}$ *and* $I_i^S := \{y_i^S\}$.

Part (i): Under the information structure I^S, since the team is static and the cost function is convex in actions, the team is convex in policies.

Parts (ii)(iii): Under the information structure I^{CS}, we have for any $\alpha \in [0,1]$ and for any policy $\gamma^{N+1} \in \Gamma_{N+1}^{CS}$

$$\begin{aligned}
&\frac{1}{N}\sum_{i=1}^N E[c(\omega_0, \alpha u^i + (1-\alpha)\hat{u}^i, \gamma^{N+1}(y_{N+1}^S, \alpha \underline{u} + (1-\alpha)\underline{\hat{u}}), \alpha\mu_N + (1-\alpha)\hat{\mu}_N)] \\
&= \frac{1}{N}\sum_{i=1}^N E[c(\omega_0, \alpha u^i + (1-\alpha)\hat{u}^i, \alpha\beta^{N+1}(y_{N+1}^S) + (1+\alpha)\hat{\beta}^{N+1}(y_{N+1}^S), \alpha\mu_N + (1-\alpha)\hat{\mu}_N)] \qquad (5.28) \\
&\leq \alpha\frac{1}{N}\sum_{i=1}^N E[c(\omega_0, u^i, \beta^{N+1}(y_{N+1}^S), \mu_N)] + (1-\alpha)\frac{1}{N}\sum_{i=1}^N E[c(\omega_0, \hat{u}^i, \hat{\beta}^{N+1}(y_{N+1}^S), \hat{\mu}_N)], \qquad (5.29)
\end{aligned}$$

where $\mu_N := \frac{1}{N}\sum_{i=1}^N u^i$, $\hat{\mu}_N := \frac{1}{N}\sum_{i=1}^N \hat{u}^i$, and (5.28) *follows from the fact that the information structure is static measurements with control-sharing, and hence, policies β^{N+1} and $\hat{\beta}^{N+1}$ can be constructed. Inequality* (5.29) *follows from convexity of the cost function in its last three arguments, and hence, the team problem under I^{CS} is convex (in policies). Similarly, we can show that the team problem under $I^{D,CS}$ is convex (in policies).*

Part (iv): For the team problem under I^D, we discuss the team problem under different conditions on R^i and K. First, let $K \neq 0$. Similar to Example 5.5.1, we can show that team problem is non-convex if $\gamma^{N+1} \in \Gamma^D_{N+1}$ is chosen such that $u^{N+1} = \sqrt[4]{(y^D_{N+1})}$. Now, let $K = 0$. If there is i for $i \in \{1, \dots, N\}$ such that $R^i \neq 0$, then since DM^{N+1} has only access to a partial information of actions of precedent DMs, μ^N, the reduction to the team problem with I^S is policy-dependent, and hence, similar to Example 5.5.1, the team problem is non-convex. However, if $K = R^i = 0$ for all $i = 1, \dots, N$, then we have

$$\begin{aligned} &E[\hat{c}(\omega_0, \gamma^{N+1}(y^S_{N+1}, \alpha\mu_N + (1-\alpha)\hat{\mu}_N), \alpha\mu_N + (1-\alpha)\hat{\mu}_N)] \\ &\leq \alpha E[\hat{c}(\omega_0, \beta^{N+1}(y^S_{N+1}), \mu_N)] + (1-\alpha)E[\hat{c}(\omega_0, \hat{\beta}^{N+1}(y^S_{N+1}), \hat{\mu}_N)], \end{aligned} \tag{5.30}$$

where $\hat{c}(\omega_0, u^{N+1}, \frac{1}{N}\sum_{i=1}^N u^i) := (\omega_0 + R^{N+1}u^{N+1} + \frac{\tilde{R}}{N}\sum_{i=1}^N u^i)^2$ and (5.30) *follows from the fact that for every μ_N and $\hat{\mu}_N$, policies β^{N+1} and $\hat{\beta}^{N+1}$ can be constructed, and hence, the reduction to the team problem with I^S is policy-independent which implies that the team problem is convex in policies. We note that in fact when $K = R^i = 0$ for all $i = 1, \dots, N$, μ_N provides enough information for DM^{N+1} to construct policies under the reduction independent of policies of DM^i for $i = 1, \dots, N$.* □

Now, in view Examples 5.5.1 and 5.5.2, we establish a result on convexity of team problems with partially nested IS under static measurements with control-sharing reduction. We note that the following result serves as a refinement and clarification of the analysis in [140, Section 3.3.2 and 3.3.3], where this distinction was not made explicit (clarification in the sense that convexity is only preserved if the IS is partially nested with control-sharing). That is, convexity strictly requires IS to be partially nested with control-sharing.

Theorem 5.5.3. *Consider dynamic team problems with a partially nested IS, where Assumption 5.4.1(a) (on convexity of the cost function) and Assumption 5.2.2 (on the invertibility condition of observations) hold. Then:*

(i) The team problem ($\mathcal{P}^D$) is not necessarily convex in policies.

(ii) The team problems ($\mathcal{P}^S$), ($\mathcal{P}^{CS}$), and ($\mathcal{P}^{D,CS}$) are convex in policies.

Proof. Part (i) is a straightforward consequence of Example 5.5.1, and for Part (ii), the proof for ($\mathcal{P}^S$) is immediate. For ($\mathcal{P}^{CS}$) and ($\mathcal{P}^{D,CS}$), since DMs can have access to actions of precedent DMs, the reduction of each DM is independent of policies of precedent DMs (Theorem 5.5.2(i)), and hence, the result follows from the fact that under the policy-dependent static reduction, the expected cost function does not change. □

5.6 Linear Quadratic Gaussian Stochastic teams with Partially Nested Information Structure

In this section, we first establish results on the connections between uniqueness of pbp optimal policies for ($\mathcal{P}^S$) and ($\mathcal{P}^D$), which is useful in particular for LQG models. The following result is a corollary to Theorems 5.4.1, 5.4.2, and 5.5.2.

Corollary 5.6.1. *Consider stochastic dynamic teams ($\mathcal{P}^D$) with partially nested IS. Assume that for all $i \in \mathcal{N}$, g_i is linear in $u^{\downarrow i}$ for all ζ (hence, Assumption 5.2.2 holds). Let Assumption 5.4.1 hold, and let $\underline{\gamma}^{S,*} \in \Gamma^S$ be the unique pbp optimal policy for ($\mathcal{P}^S$) (hence, globally optimal).*

(i) If $\underline{\gamma}^{D,*} \in \Gamma^D$ satisfying (5.11)*, is linear, then $\underline{\gamma}^{D,*}$ is the essentially unique linear pbp optimal policy for ($\mathcal{P}^D$) (unique in the class of linear policies).*

(ii) If $\underline{\hat{\gamma}}^D \in \Gamma^D$ is any nonlinear stationary (pbp optimal) policy for ($\mathcal{P}^D$) (if it exists), then $J(\underline{\gamma}^{D,*}) \leq J(\underline{\hat{\gamma}}^D)$.

(iii) If there exists a linear policy β^* for ($\mathcal{P}^{CS}$) with a representation of $\beta_i^*(y_i^{CS}) = \gamma_i^{S,*}(y_i^S)$ for $i \in \mathcal{N}$ and almost all ζ, then β^* is an essentially unique linear pbp optimal policy for ($\mathcal{P}^{CS}$) (there might exist other linear representations of the policy).

(iv) If $\hat{\beta}$ is any nonlinear pbp optimal policy for ($\mathcal{P}^{CS}$) (if it exists), then $J(\beta^*) \leq J(\hat{\beta})$.

Now, we use Corollaries 5.5.1 and 5.6.1 to re-visit a well-known results for LQG team problems ([66]).

Example 5.6.1. *Consider LQG dynamic teams with partially nested IS. Define the observation of DMs for dynamic teams as*

$$y_i^D = \left\{ y_{\downarrow i}^D, \hat{y}_i^D := H_i \zeta + \sum_{j \in \downarrow i} B_{ij} u_j^D \right\},$$

where ζ denotes all relevant random variables, and H_i and B_{ij} are matrices of appropriate dimensions. Let the IS of DM^i be $I_i^D = \{y_i^D\}$. Define the observations under the policy-dependent static reduction as follows:

$$y_i^S = \left\{ y_{\downarrow i}^S, \hat{y}_i^S := H_i \zeta \right\},$$

and let the expected cost function under $\underline{\gamma}^D$ be given as

$$E^{\underline{\gamma}^D}[c(\omega_0, u_1, \ldots, u_N)] := E^{\underline{\gamma}^D}[\zeta' M \zeta' + \underline{u}' R \underline{u}],$$

where $M \geq 0$, $R > 0$, $\underline{u} := \{u_1, \ldots, u_N\}$ and a' denotes transpose of a for $a = \zeta, \underline{u}$. Following from [66], under the policy-dependent static reduction, the globally optimal policy is unique and linear (since the cost function is strictly convex in actions). Denote this globally optimal policy by $\underline{\gamma}^{S,*} = (Q_1^*, \ldots, Q_N^*) \in \Gamma^S$. Hence, by Preposition 5.5.1, $\underline{\gamma}^{D,*} := (K_1^*, \ldots, K_N^*) \in \Gamma^D$ with $Q_i^* y_i^S = K_i^* y_i^D$ for all $i \in \mathcal{N}$ and almost all ζ (satisfying (5.11)*) is the unique globally optimal for the dynamic team, which satisfies for all $i \in \mathcal{N}$ and $K_i^* := (\{K_i^j\}_{j \in \downarrow i}, K_i^{i*})$ and $Q_i^* := (\{Q_i^j\}_{j \in \downarrow i}, Q_i^{i*})$*

$$K_i^{j*} = Q_i^{j*} - Q_i^{i*} B_{ij} K_j^* \text{ for all } j \in \downarrow i,$$
$$K_i^{i*} = Q_i^{i*}.$$

Moreover, following from Corollary 5.6.1(i), $\underline{\gamma}^{D,*}$ is the essentially unique linear pbp optimal policy for the dynamic team.

5.7 Multi-stage Team Problems: Player-wise Optimality Analysis and Reductions

In this section, we consider multi-stage stochastic dynamic team problems. We first provide examples, where the above independent-data and nested reductions apply. Then, we establish connections between player-wise/DM-wise pbp optimal policies of dynamic multi-stage teams and their static reduction, and finally, we provide a sufficient condition under which DM-wise pbp optimality implies player-wise pbp optimality, which leads us to use the results for the single-stage problems discussed in the previous sections.

Example 5.7.1. *Consider multi-stage stochastic dynamic teams ($\mathcal{P}^{Multi}$).*

(i) Assume that $x^i_{t+1} = \tilde{f}^i_t(x^{1:N}_{0:t}, u^{1:N}_{0:t}) + w^i_t$, where $\tilde{f}^i_t$ is a measurable function and that observations of each PL^i at time t is of the form $y^i_t = \tilde{h}^i_t(x^{1:N}_{0:t}, u^{1:N}_{0:t-1}) + v^i_t$ with random variables v^i_ts being independent of other exogenous random variables of dynamics and observations, and having a zero-mean Gaussian density functions N^i_t with a positive-definite covariance for all $i \in \mathcal{N}$ and $t \in \mathcal{T}$. If $I^i_t = y^i_{0:t}$, then an independent-data static reduction exists. The above holds since we can write

$$y^i_t = \hat{h}^i_t(x^{1:N}_0, w^{1:N}_{0:t-1}, v^{1:N}_{0:t-1}, u^{1:N}_{0:t-1}) + v^i_t,$$

and we can define ϕ^i_t and η^i_t as

$$\phi^i_t = \frac{N^i_t(y^i_t - \hat{h}^i_t(x^{1:N}_0, w^{1:N}_{0:t-1}, v^{1:N}_{0:t-1}, u^{1:N}_{0:t-1}))}{N^i_t(y^i_t)},\ \eta^i_t = N^i_t(y^i_t)dy^i_t.$$

(ii) Assume that $x^i_{t+1} = \tilde{f}^i_t(\omega_0, x^i_{0:t}, u^i_{0:t}) + w^i_t$ where $\tilde{f}^i_t$ is a measurable function and w^i_t admits zero-mean Gaussian density function N^i_t with positive-definite covariance. Let observations

of each PL^i at time t is of the form $y_t^i = h_t^i(x_{0:t}^i, y_{0:t-1}^i, v_{0:t}^i)$, where $\sigma(y_t^i) \subset \sigma(y_{t+1}^i)$ and $(v_t^i)_t$ are independent of disturbances of other DMs and independent of ω_0. If $I_t^i := \{y_t^i\}$ for all $i = 1, \dots, N$ and $t = 0, \dots, T-1$, then a nested reduction exists.

Corollary 5.7.1. *Consider multi-stage stochastic dynamic teams ($\mathcal{P}^{Multi}$).*

(i) Assume that there exists an independent-data static reduction. Then, $\underline{\boldsymbol{\gamma}}^*$ is a player-wise (DM-wise) pbp optimal policy for the dynamic team problems ($\mathcal{P}^{Multi}$) if and only if it is a player-wise (DM-wise) pbp optimal policy under independent-data static reductions.

(ii) Assume that there exists a nested reduction. Then, $\underline{\boldsymbol{\gamma}}^*$ is a player-wise pbp optimal policy for the dynamic team problems ($\mathcal{P}^{Multi}$) if and only if it is a player-wise pbp optimal policy under nested reductions.

We emphasize that (ii) is not necessarily true for DM-wise pbp optimal policies.

Proof. Part(i) follows from Theorem 5.3.1, and the fact that the independent-data static reduction is policy-independent. Part(ii) follows from the fact that in the nested reduction, following from Assumption 5.2.4, the team problem is static through players, and hence, every player-wise pbp optimal policy will be preserved under the reduction (since fixing policies of other players, a player-wise pbp optimal policy is globally optimal for the player through time which will be preserved under policy-independent, policy-dependent static reductions, static measurement with control-sharing reduction). □

As we discussed before, every player-wise pbp optimal policy is DM-wise pbp optimal; however, the converse statement is not true in general. In the following, we use Corollary 5.7.1 to establish a variational analysis for multi-stage stochastic dynamic teams ($\mathcal{P}^{\text{Multi}}$) under which DM-wise pbp optimal policies are player-wise pbp optimal.

Corollary 5.7.2. *Consider multi-stage stochastic dynamic teams ($\mathcal{P}^{Multi}$). Assume that there exists an independent-data static reduction. Let $\underline{\boldsymbol{\gamma}}^*$ be a (DM-wise) pbp optimal policy for ($\mathcal{P}^{Multi}$). Assume that, for every $t \in \mathcal{T}$,*

(i) $\hat{c}$ *(see* (5.21)*) is continuously differentiable in* $\underline{\boldsymbol{u}} = (\boldsymbol{u}^1, \ldots, \boldsymbol{u}^N)$*,*

(ii) for every $i \in \mathcal{N}$*,* $\hat{c}$ *is convex in* $\boldsymbol{u}^i$*, where policies of other players (for PL*^{j}*s with* $j \in \{1, \ldots, i-1, i+1, \ldots, N\}$*) are fixed to be* $\boldsymbol{\gamma}^{-i*}$*.*

If for all $i \in \mathcal{N}$ *and for all* $\boldsymbol{\gamma}^i \in \mathbf{\Gamma^i}$ *with* $E^{\boldsymbol{\gamma}^i, \boldsymbol{\gamma}^{-i*}}[\hat{c}(\cdot)] < \infty$

$$E\Bigg[\nabla_{u_t^i}\hat{c}\Big(\omega_0, x_0, \underline{\boldsymbol{w}}, \underline{\boldsymbol{v}}, u_t^i, \boldsymbol{\gamma}^{-i*}(\boldsymbol{y}^{-i}), \gamma_{0:t-1}^{i*}(y_{0:t-1}^i), \gamma_{t+1:T-1}^{i*}(y_{t+1:T-1}^i), \underline{\boldsymbol{y}}\Big)\Big|_{u_t^i=\gamma_t^{i*}(y_t^i)} \times \Big(\gamma_t^i(y_t^i) - \gamma_t^{i*}(y_t^i)\Big)\Bigg] < \infty \quad \textit{for all } t \in \mathcal{T},$$

then, $\underline{\boldsymbol{\gamma}}^*$ *is player-wise pbp optimal for (*$\mathcal{P}^{Multi}$*).*

Proof. Following from Corollary 5.7.1, $\underline{\boldsymbol{\gamma}}^*$ is also a (DM-wise) pbp optimal under an independent-data static reduction. By fixing policies $\boldsymbol{\gamma}^{-i*}$ and using convexity and regularity assumptions under the reduction, Theorem 5.8.1 implies that $\underline{\boldsymbol{\gamma}}^*$ is a player-wise pbp optimal policy under an independent-data static reduction (through showing that by fixing policies $\boldsymbol{\gamma}^{-i*}$, $\underline{\boldsymbol{\gamma}}^*$ is globally optimal for PL^i), and this completes the proof. □

Remark 5.7.1. *For dynamic teams under a nested reduction, the above variational analysis might not in general hold since if* $\underline{\boldsymbol{\gamma}}^*$ *is a (DM-wise) pbp optimal policy for the dynamic problem, it might not be (DM-wise) pbp optimal policy under a nested reduction. However, for dynamic teams under a nested reduction, since under Assumption 5.2.4, the team is static through players (observations of each player at a given time are not affected by actions of any other player), by considering joint perturbations through times of a given player (through considering a player-wise stationary policy), variational inqualities (similar to those in Theorem 5.8.1 and Corollary 5.3.1) can be obtained to show the global optimality of player-wise stationary (pbp optimal) policies. Hence, under a nested reduction, if the cost function* c *is convex and continuously differentiable in actions, then, a variational analysis guarantees the global optimality of player-wise stationary (pbp optimal) policies but not DM-wise stationary (pbp optimal) policies.*

5.8 Proofs

5.8.1 Proof of Theorem 5.3.1

We first recall sufficient conditions for *Bayes Formula* (e.g., [51, p. 216]) which is used in the proof of Theorem 5.3.1.

Lemma 5.8.1. *Consider a probability space $(\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{\mathbb{P}})$ where $\tilde{\mathbb{P}} \ll \tilde{\mathbb{Q}}$ for some probability measure $\tilde{\mathbb{Q}}$. If a σ-field $\mathcal{G} \subset \tilde{\mathcal{F}}$ and a random variable X is integrable ($E_{\tilde{\mathbb{P}}}[|X|] < \infty$), then Bayes formula holds, that is,*

$$E_{\tilde{\mathbb{P}}}[X|\mathcal{G}] = \frac{E_{\tilde{\mathbb{Q}}}[X\frac{d\tilde{\mathbb{P}}}{d\tilde{\mathbb{Q}}}|\mathcal{G}]}{E_{\tilde{\mathbb{Q}}}[\frac{d\tilde{\mathbb{P}}}{d\tilde{\mathbb{Q}}}|\mathcal{G}]}\ \tilde{\mathbb{P}} - a.s.$$

□

Proof of Theorem 5.3.1. Since policies do not change under the reduction, the proof of the result for globally and pbp optimal policies follows from (5.5). Now, we show the result for stationary policies. Let $\underline{\gamma}^*$ be a stationary policy for $(\mathcal{P}^D)$. In the following, we show that if $\underline{\gamma}^*$ satisfies (5.22), then it is also stationary under a policy-independent static reduction. Since $\underline{\gamma}^*$ is a stationary policy for $(\mathcal{P}^D)$, using Lemma 5.8.1, we have P-almost surely

$$\begin{aligned} 0 &= \nabla_{u^i} E_{\mathbb{P}}^{\gamma^{-i*}}[c(\omega_0, u^1, \ldots, u^N)|y^i]\Big|_{u^i=\gamma^{i*}(y^i)} \\ &= \nabla_{u^i}\left\{\frac{E_{\mathbb{Q}}^{\gamma^{-i*}}[\tilde{c}(\omega_0, u^1, \ldots, u^N, y^1, \ldots, y^N)|y^i]}{E_{\mathbb{Q}}^{\gamma^{-i*}}[\frac{d\mathbb{P}}{d\mathbb{Q}}|y^i]}\right\}\Big|_{u^i=\gamma^{i*}(y^i)}, \end{aligned} \tag{5.31}$$

where (5.31) follows from Lemma 5.8.1. Hence,

$$\left\{\frac{\left(\nabla_{u^i} E_{\mathbb{Q}}^{\gamma^{-i*}}[\tilde{c}(\omega_0, u^1, \ldots, u^N, y^1, \ldots, y^N)|y^i]\right) E_{\mathbb{Q}}^{\gamma^{-i*}}[\frac{d\mathbb{P}}{d\mathbb{Q}}|y^i]}{\left(E_{\mathbb{Q}}^{\gamma^{-i*}}[\frac{d\mathbb{P}}{d\mathbb{Q}}|y^i]\right)^2}\right. \tag{5.32}$$

$$- \frac{E_{\mathbb{Q}}^{\gamma^{-i*}}[\tilde{c}(\omega_0, u^1, \ldots, u^N, y^1, \ldots, y^N)|y^i]\left(\nabla_{u^i} E_{\mathbb{Q}}^{\gamma^{-i*}}[\frac{d\mathbb{P}}{d\mathbb{Q}}|y^i]\right)}{\left(E_{\mathbb{Q}}^{\gamma^{-i*}}[\frac{d\mathbb{P}}{d\mathbb{Q}}|y^i]\right)^2} \Bigg\}\Bigg|_{u^i=\gamma^{i*}(y^i)} = 0. \tag{5.33}$$

Since $\underline{\gamma}^*$ satisfies (5.22), we have P-almost surely (5.33) is equal to zero, and since $\frac{d\mathbb{P}}{d\mathbb{Q}} > 0$ P-almost surely and (5.32) must be equal to zero P-almost surely, we have P-almost surely

$$\nabla_{u^i} E_{\mathbb{Q}}^{\gamma^{-i*}}[\tilde{c}(\omega_0, u^1, \ldots, u^N, y^1, \ldots, y^N)|y^i]\Big|_{u^i=\gamma^{i*}(y^i)} = 0, \tag{5.34}$$

which implies that $\underline{\gamma}^*$ is a stationary policy for $(\mathcal{P}^D)$ under policy-independent static reductions.

For the converse statement, if a policy $\underline{\gamma}^*$ is stationary for $(\mathcal{P}^D)$ under policy-independent static reductions and satisfies (5.22), expressions (5.32) and (5.33) are equal to zero P-almost surely which implies that $\underline{\gamma}^*$ is a stationary policy for $(\mathcal{P}^D)$, and hence, this completes the proof. □

5.8.2 Proof of Corollary 5.3.1

We first recall a result which follows from [140, Theorem 3.3 and 3.4] and is used in the proof of Corollary 5.3.1.

Theorem 5.8.1. *[140, Theorem 3.3 and 3.4] Consider stochastic dynamic teams $(\mathcal{P})$ with a policy-independent static reduction* (1.3)*. Assume that*

(a) The cost function under the policy-independent static reductions, $\tilde{c}$, is convex in $u^1, \ldots, u^N$,

(b) The cost function under the policy-independent static reductions, $\tilde{c}$, is continuously differentiable in $u^1, \ldots, u^N$.

If $\underline{\gamma}^*$ is a stationary policy for the dynamic team $(\mathcal{P})$ under a policy-independent static reduction, and for all $\underline{\gamma} \in \Gamma$ with $E_{\mathbb{Q}}^{\underline{\gamma}}[\tilde{c}(\cdot)] < \infty$,

$$E_{\mathbb{Q}}\left[\nabla_{u^i}\tilde{c}(\omega_0, \underline{\gamma}^*(\underline{y}), \underline{y})\left(\gamma^i(y^{i*}) - \gamma^{i*}(y^{i*})\right)\right] < \infty \quad \textit{for all } i \in \mathcal{N}, \tag{5.35}$$

where

$$\begin{aligned}&\nabla_{u^i}\tilde{c}(\omega_0,\underline{\gamma}^*(\underline{y}),\underline{y})\\&:=\nabla_{u^i}\tilde{c}\Big(\omega_0,\gamma^{1*}(y^1),\ldots,\gamma^{(i-1)*}(y^{i-1}),u^i,\gamma^{(i+1)*}(y^{i+1}),\ldots,\gamma^{N*}(y^N),y^{1:N}\Big)\Big|_{u^i=\gamma^{i*}(y^i)},\end{aligned}$$

then $\underline{\gamma}^*$ is globally optimal for ($\mathcal{P}$). Moreover, if the cost function $\tilde{c}$ is strictly convex in $(u^1,\ldots,u^N)$, $\underline{\gamma}^*$ is the unique globally optimal policy for ($\mathcal{P}$).

□

Proof of Corollary 5.3.1. Following from Theorem 5.3.1, since stationary policy for ($\mathcal{P}$) satisfies (5.22), it is also stationary under policy-independent static reductions. In view of Theorem 5.8.1, since conditions (a) and (b) hold, we only need to show that (5.35) holds. Since the cost function c and $\frac{d\mathbb{P}}{d\mathbb{Q}}$ are differentiable functions in actions, (5.35) can be written as

$$\begin{aligned}&E_{\mathbb{Q}}\bigg[\Big(\nabla_{u^i}c(\omega_0,\underline{\gamma}^*(\underline{y}))\Big)\frac{d\mathbb{P}}{d\mathbb{Q}}\Big(\gamma^i(y^{i*})-\gamma^{i*}(y^{i*})\Big)\\&+\Big(\nabla_{u^i}\frac{d\mathbb{P}}{d\mathbb{Q}}(\omega_0,\underline{y},\underline{\gamma}^*(\underline{y}))\Big)c(\omega_0,\underline{\gamma}^*(\underline{y}^*))\Big(\gamma^i(y^{i*})-\gamma^{i*}(y^{i*})\Big)\bigg]\\&=E_{\mathbb{Q}}\bigg[\Big(\nabla_{u^i}c(\omega_0,\underline{\gamma}^*(\underline{y}))\Big)\frac{d\mathbb{P}}{d\mathbb{Q}}\Big(\gamma^i(y^{i*})-\gamma^{i*}(y^{i*})\Big)\bigg] &(5.36)\\&+E_{\mathbb{Q}}\bigg[\Big(\nabla_{u^i}\frac{d\mathbb{P}}{d\mathbb{Q}}(\omega_0,\underline{y},\underline{\gamma}^*(\underline{y}))\Big)c(\omega_0,\underline{\gamma}^*(\underline{y}^*))\Big(\gamma^i(y^{i*})-\gamma^{i*}(y^{i*})\Big)\bigg] &(5.37)\end{aligned}$$

where (5.36) is finite following from (5.23), and (5.37) is finite following from (5.24), and hence, (5.35) holds and Theorem 5.8.1 completes the proof. □

5.8.3 Proof of Theorem 5.4.2

For simplicity of our analysis, we consider 2-DM team problems; however, a similar proof can be used to show that the result holds for N-DM team problems as well. Consider a 2-DM stochastic dynamic team ($\mathcal{P}^D$) with $I^1=\{y_1^D\}$ and $I^2=\{y_2^D\}:=\{y_1^D,\hat{y}_2^D\}$, where $\hat{y}_2^D=g(\hat{y}_2^S,u^1)$, and y_1^D and $\hat{y}_2^S$ are primitive random variables.

Part (i), "⇒": We first show that if $\underline{\gamma}^{D,*}$ satisfying (C) is not a stationary policy for $(\mathcal{P}^D)$, then $\underline{\gamma}^{S,*}$ is not a stationary policy for $(\mathcal{P}^S)$.

(Step 1) If $\underline{\gamma}^{D,*}$ satisfying (C) is not a stationary policy for $(\mathcal{P}^D)$, then there is a set $B \subseteq \Omega$ with $P(B) > 0$ such that

$$\nabla_{u^2} E[c(\omega_0, \gamma_1^{D*}(y_1^D), u^2)|y_2^D]_{u^2=\gamma_2^{D*}(y_2^D)} \neq 0 \tag{5.38}$$

and/or

$$\nabla_{u^1} E[c(\omega_0, u^1, \gamma_2^{D*}(y_1^D, g(\hat{y}_2^S, u^1))|y_1^D]_{u^1=\gamma_1^{D*}(y_1^D)} \neq 0, \tag{5.39}$$

If (5.38) holds, then since (5.11) holds, and since under Assumption 5.2.2, g_i is invertable, we have on the set B

$$\nabla_{u^2} E[c(\omega_0, \gamma_1^{S*}(y_1^S), u^2)|y_2^S]_{u^2=\gamma_2^{S*}(y_2^S)} \neq 0 \tag{5.40}$$

which contradicts the assumption that $\underline{\gamma}^{S,*}$ is a stationary policy for $(\mathcal{P}^S)$.

(Step 2) Since $\underline{\gamma}^{D,*}$ satisfies (C), by the first-order approximation, we have for $\epsilon_n \in [0, 1)$ close to zero and a policy $\delta^1 \in \Gamma_1^D$

$$\begin{aligned}\gamma_2^{D*}\left(y_{2,\epsilon_n}^D\right) &:= \gamma_2^{D*}\left(y_1^D, g(\hat{y}_2^S, \gamma_1^{D*}(y_1^D) + \epsilon_n \delta^1(y_1^D))\right) \\ &= \gamma_2^{D*}\left(y_1^D, g(\hat{y}_2^S, \gamma_1^{D*}(y_1^D))\right) + \epsilon_n \tilde{\delta}\left(y_1^D, \hat{y}_2^S, \gamma_1^{D*}(y_1^D)\right) + O(\epsilon_n^2),\end{aligned} \tag{5.41}$$

where $y_{2,\epsilon_n}^D := [y_1^D, g(\hat{y}_2^S, \gamma_1^{D*}(y_1^D) + \epsilon_n \delta^1(y_1^D))]$ and $\tilde{\delta}$ satisfies

$$\begin{aligned}\tilde{\delta}\left(y_1^D, \hat{y}_2^S, \gamma_1^{D*}(y_1^D)\right) &:= D_u^{\delta^1} \gamma_2^{D*}\left(y_1^D, g(\hat{y}_2^S, u)\right)\Big|_{u=\gamma_1^{D*}(y_1^D)} \\ &:= \lim_{n\to\infty} \frac{1}{\epsilon_n}\left[\gamma_2^{D*}\left(y_1^D, g(\hat{y}_2^S, \gamma_1^{D*}(y_1^D) + \epsilon_n \delta^1)\right) - \gamma_2^{D*}\left(y_1^D, g(\hat{y}_2^S, \gamma_1^{D*}(y_1^D))\right)\right],\end{aligned}$$

where ϵ_n goes to zero as n goes to infinity and $D_u^{\delta^1}$ denotes the directional derivative with respect to u in the direction of δ^1 [89].

(Step 3) If (5.39) holds, then on the set B

$$\lim_{n\to\infty}\frac{1}{\epsilon_n}E\bigg[c\bigg(\omega_0,\gamma_1^{D*}(y_1^D)+\epsilon_n\delta^1(y_1^D),\gamma_2^{D*}(y_{2,\epsilon_n}^D)\bigg)-c\bigg(\omega_0,\gamma_1^{D*}(y_1^D),\gamma_2^{D*}(y_2^D)\bigg)\bigg|y_1^D\bigg]\neq 0, \tag{5.42}$$

for a non-zero $\delta^1(y_1^D)$, where $y_{2,\epsilon_n}^D := [y_1^D, g(\hat{y}_2^S, \gamma_1^{D*}(y_1^D)+\epsilon_n\delta^1(y_1^D))]$. Replacing (5.41) in (5.42), and using Assumption 5.4.1, we can see that the expression inside the conditional expectation (5.42) is convex and continuously differentiable in ϵ_n. Hence, similar to [78, Theorem 2 and 3], using the extended monotone convergence theorem, we can exchange the limit and the expectation. Therefore, since (5.11) holds, and also since $y_1^S = y_1^D$, on the set B

$$E\bigg[\lim_{n\to\infty}\frac{1}{\epsilon_n}c\bigg(\omega_0,\gamma_1^{S*}(y_1^S)+\epsilon_n\delta^1(y_1^S),\gamma_2^{S*}(y_2^S)+\epsilon_n\hat{\delta}(y_2^S)\bigg)-c\bigg(\omega_0,\gamma_1^{S*}(y_1^S),\gamma_2^{S*}(y_2^S)\bigg)\bigg|y_1^S\bigg]\neq 0, \tag{5.43}$$

for the policy $\hat{\delta}$ with $\hat{\delta}(y_2^S) = \tilde{\delta}(y_1^D, \hat{y}_2^S, \gamma_1^{D*}(y_1^D))$. Under Assumption 5.4.1, using the chain rule of derivative and linearity of the conditional expectation, (5.43) implies that for on the set B

$$E\bigg[\lim_{n\to\infty}\frac{1}{\epsilon_n}c\bigg(\omega_0,\gamma_1^{S*}(y_1^S)+\epsilon_n\delta^1(y_1^S),\gamma_2^{S*}(y_2^S)\bigg)-c\bigg(\omega_0,\gamma_1^{S*}(y_1^S),\gamma_2^{S*}(y_2^S)\bigg)\bigg|y_1^S\bigg] \tag{5.44}$$

$$+E\bigg[\lim_{n\to\infty}\frac{1}{\epsilon_n}c\bigg(\omega_0,\gamma_1^{S*}(y_1^S),\gamma_2^{S*}(y_2^S)+\epsilon_n\hat{\delta}(y_2^S)\bigg)-c\bigg(\omega_0,\gamma_1^{S*}(y_1^S),\gamma_2^{S*}(y_2^S)\bigg)\bigg|y_1^S\bigg]\neq 0. \tag{5.45}$$

(Step 4) If (5.44) is non-zero on the set B, then by exchanging the limit and expectation, we have on the set B

$$\nabla_{u^1}E[c(\omega_0,u^1,\gamma_2^{S*}(y_2^S))|y_1^S]_{u^1=\gamma_1^{S*}(y_1^S)}\neq 0, \tag{5.46}$$

which contradicts the assumption that $\underline{\gamma}^{S,*}$ is a stationary policy for $(\mathcal{P}^S)$. If (5.45) is non-zero on the set B, then we have on the set B

$$\nabla_{u^2} E[c(\omega_0, \gamma_1^{S*}(y_1^S), u^2)|y_2^S]_{u^2=\gamma_2^{S*}(y_2^S)} \neq 0. \tag{5.47}$$

That is because, $\sigma(y_1^S) \subseteq \sigma(y_2^S)$ and by the towering property of the conditional expectation, we have on the set B

$$\nabla_{u^2} E[E[c(\omega_0, \gamma_1^{S*}(y_1^S), u^2)|y_2^S]\hat{\delta}(y_2^S)|y_1^S]_{u^2=\gamma_2^{S*}(y_2^S)} \neq 0.$$

Hence, on the set B, (5.40) holds, which contradicts the assumption that $\underline{\gamma}^{S,*}$ is a stationary policy for $(\mathcal{P}^S)$.

"$\Leftarrow$": For the converse statement in Part(i), we can use similar steps. First, we note that if (5.40) holds on the set B on a set of positive measures B, then (5.38) holds on the set B, which contradicts the assumption that $\underline{\gamma}^{D,*}$ is a stationary policy for $(\mathcal{P}^D)$. Hence, P-almost surely

$$\nabla_{u^2} E[c(\omega_0, \gamma_1^{S*}(y_1^S), u^2)|y_2^S]_{u^2=\gamma_2^{S*}(y_2^S)} = 0. \tag{5.48}$$

Now, similar to the steps above, we can show that if $\underline{\gamma}^{D,*}$ is a stationary policy for $(\mathcal{P}^D)$, then sum of (5.44) and (5.45) is equal to zero P-almost surely. Hence, this implies that either both (5.44) and (5.45) are equal to zero P-almost surely or non of them equal to zero P-almost surely. But if (5.44) is not equal to zero on the set B, then similar to the above we can show that (5.40) holds for on the set B, which contradicts (5.48), and hence, this completes the proof of Part(i).

Part(ii): the proof follows from a similar argument as Part(i) since $\underline{\gamma}^{D*}$ satisfies (C). We note that Assumption 5.4.1(b) is not needed for proving the first claim of Part(ii) since we do not require exchanging the limit and the expectation (there is no $O(\epsilon_n^2)$ in (5.41) under the linearity assumption of $\underline{\gamma}^{D*}$). For the second claim of Part(ii), we only need to show that $c(\omega_0, u, \gamma_2^{D*}(y_1^D, g(\hat{y}_2^S, u))$ is

convex in u for every realizations of ζ, where γ_2^{D*} is linear in actions for every ζ. Let $\alpha \in [0,1]$, and consider arbitrary actions u and $\hat{u}$; then, for every ω_0,

$$
\begin{aligned}
&c\bigg(\omega_0, \alpha u + (1-\alpha)\hat{u}, \gamma_2^{D*}(y_1^D, g(\hat{y}_2^S, \alpha u + (1-\alpha)\hat{u}))\bigg) \\
&= c\bigg(\omega_0, \alpha u + (1-\alpha)\hat{u}, \alpha\gamma_2^{D*}(y_1^D, g(\hat{y}_2^S, u)) + (1-\alpha)\gamma_2^{D*}(y_1^D, g(\hat{y}_2^S, \hat{u}))\bigg) \qquad (5.49) \\
&\leq \alpha c\bigg(\omega_0, u, \gamma_2^{D*}(y_1^D, g(\hat{y}_2^S, u))\bigg) + (1-\alpha)c\bigg(\omega_0, u, \gamma_2^{D*}(y_1^D, g(\hat{y}_2^S, \hat{u}))\bigg), \qquad (5.50)
\end{aligned}
$$

where (5.49) follows from the linearity of $\underline{\gamma}^{D,*}$ and g_i in actions, and (5.50) follows from Assumption 5.4.1(b). The above convexity implies that stationary policies are pbp optimal for $(\mathcal{P}^D)$ under hypothesis of Part(ii), and hence, the proof is completed.

5.8.4 Proof of Theorem 5.5.2

Part(i): This follows from Theorem 5.5.1 since the static measurements with control-sharing reduction satisfying a relation (5.12) is policy independent, and the cost function remains unchanged the static measurements with control-sharing reduction. For the connections between stationary policies, we have P-almost surely

$$
\begin{aligned}
0 &= \nabla_{u^i} E\bigg[c\bigg(\gamma_1^{D,CS,*}(y_1^{D,CS}), \ldots, \gamma_{i-1}^{D,CS,*}(y_{i-1}^{D,CS}), u^i, \gamma_{i+1,u^i}^{D,CS*}(y_{i+1}^{D,CS}), \ldots, \gamma_N^{D,CS,*}(y_{N,u^i}^{D,CS})\bigg)\bigg| y_i^{D,CS}\bigg]\bigg|_{u^i=\gamma_i^{D,CS,*}(y_i^{D,CS})} \\
&= \nabla_{u^i} E\bigg[c\bigg(\gamma_1^{CS,*}(y_1^{CS}), \ldots, \gamma_{i-1}^{D,CS,*}(y_{i-1}^{CS}), u^i, \gamma_{i+1,u^i}^{CS*}(y_{i+1}^{CS}), \ldots, \gamma_N^{CS,*}(y_{N,u^i}^{CS})\bigg)\bigg| y_i^D, \gamma_{\downarrow i}^{CS*}(y_{\downarrow i}^{CS})\bigg]\bigg|_{u^i=\gamma_i^{CS,*}(y_i^{CS})} \qquad (5.51) \\
&= \nabla_{u^i} E\bigg[c\bigg(\gamma_1^{CS,*}(y_1^{CS}), \ldots, \gamma_{i-1}^{D,CS,*}(y_{i-1}^{CS}), u^i, \gamma_{i+1,u^i}^{CS*}(y_{i+1}^{CS}), \ldots, \gamma_N^{CS,*}(y_{N,u^i}^{CS})\bigg)\bigg| y_i^S, \gamma_{\downarrow i}^{CS*}(y_{\downarrow i}^{CS})\bigg]\bigg|_{u^i=\gamma_i^{CS,*}(y_i^{CS})} \qquad (5.52) \\
&= \nabla_{u^i} E\bigg[c\bigg(\gamma_1^{CS,*}(y_1^{CS}), \ldots, \gamma_{i-1}^{D,CS,*}(y_{i-1}^{CS}), u^i, \gamma_{i+1,u^i}^{CS*}(y_{i+1}^{CS}), \ldots, \gamma_N^{CS,*}(y_{N,u^i}^{CS})\bigg)\bigg| y_i^{CS}\bigg]\bigg|_{u^i=\gamma_i^{CS,*}(y_i^{CS})},
\end{aligned}
$$

where (5.51) follows from a relation (5.12) since the static measurements with control-sharing reduction satisfying a relation (5.12) is policy independent, and (5.52) follows from Assumption (5.2.2) since there is a bijection between y_i^D and y_i^S, and this completes the proof.

Part(ii): Let $\underline{\gamma}^{D,*}$ be a pbp optimal policy for $(\mathcal{P}^D)$, and let a policy $\underline{\gamma}^{D,CS,*} \in \Gamma^{D,CS}$ be such that for all $i \in \mathcal{N}$, $\gamma_i^{D,*}(y_i^D) = \gamma_i^{D,CS,*}(y_i^{D,CS})$ for almost all $u^{\downarrow i}$. A representation of policy $\underline{\gamma}^{D,CS,*}$ is $\underline{\gamma}^{D,*}$ itself, where for every $i \in \mathcal{N}$, the extra information $u^{\downarrow i}$ has not been used. In the following, we show that $\underline{\gamma}^{D,*}$ is also pbp optimal for $(\mathcal{P}^{D,CS})$. Assume it is not, then there is an index $i \in \mathcal{N}$ and a policy $\beta^i \in \Gamma_i^{D,CS}$ (with $(\beta^i, \gamma_{-i}^{D,*}) \in \Gamma^{D,CS}$) such that

$$E\Bigg[c\Bigg(\gamma_1^{D,*}(y_1^D), \ldots, \gamma_{i-1}^{D,*}(y_{i-1}^D), \beta^i(y_i^D, \gamma_{\downarrow i}^{D,*}(y_{\downarrow i}^D)), \gamma_{i+1}^{D,*}(y_{i+1}^D), \ldots, \gamma_N^{D,*}(y_N^D)\Bigg)\Bigg]$$
$$< E\Bigg[c\Bigg(\gamma_1^{D,*}(y_1^D), \ldots, \gamma_N^{D,*}(y_N^D)\Bigg)\Bigg]. \tag{5.53}$$

Since IS is partially nested, for a policy $(\beta^i, \gamma_{-i}^{D,*}) \in \Gamma^{D,CS}$, there exists a policy $(\hat{\gamma}_i^D, \gamma_{-i}^{D,*}) \in \Gamma^D$ such that $u^i = \beta^i(y_i^D, \gamma_{\downarrow i}^{D,*}(y_{\downarrow i}^D)) = \hat{\gamma}_i^D(y_i^D)$ for almost all ζ. We note that policies $\gamma_{-i}^{D,*}$ remains unchanged since the construction $\gamma_{-i}^{D,CS,*}$ from $\underline{\gamma}^{D,*}$ is independent of preceding policies and is only depend on preceding actions which are remained unchanged by the construction. Hence, (5.56) can be written as

$$E\Bigg[c\Bigg(\gamma_1^{D,*}(y_1^D), \ldots, \gamma_{i-1}^{D,*}(y_{i-1}^D), \hat{\gamma}_i^D(y_i^D), \gamma_{i+1}^{D,*}(y_{i+1}^D), \ldots, \gamma_N^{D,*}(y_N^D)\Bigg)\Bigg]$$
$$< E\Bigg[c\Bigg(\gamma_1^{D,*}(y_1^D), \ldots, \gamma_N^{D,*}(y_N^D)\Bigg)\Bigg],$$

which contradicts the assumption that $\underline{\gamma}^{D,*}$ is pbp optimal for $(\mathcal{P}^D)$, and hence, $\underline{\gamma}^{D,*}$ is pbp optimal for $(\mathcal{P}^{D,CS})$.

Now, we show the connections for stationary policies. Let $\underline{\gamma}^{D,*}$ be stationary for $(\mathcal{P}^D)$. Similar to the above, a representation of policy $\underline{\gamma}^{D,CS,*}$ is $\underline{\gamma}^{D,*}$ itself. In the following, we show that $\underline{\gamma}^{D,*}$

is also stationary for ($\mathcal{P}^{D,CS}$). We have P-almost surely

$$\begin{aligned}
&\nabla_{u^i} E\left[c\left(\gamma_1^{D,*}(y_1^D),\ldots,\gamma_{i-1}^{D,*}(y_{i-1}^D),u^i,\gamma_{i+1,u^i}^{D,*}(y_{i+1}^D),\ldots,\gamma_N^{D,*}(y_{N,u^i}^D)\right)\middle| y_i^{D,CS}\right]\Big|_{u^i=\gamma_i^{D,*}(y_i^D)}\\
&=\nabla_{u^i}\left[\int c\left(\gamma_1^{D,*}(y_1^D),\ldots,\gamma_{i-1}^{D,*}(y_{i-1}^D),u^i,\gamma_{i+1,u^i}^{D,*}(y_{i+1}^D),\ldots,\gamma_N^{D,*}(y_{N,u^i}^D)\right)P(d\zeta|y_i^D,\gamma_{\downarrow i}^{D,*}(y_{\downarrow i}^D))\right]\Big|_{u^i=\gamma_i^{D,*}(y_i^D)}\\
&=\nabla_{u^i}\left[\int c\left(\gamma_1^{D,*}(y_1^D),\ldots,\gamma_{i-1}^{D,*}(y_{i-1}^D),u^i,\gamma_{i+1,u^i}^{D,*}(y_{i+1}^D),\ldots,\gamma_N^{D,*}(y_{N,u^i}^D)\right)P(d\zeta|y_i^D)\right]\Big|_{u^i=\gamma_i^{D,*}(y_i^D)} \qquad (5.54)\\
&=\nabla_{u^i} E\left[c\left(\gamma_1^{D,*}(y_1^D),\ldots,\gamma_{i-1}^{D,*}(y_{i-1}^D),u^i,\gamma_{i+1,u^i}^{D,*}(y_{i+1}^D),\ldots,\gamma_N^{D,*}(y_{N,u^i}^D)\right)\middle| y_i^D\right]\Big|_{u^i=\gamma_i^{D,*}(y_i^D)}\\
&=0,
\end{aligned}$$

where (5.54) is true since P-almost surely

$$\begin{aligned}
P(d\zeta|y_i^D,\gamma_{\downarrow i}^{D,*}(y_{\downarrow i}^D)) &= \frac{P(d\zeta|y_i^D)P(d\gamma_{\downarrow i}^{D,*}(y_{\downarrow i}^D)|y_i^D,\zeta)}{P(d\gamma_{\downarrow i}^{D,*}(y_{\downarrow i}^D)|y_i^D)}\\
&= P(d\zeta|y_i^D), \qquad (5.55)
\end{aligned}$$

where (5.55) follows from the fact that y_i^D includes $y_{\downarrow i}^D$, and since $\gamma_{\downarrow i}^{D,*}$ are deterministic, and hence, given y_i^D, random variables $\gamma_{\downarrow i}^{S,*}(y_{\downarrow i}^D)$ are independent of ζ. This shows that $\underline{\gamma}^{D,*}$ is also stationary for ($\mathcal{P}^{D,CS}$), and completes the proof.

Part(iii): Let $\underline{\gamma}^{S,*}$ be pbp optimal for ($\mathcal{P}^S$), and let a policy $\underline{\gamma}^{CS,*}\in\Gamma^{CS}$ be such that for all $i\in\mathcal{N}$, $\gamma_i^{S,*}(y_i^S)=\gamma_i^{CS,*}(y_i^{CS})$ for almost all ζ. A representation of policy $\underline{\gamma}^{CS,*}$ is $\underline{\gamma}^{S,*}$ itself, where for every $i\in\mathcal{N}$, the extra information $u^{\downarrow i}$ has not been used. In the following, we show that $\underline{\gamma}^{S,*}$ is also pbp optimal for ($\mathcal{P}^{CS}$). Assume it is not, then there is an index $i\in\mathcal{N}$ and a policy $\beta^i\in\Gamma_i^{CS}$ (with $(\beta^i,\gamma_{-i}^{S,*})\in\Gamma^{CS}$) such that

$$\begin{aligned}
&E\left[c\left(\gamma_1^{S,*}(y_1^S),\ldots,\gamma_{i-1}^{S,*}(y_{i-1}^S),\beta^i(y_i^S,\gamma_{\downarrow i}^{S,*}(y_{\downarrow i}^S)),\gamma_{i+1}^{S,*}(y_{i+1}^S),\ldots,\gamma_N^{S,*}(y_N^S)\right)\right]\\
&<E\left[c\left(\gamma_1^{S,*}(y_1^S),\ldots,\gamma_N^{S,*}(y_N^S)\right)\right]. \qquad (5.56)
\end{aligned}$$

Since IS is partially nested, for a policy $(\beta^i, \gamma^{S,*}_{-i}) \in \Gamma^{CS}$, there exists a policy $(\hat{\gamma}^S_i, \gamma^{S,*}_{-i}) \in \Gamma^S$ such that $\beta^i(y^S_i, \gamma^{S,*}_{\downarrow i}(y^S_{\downarrow i})) = \hat{\gamma}^S_i(y^S_i)$ for almost all ζ. We note that policies $\gamma^{S,*}_{-i}$ remains unchanged since their representation is independent of precedent policies. Hence, (5.56) can be written as

$$E\Bigg[c\Bigg(\gamma^{S,*}_1(y^S_1), \ldots, \gamma^{S,*}_{i-1}(y^S_{i-1}), \hat{\gamma}^S_i(y^S_i), \gamma^{S,*}_{i+1}(y^S_{i+1}), \ldots, \gamma^{S,*}_N(y^S_N)\Bigg)\Bigg]$$
$$< E\Bigg[c\Bigg(\gamma^{S,*}_1(y^S_1), \ldots, \gamma^{S,*}_N(y^S_N)\Bigg)\Bigg],$$

which contradicts the assumption that $\underline{\gamma}^{S,*}$ is a pbp optimal policy for $(\mathcal{P}^S)$, and hence, $\underline{\gamma}^{S,*}$ is pbp optimal for $(\mathcal{P}^{CS})$. The proof to show that if $\underline{\gamma}^{S,*}$ is stationary for $(\mathcal{P}^S)$, then it is also stationary for $(\mathcal{P}^{CS})$, is identical to that of Part(ii).

5.8.5 Proof of Corollary 5.5.1

Part (i): The proof follows from the fact that the expected cost function is identical for policies $\underline{\gamma}^{D,*}$, $\underline{\gamma}^{S,*}$ and $\underline{\gamma}^{CS,*}$.

Part(ii): The proof follows from the fact that $\underline{\gamma}^{CS,*}$ can only be a linear function of actions of precedent DMs, and hence, we can use a similar argument as in the proof of Theorem 5.4.2(ii).

5.8.6 Proof of Corollary 5.6.1

Part (i): A policy $\underline{\gamma}^{D,*}$ and g_i are linear in actions, and hence, $\underline{\gamma}^{D,*}$ satisfies (C). Since $\underline{\gamma}^{S,*}$ is a stationary policy for $(\mathcal{P}^S)$ and $\underline{\gamma}^{D,*}$ satisfies (C), by Theorem 5.4.2, $\underline{\gamma}^{D,*}$ is a stationary policy (also pbp optimal following Theorem 5.4.2(ii)) for $(\mathcal{P}^D)$. If there exists another linear stationary policy $\underline{\tilde{\gamma}}^{D,*}$ for $(\mathcal{P}^D)$, then by Theorem 5.4.2, $\underline{\tilde{\gamma}}^{S,*}$ with $\tilde{\gamma}^{S,*}_i(y^S_i) = \tilde{\gamma}^{D,*}_i(y^D_i)$ is a stationary policy for $(\mathcal{P}^S)$, which contradicts the uniqueness of the stationary policy of $(\mathcal{P}^S)$.

Part(ii): This follows from the fact that $\underline{\gamma}^{D,*}$ is a globally optimal policy for $(\mathcal{P}^S)$ by Theorem 5.4.1.

Part(iii): This part follows from a similar argument as in Part(i), the only difference is that for every $\underline{\gamma}^{S,*}$, there exist multiple representations of $u^i = \gamma_i^{S,*}(y_i^S)$ for every $i \in \mathcal{N}$ of policies of $(\mathcal{P}^{CS})$. However, if there exists another essentially non-unique linear stationary policy for $(\mathcal{P}^{CS})$, by Theorem 5.5.1(ii), this leads to a contradiction to the uniqueness of the stationary policies of $(\mathcal{P}^S)$.

Part(iv): The proof follows from the fact that $\underline{\gamma}^{S,*}$ is a globally optimal policy for $(\mathcal{P}^S)$ and Theorem 5.5.1(i).

Chapter 6

Conclusions and Future works

In this thesis, we have focused on two main problems: (i) Characterization of existence, structural properties, and optimality conditions for large-scale stochastic static and dynamic teams and their mean-field limit; (ii) Characterization of informational properties of optimality notions under policy-independent and policy-dependent static reductions for stochastic dynamic teams.

Our ongoing work is on extending the results of Chapter 5 to game problems. We note that the specific form of an IS has been shown to have subtle impacts on different notions of equilibria as well as the existence and the uniqueness of the equilibrium for game problems (see for example [132, 9, 17, 16, 59, 94, 98]). We list below some salient features of changes in information in games:

(a) More information to a given player does not necessarily lead to better performance for that player (see e.g. [8, 22]).

(b) The team theoretic or stochastic control argument of a player *choosing to ignore the additional information* or *expansion of information structures as a recipe for optimization certification* (as in [139]) do not apply in games, even in zero-sum games.

(c) Dynamic games and their static reductions may lead to crucial differences and non-equivalence results. If information is expanded, e.g., by control-sharing, then there may be infinitely many, essentially non-unique (in that, the induced costs are different) equilibria and there

may not be one that is the most desirable for every player.

(d) On the other hand, zero-sum games have particularly appealing aspects. Under mild conditions, every zero-sum game, and accordingly every information structure, has a value (see, in particular, [68], generalizing a result of [95, Theorem 1]). Furthermore, for zero-sum games, a complete theory of ordering of information structures can be obtained (see [100] for finite space case and [68] for the standard Borel setup). In particular, additional information cannot hurt a player, though this argument is quite subtle in the game case: once again the *choose to ignore the additional information and thus no information cannot hurt* argument is not applicable, even though the implication that more information cannot hurt turns out to be correct via a more delicate argument.

Our preliminary results in [114] suggest that an equivalence of ISs (static reductions in team theory in Chapter 5), no longer holds in general for games under policy-independent and policy-dependent static reductions. In particular, we have shown that although the results for teams under policy-independent static reductions hold for games, the results for teams under policy-dependent static reductions, no longer hold for dynamic nonzero-sum games. Hence, we have established some negative and positive results on the connections between Nash equilibrium of nonzero-sum dynamic games under policy-independent and policy-dependent static reductions. At the end, thanks to the ordered interchangeability property of pairs of saddle-point equilibria, we have relaxed sufficient conditions for nonzero-sum games, through showing that similar results as those for teams hold for zero-sum games. At the end, we note that an equivalence of Information structures (static reductions), no longer holds in general for games under policy-independent and policy-dependent static reductions.

Also, study of games with non-identical subjective priors, as in [13], is also of interest and a potential future work. Our analysis in Chapter 5 and [114] does not hold for these challenging problems but we expect some of the tools developed here to be found useful in that more general context.

Bibliography

[1] D. J. Aldous, I. A. Ibragimov, and J. Jacod. *Ecole d'Ete de Probabilites de Saint-Flour XIII, 1983*, volume 1117. Springer, 1985.

[2] C. D. Aliprantis and K. C. Border. *Infinite Dimensional Analysis: A Hitchhiker-s Guide, third edition*. Springer, Berlin, 2006.

[3] M. Andersland and D. Teneketzis. Information structures, causality, and non-sequential stochastic control, I: design-independent properties. *SIAM J. Control and Optimization*, 30:1447 – 1475, 1992.

[4] M. Andersland and D. Teneketzis. Information structures, causality, and non-sequential stochastic control, II: design-dependent properties. *SIAM J. Control and Optimization*, 32:1726 – 1751, 1994.

[5] J. Arabneydi and A. Aghdam. A certainty equivalence result in team-optimal control of mean-field coupled markov chains. In *IEEE 56th Annual Conference on Decision and Control (CDC)*, pages 3125–3130, 2017.

[6] J. Arabneydi and A. Mahajan. Team-optimal solution of finite number of mean-field coupled LQG subsystems. In *IEEE 54th Annual Conference on Decision and Control (CDC)*, pages 5308–5313, 2015.

[7] A. Arapostathis, A. Biswas, and J. Carroll. On solutions of mean field games with ergodic cost. *Journal de Mathématiques Pures et Appliquées*, 107(2):205–251, 2017.

[8] T. Başar. A contradictory property of the Nash solution for two stochastic nonzero-sum games. In *Proc. 10th Allerton Conference On Circnit and System Theory*, pages 819–827, 1972.

[9] T. Başar. A counterexample in linear-quadratic games: Existence of nonlinear nash solutions. *Journal of Optimization Theory and Applications*, 14(4):425–430, 1974.

[10] T. Başar. Some thoughts on saddle-point conditions and information structures in zero-sum differential games. *Journal of Optimization Theory and Applications*, 18(1):165–170, 1976.

[11] T. Başar. Two general properties of the saddle-point solutions of dynamic games. *IEEE Transactions on Automatic Control*, 22(1):124–126, 1977.

[12] T. Başar. Decentralized multicriteria optimization of linear stochastic systems. *IEEE Transactions on Automatic Control*, 23(2):233–243, 1978.

[13] T. Başar. An equilibrium theory for multiperson decision making with multiple probabilistic models. *IEEE Transactions on Automatic Control*, 30(2):118–132, 1985.

[14] T. Başar. Time consistency and robustness of equilibria in non-cooperative dynamic games. In *Contributions to economic analysis*, volume 181, pages 9–54. Elsevier, 1989.

[15] T. Başar and Y. C. Ho. Informational properties of the Nash solutions of two stochastic nonzero-sum games. *Journal of Economic Theory*, 7(4):370–387, 1974.

[16] T. Başar and G.J. Olsder. *Dynamic Noncooperative Game Theory*. SIAM Classics in Applied Mathematics, Philadelphia, PA, 1999.

[17] T. Başar and H. Selbuz. Properties of nash solutions of a two-stage nonzero-sum game. *IEEE Transactions on Automatic Control*, 21(1):48–54, 1976.

[18] T. Banica, S. Curran, and R. Speicher. De finetti theorems for easy quantum groups. *The Annals of Probability*, 40(1):401–435, 2012.

[19] M. Bardi and M. Fischer. On non-uniqueness and uniqueness of solutions in finite-horizon mean field games. *ESAIM: Control, Optimisation and Calculus of Variations*, 25:44, 2019.

[20] M. Bardi and F. S. Priuli. Linear-quadratic N-person and mean-field games with ergodic cost. *SIAM Journal on Control and Optimization*, 52(5):3022–3052, 2014.

[21] T. Başar. On the saddle-point solution of a class of stochastic differential games. *Journal of Optimization Theory and Applications*, 33(4):539–556, 1981.

[22] B. Bassan, O. Gossner, M. Scarsini, and S. Zamir. Positive value of information in games. *International Journal of Game Theory*, 32:17–31, December 2003.

[23] E. Bayraktar and X. Zhang. On non-uniqueness in mean field games. *Proceedings of the American Mathematical Society*, 148:4091–4106, 2020.

[24] V. E. Beneš. Existence of optimal stochastic control laws. *SIAM Journal on Control*, 9(3):446–472, 1971.

[25] D. P. Bertsekas and S. Shreve. *Stochastic Optimal Control: The Discrete Time Case*. Academic Press, New York, 1978.

[26] P. Billingsley. *Convergence of Probability Measures*. Wiley, New York, 1968.

[27] A. Biswas. Mean field games with ergodic cost for discrete time markov processes. *arXiv preprint arXiv:1510.08968*, 2015.

[28] D. Blackwell. Memoryless strategies in finite-stage dynamic programming. *Annals of Mathematical Statistics*, 35:863–865, 1964.

[29] V. S. Borkar. White-noise representations in stochastic realization theory. *SIAM Journal on Control and Optimization*, 31:1093–1102, 1993.

[30] V.S. Borkar. Average cost dynamic programming equations for controlled Markov chains with partial observations. *SIAM Journal on Control and Optimization*, 39(3):673–681, 2000.

[31] V.S. Borkar. Dynamic programming for ergodic control of Markov chains under partial observations: A correction. *SIAM Journal on Control and Optimization*, 45(6):2299–2304, 2007.

[32] F. G. S. L. Brandao and A. W. Harrow. Quantum de finetti theorems under local measurements with applications. *Communications in Mathematical Physics*, 353(2):469–506, 2017.

[33] N. Brunner, D. Cavalcanti, S. Pironio, V. Scarani, and S. Wehner. Bell nonlocality. *Reviews of Modern Physics*, 86(2):419, 2014.

[34] R. Buckdahn, J. Li, and S. Peng. Nonlinear stochastic differential games involving a major player and a large number of collectively acting minor agents. *SIAM Journal on Control and Optimization*, 52(1):451–492, 2014.

[35] L. Campi and M. Fischer. Correlated equilibria and mean field games: a simple model. *arXiv preprint arXiv:2004.06185*, 2020.

[36] P. Cardaliaguet. Notes on mean field games. Technical report, Technical report, 2010.

[37] P. Cardaliaguet, F. Delarue, J. Lasry, and P. Lions. *The Master Equation and the Convergence Problem in Mean Field Games:(AMS-201)*, volume 201. Princeton University Press, 2019.

[38] P. Cardaliaguet and C. Rainer. An example of multiple mean field limits in ergodic differential games. *Nonlinear Differential Equations and Applications NoDEA*, 27(3):1–19, 2020.

[39] R. Carmona and F. Delarue. *Probabilistic Theory of Mean Field Games with Applications I-II*. Springer, 2018.

[40] R. Carmona, F. Delarue, and D. Lacker. Mean field games with common noise. *The Annals of Probability*, 44(6):3740–3803, 2016.

[41] D. Castanon and N. Sandell. Equilibrium solutions in sequential stochastic games. *IEEE Transactions on Automatic Control*, 22(4):624–627, 1977.

[42] C. M. Caves, C. A. Fuchs, and R. Schack. Unknown quantum states: the quantum de finetti representation. *Journal of Mathematical Physics*, 43(9):4537–4559, 2002.

[43] A. Cecchin and M. Fischer. Probabilistic approach to finite state mean field games. *Applied Mathematics & Optimization*, pages 1–48, 2018.

[44] C. D. Charalambous. Decentralized optimality conditions of stochastic differential decision problems via Girsanov's measure transformation. *Mathematics of Control, Signals, and Systems*, 28(3):1–55, 2016.

[45] M. Christandl and B. Toner. Finite de finetti theorem for conditional probability distributions describing physical theories. *Journal of Mathematical Physics*, 50(4):042104, 2009.

[46] M. Colombino, R. Smith, and T. H. Summers. Mutually quadratically invariant information structures in two-team stochastic dynamic games. *IEEE Transactions on Automatic Control*, 63(7):2256–2263, 2017.

[47] P. Dasgupta and E. Maskin. The existence of equilibrium in discontinuous economic games, I: Theory. *The Review of Economic Studies*, 53(1):1–26, 1986.

[48] F. Delarue and R. Tchuendom. Selection of equilibria in a linear quadratic mean-field game. *Stochastic Processes and their Applications*, 130(2):1000–1040, 2020.

[49] P. Diaconis and D. Freedman. Finite exchangeable sequences. *The Annals of Probability*, pages 745–764, 1980.

[50] R. M. Dudley. *Real Analysis and Probability*. Cambridge University Press, Cambridge, 2nd edition, 2002.

[51] R. Durrett. *Probability: theory and examples*, volume 3. Cambridge university press, 2010.

[52] E. A. Feinberg, P. O. Kasyanov, and N. V. Zadoianchuk. Fatou's lemma for weakly converging probabilities. *Theory of Probability and Its Applications*, 58(4):683–689, 2014.

[53] E. A. Feinberg, P. O. Kasyanov, and M. Z. Zgurovsky. Partially observable total-cost Markov decision processes with weakly continuous transition probabilities. *Mathematics of Operations Research*, 41(2):656–681, 2016.

[54] E. Feleqi. The derivation of ergodic mean field game equations for several populations of players. *Dynamic Games and Applications*, 3(4):523–536, 2013.

[55] A. Filippov. On certain questions in the theory of optimal control. *Journal of the Society for Industrial and Applied Mathematics, Series A: Control*, 1(1):76–84, 1962.

[56] M. Fischer. On the connection between symmetric N-player games and mean field games. *Annals of Applied Probability*, 27(2):757–810, 2017.

[57] W. H. Fleming. *Functions of Several Variables*. Addison-Wesley Publishing Co., Reading, Mass.-London, 1965.

[58] D. Fudenberg and D. Levine. Open-loop and closed-loop equilibria in dynamic games with many players. *Journal of Economic Theory*, 44(1):1–18, 1988.

[59] D. Fudenberg and J. Tirole. Game theory. *MIT press*, 1991.

[60] A. Gattami. Team decision problems with convex quadratic constraints. *IEEE Transactions on Control of Network Systems*, 4(3):555–563, 2017.

[61] I. V. Girsanov. On transforming a certain class of stochastic processes by absolutely continuous substitution of measures. *Theory of Probability & Its Applications*, 5(3):285–301, 1960.

[62] A. Gupta, S. Yüksel, T. Başar, and C. Langbort. On the existence of optimal policies for a class of static and sequential dynamic teams. *SIAM Journal on Control and Optimization*, 53:1681–1712, 2015.

[63] B. Hajek and M. Livesay. On non-unique solutions in mean field games. In *IEEE 58th Annual Conference on Decision and Control (CDC)*, pages 1219–1224, 2019.

[64] O. Hernández-Lerma and J. B. Lasserre. *Discrete-Time Markov Control Processes: Basic Optimality Criteria*. Springer, 1996.

[65] E. Hewitt and L. J. Savage. Symmetric measures on cartesian products. *Transactions of the American Mathematical Society*, 80(2):470–501, 1955.

[66] Y. C. Ho and K. C. Chu. Team decision theory and information structures in optimal control problems - part I. *IEEE Transactions on Automatic Control*, 17:15–22, February 1972.

[67] Y. C. Ho and K. C. Chu. On the equivalence of information structures in static and dynamic teams. *IEEE Transactions on Automatic Control*, 18(2):187–188, 1973.

[68] I. Hogeboom-Burr and S. Yüksel. Comparison of information structures for zero-sum games and a partial converse to Blackwell ordering in standard Borel spaces. *To appear in SIAM Journal on Control and Optimization, also arXiv:2005.06673*, 2020.

[69] M. Huang, P. E. Caines, and R. P. Malhamé. Large population stochastic dynamic games: closed-loop Mckean-Vlasov systems and the Nash certainty equivalence principle. *Communications in Information and Systems*, 6:221–251, 2006.

[70] M. Huang, P. E. Caines, and R. P. Malhamé. Large-population cost-coupled LQG problems with nonuniform agents: Individual-mass behavior and decentralized ϵ-Nash equilibria. *IEEE Transactions on Automatic Control*, 52:1560–1571, 2007.

[71] M. Huang, P. E. Caines, and R. P. Malhamé. Social optima in mean field LQG control: centralized and decentralized strategies. *IEEE Transactions on Automatic Control*, 57(7):1736–1751, 2012.

[72] M. Huang, R.P. Malhamé, and P.E. Caines. Large population stochastic dynamic games: Closed-loop McLean-Vlasov systems and the Nash certainty equivalence principle. *Special issue in honour of the 65th birthday of Tyrone Duncan, Communications in Information and Systems*, 6:221–252, 2006.

[73] M. Huang and S. L. Nguyen. Stochastic control of mean field models with mixed players. In *Proc. 18th IFAC World Congress, Milan, Italy*, 2011.

[74] M. Huang and S. L. Nguyen. Linear-quadratic mean field teams with a major agent. In *IEEE 55th Conference on Decision and Control (CDC)*, pages 6958–6963, 2016.

[75] O. Kallenberg. Canonical representations and convergence criteria for processes with interchangeable increments. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 27(1):23–36, 1973.

[76] O. Kallenberg. *Probabilistic symmetries and invariance principles*. 2006.

[77] J. F. C. Kingman. Uses of exchangeability. *The Annals of Probability*, 6(2):183–197, 1978.

[78] J. C. Krainak, J. L. Speyer, and S. I. Marcus. Static team problems – part I: Sufficient conditions and the exponential cost criterion. *IEEE Transactions on Automatic Control*, 27:839–848, April 1982.

[79] J.C. Krainak, J.S. Speyer, and S.I. Marcus. Static team problems – part II: Affine control laws, projections, algorithms, and the LEGT problem. *IEEE Transactions on Automatic Control*, 27:848–859, 1982.

[80] D. Lacker. Mean field games via controlled martingale problems: existence of markovian equilibria. *Stochastic Processes and their Applications*, 125(7):2856–2894, 2015.

[81] D. Lacker. A general characterization of the mean field limit for stochastic differential games. *Probability Theory and Related Fields*, 165(3-4):581–648, 2016.

[82] D. Lacker. Limit theory for controlled Mckean–Vlasov dynamics. *SIAM Journal on Control and Optimization*, 55(3):1641–1672, 2017.

[83] D. Lacker. On the convergence of closed-loop nash equilibria to the mean field game limit. *The Annals of Applied Probability*, 30(4):1693–1761, 2020.

[84] A. Lamperski and L. Lessard. Optimal decentralized state-feedback control with sparsity and delays. *Automatica*, 58:143–151, 2015.

[85] H. J. Langen. Convergence of dynamic programming models. *Mathematics of Operations Research*, 6(4):493–512, 1981.

[86] J. M. Lasry and P. L. Lions. Mean field games. *Japanese J. of Mathematics*, 2:229–260, 2007.

[87] L. Lessard and S. Lall. Optimal control of two-player systems with output feedback. *IEEE Transactions on Automatic Control*, 60(8):2129–2144, 2015.

[88] B. Light and G. Y. Weintraub. Mean field equilibrium: uniqueness, existence, and comparative statics. *Existence, and Comparative Statics*, 2018.

[89] D. G. Luenberger. *Optimization by Vector Space Methods*. John Wiley & Sons, New York, NY, 1969.

[90] A. Mahajan, N. C. Martins, and S. Yüksel. Static LQG teams with countably infinite players. In *IEEE 52nd Annual Conference on Decision and Control (CDC)*, pages 6765–6770, 2013.

[91] A. Mahajan, N.C. Martins, M. Rotkowitz, and S. Yüksel. Information structures in optimal decentralized control. In *IEEE Conference on Decision and Control*, Hawaii, USA, 2012.

[92] J. Marschak. Elements for a theory of teams. *Management Science*, 1:127–137, 1955.

[93] A. Mas-Colell. On a theorem of schmeidler. *Journal of Mathematical Economics*, 13(3):201–206, 1984.

[94] J. Mertens. Repeated games. In *Game Theory and Applications*, pages 77–130. Elsevier, 1990.

[95] P. Milgrom and R. Weber. Distributional strategies for games with incomplete information. *Mathematics of operations research*, 10(4):619–632, 1985.

[96] J. Nash. Non-cooperative games. *Annals of mathematics*, 1951.

[97] A. Nayyar and L. Lessard. Optimal control for LQG systems on graphs—part I: Structural results. *arXiv preprint arXiv:1408.2551*, 2014.

[98] M. Osborne and A. Rubinstein. *A course in game theory*. MIT press, 1994.

[99] Y. Ouyang, S. M. Asghari, and A. Nayyar. Optimal infinite horizon decentralized networked controllers with unreliable communication. *arXiv preprint arXiv:1806.06497*, 2018.

[100] M. Peski. Comparison of information structures in zero-sum games. *Games and Economic Behaviour*, 62:732–735, 2008.

[101] S. Popescu. Nonlocality beyond quantum mechanics. *Nature Physics*, 10(4):264, 2014.

[102] R. Radner. Team decision problems. *Annals of Mathematical Statistics*, 33:857–881, 1962.

[103] R. Renner. Symmetry of large physical systems implies independence of subsystems. *Nature Physics*, 3(9):645, 2007.

[104] M. Rotkowitz. On information structures, convexity, and linear optimality. In *IEEE 47th Annual Conference on Decision and Control (CDC)*, pages 1642–1647, 2008.

[105] M. Rotkowitz and S. Lall. A characterization of convex problems in decentralized control. *IEEE Transactions on Automatic Control*, 51:274–286, February 2006.

[106] S. Sabău and N. C. Martins. Youla-like parametrizations subject to QI subspace constraints. *IEEE transactions on Automatic Control*, 59(6):1411–1422, 2014.

[107] S. Sabău, N. C. Martins, and M. Rotkowitz. A convex characterization of multidimensional linear systems subject to SQI constraints. *IEEE Transactions on Automatic Control*, 62(6):2981–2986, 2016.

[108] N. Saldi. A topology for team policies and existence of optimal team policies in stochastic team theory. *IEEE Transactions on Automatic Control*, 65(1):310–317, 2019.

[109] N. Saldi. A topology for team policies and existence of optimal team policies in stochastic team theory. *IEEE Transactions on Automatic Control*, 65(1):310–317, 2020.

[110] N. Saldi, T. Başar, and M. Raginsky. Approximate Nash equilibria in partially observed stochastic games with mean-field interactions. *Mathematics of Operations Research*, 2019.

[111] N. Saldi and S. Yüksel. Geometry of information structures, strategic measures and associated control topologie. *arXiv*, pages arXiv–2010.07377, 2020.

[112] N. Saldi, S. Yüksel, and T. Linder. Finite model approximations and asymptotic optimality of quantized policies in decentralized stochastic control. *IEEE Transactions on Automatic Control*, 2017.

[113] S. Sanjari, T. Başar, and S. Yüksel. Optimality conditions under policy-dependent and policy-independent static reductions of stochastic dynamic teams and games—Part I: Dynamic teams. *arXiv preprint arXiv:*, 2021.

[114] S. Sanjari, T. Başar, and S. Yüksel. Optimality conditions under policy-dependent and policy-independent static reductions of stochastic dynamic teams and games—Part II: Dynamic games. *arXiv preprint arXiv:*, 2021.

[115] S. Sanjari, N. Saldi, and S. Yüksel. Independently randomized symmetric policies are optimal for exchangeable stochastic teams with infinitely many decision makers. In *2020 59th IEEE Conference on Decision and Control (CDC)*, pages 5986–5991. IEEE, 2020.

[116] S. Sanjari, N. Saldi, and S. Yüksel. Optimality of independently randomized symmetric policies for exchangeable stochastic teams with infinitely many decision makers. *arXiv preprint arXiv:2008.11570*, 2020.

[117] S. Sanjari and S. Yüksel. Optimal stochastic teams with infinitely many decision makers and Mean-field teams. In *IEEE 57th Annual Conference on Decision and Control (CDC)*, pages 4123–4128, 2018.

[118] S. Sanjari and S. Yüksel. Convex symmetric stochastic dynamic teams and their mean-field limit. In *IEEE 58th Annual Conference on Decision and Control (CDC)*, pages 4662–4667, 2019.

[119] S. Sanjari and S. Yüksel. Optimal policies for convex symmetric stochastic dynamic teams and their mean-field limit. *SIAM Journal on Control and Optimization*, 59(2):777–804, 2021.

[120] S. Sanjari and S. Yüksel. Optimal solutions to infinite-player stochastic teams and mean-field teams. *IEEE Transactions on Automatic Control*, 66(3):1071–1086, 2021.

[121] M. Schäl. Conditions for optimality in dynamic programming and for the limit of n-stage optimal policies to be optimal. *Z. Wahrscheinlichkeitsth*, 32:179–296, 1975.

[122] D. Schmeidler. Equilibrium points of nonatomic games. *Journal of Statistical Physics*, 7(4):295–300, 1973.

[123] R. Serfozo. Convergence of Lebesgue integrals with varying measures. *Sankhyā: The Indian Journal of Statistics, Series A*, pages 380–402, 1982.

[124] P. Shah and P. Parrilo. H2 optimal decentralized control over posets: A state-space solution for state-feedback. *IEEE Transactions on Automatic Control*, 58(12):3084–3096, 2013.

[125] J. Swigart and S. Lall. Optimal controller synthesis for decentralized systems over graphs via spectral factorization. *IEEE Transactions on Automatic Control*, 59(9):2311–2323, 2014.

[126] D. Teneketzis. On information structures and nonsequential stochastic control. *CWI Quarterly*, 9:241–260, 1996.

[127] B. C. Wang and J. F. Zhang. Social optima in mean field linear-quadratic-gaussian models with markov jump parameters. *SIAM Journal on Control and Optimization*, 55(1):429–456, 2017.

[128] D. B. West. *Introduction to graph theory*, volume 2. Prentice hall Upper Saddle River, NJ, 1996.

[129] H. Witsenhausen. Alternatives to the tree model for extensive games. In *The Theory and Application of Differential Games*, pages 77–84. Springer, 1975.

[130] H. S. Witsenhausen. A counterexample in stochastic optimal control. *SIAM Journal on Control and Optimization*, 6:131–147, 1968.

[131] H. S. Witsenhausen. On information structures, feedback and causality. *SIAM J. Control*, 9:149–160, May 1971.

[132] H. S. Witsenhausen. On the relations between the values of a game and its information structure. *Information and Control*, 19(3):204–215, 1971.

[133] H. S. Witsenhausen. A standard form for sequential stochastic control. *Mathematical Systems Theory*, 7:5–11, 1973.

[134] H. S. Witsenhausen. The intrinsic model for discrete stochastic control: Some open problems. *Lecture Notes in Econ. and Math. Syst., Springer-Verlag*, 107:322–335, 1975.

[135] H.S. Witsenhausen. Equivalent stochastic control problems. *Mathematics of Control, Signals and Systems*, 1(1):3–11, 1988.

[136] L. Young. Generalized curves and the existence of an attained absolute minimum in the calculus of variations. *Comptes Rendus de la Societe des Sci. et des Lettres de Varsovie*, 30:212–234, 1937.

[137] S. Yüksel. On stochastic stability of a class of non-Markovian processes and applications in quantization. *SIAM Journal on Control and Optimization*, 55:1241–1260, 2017.

[138] S. Yüksel. A universal dynamic program and refined existence results for decentralized stochastic control. *SIAM Journal on Control and Optimization*, 58(5):2711–2739, 2020.

[139] S. Yüksel and T. Başar. *Stochastic Networked Control Systems: Stabilization and Optimization under Information Constraints*. Springer, New York, 2013.

[140] S. Yüksel and N. Saldi. Convex analysis in decentralized stochastic control, strategic measures and optimal solutions. *SIAM Journal on Control and Optimization*, 55:1–28, 2017.

[141] K. Zhou, J. C. Doyle, and K. Glover. *Robust and optimal control*, volume 40. Prentice-Hall, 1996.