Sequential Decision Making in Decentralized Systems

by

Ashutosh Nayyar

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Electrical Engineering: Systems)
in The University of Michigan
2011

Doctoral Committee:

Professor Demosthenis Teneketzis, Chair
Emeritus Professor Stephen Pollock
Associate Professor Achilleas Anastasopoulos
Associate Professor Mingyan Liu
Associate Professor Mark P. Van Oyen
Emeritus Professor Pravin Varaiya, University of California, Berkeley

To my family

ACKNOWLEDGEMENTS

I started my graduate studies with a very vague idea of wanting to work on problems that would spark my interest without frustrating my abilities. I would like to express my sincerest gratitude to Professor Demosthenis Teneketzis for accepting me as a student and guiding me to a research field that has proven to be both stimulating and enjoyable. I would also like to thank him for our many discussions (not all technical) and for pointing out to me that some of my observations had more value than I initially perceived. I have thoroughly enjoyed working with him and learning from him.

I would also like to thank the members of my thesis committee for their feedback and encouragement. Their interest in my research has been a great source of pleasure and motivation for me.

I am grateful for the opportunity to work with many interesting students at Michigan. I would specially like to thank Aditya Mahajan who in many respects has been my second advisor. Our work together and our many discussions have greatly enhanced my understanding and vastly improved the quality of my work. I have also enjoyed working with David Shuman and I have greatly benefited from his tips about everything from writing, presentations to Matlab coding. I would also like to thank Ali Kakhbod, Shrutivandana Sharma and Yi Wang for many discussions and helpful suggestions.

I would like to thank my friends and roommates in Ann Arbor for all their help and support. I am specially thankful to Demos and Barbara for the many evenings I

have spent at their home and for many hours of *Jeopardy!* we have watched together.

Finally, I am grateful to my family for their love and support and for putting my happiness before their own.

TABLE OF CONTENTS

LIST OF FIGURES

Figure

LIST OF APPENDICES

Appendix

ABSTRACT

Sequential Decision Making in Decentralized Systems

by

Ashutosh Nayyar

Chair: Demosthenis Teneketzis

We study sequential decision making problems in cooperative systems where different agents with different information want to achieve a common objective. The sequential nature of the decision problem implies that all decisions can be arranged in a sequence such that the information available to make the t^{th} decision only depends on preceding decisions. Markov decision theory provides tools for addressing sequential decision making problems with classical information structures. In this thesis, we introduce a new approach for decision making problems with non-classical information structures. This approach relies on the idea of common information between decision-makers. Intuitively, common information consists of past observations and decisions that are commonly known to the current and future decision makers. We show that a common information based approach can allow us to discover new structural results of optimal decision strategies and provide a sequential decomposition of the decision-making problems. We first demonstrate this approach on two specific instances of sequential problems, namely, a real-time multi-terminal communication system and a decentralized control system with delayed sharing of information. We then show that the common information methodology applies more generally to any

sequential decision making problem. Moreover, we show that our common information methodology unifies the separate sequential decomposition results available for classical and non-classical information structures. We also present sufficient conditions for simplifying common information based sequential decompositions. This simplification relies on the concept of state sufficient for the input output map of a coordinator that only knows the common information.

CHAPTER I

Introduction

Decentralized systems are ubiquitous in the modern technological world. Communication systems like the Internet and wireless networks, information-gathering systems like sensor networks and surveillance networks, spatially distributed systems like power generation and transmission systems, transportation networks, and networked control systems are all examples of decentralized systems. Such systems are characterized by the presence of multiple agents/decision-makers that may observe their environment, communicate with each other and make decisions that affect the overall system. Two salient features of such systems are:

1. *Decentralization of Information:* Different agents have different information about the system.

2. *Decentralization of Decision-Making:* Multiple agents have to make decisions in the absence of a centralized decision-making authority.

The main focus of this thesis is to investigate the relationships between the above two features of decentralized systems: *how does decentralization of information affect decision-making by different agents?*. An agent uses the information it has to make decisions according to a *decision rule/decision strategy*. Thus, a decision rule/decision strategy is a mapping from an agent's information to its decisions. A *decision strategy* *profile* is a collection of decision rules/decision strategies for all the decision-makers. The goal of this thesis is to characterize optimal decision strategy profiles in some specific instances of decentralized systems as well as to develop general methodologies for finding optimal decision strategy profiles. In this thesis, we focus on decentralized decision-making problems that are:

- **Cooperative:** We consider decentralized decision-making problems where different decision-makers share the same objective. Such problems are called *Team Problems* (*Radner* (1962), *Marschak and Radner* (1972), *Y. C. Ho* (1980), *Ho and Chu* (1972), *Ho et al.* (1978) etc.) In contrast, Game Theory deals with decision-making problems where different decision makers may have different objectives.

- **Stochastic:** We consider decision-making problems where the decision-making environment includes uncertain inputs. We assume stochastic models of uncertainties about the decision-making environment. The presence of uncertainty implies that under any choice of decision strategy profile, the value of the objective is a random variable. In this thesis, we seek to find strategy profiles that maximize the *expected value* of the objective.

- **Sequential:** We consider decision-making problems where the decision-makers act in a pre-determined order that is independent of events in the nature or the decision strategy profile. Further, the information available to make a decision does not depend on the decisions to be made in the future.

1.1 Information Structures and Optimization Approaches

A key concept in decision-making problems is the concept of Information Structure. Information structure of a decision-making problem describes what information is available to each decision-maker. In sequential decision-making problems, information structures can be classified as classical or non-classical. In classical information structures, a decision-maker knows all the information available to all agents that acted before it. When an information structure does not satisfy this condition, it is called non-classical. Markov decision theory provides a systematic way of solving sequential decision-making problems with classical information structure [*Kumar and Varaiya* (1986a)]. This theory allows us to decompose the problem of finding optimal strategies into several smaller problems which must be solved sequentially backwards in time to obtain the optimal strategies. We refer to this simplification as **sequential decomposition** of the problem. For classical information structures, the sequential decomposition has the additional property that for each sub-problem optimal decisions for a decision-maker can be found separately for each realization of its information.

In this thesis, we will focus on sequential decision-making problems with non-classical information structures. Two different conceptual approaches have been used for such decision-making problems:

1. *Agent-by-Agent Approach:* In some decision-making problems, while the overall problem has a non-classical information structure, it is possible that when the strategies of all except one agent are fixed, then that agent's decision-making problem has a classical information structure. Consider such a problem where strategies of all agents except Agent 1 are fixed and the resulting decision problem for Agent 1 has a classical information structure. Then, for the fixed choice of other agents' strategies, we can find an optimal strategy for Agent 1 using ideas from Markov decision theory. If such an optimal strategy has qualitative features or structural properties *that are independent of the choice of strategies of other agents, then we can conclude that such qualitative features or structural properties are true for a globally optimal strategy of Agent 1.* Such an approach

to find structural results of globally optimal strategies has been used in *Walrand and Varaiya* (1983a), *Walrand and Varaiya* (1983b), *Teneketzis* (2006) etc.

2. *Designer' Approach:* The philosophy of this approach is to consider the optimization problem of a designer who has to select a sequence of decision rules. The designer sequentially decides a decision rule for each time. The designer's problem can be thought of as a problem with (trivially) classical information structure. This approach can be used to decompose the designer's problem of choosing a sequence of decision rules into several sub-problems that must be solved sequentially backwards in time. In each of these sub-problems, the designer has to optimize over one decision rule (instead of the whole sequence). Such an approach for finding a sequential decomposition of the problem of finding optimal strategies has been described in detail in *Witsenhausen* (1973) and *Mahajan* (2008).

In this thesis, we introduce a new approach for addressing sequential decision making problems with non-classical information structures. Our new approach relies on the idea of common information among decision-makers. Intuitively, common information consists of past observations and decisions that are commonly known to the current and future decision makers. We show that decision makers can use the common information to coordinate how they make decisions. Such a coordination approach can allow us to:

1. Find structural results of optimal decision strategies that cannot be identified by the agent-by-agent approach.

2. Provide a sequential decomposition of the decision-making problem that is distinct from (and often simpler than) the decomposition obtained by the designer's approach.

1.2 Structural Results Using Common Information

In this thesis, we consider two specific sequential decision-making problems with non-classical information structures. In both these problems, we use the idea of common information to find structural results for optimal decision strategies.

1.2.1 Delayed Sharing Information Structure in Decentralized Control

An interesting special case of non-classical information structures is the delayed sharing information structure. This information structure consists of K controllers that at each time make a private observation of the controlled system and take an action that affects the evolution of the control system as well as the control objective. While each controller perfectly remembers its own observation and actions of the past, it has access to only a *delayed* version of other controller's observation and action history. In a system with n-step delayed sharing, at each time t, every controller knows prior observations and control actions of all other controllers up to time $t-n$. This information structure was proposed in *Witsenhausen* (1971). Witsenhausen asserted a structural result for optimal control laws for this model without any proof. *Varaiya and Walrand* (1978) proved that Witsenhausen's assertion was true for $n = 1$ but false for $n > 1$.

In this thesis, we resolve Witsenhausen's conjecture; we prove two structural results of the optimal control laws for the delayed sharing information structure. Our structural results critically rely on the observation that at any time t, the controllers have some information that is known to all controllers at that time. We call this the **common information** at time t. A central idea in our proofs is that when this common information is increasing with time (as it is in the delayed sharing information structure), the controllers can use this common information to coordinate how they use their private information to choose their control action. This approach allows us to view the control problem from the perspective of a coordinator who knows the common information and who has to provide prescriptions to each controller on how to use their private information. Since the common information keeps increasing with time, the coordinator's problem then becomes a problem with classical information structure.

In order to address the coordinator's problem, we identify a state sufficient for input-output mapping of the controlled system from the coordinator's perspective. This allows us to describe the coordinator's problem as the problem of controlling the evolution of a Markov chain by means of the prescriptions that the coordinator provides to the controllers. Using Markov decision theory [*Kumar and Varaiya* (1986a)] allows to find structural results as well as a sequential decomposition for finding optimal strategy for the coordinator. These results can then be carried over to structural results and sequential decomposition for finding optimal strategies for each controller.

1.2.2 A Real-Time Multi-Terminal Communication System

The second instance of a problem where common information can be used to find structural results for optimal decision strategies is a real-time multi-terminal communication system with two encoders communicating with a single receiver. The two encoders make distinct partial observations of a Markov source. Each encoder must encode its observations into a sequence of discrete symbols. The symbols are transmitted over noisy/noiseless channels to a receiver that attempts to reconstruct some function of the state of the Markov source. Real-time constraint implies that at each time, the receiver must produce an estimate of the current state of the Markov source. We view this problem as a sequential decision-making problem with the encoders and receivers as the decision-makers. We first obtain structural results of optimal strategies of encoders and receivers using an agent-by-agent approach. We then show that our common information approach can be used to improve the structural results found by the agent-by-agent approach.

1.3 Common Information and General Sequential Problems

As mentioned above, our results for delayed sharing information structures and the real-time multi-terminal communication system rely crucially on using the concept of common information as the basis for coordination among decision-makers. A natural question then is the following: *What is the most general class of decision-making problems where the concept of common information can be used to find a structural results of optimal decision strategies and/or sequential decomposition of the problem of finding optimal decision strategies?* In Chapter IV, we show that one can employ a common information based methodology to find a sequential decomposition of the problem of finding optimal decision strategies in **any sequential decision-making problem** with finite spaces of observation and decisions and finite number of decisions.

Witsenhausen (1973) showed that any sequential decision-making problem (with finite number of decisions) can be converted into a standard form and provided a sequential decomposition of any problem in standard form; thereby providing a a sequential decomposition of any sequential decision-making problem. To the best of our knowledge, our result in Chapter IV is the *only other methodology to provide a sequential decomposition for any sequential decision-making problem.* In general, our sequential decomposition differs from that of Witsehnausen's because of its dependence on common information. Our sequential decomposition coincides with Witsenhausen's only in problems where common information is absent. Further, our sequential decomposition specializes to the classical dynamic program if the decision-making problem has a classical information structure.

In addition to the concept of common information, our result in delayed sharing information structures relies on the identification of a state sufficient for input-output map of the system from the coordinator's point of view. In Chapter V, we investigate if such a state can be found for general sequential decision-making problems. In broad terms, such a state should be a summary of past data that is sufficient for an input-output description of the system from the coordinator's point of view. Although we do not have a algorithmic way of identifying such a state, we present sufficient conditions for a function of past data to be a state sufficient for the coordinator's input output map and provide a simplified sequential decomposition when such a state can be found.

1.4 Computational Aspects

The sequential decompositions that we find in the delayed sharing information structures in Chapter III and for general sequential problems in V are both similar to dynamic programs for partially observable Markov decision problems (POMDPs). As in POMDPs, our sequential decompositions involve value functions that are functions defined on the continuous space of probability measures on a finite state. Also, each step of sequential decomposition requires the maximization over a finite set. One can further show that value functions in our sequential decompositions are piecewise linear and convex/concave. Such characterization of value functions as piecewise linear and concave is utilized to find computationally efficient algorithms for POMDPs. Because of the similarity of our sequential decomposition with the dynamic programs of POMDPS, such algorithmic solutions to general POMDPs can be employed for sequential decompositions as well. Thus, our results provide a conceptual link between sequential decision-making problems with non-classical information structures and algorithmic solutions of POMDPs.

1.5 Contribution of the Thesis

The main contributions of this thesis can be summarized as follows:

1. We demonstrate the use of common information in finding structural results of optimal decision strategies in problems from real-time communication as well as decentralized control. In particular, our structural results for delayed sharing information structures resolve a long-standing open problem in decentralized control.

2. We show that the concept of common information can be used to find a sequential decomposition of the problem of finding optimal decision strategies in any sequential decision-making problem with finite spaces of observation and decisions and finite number of decisions. This result unifies the dynamic programming results of classical information structures and Witsenhausen's sequential decomposition of general sequential problems.

3. We identify sufficient conditions for identifying a state sufficient for input-output map from the perspective of a coordinator that knows the common information and selects prescriptions for decision-makers. We show that the existence of such a state allows us to find a simplified sequential decomposition in a general sequential decision-making problem.

4. Finally, our results establish a conceptual link between sequential decision-making problems with non-classical information structures and algorithmic solutions of Partially Observable Markov decision problems. We believe this will be crucial in finding algorithmic solutions of decision problems with non-classical information structures.

1.6 Organization of the Thesis

This thesis is organized as follows. In Chapter II, we investigate a real-time multi-terminal communication problem as a decision-making problem and find structural results for optimal decision strategies for the encoders and the receiver. In Chapter III, we provide structural results and sequential decomposition for delayed sharing information structures. In Chapter IV, we consider a general sequential decision-making problem and finds a sequential decomposition based on common information. We also show the relationship of our decomposition with the classical dynamic program and Witsenhausen's sequential decomposition. In Chapter V, we provide conditions for finding simpler sequential decompositions in a general sequential problem by using the ideas of common information and a state sufficient for input-output map. We conclude in Chapter VI.

1.7 Notation

Random variables are denoted by upper case letters; their realization by the corresponding lower case letter. For function valued random variables, a tilde ($\tilde{}$) denotes their realization (for example, $\tilde{\gamma}_t^k$). $X_{a:b}$ is a short hand for the vector $(X_a, X_{a+1}, \dots, X_b)$ while $X^{c:d}$ is a short hand for the vector $(X^c, X^{c+1}, \dots, X^d)$. The combined notation $X_{a:b}^{c:d}$ is a short hand for the vector $(X_i^j : i = a, a+1, \dots, b, j = c, c+1, \dots, d)$. $\mathbb{E}$ denotes the expectation operator.

For two random variables X and Y taking values in $\mathcal{X}$ and $\mathcal{Y}$, $\mathbb{P}(X = x|Y)$ denotes the conditional probability of the event $\{X = x\}$ given Y and $\mathbb{P}(X|Y)$ denotes the conditional PMF (probability mass function) of X given Y, that is, it denotes the collection of conditional probabilities $\{\mathbb{P}(X = x|Y), x \in \mathcal{X}\}$. Note that $\mathbb{P}(X = x|Y)$ (resp. $\mathbb{P}(X|Y)$) are random variables (resp. random vectors) with realizations $\mathbb{P}(X = x|Y = y)$ (resp. $\mathbb{P}(X|Y = y)$), $y \in \mathcal{Y}$. Finally, all equalities involving conditional probabilities or conditional expectations are to be interpreted as almost sure equalities (that is, they hold with probability one).

CHAPTER II

A Multi-Terminal Communication System

2.1 Introduction

A large variety of decentralized systems require communication between various devices or agents. In general, since such systems may have multiple senders and receivers of information, the models of point-to-point communication are not sufficient. Further, in many decentralized systems, the purpose of communication is to achieve a higher system objective. Examples include networked control systems where the overall objective of communication between various sensors and controllers is to control the plant in order to achieve a performance objective, or sensor networks where the goal of communication between sensors and a fusion center may be to quickly estimate a physical variable or to track in real-time the evolution of a physical phenomenon. In such systems, agents (sensors, controllers etc.) have to make decisions that affect the overall system performance based only on information they currently have gathered from the environment or from other agents through the underlying communication system. The communication problem therefore should not only address what information can be made available to each agent but also when is this information available. Thus, the overall system objectives may impose constraints on the time delay associated with communication.

In the presence of strict delay constraints on information transmission, the communication problem becomes drastically different from the classical information-theoretic formulations. Information theory deals with encoding and decoding of long sequences which inevitably results in large undesirable delays. For systems with fixed (and typically small) delay requirements, the ideas of asymptotic typicality can not be used. Moreover, information-theoretic bounds on the trade-off between delay and reliability are only asymptotically tight and are of limited value for short sequences (*Gallager* (1968)).

In this chapter we address some issues in multi-terminal communication systems under the real-time constraint. Specifically, we look at problems with multiple senders (encoders) communicating with a single receiver (decoder). We analyze systems with two encoders, although our results generalize to n encoders ($n > 2$) and a single receiver. The two encoders make distinct partial observations of a discrete-time Markov source. Each encoder must encode in real-time its observations into a sequence of discrete variables that are transmitted over separate noisy channels to a common receiver. The receiver must estimate, in real-time, a given function of the state of the Markov source. The main feature of this multi-terminal problem that distinguishes it from a point to point communication problem is the presence of coupling between the encoders (that is, each encoder must take into account what other encoder is doing). This coupling arises because of the following reasons : 1) The encoders' observations are correlated with each other. 2) The encoding problems are further coupled because the receiver wants to minimize a non-separable distortion metric. That is, the distortion metric cannot be simplified into two separate functions each one of which depends only on one encoder's observations. The nature of optimal strategies strongly depends on the nature and extent of the coupling between the encoders.

Our model involves real-time distributed coding of correlated observations that are to be transmitted over noisy channels. Information-theoretic results on asymptotically achievable rate-regions have been known for some distributed coding problems.

The first available results on distributed coding of correlated memoryless sources appear in *Slepian and Wolf* (1973a) and *Slepian and Wolf* (1973b). Multiple access channels with arbitrarily correlated sources were considered in *Cover et al.* (1980). In *Flynn and Gray* (1987), the encoders make noisy observations of an i.i.d source. The authors in *Flynn and Gray* (1987) characterize the achievable rates and distortions, and propose two specific distributed source coding techniques. Coding of dependent sources with separated encoders was considered in *Wyner* (1974). Constructive methods for distributed source coding were presented in *Zhao and Effros* (2001), *A. Kh. Al Jabri and Al-Issa* (1997) and *Pradhan and Ramchandran* (1999). In particular, *Zhao and Effros* (2001) address lossless and nearly lossless source coding for the multiple access system, and *A. Kh. Al Jabri and Al-Issa* (1997) addresses zero-error distributed source coding. The CEO problem, where a number of encoders make conditionally independent observations of an i.i.d source, was presented in *Berger et al.* (1996). The case where the number of encoders tends to infinity was investigated there. The quadratic Gaussian case of the CEO problem has been investigated in *Viswanathan and Berger* (1997), *Oohama* (1998) and *Draper and Wornell* (2002). Bounds on the achievable rate-regions for finitely many encoders were found in *Chen et al.* (2004). A lossy extension of the Slepian-Wolf problem was analyzed in *Berger and S.Tung* (1975) and *Zamir and Berger* (1999). Multi-terminal source coding for memoryless Gaussian sources was considered in *Oohama* (1997).

In *Krithivasan and Pradhan* (2007), *Gelfand and Pinsker* (1979), *Korner and Marton* (1979), *Han and Kobayashi* (1987), *Ahlswede and Han* (1983) and *Csiszar and Korner* (1981), distributed source coding problems with the objective of reconstructing a function of the source are investigated. In *Krithivasan and Pradhan* (2007), the authors consider distributed source coding of a pair of correlated Gaussian sources. The objective is to reconstruct a linear combination of the two sources. The authors discover an inner bound on the optimal rate-distortion region and provide a coding scheme that achieves a portion of this inner bound. The problem of distributed source coding to reconstruct a function of the sources losslessly was considered in *Gelfand and Pinsker* (1979). An inner bound was obtained for the performance limit which was shown to be optimal if the sources are conditionally independent given the function. The case of lossless reconstruction of the modulo-2 sum of two correlated binary sources was considered in *Korner and Marton* (1979). These results were extended in *Csiszar and Korner* (1981) (see Problem 23 on page 400) and *Han and Kobayashi* (1987). An improved inner bound for the problem in *Korner and Marton* (1979) was provided in *Ahlswede and Han* (1983).

The real-time constraint of our problem differentiates it from the information-theoretic results mentioned above. Real-time communication problems for point-to-point systems have been studied using a decision-theoretic/stochastic control perspective. In general, two types of results have been obtained for point to point systems. One type of results establish qualitative properties of optimal encoding and decoding strategies. The central idea here has been to consider the encoders and the decoders as control agents/decision-makers in a team trying to optimize a common objective of minimizing a distortion metric between the source and its estimates at the receiver. Such sequential dynamic teams - where the agents sequentially make multiple decisions in time and may influence each other's information - involve the solution of non-convex functional optimization to find the best strategies for the agents (*Y. C. Ho* (1980),*Witsenhausen* (1968)). However, if the strategies of all but one of the agents are fixed, the resulting problem of optimizing a single agent's strategy can, in many cases, be posed in the framework of Markov decision theory. This approach can explain some of the structural results obtained in *Witsenhausen* (1978), *Teneketzis* (2006), *Walrand and Varaiya* (1983a), *Borkar et al.* (2001), *Yuksel and Basar* (2008). Another class of results establish a decomposition of the problem of choosing a sequence of globally optimal encoding and decoding functions. In the resulting decomposition, at each step, the optimization is over one encoding and decoding functions instead of a sequence of functions. This optimization, however, must be repeated for all realizations of an information state that captures the effect of past encoding/decoding functions (see *Walrand and Varaiya* (1983a), *Borkar et al.* (2001), *Mahajan and Teneketzis* (2008), *Mahajan* (2008)).

Inspired by the decision-theoretic approach to real-time point-to-point systems, we look at our problem from a decentralized stochastic control/team-theoretic perspective with the encoders and the receiver as our control agents/decision makers. *We are primarily interested in discovering the structure of optimal real-time encoding and decoding functions.* In other words, given all the observations available to an agent (i.e, an encoder or the receiver), what is a sufficient statistic to decide its action (i.e, the symbol to be transmitted in case of the encoders and the best estimate in case of the receiver)? The structure of optimal real-time encoding and decoding strategies provides insights into their essential complexity (for example, the memory requirements at the encoders and the receiver for finite and infinite time horizon communication problems) as well as the effect of the coupling between the encoders mentioned earlier.

A universal approach for discovering the structure of optimal real-time encoding/decoding strategies in a multi-terminal system with any general form of correlation between the encoders' observations has so far remained elusive. In this chapter, we restrict ourselves to a simple model for the encoders' observations. For such a model (described in Section 2.3), we obtain results on the structure of optimal real-time encoding strategies when the receiver is assumed to a have a finite memory. Our results reveal that for any time horizon, however large (or even infinite), there exists a finite dimensional sufficient statistic for the encoders. This implies that an encoder with a memory that can store a fixed finite number of real-numbers can perform as well as encoders with arbitrarily large memories. Subsequently, we consider communication with noiseless channels and remove the assumption of having limited receiver memory. For this problem, the approach in Section 2.3 results in sufficient statistics for the encoders that belong to spaces which keep increasing with time. This is undesirable if one wishes to look at problems with large/infinite time-horizons. In order to obtain a sufficient statistic with time-invariant domain, we invent a new methodology for decentralized decision-making problems. This methodology highlights the importance of common information/ common knowledge (in the sense of *Aumann* (1976)), in determining structural properties of decision makers in a team. In general, the resulting sufficient statistic belongs to an infinite dimensional space. However, we present special cases where a finite dimensional representation is possible. Moreover, we believe that the infinite dimensional sufficient statistic may be intelligently approximated to obtain real-time finite-memory encoding strategies with good performance.

The rest of this chapter is organized as follows: In Section 2.2 we present a real-time multi-terminal communication system and formulate the optimization problem. In Section 2.3 we present our assumptions on the nature of the source and the receiver and obtain structural results for optimal real-time encoding and decoding functions. In Section 2.4 we consider the problem with noiseless channels and perfect receiver memory. We develop a new methodology to find structural results for optimal real-time encoders for this case. We look at some extensions and special cases of our results in Section 2.5. We conclude in Section 2.6.

2.2 A Real-Time Multi-terminal Communication Problem

Consider the real-time communication system shown in Figure 2.1. We have two encoders that partially observe a Markov source and communicate it to a single receiver over separate noisy channels. The receiver may be interested in estimating the state of the Markov source or some function of the state of the source. We wish to find sufficient statistics for the encoders and the receiver and/or qualitative properties for the encoding and decoding functions. Below, we elaborate on the model and the optimization problem.

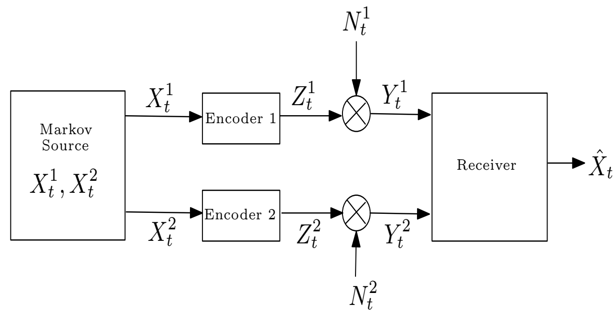

Figure 2.1: A Multi-terminal Communication System

2.2.1 Problem Formulation

1) The Model: The state of the Markov source at time t is described as

$$X_t = (X_t^1, X_t^2)$$

where $X_t^i \in \mathcal{X}^i$, $i = 1, 2$ and $\mathcal{X}^1, \mathcal{X}^2$ are finite spaces. The time-evolution of the source is given by the following equation

$$X_{t+1} = F_t(X_t, W_t) \tag{2.1}$$

where $W_t, t = 1, 2, ..$ is a sequence of independent random variables that are independent of the initial state X_1.

Two encoders make partial observations of the source. In particular, at time t, encoder 1 observes X_t^1 and encoder 2 observes X_t^2. The encoders have perfect memory, that is, they remember all their past observations and actions. At each time t, encoder 1 sends a symbol Z_t^1 belonging to a finite alphabet $\mathcal{Z}^1$ to the receiver. The encoders operate in real-time, that is, each encoder can select the symbol to be sent at time t, based only on the information available to it till that time. That is, the encoding rule at time t must be of the form:

$$Z_t^1 = f_t^1(X_{1:t}^1, Z_{1:t-1}^1) \tag{2.2}$$

where $X_{1:t}^1$ represents the sequence $X_1^1, X_2^1, \ldots, X_t^1$ and $Z_{1:t-1}^1$ represents the sequence $Z_1^1, Z_2^1, \ldots, Z_{t-1}^1$. In general, one can allow randomized encoding rules instead of deterministic encoding functions. That is, for each realization of its observations till time t, encoder 1 selects a probability distribution on $\mathcal{Z}^1$ and then transmits a random symbol generated according to the selected distribution. We will show later that, under our assumptions on the model, such randomized encoding rules cannot provide any performance gain and we can restrict our attention to deterministic encoding functions.

Encoder 2 operates in a similar fashion as encoder 1. Thus, encoding rules of encoder 2 are functions of the form:

$$Z_t^2 = f_t^2(X_{1:t}^2, Z_{1:t-1}^2) \tag{2.3}$$

where Z_t^2 belongs to finite alphabet $\mathcal{Z}^2$.

The symbols Z_t^1 and Z_t^2 are transmitted over separate noisy channels to a single receiver. The channel noises at time t are mutually independent random variables N_t^1 and N_t^2 belonging to finite alphabets $\mathcal{N}^1$ and $\mathcal{N}^2$ respectively. The noise variables $(N_1^1, N_1^2, N_2^1, N_2^2, \ldots, N_t^1, N_t^2, \ldots)$ form a collection of independent random variables that are independent of the source process $X_t, t = 1, 2, ...$

The receiver receives Y_t^1 and Y_t^2 which belong to finite alphabets $\mathcal{Y}^1$ and $\mathcal{Y}^2$ respectively. The received symbols are noisy versions of the transmitted symbols according to known channel functions h_t^1 and h_t^2, that is,

$$Y_t^i = h_t^i(Z_t^i, N_t^i) \tag{2.4}$$

for $i = 1, 2$.

At each time t, the receiver produces an estimate of the source $\hat{X}_t$ based on the symbols received till time t, i.e.,

$$\hat{X}_t = g_t(Y^1_{1:t}, Y^2_{1:t}) \tag{2.5}$$

A non-negative distortion function $\rho_t(X_t, \hat{X}_t)$ measures the instantaneous distortion between the source and the estimate at time t. (Note that the distortion function may take into account that the receiver only needs to estimate a function of X^1_t and X^2_t)

2) The Optimization Problem P: Given the source and noise statistics, the encoding alphabets, the channel functions h^1_t, h^2_t, the distortion functions ρ_t and a time horizon T, the objective is to find globally optimal encoding and decoding functions $f^1_{1:T}, f^2_{1:T}, g_{1:T}$ so as to minimize

$$J(f^1_{1:T}, f^2_{1:T}, g_{1:T}) = \mathbb{E}\{\sum_{t=1}^{T} \rho_t(X_t, \hat{X}_t)\} \tag{2.6}$$

where the expectation in (2.6) is over the joint distribution of $X_{1:T}$ and $\hat{X}_{1:T}$ which is determined by the given source and noise statistics and the choice of encoding and decoding functions $f^1_{1:T}, f^2_{1:T}, g_{1:T}$.

We refer to the collection of functions $f^i_{1:T}$ as encoder i's strategy ($i = 1, 2$). The collection of functions $g_{1:T}$ is the decoding strategy.

Remarks: 1. Since we consider only finite alphabets for the source, the encoded symbols, the channel noise, the received symbols and a finite time horizon, the number of possible choices of encoding and decoding functions is finite. Therefore, an optimal choice of strategies $(f^1_{1:T}, f^2_{1:T}, g_{1:T})$ always exists.

2. A brute force search method to find the optimal can always be used in principle. It is clear however that even for small time-horizons, the number of possible choices

would be large enough to make such a search inefficient. Moreover, such a scheme would not be able to identify any characteristics of optimal encoding and decoding functions.

The encoding functions and the decoding functions in equations (2.2), (2.3) and (2.5) require the encoders and the receiver to store entire sequences of their past observations and actions. For large time-horizons storing all past data becomes prohibitive. Therefore, one must decide what part of the information contained in these arbitrarily large sequences is sufficient for decision-making at the encoders and the receiver. In particular, we are interested in addressing the following questions:

1. Is there a sufficient statistic for the encoders and the decoder that belongs to a time-invariant space? (Clearly, all the past data available at an agent is a sufficient statistic but it belongs to a space that keeps increasing with time.) If such a sufficient statistic exists, one can potentially look at problems with large (or infinite) time-horizons.

2. Is there a finite-dimensional sufficient statistic for the encoders and the receiver? If such a sufficient statistic exists, then we can replace the requirement of storing arbitrarily long sequences of past observations/messages with storing a fixed finite number of real numbers at the encoders and the receiver.

The above communication problem can be viewed as a sequential team problem where the encoders and the receiver are the decision-making agents that are sequentially making decisions to optimize a common objective. The communication problem is a dynamic team problem since the encoders' decisions influence the information available to the receiver. Dynamic team problems are known to be hard. For dynamic teams, a general answer to the questions on the existence of sufficient statistics that either have time-invariant domains or are finite-dimensional is not known. In the next section we will make simplifying assumptions on the nature of the source and the receiver and present sufficient statistics for the encoders and the receiver.

2.3 Problem P1

We consider the optimization problem (Problem P) formulated in the previous section under the following assumptions on the source and the receiver.

1. Assumption A1 on the Source: We assume that the time-evolution of the source can be described by the following model:

$$X^1_{t+1} = F^1_t(X^1_t, A, W^1_t) \tag{2.7a}$$

$$X^2_{t+1} = F^2_t(X^2_t, A, W^2_t) \tag{2.7b}$$

where A is a random-variable taking values in the finite set $\mathcal{A}$ and $W^1_t, t = 1, 2, ...$ and $W^2_t, t = 1, 2...$ are two independent noise processes (that is, sequences of independent random variables) that are independent of the initial state (X^1_1, X^2_1) and A as well. Thus, the transition probabilities satisfy:

$$\begin{aligned} &P(X^1_{t+1}, X^2_{t+1}|X^1_t, X^2_t, A) \\ =&P(X^1_{t+1}|X^1_t, A) \cdot P(X^2_{t+1}|X^2_t, A) \end{aligned} \tag{2.8}$$

The initial state of the Markov source has known statistics that satisfy the following equation :

$$\begin{aligned} P(X^1_1, X^2_1, A) =&P(X^1_1, X^2_1|A) \cdot P(A) \\ =&P(X^1_1|A) \cdot P(X^2_1|A) \cdot P(A) \end{aligned} \tag{2.9}$$

Thus, A is a time-invariant random variable that couples the evolution of X_t^1 and X_t^2. Note that conditioned on A, X_t^1 and X_t^2 form two conditionally independent Markov chains. We define

$$X_t := (X_t^1, X_t^2, A) \tag{2.10}$$

which belongs to the space $\mathcal{X} := \mathcal{X}^1 \times \mathcal{X}^2 \times \mathcal{A}$.

The encoders' model is same as before. Thus encoder 1 observes X_t^1 and encoder 2 observes X_t^2. Note that the random variable A is not observed by any encoder. The encoders have perfect memories and the encoding functions are given by equations (2.2) and (2.3).

2. Assumption A2 on the Receiver: We have a finite memory receiver that maintains a separate memory for symbols received from each channel. This memory is updated as follows:

$$M_1^i = l_1^i(Y_1^i), i = 1, 2 \tag{2.11a}$$

$$M_t^i = l_t^i(M_{t-1}^i, Y_t^i), i = 1, 2 \tag{2.11b}$$

where M_t^i belongs to finite alphabet $\mathcal{M}^i$, $i = 1, 2$ and l_t^i are the memory update functions at time t for $i = 1, 2$. For notational convenience, we define $M_0^i := 0$ for $i = 1, 2$. The receiver produces an estimate of the source $\hat{X}_t$ based on its memory contents at time $t - 1$ and the symbols received at time t, that is,

$$\hat{X}_t = g_t(Y_t^1, Y_t^2, M_{t-1}^1, M_{t-1}^2) \tag{2.12}$$

We now formulate the following problem.

Problem P1: With assumptions A1 and A2 as above, and given source and noise statistics, the encoding alphabets, the channel functions h_t^1, h_t^2, the distortion functions ρ_t and a time horizon T, the objective is to find globally optimal encoding, decoding and memory update functions $f^1_{1:T}, f^2_{1:T}, g_{1:T}, l^1_{1:T}, l^2_{1:T}$ so as to minimize

$$J(f^1_{1:T}, f^2_{1:T}, g_{1:T}, l^1_{1:T}, l^2_{1:T}) = \mathbb{E}\{\sum_{t=1}^{T} \rho_t(X_t, \hat{X}_t)\} \tag{2.13}$$

where the expectation in (2.13) is over the joint distribution of $X_{1:T}$ and $\hat{X}_{1:T}$ which is determined by the given source and noise statistics and the choice of encoding, decoding and memory update functions $f^1_{1:T}, f^2_{1:T}, g_{1:T}, l^1_{1:T}, l^2_{1:T}$.

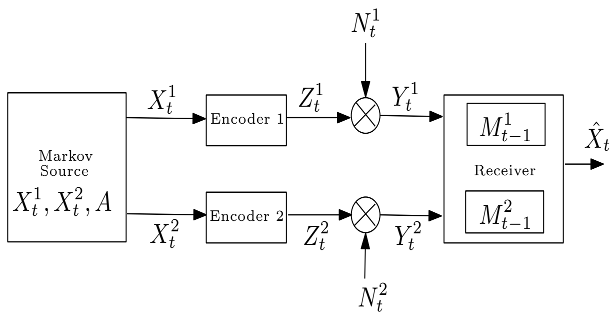

Figure 2.2: Problem P1

2.3.1 Features of the Model

We discuss situations that give rise to models similar to that of Problem P1.

1. A Sensor Network: Consider a sensor network where the sensors' observations are influenced by a slowly varying global parameter and varying local phenomena. Our model is an approximation of this situation where A models the global parameter that is constant over the time-horizon T and X_t^i are the local factors at the location of the i^{th} sensor at time t. A finite memory assumption on the receiver may be justified in situations where the receiver is itself a node in the network and is coordinating the individual sensors. We will show that this assumption implies that the sensors (encoders in our model) themselves can operate on finite-dimensional sufficient statistics without losing any optimality with respect to sensors with perfect (infinite) memory.

2. Decentralized Detection/Estimation Problem: Consider the following scenario of a decentralized detection problem; Sensors make noisy observations X_t^i on the state A of environment. Sensors must encode their information in real-time and send it to a fusion center. Assuming that sensor noises are independent, we have that, conditioned on A, the sensor observations are independent. (Typically, the observations are also assumed to be i.i.d in time conditioned on the state of the environment, but we allow them to be a Markov process.) Thus, the encoding rule for the i^{th} sensor must be of the form:

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

Consider the case where Z_t^i can either be "blank" or a value from the set $\mathcal{A}$. Each sensor is restricted to send only one non-blank message, and within a fixed time-horizon each sensor must send its final non-blank message. When a sensor sends a non-blank message Z_t^i, the fusion center receives a noisy version Y_t^i of this message. As long as the fusion center does not receive final (non-blank) messages from all sensors, its decision is $\hat{X}_t =$ "no decision". If all sensors have sent a non-blank message, the fusion center produces an estimate $\hat{X}_t \in \mathcal{A}$ as its final estimate on A and incurs a distortion cost $\rho(A, \hat{X}_t)$. Thus, we can view the receiver as maintaining a separated memory for messages from each sensor which is initialized to "blank" and updated as follows:

$$M_t^i = \begin{cases} Y_t^i & \text{if } M_{t-1}^i \text{ was ``blank''} \\ M_{t-1}^i & \text{otherwise} \end{cases} \tag{2.14}$$

The receiver's decision is $\hat{X}_t =$ "no decision", if $Y_t^i = M_{t-1}^i =$ "blank" for some sensor i, else the receiver uses a function g_t to find an estimate

$$\hat{X}_t = g_t(Y_t^1, Y_t^2, M_{t-1}^1, M_{t-1}^2) \tag{2.15}$$

The above detection problem therefore is a special case of our model with fixed memory update rules from (2.14).

Clearly, our model also includes the case when the encoders' observations are independent Markov chains (not just conditionally independent). In this case, the coupling between encoders is only due to the fact the receiver may be interested in estimating some function of the state of the two Markov chains and not their respective individual states.

2.3.2 Structure Result for Encoding Functions

We define the following probability mass functions (pmf) for encoder i, $(i = 1, 2)$:

Definition II.1. For $t = 1, 2, \ldots, T$ and $a \in \mathcal{A}$,

$$b_t^i(a) := P(A = a | X_{1:t}^i)$$

Definition II.2. For $t = 2, 3, \ldots, T$ and $m \in \mathcal{M}^i$,

$$\mu_t^i(m) := P(M_{t-1}^i = m | Z_{1:t-1}^i, l_{1:t-1}^i)$$

where $l_{1:t-1}^i$ in the conditioning indicate that μ_t^i is defined for a fixed choice of the memory update rules $l_{1:t-1}^i$. For notational convenience, we also define for each $m \in \mathcal{M}^i, i = 1, 2,$

$$\mu_1^i(m) := 0$$

Theorem II.3. *There exist globally optimal encoding rules of the form :*

$$Z_t^i = f_t^i(X_t^i, b_t^i, \mu_t^i) \tag{2.16}$$

where f_t^i are deterministic functions for $t = 1, 2, \ldots, T$ and $i = 1, 2$.

Discussion: In contrast to equation (2.2), Theorem II.3 says that an optimal encoder 1 only needs to use the current observation X_t^1 and the probability mass functions b_t^1, μ_t^1 that act as a compressed representation of the past observations $X_{1:t-1}^1$ and $Z_{1:t-1}^1$. These pmfs represent the encoder 1's belief on A and M_{t-1}^1.

To obtain the result of Theorem II.3 for the encoder 1, we fix arbitrary encoding rules for the encoder 2 of the form in (2.3), arbitrary memory update rules of the form in (2.11) and arbitrary decoding rules of the form in (2.12). Given these functions, we consider the problem of selecting optimal encoding rules for encoder 1. We identify a structural property of the optimal encoding rules of encoder 1 that is independent of the arbitrary choice of strategies for encoder 2 and the receiver. We conclude that the identified structure of optimal rules of encoder 1 must also be true when encoder 2 and the receiver are using the globally optimal strategies. Hence, the identified structure is true for globally optimal encoding rules of encoder 1. We now present this argument in detail.

Consider arbitrary (but fixed) encoding rules for encoder 2 of the of the form in (2.3), arbitrary memory update rules for the receiver of the form in (2.11) and arbitrary decoding rules of the form in (2.12). We will prove Theorem II.3 using the following lemmas.

Lemma II.4. *The belief of encoder 1 about the random variable A can be updated as follows:*

$$b_t^1 = \alpha_t^1(b_{t-1}^1, X_t^1, X_{t-1}^1) \tag{2.17}$$

where $\alpha_t^1, t = 2, 3, \ldots, T$ are deterministic functions.

Proof. See Appendix A. □

Lemma II.5. *The belief of encoder 1 about the receiver memory M_{t-1}^1 can be updated as follows:*

$$\mu_t^1 = \beta_t^1(\mu_{t-1}^1, Z_{t-1}^1) \tag{2.18}$$

where $\beta_t^1, t = 2, 3, \ldots, T$ *are deterministic functions.*

Proof. See Appendix A. $\square$

We now define the following random variables:

$$R_t^1 := (X_t^1, b_t^1, \mu_t^1), \tag{2.19}$$

for $t = 1, 2, \ldots, T$.

Observe that R_t^1 is a function of encoder 1's observations till time t, that is, $X_{1:t}^1, Z_{1:t-1}^1$. Moreover, any encoding rule of the form in (2.2) can also be written as

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

Lemma II.6. $R_t^1, t = 1, 2, ..., T$ *is a perfectly observed controlled Markov process for encoder 1 with* Z_t^1 *as the control action at time* t.

Proof. Since R_t^1 is a function of encoder 1's observations till time t, that is, $X_{1:t}^1, Z_{1:t-1}^1$, it is perfectly observed at encoder 1.

Let $x_{1:t}^1, z_{1:t-1}^1$ be a realization of the encoder 1's observations $X_{1:t}^1, Z_{1:t-1}^1$. Similarly, let r_t^1 be a realization of R_t^1 and $\tilde{b}_t^1$ and $\tilde{\mu}_t^1$ be realizations of b_t^1 and μ_t^1 respectively. Then,

$$\begin{aligned}
&P(R_{t+1}^1 = (x_{t+1}^1, \tilde{b}_{t+1}^1, \tilde{\mu}_{t+1}^1)|r_{1:t}^1, z_{1:t}^1) \\
&= P(x_{t+1}^1, \tilde{b}_{t+1}^1, \tilde{\mu}_{t+1}^1|x_{1:t}^1, \tilde{b}_{1:t}^1, \tilde{\mu}_{1:t}^1, z_{1:t}^1) \\
&= P(\tilde{b}_{t+1}^1, \tilde{\mu}_{t+1}^1|x_{t+1}^1, x_{1:t}^1, \tilde{b}_{1:t}^1, \tilde{\mu}_{1:t}^1, z_{1:t}^1) \\
&\quad \times P(x_{t+1}^1|x_{1:t}^1, \tilde{b}_{1:t}^1, \tilde{\mu}_{1:t}^1, z_{1:t}^1) \qquad (2.20) \\
&= P(\tilde{b}_{t+1}^1, \tilde{\mu}_{t+1}^1|x_{t+1}^1, x_t^1, \tilde{b}_t^1, \tilde{\mu}_t^1, z_t^1) \\
&\quad \times P(x_{t+1}^1|x_{1:t}^1, \tilde{b}_{1:t}^1, \tilde{\mu}_{1:t}^1, z_{1:t}^1) \qquad (2.21)
\end{aligned}$$

where the first term in (2.21) is true because of Lemma II.4 and Lemma II.5. Consider the second term in (2.21). It can be expressed as follows:

$$
\begin{aligned}
&P(x^1_{t+1}|x^1_{1:t}, \tilde{b}^1_{1:t}, \tilde{\mu}^1_{1:t}, z^1_{1:t}) \\
&= \sum_{a\in\mathcal{A}} P(x^1_{t+1}, A = a|x^1_{1:t}, \tilde{b}^1_{1:t}, \tilde{\mu}^1_{1:t}, z^1_{1:t}) && (2.22) \\
&= \sum_{a\in\mathcal{A}} P(x^1_{t+1}|A = a, x^1_{1:t}, \tilde{b}^1_{1:t}, \tilde{\mu}^1_{1:t}, z^1_{1:t}) \\
&\quad \times P(A = a|x^1_{1:t}, \tilde{b}^1_{1:t}, \tilde{\mu}^1_{1:t}, z^1_{1:t}) && (2.23) \\
&= \sum_{a\in\mathcal{A}} P(x^1_{t+1}|A = a, x^1_t).\tilde{b}^1_t(a) && (2.24)
\end{aligned}
$$

where the first term in (2.24) is true because of the Markov property of X^1_t when conditioned on A. Therefore, substituting (2.24) in (2.21), we get

$$
\begin{aligned}
&P(R^1_{t+1} = (x^1_{t+1}, \tilde{b}^1_{t+1}, \tilde{\mu}^1_{t+1})|x^1_{1:t}, \tilde{b}^1_{1:t}, \tilde{\mu}^1_{1:t}, z^1_{1:t}) \\
=&P(\tilde{b}^1_{t+1}, \tilde{\mu}^1_{t+1}|x^1_{t+1}, x^1_t, \tilde{b}^1_t, \tilde{\mu}^1_t, z^1_t) \\
&\times \sum_{a\in\mathcal{A}} [P(x^1_{t+1}|A = a, x^1_t) \times \tilde{b}^1_t(a)] && (2.25)
\end{aligned}
$$

The right hand side of (2.25) depends only on x^1_t,$\tilde{b}^1_t$, $\tilde{\mu}^1_t$ and z^1_t from the entire collection of conditioning variables in the left hand side of (2.25). Hence,

$$
\begin{aligned}
P(R^1_{t+1}|r^1_{1:t}, z^1_{1:t}) =&P(R^1_{t+1}|x^1_{1:t}, \tilde{b}^1_{1:t}, \tilde{\mu}^1_{1:t}, z^1_{1:t}) \\
=&P(R^1_{t+1}|x^1_t, \tilde{b}^1_t, \tilde{\mu}^1_t, z^1_t) \\
=&P(R^1_{t+1}|r^1_t, z^1_t) && (2.26)
\end{aligned}
$$

This establishes the lemma. □

Lemma II.7. *The expected instantaneous distortion cost for encoder 1 can be ex-*

pressed as :

$$\mathbb{E}\{\rho_t(X_t, \hat{X}_t)|X^1_{1:t}, Z^1_{1:t}\} = \hat{\rho}_t(R^1_t, Z^1_t) \tag{2.27}$$

where $\hat{\rho}_t, t = 1, 2, \ldots, T$ *are deterministic functions.*

Proof. For any realization $x^1_{1:t}, z^1_{1:t}$ of $X^1_{1:t}, Z^1_{1:t}$, we have

$$\begin{aligned}
&\mathbb{E}\{\rho_t(X_t, \hat{X}_t)|x^1_{1:t}, z^1_{1:t}\} \\
=&\mathbb{E}\{\rho_t(x^1_t, X^2_t, A, g_t(Y^1_t, Y^2_t, M^1_{t-1}, M^2_{t-1})|x^1_{1:t}, z^1_{1:t})\}
\end{aligned} \tag{2.28}$$

The expectation in (2.28) depends on x^1_t (appearing in the argument of ρ_t) and the conditional probability: $P(X^2_t, A, Y^1_t, Y^2_t, M^1_{t-1}, M^2_{t-1}|x^1_{1:t}, z^1_{1:t})$. We can evaluate this conditional probability as follows:

$$P(X^2_t = x^2_t, A = a, Y^1_t = y^1_t, Y^2_t = y^2_t, M^1_{t-1} = m^1_{t-1}, M^2_{t-1} = m^2_{t-1}|x^1_{1:t}, z^1_{1:t}) \tag{2.29}$$

$$\begin{aligned}
=&P(X^2_t = x^2_t, Y^2_t = y^2_t, M^2_{t-1} = m^2_{t-1}|A = a, Y^1_t = y^1_t, M^1_{t-1} = m^1_{t-1}, x^1_{1:t}, z^1_{1:t}) \\
&\times P(Y^1_t|A = a, M^1_{t-1} = m^1_{t-1}, x^1_{1:t}, z^1_{1:t}) \\
&\times P(M^1_{t-1} = m^1_{t-1}|A = a, x^1_{1:t}, z^1_{1:t}) \\
&\times P(A = a|x^1_{1:t}, z^1_{1:t})
\end{aligned} \tag{2.30}$$

$$\begin{aligned}
=&P(X^2_t = x^2_t, Y^2_t = y^2_t, M^2_{t-1} = m^2_{t-1}|A = a)\times \\
&P(Y^1_t = y^1_t|z^1_t) \times P(M^1_{t-1} = m^1_{t-1}|z^1_{1:t}) \times P(A = a|x^1_{1:t})
\end{aligned} \tag{2.31}$$

$$\begin{aligned}
=&P(X^2_t = x^2_t, Y^2_t = y^2_t, M^2_{t-1} = m^2_{t-1}|A = a)\times \\
&P(Y^1_t = y^1_t|z^1_t) \times \tilde{\mu}^1_t(m^1_{t-1}) \times \tilde{b}^1_t(a)
\end{aligned} \tag{2.32}$$

In the first term of (2.31), we used the fact that conditioned on A, the observations of encoder 2 and received messages from the second channel are independent of the observations of encoder 1 and the messages received from the first channel. We used the fact that the noise variables N^1_t are i.i.d and independent of the source in the second and third term of (2.31). Thus, the conditional probability in (2.29) depends only on z^1_t, $\tilde{\mu}^1_t$ and $\tilde{b}^1_t$. Therefore, the expectation in (2.28) is a function of $x^1_t, z^1_t, \tilde{\mu}^1_t, \tilde{b}^1_t$. That is,

$$\mathbb{E}\{\rho_t(X_t, \hat{X}_t)|x^1_{1:t}, z^1_{1:t}\} = \hat{\rho}_t(x^1_t, z^1_t, \tilde{\mu}^1_t, \tilde{b}^1_t) \tag{2.33}$$

$$= \hat{\rho}_t(r^1_t, z^1_t) \tag{2.34}$$

□

Proof of Theorem II.3. From Lemma II.6 and Lemma II.7, we conclude that the optimization problem for encoder 1, when the strategies of encoder 2 and the receiver have been fixed, is equivalent to controlling the transition probabilities of the controlled Markov chain R^1_t through the choice of the control actions Z^1_t (where Z^1_t can be any function of $R^1_{1:t}$ and $Z^1_{1:t-1}$) in order to minimize $\sum_{t=1}^{T} \mathbb{E}\{\hat{\rho}_t(R^1_t, Z^1_t)\}$. It is a well-known result of Markov decision theory (*Kumar and Varaiya* (1986b), Chapter 6) that there is an optimal control law of the form:

$$Z^1_t = f^1_t(R^1_t)$$

or equivalently,

$$Z^1_t = f^1_t(X^1_t, b^1_t, \mu^1_t)$$

Moreover, it also follows from Markov decision theory that allowing randomized control policies for encoder 1 cannot provide any performance gain. Since the above structure of the optimal choice of encoder 1's strategy is true for any arbitrary choice of encoder 2's and the receiver's strategies, we conclude that the above structure of optimal encoder 1 is true when the encoder 2 and the receiver are using their globally optimal choices as well. Therefore, the above structure is true for globally optimal strategy of encoder 1 as well. This completes the proof of Theorem II.3. Structural result for encoder 2 follows from the same arguments simply by interchanging the roles of encoder 1 and encoder 2. □

2.3.3 Structural result for Decoding Functions

We now present the structure of an optimal decoding strategy. Consider fixed encoding rules of the form in (2.2) and (2.3) and fixed memory update rules of the form in (2.11). We define the following probability mass function for the receiver :

Definition II.8. For $x \in \mathcal{X}$ and $t = 1, 2, \ldots, T$,

$$\psi_t(x) := P(X_t = x | Y_t^1, Y_t^2, M_{t-1}^1, M_{t-1}^2, f_{1:t}^1, f_{1:t}^2, l_{1:t}^1, l_{1:t}^2)$$

where the functions $f_{1:t}^1, f_{1:t}^2, l_{1:t}^1, l_{1:t}^2$ in the conditioning indicate that ψ_t is defined for a fixed choice of encoding and memory update strategies.

Let $\Delta(\mathcal{X})$ denote the set of probability mass functions on the finite set $\mathcal{X}$. We define the following functions on $\Delta(\mathcal{X})$.

Definition II.9. For any $\psi \in \Delta(\mathcal{X})$ and $t = 1, 2, \ldots, T$,

$$\tau_t(\psi) = \operatorname*{argmin}_{s \in \mathcal{X}} \sum_{x \in \mathcal{X}} \psi(x) \rho_t(x, s)$$

With the above definitions, we can present the result on the structure of a globally optimal decoding rule.

Theorem II.10. *For any fixed encoding rules of the form in (2.2) and (2.3) and memory update rules of the form in (2.11), there is an optimal decoding rule of the form*

$$\hat{X}_t = \tau_t(\psi_t) \tag{2.35}$$

where the belief ψ_t is formed using the fixed encoding and memory update rules. In particular, equation (2.35) is true for a globally optimal receiver, when the fixed encoding rules and memory update rules are the globally optimal rules.

Proof. The result follows from standard statistical decision theory *DeGroot* (1970). □

2.3.4 Discussion of the Result

Theorem II.3 identifies sufficient statistics for the encoders. Instead of storing all past observations and transmitted messages, each encoder may store only the probability mass functions (pmf) on the finite sets $\mathcal{A}$ and $\mathcal{M}^i$ generated from past observations and transmitted messages. Thus we have finite-dimensional sufficient statistics for the encoders that belong to time-invariant spaces (the space of pmfs on $\mathcal{A}$ and $\mathcal{M}^i$). Clearly, this amounts to storing a fixed number of real-numbers in the memory of each encoder instead of arbitrarily large sequences of past observations and past transmitted symbols. However, the encoders now have to incur an additional computational burden involved in updating their beliefs on A and the receiver memory.

We would like to emphasize that the presence of a finite dimensional sufficient statistic that belong to time-invariant spaces is strongly dependent on the nature of the source and the receiver. Indeed, without the conditionally independent nature of the encoders' observations or the separated finite memories at the receiver, we have not been able to identify a sufficient statistic whose domain does not keep increasing with time. For example, if the finite memory receiver maintained a coupled memory which is updated as:

$$M_t = l_t(M_{t-1}, Y_t^1, Y_t^2)$$

then one may conjecture that the encoder could use a belief on M_{t-1} as a sufficient representation of past transmitted symbols, analogous to μ_t^1 in Theorem II.3. However, such a statistic cannot be updated without remembering all past data, that is, an update equation analogous to Lemma II.5 for μ_t^1 does not hold. This implies that the Markov decision-theoretic arguments of Theorem II.3 do not work for this case. In the case when encoders' observations have a more general correlation structure, a finite dimensional statistic like b_t^1 that compresses all the past observations seems unlikely. It appears that in the absence of the assumptions mentioned above, the optimal encoders should remember all their past information.

If the receiver has perfect memory, that is, it remembers all past messages received, ($M_{t-1}^i = Y_{1:t-1}^i, i = 1, 2$), Theorem II.3 implies $\mu_t^i = P(Y_{1:t-1}^i | Z_{1:t-1}^i)$ as a part of the sufficient statistic for encoder i. Thus, Theorem II.3 says that each encoder needs to store beliefs on the increasing space of all past observations at the receiver. This sufficient statistic does not belong to a time-invariant space. In the next section, we will consider this problem with noiseless channels and show that for noiseless channels there is in fact a sufficient statistic that belongs to a time-invariant space. However, this sufficient statistic is no longer finite dimensional and for implementation purposes, one would have to come up with approximate representations of it.

2.4 Problem P2

We now look at the Problem P1 with noiseless channels. Firstly, we assume the same model for the nature of the source and the separated memories at the receiver as in Problem P1. The result of Theorem II.3 then holds with the belief on M_{t-1}^i replaced by the true value of M_{t-1}^i. The presence of noiseless channels implies that encoder i and the receiver have some common information. That is, at time t they both know the state of M_{t-1}^i.In this section, we will show that the presence of common information allows us to explore the case when the receiver may have perfect memory. We will present a new methodology that exploits the presence of common information between the encoder and the receiver to find sufficient statistics for the encoders that belong to time-invariant spaces.

2.4.1 Problem Formulation

1. *The Model:* We consider the same model as in P1 with following two modifications:

   (a) The channels are noiseless; thus the received symbol Y_t^i is same as the transmitted symbol Z_t^i, for $i = 1, 2$ and $t = 1, 2, \ldots, T$.

   (b) The receiver has perfect memory, that is, it remembers all the past received symbols. Thus, $M_{t-1}^i = Z_{1:t-1}^i$, for $i = 1, 2$ and $t = 2, 3, \ldots, T$. (See Fig. 2.3)

2. *The Optimization Problem, P2*: Given the source statistics, the encoding alphabets, the time horizon T, the distortion functions ρ_t, the objective is to find globally optimal encoding and decoding functions $f_{1:T}^1, f_{1:T}^2, g_{1:T}$ so as to minimize

$$J(f_{1:T}^1, f_{1:T}^2, g_{1:T}) = \mathbb{E}[\sum_{t=1}^{T} \rho_t(X_t, \hat{X}_t)] \tag{2.36}$$

   where the expectation in (2.36) is over the joint distribution of $X_{1:T}$ and $\hat{X}_{1:T}$ which is determined by the given source statistics and the choice of encoding and decoding functions $f_{1:T}^1, f_{1:T}^2, g_{1:T}$.

2.4.2 Structure of the Receiver

Clearly, problem P2 is a special case of problem P1. The decoder structure of P1 can now be restated for P2 as follows: For fixed encoding rules of the form in (2.2) and (2.3), we can define the receiver's belief on the source as:

$$\psi_t(x) := P(X_t = x | Z_{1:t}^1, Z_{1:t}^2, f_{1:t}^1, f_{1:t}^2)$$

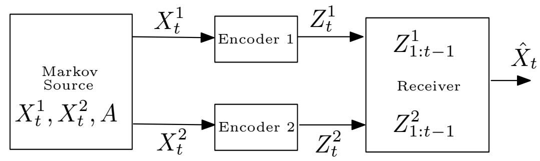

Figure 2.3: Problem P2

for $x \in \mathcal{X}$ and $t = 1, 2, \ldots, T$.

Theorem II.11. *For any fixed encoding rules of the form in (2.2) and (2.3), there is an optimal decoding rule of the form*

$$\hat{X}_t = \tau_t(\psi_t) \tag{2.37}$$

where the belief ψ_t is formed using the fixed encoding rules and τ_t is as defined in Definition II.9. In particular, equation (2.37) is true for a globally optimal receiver, when the fixed encoding rules are globally optimal rules.

2.4.3 Structural Result for Encoding Functions

For a fixed realization of $Z^i_{1:t-1}$, encoder i's belief on the receiver memory M^i_{t-1} is simply:

$$\tilde{\mu}^i_t(m) = P(M^i_{t-1} = m | z^i_{1:t-1}) = \begin{cases} 1 & \text{if } m = z^1_{1:t-1} \\ 0 & \text{otherwise} \end{cases} \tag{2.38}$$

Therefore, using Theorem II.3, we conclude that there is a globally optimal encoder of the form:

$$Z^i_t = f^i_t(X^i_t, b^i_t, \mu^i_t)$$

for $t = 1, 2, \ldots, T$ and $i = 1, 2$.

Or equivalently,

$$Z_t^i = f_t^i(X_t^i, b_t^i, Z_{1:t-1}^i) \tag{2.39}$$

Observe that the domain of the encoding functions in (2.39) keeps increasing with time since it includes all past transmitted symbols $Z_{1:t-1}^1$. We would like to find a sufficient statistic that belongs to a time-invariant space. Such a statistic would allow us to address problems with large (or infinite) time horizons.

For that matter, let us first review the approach used for obtaining the first structural result for the encoders (Theorem II.3). We fixed the strategy of encoder 2 and the receiver to any arbitrary choice and looked at the optimization problem P1 from encoder 1's perspective. Essentially, we addressed the following question: if encoder 2 and the receiver have fixed their strategies, how can we characterize the best strategy of encoder 1 in response to the other agents' fixed strategies? In other words, with $f_{1:T}^2$ and $g_{1:T}$ as fixed, what kind of strategies of encoder 1 ($f_{1:T}^1$) minimize the objective in equation (2.13)? This approach allowed us to formulate a Markov decision problem for encoder 1. The Markov decision problem gave us a sufficient statistic for encoder 1 that holds for any choice of strategies of encoder 2 and the receiver and this led to the result of Theorem II.3. In problem P2, such an approach gives the result of equation (2.39) - which implies a sufficient statistic whose domain keeps increasing with time.

To proceed further, we need to adopt a different approach. As before, we will make an arbitrary choice of encoder 2's strategy of the form in (2.3). Given this fixed encoder 2, we will now ask, what are the jointly optimal strategies for encoder 1 and the receiver? That is, assuming $f_{1:T}^2$ is fixed, what choice of $f_{1:T}^1$ and $g_{1:T}$ together minimize the objective in equation (2.36)? From our previous structural results, we know that we can restrict to encoding rules $f_{1:T}^1$ of the form in (2.39) and decoding rules from Theorem II.11 without any loss of optimality. We thus have the following problem:

<u>*Problem P2'*</u>*:* In Problem P2, with encoder 2's strategy fixed to an arbitrary choice $f'^2_{1:T}$, find the jointly optimal strategies of encoder 1 of the form in (2.39) and of the receiver in Theorem II.11 to minimize

$$J(f^1_{1:T}, f'^2_{1:T}, g_{1:T}) = \mathbb{E}[\sum_{t=1}^{T} \rho_t(X_t, \hat{X}_t)]$$

Problem P2' is in some sense a real-time point-to-point communication problem with side information at the receiver. This is now a decentralized team problem with the first encoder and the receiver as the two agents. Note that encoder 1 influences the decisions at the receiver not only by the symbols Z^1_t it sends but by the entire encoding functions it employs (since the receiver's belief ψ_t depends on the choice of encoding functions $f^1_{1:t}$). A general way to solve such dynamic team problems is to search through the space of all strategies to identify the best choice. For our problem (and for many team problems), this is not a useful approach for two reasons: 1) Complexity - the space of all strategies is clearly too large even for small time horizons, thus making a brute force search prohibitive. 2) More importantly, such a method does not reveal any characteristic of the optimal strategies and does not lead to the identification of a sufficient statistic. We will therefore adopt a different philosophy to address our problem.

Our approach is to first consider a modified version of problem P2'. We will construct this modified problem in such a way so as to ensure that:

(a) The new problem is a single agent problem instead of a team problem. Single agent centralized problems (in certain cases) can be studied through the framework of Markov decision theory and dynamic programming.

(b) The new problem is *equivalent* to the original team problem. We will show that the conclusions from the modified problem remain true for the problem P2' as well.

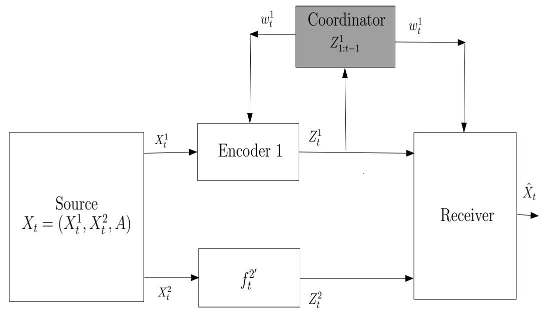

Figure 2.4: Coordinator's Problem P2"

We proceed as follows:

Step 1: We introduce a centralized stochastic control problem from the point of view of a fictitious agent who knows the "common information" between encoder 1 and the receiver.

Step 2: We argue that the centralized problem of Step 1 is equivalent to the original decentralized team problem.

Step 3: We solve the centralized stochastic control problem by identifying an information state and employing dynamic programming arguments. The solution of this problem will reveal a sufficient statistic with a time-invariant domain for encoder 1.

Below, we elaborate on these steps.

Step 1: We observe that the first encoder and the receiver have some common information. At time t, they both know $Z^1_{1:t-1}$. We now formulate a centralized problem from the perspective of a fictitious agent that knows just the common information $Z^1_{1:t-1}$. We call this fictitious agent the "coordinator" (See Fig. 2.4).

The system operates as follows in this new problem: Based on $Z^1_{1:t-1}$, the coordinator selects a partial-encoding function

$$w^1_t : \mathcal{X}^1 \times \Delta(\mathcal{A}) \longrightarrow \mathcal{Z}^1$$

An encoding function of the form in (2.39) can be thought of as a collection of mappings from $\mathcal{X}^1 \times \Delta(\mathcal{A})$ to $\mathcal{Z}^1$ - one for each realization of $Z^1_{1:t-1}$. Clearly, w^1_t

represents one such mapping corresponding to the true realization of $Z^1_{1:t-1}$ that was observed by the coordinator. (At $t = 1$, since there is no past common information, the partial-encoding rule w^1_1 is simply f^1_1 which is a mapping from $\mathcal{X}^1 \times \Delta(\mathcal{A})$ to $\mathcal{Z}^1$.) The coordinator informs the encoder 1 of its choice w^1_t. The encoder 1 then uses w^1_t on its observations X^1_t and b^1_t to find the symbol to be transmitted, i.e,

$$Z^1_t = w^1_t(X^1_t, b^1_t) \tag{2.40}$$

The coordinator also informs the receiver of the partial-encoding function. The receiver at each time t, forms its belief on the state of the source based on the received symbols, the partial-encoding functions and the fixed strategy of encoder 2. This belief is

$$\psi_t(x) := P(X_t = x | z^1_{1:t}, z^2_{1:t}, w^1_{1:t}, f'^2_{1:t})$$

for $x \in \mathcal{X}$. The receiver's optimal estimate at time t is then given as:

$$\hat{X}_t = \operatorname*{argmin}_{s \in \mathcal{X}} \sum_{x \in \mathcal{X}} \psi_t(x) \rho_t(x, s) \tag{2.41}$$

The coordinator then observes the symbol Z^1_t sent from encoder 1 to the receiver and then selects the partial-encoding function for time $t+1$ (w^1_{t+1}). The system continues like this from time $t = 1$ to T. The objective of the coordinator is to minimize the performance criterion of equation (2.36), that is, to minimize

$$\mathbb{E}[\sum_{t=1}^{T} \rho_t(X_t, \hat{X}_t)]$$

We then have the following problem:

Problem P2": In Problem P2, with encoder 2's strategy fixed to the same choice $f'^2_{1:T}$ as in P2' and with a coordinator between encoder 1 and the receiver as described above, find an optimal selection rule for the coordinator, that is find the mappings

$\Lambda_t, t = 1, 2, ..., T$ that map the coordinator's information to its decision

$$w_t^1 = \Lambda_t(Z_{1:t-1}^1, w_{1:t-1}^1)$$

so as to minimize the total expected distortion over time T.
(Note we have included the past decisions ($w_{1:t-1}^1$) of the coordinator in the argument of Λ_t since they themselves are functions of the past observations $Z_{1:t-1}^1$).
Remark: Under a given selection rule for the coordinator, the function w_t^1 is a random variable whose realization depends on the realization of past Z_{t-1}^1 which, in turn, depends on the realization of the source process and the past partial-encoding functions.

Step 2: We now argue that the original team problem P2' is equivalent to the problem in the presence of the coordinator (Problem P2"). Specifically, we show that any achievable value of the objective (that is, the total expected distortion over time T) in problem P2' can also be achieved in problem P2" and vice versa. Consider first any selection rule $\Lambda_t, t = 1, 2, ..., T$ for the coordinator. Since the coordinator in Step 1 only knows the common information between encoder 1 and the receiver, it implies that all information available at the coordinator is in fact available to both encoder 1 and the receiver. Thus, the selection rule Λ_t of the coordinator can be used by both encoder 1 and the receiver to determine the partial-encoding function, w_t^1, to be used at time t even when the coordinator is not actually present. Therefore, the coordinator can effectively be simulated by encoder 1 and the receiver, and hence any achievable value of the objective in Problem P2" with the coordinator can be achieved even in the absence of a physical coordinator.

Conversely, in Problem P2' consider any strategy $f_{1:T}^1$ of encoder 1 and the corresponding optimal receiver given by Theorem II.11. Now consider the following selection rule for the coordinator in P2": At each time t, after having observed $z_{1:t-1}^1$, the coordinator selects the following partial encoding function.

$$w_t^1(\cdot) = f_t^1(\cdot, z_{t-1}^1)$$

Then it is clear that for every realization of the source, encoder 1 in Problem P2" will produce the same realization of encoded symbols as encoder 1 of Problem P2'. Consequently the above selection rule of the coordinator will induce the same joint distribution $P(X_{1:T}, Z_{1:T}^1, Z_{1:T}^2)$ as the encoding rules $f_{1:T}^1$ for encoder 1 in problem P2'. Then the receivers in Problem P2' and Problem P2" will have the same conditional belief ψ_t and will make the same estimates (given by Theorem II.11 and equation (2.41) respectively). Thus any achievable value of the objective in Problem P2' can also be achieved in Problem P2".

The above equivalence allows us to focus on the coordinator's problem to solve the original problem P2'. We now argue that the coordinator's problem is in fact a single agent centralized problem for which Markov decision-theoretic results can be employed.

Step 3: To further describe the coordinator's problem we need the following definition and lemma.

Definition II.12. For $t = 1, 2, \ldots, T$, let ξ_t^1 be the coordinator's belief on X_t^1, b_t^1. That is,

$$\xi_t^1(x_t^1, \tilde{b}_t^1) := P(X_t^1 = x_t^1, b_t^1 = \tilde{b}_t^1 | Z_{1:t}^1, w_{1:t}^1)$$

for $x_t^1 \in \mathcal{X}^1$ and $\tilde{b}_t^1 \in \Delta(\mathcal{A})$.

For notational convenience, we define $\xi_0^1 := 0$.

Remark: Recall that b_t^1 is encoder 1's posterior belief on A given its observations X_t^1. Due to finiteness of the space $\mathcal{X}^1$, b_t^1 can only take one of finitely many values from the set $\Delta(\mathcal{A})$. Hence, $\xi(x_t^1, \cdot)$ is a purely atomic measure on the space $\Delta(\mathcal{A})$ with finite number of atoms.

Lemma II.13. *For a fixed strategy of encoder 2, there is an optimal decoding rule of the form:*

$$\hat{X}_t = \tau_t(\psi_t) = \tau_t(\delta_t(\xi_t^1, Z_{1:t}^2)) \tag{2.42}$$

where $\delta_t, t = 1, 2, \ldots, T$ are fixed transformations that depend only on source statistics and the fixed strategy of encoder 2 and $\tau_t, t = 1, 2, \ldots, T$ are the decoding functions as defined in Definition II.9.

Proof. See Appendix A. □

From equations (2.40) and (2.42), it follows that in the coordinator's problem P2", encoder 1 and the receiver are simply implementors of fixed transformations. They do not make any decisions. Thus, in this formulation, the coordinator is the sole decision maker. We now analyze the centralized problem for the coordinator.

Firstly, observe that at time t, the coordinator knows its observations so far - $Z_{1:t-1}^1$ and the partial encoding functions it used so far - $w_{1:t-1}^1$; it then selects an "action" w_t^1 and makes the next "observation" Z_t^1. In particular, note that the coordinator has perfect recall, that is, it remembers all its past observations and actions-this is a critical characteristic of classical centralized problems for which Markov decision-theoretic results hold.

We can now prove the following lemma :

Lemma II.14. *1. With a fixed strategy of the second encoder, ξ_t^1 can be updated as follows:*

$$\xi_t^1 = \gamma_t^1(\xi_{t-1}^1, Z_t^1, w_t^1) \tag{2.43}$$

where $\gamma_t^1, t = 2, \ldots, T$ are fixed transformations that depend only on the source statistics.

2. For a fixed strategy of the second encoder, the expected instantaneous cost from *the coordinator's perspective can be written as:*

$$\mathbb{E}\{\rho_t(X_t, \hat{X}_t)|Z^1_{1:t}, w^1_{1:t}\} = \overline{\rho}_t(\xi^1_t) \tag{2.44}$$

for $t = 1, 2, \ldots, T$, where $\overline{\rho}_t$ are deterministic functions.

Proof. See Appendix A. □

Based on Lemma II.14, we obtain the following result on the coordinator's optimization problem.

Theorem II.15. *For any given selection rule $\Lambda_t, t = 1, 2..., T$ for the coordinator, there exists another selection rule $G_t, t = 1, 2, ..., T$ that selects the partial-encoding function to be used at time t, (w^1_t) based only on ξ^1_{t-1} and whose performance is no worse than that of $\Lambda_t, t = 1, 2, ..., T$. Therefore, one can optimally restrict to selection rules for the coordinator of the form:*

$$w^1_t = G_t(\xi^1_{t-1}) \tag{2.45}$$

Proof. Because of Lemma II.14, the optimization problem for the coordinator is to control the evolution of ξ^1_t (given by (2.43)) through its actions w^1_t, when the instantaneous cost depends only on ξ^1_t. Since ξ^1_t is known to the coordinator, this problem is similar to the control of a perfectly observed Markov process. This observation essentially implies the result of the theorem, as it follows from Markov decision theory (*Kumar and Varaiya* (1986b), Chapter 6) that to control a perfectly observed Markov process one can restrict attention to policies that depend only on the current state of the Markov process without any loss of optimality.

□

We have therefore identified the structure of the coordinator's selection rule. The coordinator does not need to remember all of its information - $Z^1_{1:t-1}$ and $w^1_{1:t-1}$. It can operate optimally by just using ξ^1_{t-1}. We can thus conclude the following result.

Theorem II.16. *In Problem P2, there is no loss of optimality in considering decoding rules of the form in Theorem II.11 with encoders that operate as follows:*
For $i = 1, 2,$ *define* $\xi^i_0 := 0$ *and for* $t = 1, 2, ...T$,

$$Z^i_t = f^i_t(X^i_t, b^i_t, \xi^i_{t-1}) \tag{2.46}$$

and

$$\xi^i_t = \gamma^i_t(\xi^i_{t-1}, Z^i_t, f^i_t(\cdot, \xi^i_{t-1})) \tag{2.47}$$

where γ^i_t *are fixed transformations (Lemma II.14).*

Proof. The assertion of the the theorem follows from Theorem II.15 and the equivalence between problem P2' and P2" established in Step 2. The coordinator (either real or simulated by encoder 1 and receiver) can select the partial encoding functions by a selection rule of the form:

$$w^1_t = G_t(\xi^1_{t-1})$$

and the encoder 1's symbol to be transmitted at time t is given as:

$$Z^1_t = w^1_t(X^1_t, b^1_t)$$

Thus, Z^1_t is a function of X^1_t, b^1_t and ξ^1_{t-1} that was used to select w^1_t. That is,

$$Z^1_t = f^1_t(X^1_t, b^1_t, \xi^1_{t-1})$$

where $w^1_t(\cdot) = f^1_t(\cdot, \xi^1_{t-1})$. The coordinator (real or simulated) then updates ξ^1_{t-1}

according to Lemma II.14 as:

$$\xi_t^1 = \gamma_t^1(\xi_{t-1}^1, Z_t^1, w_t^1)$$

The same argument holds for encoder 2 as well. $\square$

2.4.4 Discussion

Observe that $Z_{1:t-1}^1$ appearing in the argument of optimal encoding functions in (2.39) have been replaced by ξ_{t-1}^1. By definition, ξ_t^1 is a joint belief on $\mathcal{X}^1$ and $\Delta(\mathcal{A})$, therefore, ξ_t^1 belongs to a time-invariant space, namely, the space of joint beliefs on $\mathcal{X}^1$ and $\Delta(\mathcal{A})$. Thus the domain of the optimal encoding functions in (2.46) is time-invariant. However, ξ_t^1 above is a joint belief on a finitely-valued random variable (X_t^1) and a real-valued vector (b_t^1). Thus, we have an infinite-dimensional sufficient statistic for the encoder. Clearly, such a sufficient statistic can not be directly used for implementation. However, it may still be used in identifying good approximations to optimal encoders. Below, we present some cases where the above structural result may suggest finite-dimensional representations of the sufficient statistic.

2.4.5 Special Cases

2.4.5.1 A observed at the Encoders

Consider the case when the encoder's observations at time $t = 1$ include the realization of the random variable A. Clearly, the encoder's belief on A, (b_t^i) can be replaced by the true value of A in Theorem II.16. Thus, for problem P2, there is an optimal encoding rule of the form:

$$Z_t^1 = f_t^1(X_t^1, A, P(X_{t-1}^1, A | Z_{1:t-1}^1, f_{1:t-1}^1)) \tag{2.48}$$

Since A belongs to a finite set, the domain of the encoding functions in (2.48) consists of the scalars X_t^1 and A and a belief on the finite space $\mathcal{X}^1 \times \mathcal{A}$. Thus when A is observed at the encoders, we have a finite dimensional sufficient statistic for each encoder.

2.4.5.2 Independent Observations at Encoders

Consider the case when the encoders' observations are independent Markov chains. This is essentially the case when A is constant with probability 1. Then, effectively, all agents know A. In this case, the result of (2.48) reduces to

$$Z_t^1 = f_t^1(X_t^1, P(X_{t-1}^1 | Z_{1:t-1}^1, f_{1:t-1}^1)) \tag{2.49}$$

and we have a finite dimensional sufficient statistic for the encoders.

2.5 Extensions

We apply our results for Problems P1 to P2 to other related problems in this section.

2.5.1 Multiple (n) encoders and single receiver problem

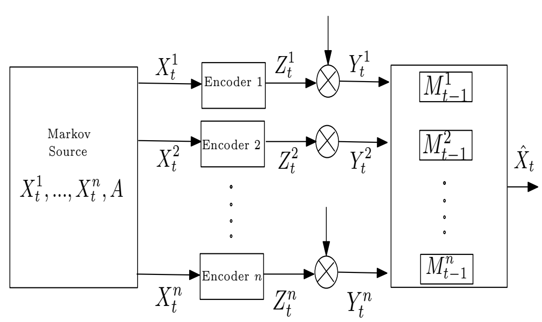

Figure 2.5: Problem with n encoders

Consider the model of Figure 2.5 where we have n ($n > 2$) encoders that partially observe a Markov source and encode their observations, in real-time, into sequences of discrete symbols that are transmitted over separate noisy channels (with independent noise) to a common receiver. We make assumptions analogous to assumptions A1 and A2 for Problem P1, that is,

1. Assumption 1: The state of the Markov source is given as :

$$X_t := (X_t^1, X_t^2, ..., X_t^n, A)$$

where A is a time-invariant random variable and conditioned on A, $X_t^1, X_t^2, ..., X_t^n$ are conditionally independent Markov chains. The i^{th} encoder observes the process $X_t^i, t = 1, 2, ...$ and uses encoding functions of the form :

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

for $i = 1, 2, ..., n$.

2. Assumption 2: We have a finite memory receiver that maintains a separate memory for symbols received from each channel. This memory is updated as follows:

$$M_1^i = l_1^i(Y_1^i), i = 1, 2, ..., n \tag{2.50a}$$

$$M_t^i = l_t^i(M_{t-1}^i, Y_t^i), i = 1, 2, ..., n \tag{2.50b}$$

where M_t^i belongs to finite alphabet $\mathcal{M}^i$, and l_t^i are the memory update functions at time t for $i = 1, 2, ..., n$. The receiver produces an estimate of the source $\hat{X}_t$ based on its memory contents at time $t - 1$ and the symbols received at time t, that is,

$$\hat{X}_t = g_t(Y_t^1, Y_t^2, ..., Y_t^n, M_{t-1}^1, M_{t-1}^2, ..., M_{t-1}^n) \tag{2.51}$$

A non-negative distortion function $\rho_t(X_t, \hat{X}_t)$ measures the instantaneous distortion at time t. We can now formulate the following problem.

Problem P3: With the assumptions 1 and 2 as above, and given source and channel statistics, the encoding alphabets, the distortion functions ρ_t and a time horizon T, the objective is to find globally optimal encoding, decoding and memory update functions $f^1_{1:T}, f^2_{1:T}, ..., f^n_{1:T}, g_{1:T}, l^1_{1:T}, l^2_{1:T}, ..., l^n_{1:T}$ so as to minimize

$$\mathbb{E}\{\sum_{t=1}^{T} \rho_t(X_t, \hat{X}_t)\} \tag{2.52}$$

For this problem we can establish, by arguments similar to those used in the problems with two encoders, the following results that are analogous to Theorem II.3 and Theorem II.16 respectively.

Theorem II.17. *There exist globally optimal encoding rules of the form :*

$$Z^i_t = f^i_t(X^i_t, b^i_t, \mu^i_t) \tag{2.53}$$

where $b^i_t := P(A|X^i_{1:t})$ *and* $\mu^i_t := P(M^i_{t-1}|Z^i_{1:t-1}, l^i_{1:t-1})$. *The optimal decoding rules are of the form:*

$$\hat{X}_t = \tau_t(\psi_t) \tag{2.54}$$

where $\psi_t := P(X_t|Y^1_t, Y^2_t, ..., Y^n_t, M^1_{t-1}, M^2_{t-1}, ..., M^n_{t-1})$ *and* τ_t *is as defined in* Definition II.9.

Proof. Consider any arbitrary choice of encoding functions for encoder 2 through encoder n and arbitrary choice of the decoding and memory update functions at the receiver. Then the problem for encoder 1 is essentially same as in the case when $n = 2$. □

Theorem II.18. *Consider Problem P3 with noiseless channel (that is,* $Y^i_t = Z^i_t$*) and perfect receiver memory (that is ,* $M^i_{t-1} = Z^i_{1:t-1}$*). Then there is no loss of optimality in considering decoding rules of the form* $\hat{X}_t = \tau_t(\psi_t)$ *where* $\psi_t = P(X_t|Z^1_{1:t}, ..., Z^n_{1:t})$

with encoders that operate as follows:

For $i = 1, 2, ..., n$, define $\xi_0^i := 0$ and for $t = 1, 2, ...T$,

$$Z_t^i = f_t^i(X_t^i, b_t^i, \xi_{t-1}^i) \tag{2.55}$$

and

$$\xi_t^i = \gamma_t^i(\xi_{t-1}^i, Z_t^i, f_t^i(\cdot, \xi_{t-1}^i)) \tag{2.56}$$

where γ_t^i are fixed transformations (Lemma II.14).

Proof. The result follows from Theorem II.16 using similar arguments as in the proof of Theorem II.17. □

2.5.2 Point-to-Point Systems

2.5.2.1 A Side Information Channel

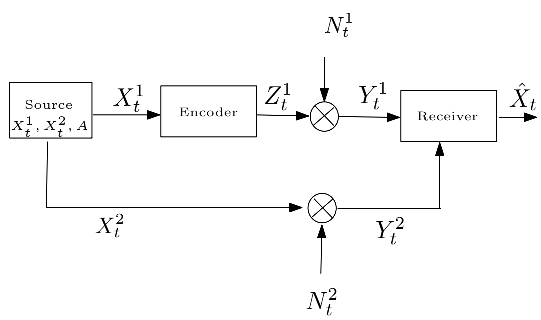

Figure 2.6: Side-Information Problem

Consider Problem P1 or P2 with encoder 2's strategy fixed as follows:

$$Z_t^2 = X_t^2$$

Then the multi-terminal communication problems reduce to a point-to-point communication problems with side-information available at the receiver (See, for example, Fig 2.6). It is clear that the results of Theorem II.3 and Theorem II.10, for noisy channels, and Theorem II.16, for noiseless channels, remain valid for these side-information problems as well (since they are true for any arbitrary choice of encoder 2's strategy).

2.5.2.2 Unknown Transition Matrix

Consider a point-to-point communication system where an encoder is communicating its observations of a Markov source X_t to a receiver (Fig. 2.7). The channel may be noisy or noiseless, the receiver may have finite memory or perfect recall. Structural results for optimal real-time encoding rules have been obtained in cases when the transition probabilities of the Markov source are known (*Witsenhausen* (1978), *Teneketzis* (2006)). Consider now the case where the encoder observes a Markov chain X_t whose transition matrix is not known. However, the set of possible transition matrices is parameterized by a parameter A with a known prior distribution over a finite set $\mathcal{A}$. The encoding functions are of the form:

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

where Z_t is the transmitted symbol at time t. The receiver receives a noisy version of Z_t given by

$$Y_t = h_t(Z_t, N_t)$$

where N_t is the noise in the channel. The receiver maintains a finite memory that is updated as follows:

$$M_1 = l_1(Y_1)$$

$$M_t = l_t(Y_t, M_{t-1})$$

where $M_t \in \mathcal{M}$, $\forall t$. The receiver's estimate at time t is given as:

$$\hat{X}_t = g_t(Y_t, M_{t-1})$$

A non-negative distortion function $\rho_t(X_t, \hat{X}_t)$ measures the instantaneous distortion at time t. We consider the following problem:

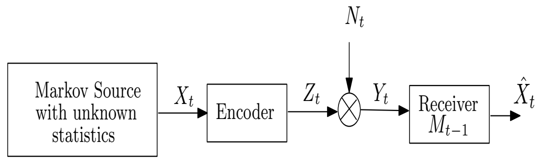

Figure 2.7: Point-to-point system with unknown source statistics

Problem P4: Given the source and receiver model as above and the noise statistics, the encoding alphabets, the channel functions h_t, the distortion functions ρ_t and a time horizon T, the objective is to find globally optimal encoding, decoding and memory update functions $f_{1:T}, g_{1:T}, l_{1:T}$ so as to minimize

$$J(f_{1:T}, g_{1:T}, l_{1:T}) = \mathbb{E}\{\sum_{t=1}^{T} \rho_t(X_t, \hat{X}_t)\} \tag{2.57}$$

The methodology employed for the analysis of Problem P1 can be used to establish the following result.

Theorem II.19. *There exist globally optimal encoding rules of the form :*

$$Z_t = f_t(X_t, b_t, \mu_t) \tag{2.58}$$

where $b_t := P(A|X_{1:t})$ *and* $\mu_t := P(M_{t-1}|Z_{1:t-1}, l_{1:t-1})$. *The optimal decoding rules are of the form:*

$$\hat{X}_t = \tau_t(\psi_t) \tag{2.59}$$

where $\psi_t := P(X_t|Y_t, M_{t-1}, f_{1:t}, l_{1:t})$ *and* τ_t *is as defined in* Definition II.9.

Proof. We can view the optimization problem P4 as a special case of Problem P1 with an imaginary second encoder that makes no observations of the source and sends no message to the receiver (that is, the set $\mathcal{X}^2$ and $\mathcal{Z}^2$ are empty). Thus, the results of the above theorem follow from Theorem II.3 and Theorem II.10. □

The methodology developed for the analysis of Problem P2 can be used to obtain the following result.

Theorem II.20. *Consider Problem P4 with noiseless channel (that is, $Y_t = Z_t$) and perfect receiver memory (that is , $M_{t-1} = Z_{1:t-1}$). Then there is no loss of optimality in considering encoding rules of the form:*

$$Z_t = f_t(X_t, b_t, \xi_{t-1})$$

where $b_t := P(A|X_{1:t})$ and

$$\xi_{t-1} := P(X_{t-1}, b_{t-1}|Z_{1:t-1})$$

with decoding rules of the form:

$$\hat{X}_t = \tau_t(\psi_t) \tag{2.60}$$

where $\psi_t := P(X_t = x|Z_{1:t})$ and τ_t is as defined in Definition II.9.

Proof. The result follows from Theorem II.16 using similar arguments as in the proof of Theorem II.20. □

2.5.3 kth order Markov Source

Consider Problem P1 or P2 with a source model given by the following equations:

$$X^1_{t+1} = F^1_t(X^1_t, X^1_{t-1}, .., X^1_{t+1-k}, A, W^1_t) \tag{2.61a}$$

$$X^2_{t+1} = F^2_t(X^2_t, X^2_{t-1}, .., X^2_{t+1-k}, A, W^2_t) \tag{2.61b}$$

Thus, conditioned on a global, time-invariant random variable A, X^1_t and X^2_t are conditionally independent kth order Markov processes. It is straightforward to consider a Markovian reformulation of the source by defining

$$B^i_t := (X^i_1, X^i_2, ..., X^i_t)$$

for $i = 1, 2$ and $t \leq k$ and

$$B^i_t := (X^i_t, X^i_{t-1}, .., X^i_{t+1-k})$$

for $i = 1, 2$ and $t > k$. We then have that

$$B^i_{t+1} = \tilde{F}^1_t(B^i_t, A, W^1_t) \tag{2.62}$$

for $i = 1, 2$. Thus, we now have a Markov system (when conditioned on A) - with B^i_t as the encoder i's observations - for which our structural results directly apply.

2.5.4 Communication with Finite Delay

Consider the models of Problem P1 or P2 with the following objective function:

$$\sum_{t=d+1}^{T+d} \mathbb{E}[\rho_t(X_{t-d}, \hat{X}_t)]$$

The above objective can be interpreted as the total expected distortion incurred when the receiver can allow a small finite delay, d, before making its final estimate on the state of the source. Thus, the receiver produces a sequence of source estimates $\hat{X}_{d+1}, \hat{X}_{d+2}, ..., \hat{X}_{T+d}$, and incurs a distortion $\sum_{t=d+1}^{T+d} \mathbb{E}[\rho_t(X_{t-d}, \hat{X}_t)]$. We can transform this problem to our problem by the following regrouping of variables.
For $i = 1, 2$ and $t = 1, 2, .., d$ define

$$B^i_t := (X^i_1, X^i_2, .., X^i_t) \tag{2.63}$$

For $t = d+1, ..., T$, define

$$B^i_t := (X^i_{t-d}, X^i_{t-d+1}, .., X^i_t) \tag{2.64}$$

and for $t = T+1, T+2, ..T+d$

$$B^i_t := (X^i_{t-d}, X^i_{t-d+1}, .., X^i_T) \tag{2.65}$$

Then, it is easily seen that conditioned on A, B^1_t and B^2_t are two conditionally independent Markov chains. Moreover, the distortion function $\rho_t(X_{t-d}, \hat{X}_t)$ can be expressed as $\tilde{\rho}(B^1_t, B^2_t, A, \hat{X}_t)$. Thus, we have modified the problem to an instance of Problem P1 or P2 with B^i_t as the encoder i's observation.

2.6 Conclusion

We considered a real-time communication problem where two encoders make distinct partial observations of a discrete-time Markov source and communicate in real-time with a common receiver which needs to estimate some function of the state of the Markov source in real-time. We assumed a specific model for the source that arises in some applications of interest. In this model, the encoders' observations are conditionally independent Markov chains given an unobserved, time-invariant random variable. We formulated a communication problem with separate noisy channels between each encoder and the receiver and a separated finite memory at the receiver. We obtained finite-dimensional sufficient statistics for the encoders in this problem. The structure of the source and the receiver played a critical role in obtaining these results.

We then considered the communication problem over noiseless channels and perfect receiver memory. We used the presence of common information between an encoder and the receiver to identify a sufficient statistic of the decoder that has a time-invariant domain.

The sufficient statistics we found for the encoders are analogous to the belief state of partially observable Markov decision problems *Kumar and Varaiya* (1986b). For such problems, under suitable conditions, it is known that the expected future cost of an action is a continuous function of the belief state *Smallwood and Sondik* (1973). We believe such results also hold for the encoder's decision problem in Problem P1 and the coordinator's decision problem in Problem P2. This would suggest that sufficient statistics that are "close to" each other will have similar future costs and the quantization of sufficient statistics may not result in large performance loss. Characterizing the effect of such quantization on system performance as well as the robustness of system performance with respect to quantization are currently open problems.

We have also not addressed the problem of finding globally optimal real-time encoding and decoding strategies in this paper. A sequential decomposition of the global optimization problem, for a special case of the problems formulated here, appears in *Nayyar and Teneketzis* (2008).

CHAPTER III

Delayed Sharing Information Structures

3.1 Introduction

3.1.1 Motivation

One of the difficulties in optimal design of decentralized control systems is handling the increase of data at the control stations with time. This increase in data means that the domain of control laws increases with time which, in turn, creates two difficulties. Firstly, the number of control strategies increases doubly exponentially with time; this makes it harder to search for an optimal strategy. Secondly, even if an optimal strategy is found, implementing functions with time increasing domain is difficult.

In centralized stochastic control (*Kumar and Varaiya* (1986a)), these difficulties can be circumvented by using the conditional probability of the state given the data available at the control station as a sufficient statistic (where the data available to a control station comprises of all observations and control actions till the current time). This conditional probability, called *information state*, takes values in a time-invariant space. Consequently, we can restrict attention to control laws with time-invariant domain. Such results, where data that is increasing with time is "compressed" to a sufficient statistic taking values in a time-invariant space, are called *structural results*. While the information state and structural result for centralized stochastic control problems is well known, no general methodology to find such information states or structural results exists for decentralized stochastic control problems.

The structural results in centralized stochastic control are related to the concept of separation. In centralized stochastic control, the information state, which is conditional probability of the state given all the available data, does not depend on the control strategy (which is the collection of control laws used at different time instants). This has been called a one-way separation between estimation and control. An important consequence of this separation is that for any given choice of control laws till time $t-1$ and a given realization of the system variables till time t, the information states at future times do not depend on the choice of the control law at time t but only on the realization of control action at time t. Thus, the future information states are *separated* from the choice of the current control law. This fact is crucial for the formulation of the classical dynamic program where at each step the optimization problem is to find the best control action for a given realization of the information state. No analogous separation results are known for general decentralized systems.

In this chapter, we find structural results for decentralized control systems with delayed sharing information structures. In a system with n-step delayed sharing, every control station knows the n-step prior observations and control actions of all other control stations. This information structure, proposed by Witsenhausen in *Witsenhausen* (1971), is a link between the classical information structures, where information is shared perfectly among the controllers, and the non-classical information structures, where there is no "lateral" sharing of information among the controllers. Witsenhausen asserted a structural result for this model without any proof in his seminal paper *Witsenhausen* (1971). Varaiya and Walrand *Varaiya and Walrand* (1978) proved that Witsenhausen's assertion was true for $n = 1$ but false for $n > 1$. For $n > 1$, Kurtaran *Kurtaran* (1979) proposed another structural result. However, Kurtaran proved his result only for the terminal time step (that is, the last time step in a finite horizon problem); for non-terminal time steps, he gave an abbreviated argument, which we believe is incomplete. (The details are given in Section 3.5 of the paper).

We prove two structural results of the optimal control laws for the delayed sharing information structure. We compare our results to those conjectured by Witsenhausen and show that our structural results for n-step delay sharing information structure simplify to that of Witsenhausen for $n = 1$; for $n > 1$, our results are different from the result proposed by Kurtaran.

We note that our structural results do not have the separated nature of centralized stochastic control. That is, for any given realization of the system variables till time t, the realization of information states at future times depend on the choice of the control law at time t. However, our second structural result shows that this dependence only propagates to the next $n - 1$ time steps. Thus, the information states from time $t+n-1$ onwards are separated from the choice of control laws before time t. We call this a *delayed* separation between information states and control laws.

The absence of classical separation rules out the possibility of a classical dynamic program to find the optimum control laws. However, optimal control laws can still be found in a sequential manner. Based on the two structural results, we present two sequential methodologies to find optimal control laws. Unlike classical dynamic programs, each step in our sequential decomposition involves optimization over a space of functions instead of the space of control actions.

3.1.2 Model

Consider a system consisting of a plant and K controllers (control stations) with decentralized information. At time t, $t = 1, \dots, T$, the state of the plant X_t takes values in a finite set $\mathcal{X}$; the control action U_t^k at control station k, $k = 1, \dots, K$, takes values in a finite set $\mathcal{U}^k$. The initial state X_0 of the plant is a random variable taking value in $\mathcal{X}$. With time, the plant evolves according to

$$X_t = f_t(X_{t-1}, U_t^{1:K}, V_t) \tag{3.1}$$

where V_t is a random variable taking values in a finite set $\mathcal{V}$. $\{V_t; t = 1, \dots, T\}$ is a sequence of independent random variables that are also independent of X_0.

The system has K observation posts. At time t, $t = 1, \dots, T$, the observation Y_t^k of post k, $k = 1, \dots, K$, takes values in a finite set $\mathcal{Y}^k$. These observations are generated according to

$$Y_t^k = h_t^k(X_{t-1}, W_t^k) \tag{3.2}$$

where W_t^k are random variables taking values in a finite set $\mathcal{W}^k$. $\{W_t^k; t = 1, \dots, T; k = 1, \dots, K\}$ are independent random variables that are also independent of X_0 and $\{V_t; t = 1, \dots, T\}$.

The system has n-step delayed sharing. This means that at time t, control station k observes the current observation Y_t^k of observation post k, the n steps old observations $Y_{t-n}^{1:K}$ of all posts, and the n steps old actions $U_{t-n}^{1:K}$ of all stations. Each station has perfect recall; so, it remembers everything that it has seen and done in the past. Thus, at time t, data available at station k can be written as (C_t, P_t^k), where

$$C_t \coloneqq (Y_{1:t-n}^{1:K}, U_{1:t-n}^{1:K})$$

is the data known to all stations and

$$P_t^k \coloneqq (Y_{t-n+1:t}^k, U_{t-n+1:t-1}^k)$$

is the additional data known at station k, $k = 1, \dots, K$. Let $\mathcal{C}_t$ be the space of all possible realizations of C_t; and $\mathcal{P}^k$ be the space of all possible realizations of P_t^k.

Table 3.1: Summary of the control laws in the model for $K = 2$.

	Controller 1	Controller 2
Observations (actual)	$\begin{pmatrix} Y^1_{1:t} & Y^2_{1:t-n} \\ U^1_{1:t-1} & U^2_{1:t-n} \end{pmatrix}$	$\begin{pmatrix} Y^1_{1:t-n} & Y^2_{1:t} \\ U^1_{1:t-n} & U^2_{1:t-1} \end{pmatrix}$
Observations (shorthand)	(C_t, P^1_t)	(C_t, P^2_t)
Control action	U^1_t	U^2_t
Control laws	g^1_t	g^2_t

Station k chooses action U^k_t according to a control law g^k_t, i.e.,

$$U^k_t = g^k_t(P^k_t, C_t). \tag{3.3}$$

The choice of $\boldsymbol{g} = \{g^k_t; k = 1, \dots, K; t = 1, \dots, T\}$ is called a *design* or a *control strategy*. $\mathcal{G}$ denotes the class of all possible designs. At time t, a cost $R_t(X_t, U^1_t, \dots, U^K_t)$ is incurred. The performance $\mathcal{J}(\boldsymbol{g})$ of a design is given by the expected total cost under it, i.e.,

$$\mathcal{J}(\boldsymbol{g}) = \mathbb{E}^{\boldsymbol{g}}\{\sum_{t=1}^{T} R_t(X_t, U^{1:K}_t)\} \tag{3.4}$$

where the expectation is with respect to the joint measure on all the system variables induced by the choice of $\boldsymbol{g}$. For reference, we summarize the notation of this model in Table 3.1. We consider the following problem.

Problem 1. *Given the statistics of the primitive random variables* X_0, $\{V_t; t = 1, \dots, T\}$, $\{W^k_t; k = 1, \dots, K; t = 1, \dots, T\}$, *the plant functions* $\{f_t; t = 1, \dots, T\}$, *the observation functions* $\{h^k_t; k = 1, \dots, K; t = 1, \dots, T\}$, *and the cost functions* $\{R_t; t = 1, \dots, T\}$ *choose a design* $\boldsymbol{g}^*$ *from* $\mathcal{G}$ *that minimizes the expected cost given by* (3.4).

Remarks on the Model

1. We assumed that all primitive random variables and all control actions take values in finite sets for convenience of exposition. Similar results can be obtained with uncountable sets under suitable technical conditions.

2. In the standard stochastic control literature, the dynamics and observations equations are defined in a different manner than (3.1) and (3.2). The usual model is

$$X_{t+1} = f_t(X_t, U_t^{1:K}, V_t) \tag{3.5}$$

$$Y_t^k = h_t^k(X_t, W_t^k) \tag{3.6}$$

   However, Witsenhausen *Witsenhausen* (1971) as well as Varaiya and Walrand *Varaiya and Walrand* (1978) used the model of (3.1) and (3.2) in their papers. We use the same model so that our results can be directly compared with earlier conjectures and results. The arguments of this chapter can be used for the dynamics and observation model of (3.5) and (3.6) with minor changes.

3.1.3 The structural results

Witsenhausen *Witsenhausen* (1971) asserted the following structural result for Problem 1. *Structural Result (Witsenhausen's Conjecture Witsenhausen (1971)):* In Problem 1, without loss of optimality we can restrict attention to control strategies of the form

$$U_t^k = g_t^k(P_t^k, \mathbb{P}(X_{t-n}|C_t)). \tag{3.7}$$

Witsenhausen's structural result claims that all control stations can "compress" the common information C_t to a sufficient statistic $\mathbb{P}(X_{t-n}|C_t)$. Unlike C_t, the size of $\mathbb{P}(X_{t-n}|C_t)$ does not increase with time.

As mentioned earlier, Witsenhausen asserted this result without a proof. Varaiya and Walrand *Varaiya and Walrand* (1978) proved that the above separation result is true for $n = 1$ but false for $n > 1$. Kurtaran *Kurtaran* (1979) proposed an alternate structural result for $n > 1$.

Structural Result (Kurtaran Kurtaran (1979)): In Problem 1, without loss of optimality we can restrict attention to control strategies of the form

$$U_t^k = g_t^k\big(Y_{t-n+1:t}^k, \mathbb{P}^{g_{1:t-1}^{1:K}}(X_{t-n}, U_{t-n+1:t-1}^{1:K} | C_t)\big). \tag{3.8}$$

Kurtaran used a different labeling of the time indices, so the statement of the result in his paper is slightly different from what we have stated above.

Kurtaran's result claims that all control stations can "compress" the common information C_t to a sufficient statistic $\mathbb{P}^{g_{1:t-1}^{1:K}}(X_{t-n}, U_{t-n+1:t-1}^{1:K} | C_t)$, whose size does not increase with time.

Kurtaran proved his result for only the terminal time-step and gave an abbreviated argument for non-terminal time-steps. We believe that his proof is incomplete for reasons that we point out in Section 3.5. In this chapter, we prove two alternative structural results.

First Structural Result: In Problem 1, without loss of optimality we can restrict attention to control strategies of the form

$$U_t^k = g_t^k\big(P_t^k, \mathbb{P}^{g_{1:t-1}^{1:K}}(X_{t-1}, P_t^{1:K} | C_t)\big). \tag{3.9}$$

This result claims that all control stations can "compress" the common information C_t to a sufficient statistic $\mathbb{P}^{g_{1:t-1}^{1:K}}(X_{t-1}, P_t^{1:K} | C_t)$, whose size does not increase with time.

Second Structural Result: In Problem 1, without loss of optimality we can restrict attention to control strategies of the form

$$U_t^k = g_t^k\bigl(P_t^k, \mathbb{P}(X_{t-n}|C_t), r_t^{1:K}\bigr). \tag{3.10}$$

where $r_t^{1:K}$ is a collection of partial functions of the previous $n-1$ control laws of each controller,

$$r_t^k \coloneqq \{(g_m^k(\cdot, Y_{m-n+1:t-n}^k, U_{m-n+1:t-n}^k, C_m), m = t-n+1, t-n+2, \dots, t-1\},$$

for $k = 1, 2, \dots, K$. Observe that r_t^k depends only on the previous $n-1$ control laws ($g_{t-n+1:t-1}^k$) and the realization of C_t (which consists of $Y_{1:t-n}^{1:K}, U_{1:t-n}^{1:K}$). This result claims that the belief $\mathbb{P}(X_{t-n}|C_t)$ and the realization of the partial functions $r_t^{1:K}$ form a sufficient representation of C_t in order to optimally select the control action at time t.

Our structural results cannot be derived from Kurtaran's result and vice-versa. At present, we are not sure of the correctness of Kurtaran's result. As we mentioned before, we believe that the proof given by Kurtaran is incomplete. We have not been able to complete Kurtaran's proof; neither have we been able to find a counterexample to his result.

Kurtaran's and our structural results differ from those asserted by Witsenhausen in a fundamental way. The sufficient statistic (also called information state) $\mathbb{P}(X_{t-n}|C_t)$ of Witsenhausen's assertion does not depend on the control strategy. That is, for any realization c_t of C_t, the knowledge of control laws is not required in evaluating the conditional probabilities $\mathbb{P}(X_{t-n} = x|c_t)$. The sufficient statistics $\mathbb{P}^{g_{1:t-1}^{1:K}}(X_{t-n}, U_{t-n+1:t-1}^{1:K}|C_t)$ of Kurtaran's result and $\mathbb{P}^{g_{1:t-1}^{1:K}}(X_{t-1}, P_t^{1:K}|C_t)$ of our first result *depend on the control laws used before time* t. Thus, for a given realization c_t of C_t, the realization of information state depends on the choice of control laws before time t. On the other hand, in our second structural result, the belief $\mathbb{P}(X_{t-n}|C_t)$ is indeed independent of the control strategy, however information about the previous $n-1$ control laws is still needed in the form of the partial functions $r_t^{1:K}$. Since the partial functions $r_t^{1:K}$ do not depend on control laws used before time $t-n+1$, we conclude that the information state at time t is separated from the choice of control laws before time $t-n+1$. We call this a delayed separation between information states and control laws.

The rest of this chapter is organized as follows. We prove our first structural result in Section 3.2. Then, in Section 3.3 we derive our second structural result. We discuss a special case of delayed sharing information structures in Section 3.4. We discuss Kurtaran's structural result in Section 3.5 and conclude in Section 3.6.

3.2 Proof of the first structural result

In this section, we prove the structural result (3.9) for optimal strategies of the K control stations. For the ease of notation, we first prove the result for $K=2$, and then show how to extend it for general K.

3.2.1 Two Controller system ($K=2$)

The proof for $K=2$ proceeds as follows:

1. First, we formulate a centralized stochastic control problem from the point of view of a coordinator who observes the shared information C_t, but does not observe the private information (P_t^1, P_t^2) of the two controllers.

2. Next, we argue that any strategy for the coordinator's problem can be implemented in the original problem and vice versa. Hence, the two problems are equivalent.

3. Then, we identify states sufficient for input-output mapping for the coordinator's problem.

4. Finally, we transform the coordinator's problem into a MDP (Markov decision process), and obtain a structural result for the coordinator's problem. This structural result is also a structural result for the delayed sharing information strucutres due to the equivalence between the two problems.

Below, we elaborate on each of these stages.

Stage 1

We consider the following modified problem. In the model described in Section 3.1.2, in addition to the two controllers, a coordinator that knows the common (shared) information C_t available to both controllers at time t is present. At time t, the coordinator decides the *partial functions*

$$\gamma_t^k : \mathcal{P}^k \mapsto \mathcal{U}^k$$

for each controller k, $k = 1, 2$. The choice of the partial functions at time t is based on the realization of the common (shared) information and the partial functions selected before time t. These functions map each controller's *private information* P_t^k to its control action U_t^k at time t. The coordinator then informs all controllers of all the partial functions it selected at time t. Each controller then uses its assigned partial function to generate a control action as follows.

$$U_t^k = \gamma_t^k(P_t^k). \tag{3.11}$$

The system dynamics and the cost are same as in the original problem. At next time step, the coordinator observes the new common observation

$$Z_{t+1} := \{Y_{t-n+1}^1, Y_{t-n+1}^2, U_{t-n+1}^1, U_{t-n+1}^2\}. \tag{3.12}$$

Table 3.2: Summary of the model with a coordinator.

	Coordinator	Controller k (passive)
Observations (actual)	$(Y^{1:K}_{1:t-n}, U^{1:K}_{1:t-n})$	$(Y^k_{t-n+1:t}, U^k_{t-n+1:t-1})$
Observations (shorthand)	C_t	P^k_t
Control action	$\gamma^{1:K}_t$	U^k_t
Control laws	ψ_t	γ^k_t

Thus at the next time, the coordinator knows $C_{t+1} = (Z_{t+1}, C_t)$ and its choice of all past partial functions and it selects the next partial functions for each controller. The system proceeds sequentially in this manner until time horizon T.

In the above formulation, the only decision maker is the coordinator: the individual controllers simply carry out the necessary evaluations prescribed by (3.11). At time t, the coordinator knows the common (shared) information C_t and all past partial functions $\gamma^1_{1:t-1}$ and $\gamma^2_{1:t-1}$. The coordinator uses a decision rule ψ_t to map this information to its decision, that is,

$$(\gamma^1_t, \gamma^2_t) = \psi_t(C_t, \gamma^1_{1:t-1}, \gamma^2_{1:t-1}), \tag{3.13}$$

or equivalently,

$$\gamma^k_t = \psi^k_t(C_t, \gamma^1_{1:t-1}, \gamma^2_{1:t-1}), \quad k = 1, 2. \tag{3.14}$$

For reference, we summarize the notation of this model in Table 3.2.

The choice of $\boldsymbol{\psi} = \{\psi_t; t = 1, \dots, T\}$ is called a *coordination strategy*. Ψ denotes the class of all possible coordination strategies. The performance of a coordinating strategy is given by the expected total cost under that strategy, that is,

$$\hat{\mathcal{J}}(\boldsymbol{\psi}) = \mathbb{E}^{\boldsymbol{\psi}}\{\sum_{t=1}^{T} R_t(X_t, U^1_t, U^2_t)\} \tag{3.15}$$

where the expectation is with respect to the joint measure on all the system variables induced by the choice of $\boldsymbol{\psi}$. The coordinator has to solve the following optimization problem.

Problem 2 (The Coordinator's Optimization Problem)**.** *Given the system model of Problem 1, choose a coordination strategy* $\boldsymbol{\psi}^*$ *from* Ψ *that minimizes the expected cost given by* (3.15)*.*

Stage 2

We now show that the Problem 2 is equivalent to Problem 1. Specifically, we will show that any design $\boldsymbol{g}$ for Problem 1 can be implemented by the coordinator in Problem 2 with the same value of the problem objective. Conversely, any coordination strategy $\boldsymbol{\psi}$ in Problem 2 can be implemented in Problem 1 with the same value of the performance objective.

Any design $\boldsymbol{g}$ for Problem 1 can be implemented by the coordinator in Problem 2 as follows. At time t the coordinator selects partial functions (γ_t^1, γ_t^2) using the common (shared) information c_t as follows.

$$\gamma_t^k(\cdot) = g_t^k(\cdot, c_t) =: \psi_t^k(c_t), \quad k = 1, 2. \tag{3.16}$$

Consider Problems 1 and 2. Use design $\boldsymbol{g}$ in Problem 1 and coordination strategy $\boldsymbol{\psi}$ given by (3.16) in Problem 2. Fix a specific realization of the initial state X_0, the plant disturbance $\{V_t; t = 1, \dots, T\}$, and the observation noise $\{W_t^1, W_t^2; t = 1, \dots, T\}$. Then, the choice of $\boldsymbol{\psi}$ according to (3.16) implies that the realization of the state $\{X_t; t = 1, \dots, T\}$, the observations $\{Y_t^1, Y_t^2; t = 1, \dots, T\}$, and the control actions $\{U_t^1, U_t^2; t = 1, \dots, T\}$ are identical in Problem 1 and 2. Thus, any design $\boldsymbol{g}$ for Problem 1 can be implemented by the coordinator in Problem 2 by using a coordination strategy given by (3.16) and the total expected cost under $\boldsymbol{g}$ in Problem 1 is same as the total expected cost under the coordination strategy given by (3.16) in Problem 2.

By a similar argument, any coordination strategy $\boldsymbol{\psi}$ for Problem 2 can be implemented by the control stations in Problem 1 as follows. At time 1, both stations know c_1; so, all of them can compute $\gamma_1^1 = \psi_1^1(c_1)$, $\gamma_1^2 = \psi_1^2(c_1)$. Then station k chooses action $u_1^k = \gamma_1^k(p_1^k)$. Thus,

$$g_1^k(p_1^k, c_1) = \psi_1^k(c_1)(p_1^k), \quad k = 1, 2. \tag{3.17a}$$

At time 2, both stations know c_2 and γ_1^1, γ_1^2, so both of them can compute $\gamma_2^k = \psi_2^k(c_2, \gamma_1^1, \gamma_1^2)$, $k = 1, 2$. Then station k chooses action $u_2^k = \gamma_2^k(p_2^k)$. Thus,

$$g_2^k(p_2^k, c_2) = \psi_2^k(c_2, \gamma_1^1, \gamma_1^2)(p_2^k), \quad k = 1, 2. \tag{3.17b}$$

Proceeding this way, at time t both stations know c_t and $\gamma_{1:t-1}^1$ and $\gamma_{1:t-1}^2$, so both of them can compute $(\gamma_{1:t}^1, \gamma_{1:t}^2) = \psi_t(c_t, \gamma_{1:t-1}^1, \gamma_{1:t-1}^2)$. Then, station k chooses action $u_t^k = \gamma_t^k(p_t^k)$. Thus,

$$g_t^k(p_t^k, c_t) = \psi_t^k(c_t, \gamma_{1:t-1}^1, \gamma_{1:t-1}^2)(p_t^k), \quad k = 1, 2. \tag{3.17c}$$

Now consider Problems 2 and 1. Use coordinator strategy $\boldsymbol{\psi}$ in Problem 2 and design $\boldsymbol{g}$ given by (3.17) in Problem 1. Fix a specific realization of the initial state X_0, the plant disturbance $\{V_t; t = 1, \dots, T\}$, and the observation noise $\{W_t^1, W_t^2; t = 1, \dots, T\}$. Then, the choice of $\boldsymbol{g}$ according to (3.17) implies that the realization of the state $\{X_t; t = 1, \dots, T\}$, the observations $\{Y_t^1, Y_t^2; t = 1, \dots, T\}$, and the control actions $\{U_t^1, U_t^2; t = 1, \dots, T\}$ are identical in Problem 2 and 1. Hence, any coordination strategy $\boldsymbol{\psi}$ for Problem 2 can be implemented by the stations in Problem 1 by using a design given by (3.17) and the total expected cost under $\boldsymbol{\psi}$ in Problem 2 is same as the total expected cost under the design given by (3.17) in

Problem 1.

Since Problems 1 and 2 are equivalent, we derive structural results for the latter problem. Unlike, Problem 1, where we have multiple control stations, the coordinator is the only decision maker in Problem 2.

Stage 3

We now look at Problem 2 as a controlled input-output system from the point of view of the coordinator and identify a state sufficient for input-output mapping. From the coordinator's viewpoint, the input at time t has two components: a stochastic input that consists of the plant disturbance V_t and observation noises W_t^1, W_t^2; and a controlled input that consists of the partial functions γ_t^1, γ_t^2. The output is the observations Z_{t+1} given by (3.12). The cost is given by $R_t(X_t, U_t^1, U_t^2)$. We want to identify a state sufficient for input-output mapping for this system. A variable is a state sufficient for input output mapping of a control system if it satisfies the following properties (see *Witsenhausen* (1976)).

1. P1: The next state is a function of the current state and the current inputs.
2. P2: The current output is function of the current state and the current inputs.
3. P3: The instantaneous cost is a function of the current state, the current control inputs, and the next state.

We claim that such a state for Problem 2 is the following.

Definition III.1. For each t define

$$S_t \coloneqq (X_{t-1}, P_t^1, P_t^2) \tag{3.18}$$

Next we show that S_t, $t = 1, 2, \dots, T+1$, satisfy properties (P1)–(P3). Specifically, we have the following.

Proposition III.2.

1. *There exist functions* $\hat{f}_t$, $t = 2, \ldots, T$ *such that*

$$S_{t+1} = \hat{f}_{t+1}(S_t, V_t, W^1_{t+1}, W^2_{t+1}, \gamma^1_t, \gamma^2_t). \tag{3.19}$$

2. *There exist functions* $\hat{h}_t$, $t = 2, \ldots, T$ *such that*

$$Z_t = \hat{h}_t(S_{t-1}). \tag{3.20}$$

3. *There exist functions* $\hat{c}_t$, $t = 1, \ldots, T$ *such that*

$$R_t(X_t, U^1_t, U^2_t) = \hat{R}_t(S_t, \gamma^1_t, \gamma^2_t, S_{t+1}). \tag{3.21}$$

Proof. Part 1 is an immediate consequence of the definitions of S_t and P^k_t, the dynamics of the system given by (3.1), and the evaluations carried out by the control stations according to (3.11). Part 2 is an immediate consequence of the definitions of state S_t, observation Z_t, and private information P^k_t. Part 3 is an immediate consequence of the definition of state and the evaluations carried out by the control stations according to (3.11). □

Stage 4

Proposition III.2 establishes S_t as the state sufficient for input-output mapping for the coordinator's problem. We now define information states for the coordinator.

Definition III.3 (Information States)**.** For a coordination strategy $\boldsymbol{\psi}$, define *information states* Π_t as

$$\Pi_t(s_t) := \mathbb{P}^{\boldsymbol{\psi}}(S_t = s_t | C_t, \gamma^1_{1:t-1}, \gamma^2_{1:t-1}). \tag{3.22}$$

As shown in Proposition III.2, the state evolution of S_t depends on the controlled inputs (γ_t^1, γ_t^2) and the random noise $(V_t, W_{t+1}^1, W_{t+1}^2)$. This random noise is independent across time. Consequently, Π_t evolves in a controlled Markovian manner as below.

Proposition III.4. *For $t = 1, \dots, T-1$, there exists functions F_t (which do not depend on the coordinator's strategy) such that*

$$\Pi_{t+1} = F_{t+1}(\Pi_t, \gamma_t^1, \gamma_t^2, Z_{t+1}). \tag{3.23}$$

Proof. See Appendix B. □

At $t = 1$, since there is no shared information, Π_1 is simply the unconditional probability $\mathbb{P}(S_1) = \mathbb{P}(X_0, Y_1^1, Y_1^2)$. Thus, Π_1 is fixed a priori from the joint distribution of the primitive random variables and does not depend on the choice of coordinator's strategy ψ. Proposition III.4 shows that at $t = 2, \dots, T$, Π_t depends on the strategy $\boldsymbol{\psi}$ only through the choices of $\gamma_{1:t-1}^1$ and $\gamma_{1:t-1}^2$. Moreover, as shown in Proposition III.2, the instantaneous cost at time t can be written in terms of the current and next states (S_t, S_{t+1}) and the control inputs (γ_t^1, γ_t^2). Combining the above two properties, we get the following:

Proposition III.5. *The process Π_t, $t = 1, 2, \dots, T$ is a controlled Markov chain with γ_t^1, γ_t^2 as the control actions at time t, i.e.,*

$$\begin{aligned} \mathbb{P}(\Pi_{t+1} | C_t, \Pi_{1:t}, \gamma_{1:t}^1, \gamma_{1:t}^2) &= \mathbb{P}(\Pi_{t+1} | \Pi_{1:t}, \gamma_{1:t}^1, \gamma_{1:t}^2) \\ &= \mathbb{P}(\Pi_{t+1} | \Pi_t, \gamma_t^1, \gamma_t^2). \end{aligned} \tag{3.24}$$

Furthermore, there exists a deterministic function C_t such that

$$\mathbb{E}\{\hat{R}_t(S_t, \gamma_t^1, \gamma_t^2, S_{t+1}) | C_t, \Pi_{1:t}, \gamma_{1:t}^1, \gamma_{1:t}^2\} = \tilde{R}_t(\Pi_t, \gamma_1^1, \gamma_t^2). \tag{3.25}$$

Proof. See Appendix B. $\square$

The controlled Markov property of the process $\{\Pi_t, t = 1, \ldots, T\}$ immediately gives rise to the following structural result.

Theorem III.6. *In Problem 2, without loss of optimality we can restrict attention to coordination strategies of the form*

$$(\gamma_t^1, \gamma_t^2) = \psi_t(\Pi_t), \quad t = 1, \ldots, T. \tag{3.26}$$

Proof. From Proposition III.5, we conclude that the optimization problem for the coordinator is to control the evolution of the controlled Markov process $\{\Pi_t,\ t = 1, 2, \ldots, T\}$ by selecting the partial functions $\{\gamma_t^1, \gamma_t^2,\, t = 1, 2, \ldots, T\}$ in order to minimize $\sum_{t=1}^{T} \mathbb{E}\{\tilde{R}_t(\Pi_t, \gamma_t^1, \gamma_t^2)\}$. This is an instance of the well-known Markov decision problems where it is known that the optimal strategy is a function of the current state. Thus, the structural result follows from Markov decision theory *Kumar and Varaiya* (1986a). $\square$

The above result can also be stated in terms of the original problem.

Theorem III.7 (Structural Result)**.** *In Problem 1 with $K = 2$, without loss of optimality we can restrict attention to coordination strategies of the form*

$$U_t^k = g_t^k(P_t^k, \Pi_t), \quad k = 1, 2. \tag{3.27}$$

where

$$\Pi_t = \mathbb{P}^{(g^1_{1:t-1}, g^2_{1:t-1})}(X_{t-1}, P_t^1, P_t^2 | C_t) \tag{3.28}$$

where $\Pi_1 = \mathbb{P}(X_0, Y_1^1, Y_1^2)$ and for $t = 2, \ldots, T$, Π_t is evaluated as follows:

$$\Pi_{t+1} = F_{t+1}(\Pi_t, g_t^1(\cdot, \Pi_t), g_t^2(\cdot, \Pi_t), Z_{t+1}) \tag{3.29}$$

Proof. Theorem III.6 established the structure of the optimal coordination strategy. As we argued in Stage 2, this optimal coordination strategy can be implemented in Problem 1 and is optimal for the objective (3.4). At $t = 1$, $\Pi_1 = \mathbb{P}(X_0, Y_1^1, Y_1^2)$ is known to both controllers and they can use the optimal coordination strategy to select partial functions according to:

$$(\gamma_1^1, \gamma_1^2) = \psi_1(\Pi_1)$$

Thus,

$$U_1^k = \gamma_1^k(P_1^k) = \psi_1^k(\Pi_1)(P_1^k) =: g_1^k(P_1^k, \Pi_1), \quad k = 1, 2. \tag{3.30}$$

At time instant $t + 1$, both controllers know Π_t and the common observations $Z_{t+1} = (Y_{t-n+1}^1, Y_{t-n+1}^2, U_{t-n+1}^1, U_{t-n+1}^2)$; they use the partial functions $(g_t^1(\cdot, \Pi_t), g_t^2(\cdot, \Pi_t))$ in equation (3.23) to evaluate Π_{t+1}. The control actions at time $t + 1$ are given as:

$$\begin{aligned} U_{t+1}^k = \gamma_{t+1}^k(P_{t+1}^k) &= \psi_{t+1}(\Pi_{t+1})(P_{t+1}^k) \\ &=: g_{t+1}^k(P_{t+1}^k, \Pi_{t+1}), \quad k = 1, 2. \end{aligned} \tag{3.31}$$

Moreover, using the design $\boldsymbol{g}$ defined according to (3.31), the coordinator's information state Π_t can also be written as:

$$\begin{aligned} \Pi_t &= \mathbb{P}^{\boldsymbol{\psi}}(X_{t-1}, P_t^1, P_t^2 | C_t, \gamma_{1:t-1}^1, \gamma_{1:t-1}^2) \\ &= \mathbb{P}^{\boldsymbol{g}}(X_{t-1}, P_t^1, P_t^2 | C_t, g_1^{1:2}(\cdot, \Pi_1), \ldots, g_{t-1}^{1:2}(\cdot, \Pi_{t-1})) \\ &= \mathbb{P}^{(g_{1:t-1}^1, g_{1:t-1}^2)}(X_{t-1}, P_t^1, P_t^2 | C_t) \end{aligned} \tag{3.32}$$

where we dropped the partial functions from the conditioning terms in (3.32) because under the given control laws $(g_{1:t-1}^1, g_{1:t-1}^2)$, the partial functions used from time 1 to $t - 1$ can be evaluated from C_t (by using Proposition III.4 to evaluate $\Pi_{1:t-1}$). $\square$

Theorem III.7 establishes the first structural result stated in Section 3.1.3 for $K = 2$. In the next section, we show how to extend the result for general K.

3.2.2 Extension to General K

Theorem III.7 for two controllers ($K = 2$) can be easily extended to general K by following the same sequence of arguments as in stages 1 to 4 above. Thus, at time t, the coordinator introduced in Stage 1 now selects partial functions $\gamma_t^k : \mathcal{P}^k \mapsto \mathcal{U}^k$, for $k = 1, 2, \ldots, K$. The state sufficient for input output mapping from the coordinator's perspective is given as $S_t := (X_{t-1}, P_t^{1:K})$ and the information state Π_t for the coordinator is

$$\Pi_t(s_t) := \mathbb{P}^{\psi}(S_t = s_t | C_t, \gamma_{1:t-1}^{1:K}). \tag{3.33}$$

Results analogous to Propositions III.2–III.5 can now be used to conclude the structural result of Theorem III.7 for general K.

3.2.3 Sequential Decomposition

In addition to obtaining the structural result of Theorem III.7, the coordinator's problem also allows us to write a dynamic program for finding the optimal control strategies as shown below. We first focus on the two controller case ($K = 2$) and then extend the result to general K.

Theorem III.8. *The optimal coordination strategy can be found by the following dynamic program: For $t = 1, \ldots, T$, define the functions $J_t : \mathcal{P}\{\mathcal{S}\} \mapsto \mathbb{R}$ as follows. For $\pi \in \mathcal{P}\{\mathcal{S}\}$ let*

$$J_T(\pi) = \inf_{\tilde{\gamma}^1, \tilde{\gamma}^2} \tilde{R}_T(\pi, \tilde{\gamma}^1, \tilde{\gamma}^2). \tag{3.34}$$

For $t = 1, \ldots, T-1$, and $\pi \in \mathcal{P}\{\mathcal{S}\}$ let

$$J_t(\pi) = \inf_{\tilde{\gamma}^1, \tilde{\gamma}^2} \Big[\tilde{R}_t(\pi, \tilde{\gamma}^1, \tilde{\gamma}^2) + \mathbb{E}\{J_{t+1}(\Pi_{t+1}) | \Pi_t = \pi, \gamma_t^{1:2} = \tilde{\gamma}^{1:2}\} \Big]. \tag{3.35}$$

The arg inf $(\gamma_t^{*,1}, \gamma_t^{*,2})$ *in the RHS of* $J_t(\pi)$ *is the optimal action for the coordinator at time* t *then* $\Pi_t = \pi$*. Thus,*

$$(\gamma_t^{*,1}, \gamma_t^{*,2}) = \phi_t^*(\pi)$$

The corresponding control strategy for Problem 1, given by (3.17) *is optimal for Problem 1.*

Proof. As in Theorem III.6, we use the fact that the coordinator's optimization problem can be viewed as a Markov decision problem with Π_t as the state of the Markov process. The dynamic program follows from standard results in Markov decision theory *Kumar and Varaiya* (1986a). The optimality of the corresponding control strategy for Problem 1 follows from the equivalence between the two problems. □

The dynamic program of Theorem III.8 can be extended to general K in a manner similar to Section 3.2.2.

3.2.4 Computational Aspects

In the dynamic program for the coordinator in Theorem III.8, the value functions at each time are functions defined on the continuous space $\mathcal{P}\{\mathcal{S}\}$, whereas the minimization at each time step is over the finite set of functions from the space of realizations of the private information of controllers ($\mathcal{P}^k$, $k = 1, 2$) to the space of control actions ($\mathcal{U}^k$, $k = 1, 2$). While dynamic programs with continuous state space can be hard to solve, we note that our dynamic program resembles the dynamic program for partially observable Markov decision problems (POMDP). In particular, just as in POMDP, the value-function at time T is piecewise linear in Π_T and by standard backward recursion, it can be shown that value-function at time t is piecewise linear and concave function of Π_t. (See Appendix B). Indeed, the coordinator's problem can be viewed as a POMDP, with S_t as the underlying partially observed state and the belief

Π_t as the information state of the POMDP. The characterization of value functions as piecewise linear and concave is utilized to find computationally efficient algorithms for POMDPs. Such algorithmic solutions to general POMDPs are well-studied and can be employed here. We refer the reader to *Zhang* (2009) and references therein for a review of algorithms to solve POMDPs.

3.2.5 One-step Delay

We now focus on the one-step delayed sharing information structure, i.e., when $n = 1$. For this case, the structural result (3.7) asserted by Witsenhausen is correct *Varaiya and Walrand* (1978). At first glance, that structural result looks different from our structural result (3.9) for $n = 1$. In this section, we show that for $n = 1$, these two structural results are equivalent.

As before, we consider the two-controller system ($K = 2$). When delay $n = 1$, we have

$$C_t = (Y^1_{1:t-1}, Y^2_{1:t-1}, U^1_{1:t-1}, U^2_{1:t-1}),$$
$$P^1_t = (Y^1_t), \quad P^2_t = (Y^2_t),$$

and

$$Z_{t+1} = (Y^1_t, Y^2_t, U^1_t, U^2_t).$$

The result of Theorem III.7 can now be restated for this case as follows:

Corollary III.9. *In Problem 1 with $K = 2$ and $n = 1$, without loss of optimality we can restrict attention to control strategies of the form:*

$$U^k_t = g^k_t(Y^k_t, \Pi_t), \quad k = 1, 2. \tag{3.36}$$

where

$$\Pi_t := \mathbb{P}^{(g^1_{1:t-1}, g^2_{1:t-1})}(X_{t-1}, Y^1_t, Y^2_t | C_t) \tag{3.37}$$

We can now compare our result for one-step delay with the structural result (3.7), asserted in *Witsenhausen* (1971) and proved in *Varaiya and Walrand* (1978). For $n = 1$, this result states that without loss of optimality, we can restrict attention to control laws of the form:

$$U^k_t = g^k_t(Y^k_t, \mathbb{P}(X_{t-1}|C_t)), \quad k = 1, 2. \tag{3.38}$$

The above structural result can be recovered from (3.37) by observing that there is a one-to-one correspondence between Π_t and the belief $\mathbb{P}(X_{t-1}|C_t)$. We first note that

$$\begin{aligned}\Pi_t &= \mathbb{P}^{(g^1_{1:t-1}, g^2_{1:t-1})}(X_{t-1}, Y^1_t, Y^2_t | C_t) \\ &= \mathbb{P}(Y^1_t | X_{t-1}) \cdot \mathbb{P}(Y^2_t | X_{t-1}) \\ &\quad \cdot \mathbb{P}^{(g^1_{1:t-1}, g^2_{1:t-1})}(X_{t-1} | C_t)\end{aligned} \tag{3.39}$$

As pointed out in *Witsenhausen* (1971); *Varaiya and Walrand* (1978) (and proved later in this chapter in Proposition III.11), the last probability does not depend on the functions $(g^1_{1:t-1}, g^2_{1:t-1})$. Therefore,

$$\Pi_t = \mathbb{P}(Y^1_t | X_{t-1}) \cdot \mathbb{P}(Y^2_t | X_{t-1}) \cdot \mathbb{P}(X_{t-1} | C_t) \tag{3.40}$$

Clearly, the belief $\mathbb{P}(X_{t-1}|C_t)$ is a marginal of Π_t and therefore can be evaluated from Π_t. Moreover, given the belief $\mathbb{P}(X_{t-1}|C_t)$, one can evaluate Π_t using equation (3.40). This one-to-one correspondence between Π_t and $\mathbb{P}(X_{t-1}|C_t)$ means that the structural result proposed in this chapter for $n = 1$ is effectively equivalent to the one proved in *Varaiya and Walrand* (1978).

3.3 Proof of the second structural result

In this section we prove the second structural result (3.10). As in Section 3.2, we prove the result for $K = 2$ and then show how to extend it for general K. To prove the result, we reconsider the coordinator's problem at Stage 3 of Section 3.2 and present an alternative characterization for the coordinator's optimal strategy in Problem 2. The main idea in this section is to use the dynamics of the system evolution and the observation equations (equations (3.1) and (3.2)) to find an equivalent representation of the coordinator's information state. We also contrast this information state with that proposed by Witsenhausen.

3.3.1 Two controller system ($K = 2$)

Consider the coordinator's problem with $K = 2$. Recall that γ_t^1 and γ_t^2 are the coordinator's actions at time t. γ_t^k maps the private information of the k^{th} controller $(Y_{t-n+1:t}^k, U_{t-n+1:t-1}^k)$ to its action U_t^k. In order to find an alternate characterization of coordinator's optimal strategy, we need the following definitions:

Definition III.10. For a coordination strategy $\boldsymbol{\psi}$, and for $t = 1, 2, \ldots, T$ we define the following:

1. $\Theta_t := \mathbb{P}(X_{t-n} | C_t)$

2. For $k = 1, 2$, define the following partial functions of γ_m^k

$$r_{m,t}^k(\cdot) := \gamma_m^k(\cdot, Y_{m-n+1:t-n}^k, U_{m-n+1:t-n}^k),$$
$$m = t-n+1, t-n+2, \ldots, t-1 \quad (3.41)$$

Since γ_m^k is a function that maps $(Y_{m-n+1:m}^k, U_{m-n+1:m-1}^k)$ to U_m^k, $r_{m,t}^k(\cdot)$ is a function that maps $(Y_{t-n+1:m}^k, U_{t-n+1:m-1}^k)$ to U_m^k. We define a collection of these partial functions as follows:

$$r_t^k := (r_{m,t}^k(\cdot), m = t-n+1, t-n+2, \ldots, t-1) \tag{3.42}$$

Note that for $n = 1$, r_t^k is empty.

We need the following results to address the coordinator's problem:

Proposition III.11. *1. For $t = 1, \ldots, T-1$, there exists functions Q_t, Q_t^k, $k = 1, 2$, (which do not depend on the coordinator's strategy) such that*

$$\begin{aligned} \Theta_{t+1} &= Q_t(\Theta_t, Z_{t+1}) \\ r_{t+1}^k &= Q_t^k(r_t^k, Z_{t+1}, \gamma_t^k) \end{aligned} \tag{3.43}$$

2. The coordinator's information state Π_t is a function of (Θ_t, r_t^1, r_t^2). Consequently, for $t = 1, \ldots, T$, there exist functions $\tilde{R}_t$ (which do not depend on the coordinator's strategy) such that

$$\mathbb{E}\{\hat{R}_t(S_t, \gamma_t^1, \gamma_t^2, S_{t+1}) | C_t, \Pi_{1:t}, \gamma_{1:t}^1, \gamma_{1:t}^2\} = \tilde{R}_t(\Theta_t, r_t^1, r_t^2, \gamma_t^1, \gamma_t^2) \tag{3.44}$$

3. The process (Θ_t, r_t^1, r_t^2), $t = 1, 2, \ldots, T$ is a controlled Markov chain with γ_t^1, γ_t^2 as the control actions at time t, i.e.,

$$\begin{aligned} &\mathbb{P}(\Theta_{t+1}, r_{t+1}^1, r_{t+1}^2 | C_t, \Theta_{1:t}, r_{1:t}^1, r_{1:t}^2, \gamma_{1:t}^1, \gamma_{1:t}^2) \\ &\quad = \mathbb{P}(\Theta_{t+1}, r_{t+1}^1, r_{t+1}^2 | \Theta_{1:t}, r_{1:t}^1, r_{1:t}^2, \gamma_{1:t}^1, \gamma_{1:t}^2) \\ &\quad = \mathbb{P}(\Theta_{t+1}, r_{t+1}^1, r_{t+1}^2 | \Theta_t, r_t^1, r_t^2, \gamma_t^1, \gamma_t^2). \end{aligned} \tag{3.45}$$

Proof. See Appendix B. □

At $t = 1$, since there is no sharing of information, Θ_1 is simply the unconditioned probability $\mathbb{P}(X_0)$. Thus, Θ_1 is fixed a priori from the joint distribution of the primitive random variables and does not depend on the choice of the coordinator's strategy ψ. Proposition 4 shows that the update of Θ_t depends only on Z_{t+1} and not on the coordinator's strategy. Consequently, the belief Θ_t depends only on the distribution of the primitive random variables and the realizations of $Z_{1:t}$. We can now show that the coordinator's optimization problem can be viewed as an MDP with (Θ_t, r_t^1, r_t^2), $t = 1, 2, \ldots, T$ as the underlying Markov process.

Theorem III.12. *(Θ_t, r_t^1, r_t^2) is an information state for the coordinator. That is, there is an optimal coordination strategy of the form:*

$$(\gamma_t^1, \gamma_t^2) = \psi_t(\Theta_t, r_t^1, r_t^2), \quad t = 1, \ldots, T. \tag{3.46}$$

Moreover, this optimal coordination strategy can be found by the following dynamic program:

$$J_T(\theta, \tilde{r}^1, \tilde{r}^2) = \inf_{\tilde{\gamma}^1, \tilde{\gamma}^2} \mathbb{E}\big\{ \tilde{R}_T(\Theta_T, r_T^1, r_T^2, \gamma_T^1, \gamma_T^2) \,\big|\, \Theta_T = \theta, r_T^{1:2} = \tilde{r}^{1:2}, \gamma_T^{1:2} = \tilde{\gamma}^{1:2} \big\}. \tag{3.47}$$

For $t = 1, \ldots, T-1$, let

$$J_t(\theta, \tilde{r}^1, \tilde{r}^2) = \inf_{\tilde{\gamma}^1, \tilde{\gamma}^2} \mathbb{E}\big\{ \tilde{R}_t(\Theta_t, r_t^1, r_t^2, \gamma_1^1, \gamma_t^2) + J_{t+1}(\Theta_{t+1}, r_{t+1}^1, r_{t+1}^2) \,\big|\, \Theta_t, = \theta, r_t^{1:2} = \tilde{r}^{1:2}, \gamma_t^{1:2} = \tilde{\gamma}^{1:2} \big\} \tag{3.48}$$

where $\theta \in \mathcal{P}(\mathcal{X})$, and $\tilde{r}^1, \tilde{r}^2$ are realizations of partial functions defined in (3.41) *and* (3.42)*. The arg inf* $(\gamma_t^{*,1}, \gamma_t^{*,2})$ *in the RHS of* (3.48) *is the optimal action for the coordinator at time* t *when* $(\Theta_t, r_t^1, r_t^2) = (\theta, \tilde{r}^1, \tilde{r}^2)$*. Thus,*

$$(\gamma_t^{*,1}, \gamma_t^{*,2}) = \psi_t^*(\Theta_t, r_t^1, r_t^2)$$

The corresponding control strategy for Problem 1, given by (3.17) *is optimal for Problem 1.*

Proof. Proposition III.11 implies that the coordinator's optimization problem can be viewed as an MDP with (Θ_t, r_t^1, r_t^2), $t = 1, 2, \ldots, T$ as the underlying Markov process and $\tilde{R}_t(\Theta_t, r_t^1, r_t^2, \gamma_t^1, \gamma_t^2)$ as the instantaneous cost. The MDP formulation implies the result of the theorem. □

The following result follows from Theorem III.12.

Theorem III.13 (Second Structural Result)**.** *In Problem 1 with* $K = 2$*, without loss of optimality we can restrict attention to coordination strategies of the form*

$$U_t^k = g_t^k(P_t^k, \Theta_t, r_t^1, r_t^2), \quad k = 1, 2. \tag{3.49}$$

where

$$\Theta_t = \mathbb{P}(X_{t-n} | C_t) \tag{3.50}$$

and

$$r_t^k = \{(g_m^k(\cdot, Y_{m-n+1:t-n}^k, U_{m-n+1:t-n}^k, C_m), t - n + 1 \leq m \leq t - 1\} \tag{3.51}$$

Proof. As in Theorem III.7, equations (3.17) can be used to identify an optimal control strategy for each controller from the optimal coordination strategy given in Theorem III.12. □

Theorem III.12 and Theorem III.13 can be easily extended for K controllers by identifying $(\Theta_t, r_t^{1:K})$ as the information state for the coordinator.

3.3.2 Comparison to Witsenhausen's Result

We now compare the result of Theorem III.12 to Witsenhausen's conjecture which states that there exist optimal control strategies of the form:

$$U_t^k = g_t^k(P_t^k, \mathbb{P}(X_{t-n}|C_t)). \tag{3.52}$$

Recall that Witsenhausen's conjecture is true for $n = 1$ but false for $n > 1$. Therefore, we consider the cases $n = 1$ and $n > 1$ separately:

Delay $n = 1$

For a two-controller system with $n = 1$, we have

$$C_t = (Y_{1:t-1}^1, Y_{1:t-1}^2, U_{1:t-1}^1, U_{1:t-1}^2),$$

$$P_t^1 = (Y_t^1), \quad P_t^2 = (Y_t^2),$$

and

$$r_t^1 = \emptyset, \quad r_t^2 = \emptyset$$

Therefore, for $n = 1$, Theorem III.13 implies that there exist optimal control strategies of the form:

$$U_t^k = g_t^k(P_t^k, \mathbb{P}(X_{t-n}|C_t)), \quad k = 1, 2. \tag{3.53}$$

Equation (3.53) is the same as equation (3.52) for $n = 1$. Thus, for $n = 1$, the result of Theorem III.12 coincides with Witsenhausen's conjecture which was proved in *Varaiya and Walrand* (1978).

Delay $n > 1$

Witsenhausen's conjecture implied that the controller k at time t can choose its action based only on the knowledge of P_t^k and $\mathbb{P}(X_{t-n}|C_t)$, without any dependence on the choice of previous control laws ($g_{1:t-1}^{1:2}$). In other words, the argument of the control law g_t^k (that is, the information state at time t) is separated from $g_{1:t-1}^{1:2}$. However, as Theorem III.13 shows, such a separation is not true because of the presence of the collection of partial functions r_t^1, r_t^2 in the argument of the optimal control law at time t. These partial functions depend on the choice of previous $n-1$ control laws. Thus, the argument of g_t^k depends on the choice of $g_{t-n+1:t-1}^{1:2}$. One may argue that Theorem III.13 can be viewed as a *delayed or partial* separation since the information state for the control law g_t^k is separated from the choice of control laws before time $t-n+1$.

Witsenhausen's conjecture implied that controllers employ common information only to form a belief on the state X_{t-n}; the controllers do not need to use the common information to guess each other's behavior from $t-n+1$ to the current time t. Our result disproves this statement. We show that in addition to forming the belief on X_{t-n}, each controller should use the common information to predict the actions of other controllers by means of the partial functions r_t^1, r_t^2.

3.4 A Special Case of Delayed Sharing Information Structure

Many decentralized systems consist of coupled subsystems, where each subsystem has a controller that perfectly observes the state of the subsystem. If all controllers can exchange their observations and actions with a delay of n steps, then the system is a special case of the n-step delayed sharing information structure with the following assumptions:

1. *Assumption 1:* At time $t = 1, \dots, T$, the state of the system is given as the

vector $X_t := (X_t^{1:K})$, where X_t^i is the state of subsystem i.

2. *Assumption 2:* The observation equation of the k^{th} controller is given as:

$$Y_t^k = X_{t-1}^k \tag{3.54}$$

This model is similar to the model considered in *Aicardi et al.* (1987). Clearly, the first structural result and the sequential decomposition of Section 3.2 apply here as well with the observations Y_t^k being replaced by X_t^k. Our second structural result simplifies when specialized to this model. Observe that in this model

$$C_t = (Y_{1:t-n}^{1:K}, U_{1:t-n}^{1:K}) = (X_{1:t-n-1}, U_{1:t-n}^{1:K}) \tag{3.55}$$

and therefore the belief,

$$\Theta_t = \mathbb{P}(X_{t-n}|C_t) = \mathbb{P}(X_{t-n}|X_{t-n-1}, U_{t-n}^{1:K}) \tag{3.56}$$

where we used the controlled Markov nature of the system dynamics in the second equality in (3.56). Thus, Θ_t is a function only of $X_{t-n-1}, U_{t-n}^{1:K}$. The result of Theorem III.12 can now be restated for this case as follows:

Corollary III.14. *In Problem 1 with assumptions 1 and 2, there is an optimal coordination strategy of the form:*

$$(\gamma_t^1, \gamma_t^2) = \psi_t(X_{t-n-1}, U_{t-n}^1, U_{t-n}^2, r_t^1, r_t^2), \quad t = 1, \ldots, T. \tag{3.57}$$

Moreover, this optimal coordination strategy can be found by the following dynamic *program:*

$$
\begin{aligned}
&J_T(x, u^1, u^2, \tilde{r}^1, \tilde{r}^2) \\
&\qquad = \inf_{\tilde{\gamma}^1, \tilde{\gamma}^2} \mathbb{E}\big\{ \tilde{R}_T(X_{T-n}, r^1_T, r^2_T, \gamma^1_T, \gamma^2_T) \,\big|\, X_{T-n-1} = x, \\
&\qquad\qquad U^{1:2}_{T-n} = u^{1:2}, r^{1:2}_T = \tilde{r}^{1:2}, \gamma^{1:2}_T = \tilde{\gamma}^{1:2} \big\}. \qquad (3.58)
\end{aligned}
$$

For $t = 1, \dots, T-1$, *let*

$$
\begin{aligned}
J_t(x, u^1, u^2, \tilde{r}^1, \tilde{r}^2) &= \inf_{\tilde{\gamma}^1, \tilde{\gamma}^2} \mathbb{E}\{ \tilde{R}_t(X_{t-n}, r^1_t, r^2_t, \gamma^1_1, \gamma^2_t) \\
&\qquad + J_{t+1}(X_{t-n+1}, r^1_{t+1}, r^2_{t+1}) \,\big|\, X_{t-n-1} = x, \\
&\qquad U^{1:2}_{t-n} = u^{1:2}, r^{1:2}_t = \tilde{r}^{1:2}, \gamma^{1:2}_t = \tilde{\gamma}^{1:2} \big\}. \qquad (3.59)
\end{aligned}
$$

We note that the structural result and the sequential decomposition in the corollary above is analogous to (*Aicardi et al.*, 1987, Theorem 1).

3.4.1 An Example

We consider a simple example of a delayed sharing information structure with two controllers ($K = 2$), a time horizon $T = 3$ and delay $n = 2$. Varaiya and Walrand *Varaiya and Walrand* (1978) used this example to show that Witsenhausen's proposed structure was suboptimal.

The system dynamics are given by

$$
\begin{aligned}
X_0 &= (X_0^1, X_0^2)\\
X_1 &= (X_1^1, X_1^2) = (X_0^1 + X_0^2, 0)\\
X_2 &= (X_2^1, X_2^2) = (X_1^1, U_2^2) = (X_0^1 + X_0^2, U_2^2)\\
X_3 &= (X_3^1, X_3^2) = (X_2^1 - X_2^2 - U_3^1, 0)\\
&= (X_0^1 + X_0^2 - U_2^2 - U_3^1, 0)
\end{aligned}
$$

X_0^1, X_0^2 are zero-mean, jointly Gaussian random variables with variance 1 and covariance -0.5. The observation equations are:

$$Y_t^k = X_{t-1}^k$$

and the total cost function is

$$\mathcal{J}(\boldsymbol{g}) = \mathbb{E}^{\boldsymbol{g}}\{(X_3^1)^2 + (U_3^1)^2\} \tag{3.60}$$

We can now specify the common and private informations. *Common Information:*

$$
\begin{gathered}
C_1 = \emptyset, \quad C_2 = \emptyset,\\
C_3 = \{Y_1^1, Y_1^2, U_1^1, U_1^2\} = \{X_0^1, X_0^2, U_1^1, U_1^2\}
\end{gathered}
$$

Private Information for Controller 1:

$$
\begin{aligned}
P_1^1 &= \{Y_1^1\} = \{X_0^1\},\\
P_2^1 &= \{Y_1^1, Y_2^1, U_1^1\} = \{X_0^1, (X_0^1 + X_0^2), U_1^1\}\\
P_3^1 &= \{Y_2^1, Y_3^1, U_2^1\} = \{(X_0^1 + X_0^2), (X_0^1 + X_0^2), U_2^1\}
\end{aligned}
$$

Private Information for Controller 2:

$$
\begin{aligned}
P_1^2 &= \{Y_1^2\} = \{X_0^2\}, \\
P_2^2 &= \{Y_1^2, Y_2^2, U_1^2\} = \{X_0^2, 0, U_1^2\} \\
P_3^2 &= \{Y_2^2, Y_3^2, U_2^2\} = \{0, U_2^2, U_2^2\}
\end{aligned}
$$

The total cost can be written as:

$$
\mathcal{J}(\boldsymbol{g}) = \mathbb{E}^{\boldsymbol{g}}\{(X_0^1 + X_0^2 - U_2^2 - U_3^1)^2 + (U_3^1)^2\} \tag{3.61}
$$

Thus, the only control actions that affect the cost are U_2^2 and U_3^1 . Hence, we can even assume that all other control laws are constant functions with value 0 and the performance of a design is completely characterized by control laws g_2^2 and g_3^1. Using the fact that all control actions other than U_2^2 and U_3^1 are 0, we get the following simplified control laws:

$$
\begin{aligned}
U_2^2 &= g_2^2(P_2^2, C_2) = g_2^2(X_0^2) \\
U_3^1 &= g_3^1(P_3^1, C_3) = g_3^1((X_0^1 + X_0^2), X_0^1, X_0^2) \\
&= g_3^1(X_0^1, X_0^2)
\end{aligned}
$$

Now consider control laws of the form given in Theorem III.13 given by

$$
U_t^k = g_t^k(P_t^k, \Theta_t, r_t^1, r_t^2) \tag{3.62}
$$

For $k = 2$ and $t = 2$, Θ_2 is a fixed prior distribution of X_0, while r_2^1, r_2^2 are constant functions. Hence, Θ_t, r_t^1, r_t^2 provide no new information and the structure of equation (3.62) boils down to

$$
U_2^2 = g_2^2(P_2^2) = g_2^2(X_0^2) \tag{3.63}
$$

For $k = 1$ and $t = 3$,

$$\Theta_3 = \mathbb{P}(X_1|C_3) = \mathbb{P}((X_0^1 + X_0^2, 0)|X_0^1, X_0^2)$$

and

$$\begin{aligned} r_3^2 &= \{(g_m^2(\cdot, Y_{m-1:1}^2, U_{m-1:1}^2, C_m), 2 \leq m \leq 2\} \\ &= \{(g_2^2(\cdot, Y_1^2, U_1^2, C_2)\} \\ &= \{(g_2^2(\cdot, X_0^2)\} = U_2^2, \end{aligned}$$

while r_3^1 are partial functions of constant functions. Therefore, equation (3.62) can now be written as:

$$\begin{aligned} U_3^1 &= g_3^1((X_0^1 + X_0^2), \mathbb{P}((X_0^1 + X_0^2, 0)|X_0^1, X_0^2), U_2^2) \\ &= g_3^1((X_0^1 + X_0^2), (X_0^1 + X_0^2), U_2^2) && (3.64) \\ &= g_3^1((X_0^1 + X_0^2), U_2^2) && (3.65) \end{aligned}$$

where we used the fact that knowing $\mathbb{P}((X_0^1 + X_0^2, 0)|X_0^1, X_0^2)$ is same as knowing the value of $(X_0^1 + X_0^2)$ in (3.64).

The optimal control laws can be obtained by solving the coordinator's dynamic program given in Theorem III.12. Observe that $\Theta_3 = \mathbb{P}((X_0^1 + X_0^2, 0)|X_0^1, X_0^2)$ is equivalent to $(X_0^1 + X_0^2)$ and that r_3^2 is equivalent to U_2^2. Thus, the dynamic program can be simplified to:

$$\begin{aligned} &J_3((x_0^1 + x_0^2), u_2^2) \\ &= \inf_{\tilde{\gamma}^1} \mathbb{E}\{(X_3^1)^2 + (U_3^1)^2 | \begin{array}{c} (X_0^1 + X_0^2) = (x_0^1 + x_0^2), \\ U_2^2 = u_2^2, \gamma_3^1 = \tilde{\gamma}^1 \end{array} \} \end{aligned}$$

where, for the given realization of $((x_0^1 + x_0^2), u_2^2)$, $\tilde{\gamma}^1$ maps $P_3^1 = (X_0^1 + X_0^2)$ to U_3^1. Further simplification yields:

$$
\begin{aligned}
&J_3((x_0^1 + x_0^2), u_2^2) \\
&= \inf_{\tilde{\gamma}^1} \mathbb{E}\Big\{ (X_0^1 + X_0^2 - U_2^2 - U_3^1)^2 + (U_3^1)^2 \ \Big| \\
&\qquad (X_0^1 + X_0^2) = (x_0^1 + x_0^2), U_2^2 = u_2^2, \gamma_3^1 = \tilde{\gamma}^1 \Big\} \\
&\geq (x_0^1 + x_0^2 - u_2^2)^2/2, \qquad (3.66)
\end{aligned}
$$

where the right hand side in (3.66) is the lower bound on the expression $(x_0^1 + x_0^2 - u_2^2 - u_3^1)^2 + (u_3^1)^2$ for any u_3^1. Given the fixed realization of $((x_0^1 + x_0^2), u_2^2)$, choosing γ^1 as a constant function with value $(x_0^1 + x_0^2 - u_2^2)/2$ achieves the lower bound in (3.66). For $t = 2$, the coordinator has no information and the value function at time $t = 2$ is

$$
\begin{aligned}
J_2 &= \inf_{\tilde{\gamma}^2} \mathbb{E}\{J_3((X_0^1 + X_0^2), U_2^2) | \gamma_2^2 = \tilde{\gamma}^2\} \\
&= \inf_{\tilde{\gamma}^2} \mathbb{E}\{(X_0^1 + X_0^2 - U_2^2)^2/2 | \gamma_2^2 = \tilde{\gamma}^2\} \qquad (3.67)
\end{aligned}
$$

where $\tilde{\gamma}^2$ maps $P_2^1 = (X_0^2)$ to U_2^2. The optimization problem in (3.67) is to choose, for each value of x_0^2, the best estimate (in a mean squared error sense) of $(X_0^1 + X_0^2)$. Given the Gaussian statistics, the optimal choice of $\tilde{\gamma}^2$ can be easily shown to be $\gamma^2(x_0^2) = x_0^2/2$.

Thus, the optimal strategy for the coordinator is to choose $\gamma_2^2(x_0^2) = x_0^2/2$ at time $t = 2$, and at $t = 3$, given the fixed realization of $((x_0^1 + x_0^2), u_2^2)$, choose $\gamma^1(\cdot) =$

$(x_0^1 + x_0^2 - u_2^2)/2$. Thus, the optimal control laws are:

$$U_2^2 = g_2^2(X_0^2) = X_0^2/2 \tag{3.68}$$

$$\begin{aligned} U_3^1 &= g_3^1((X_0^1 + X_0^2), U_2^2) \\ &= (X_0^1 + X_0^2 - U_2^2)/2 \end{aligned} \tag{3.69}$$

These are same as the unique optimal control laws identified in *Varaiya and Walrand* (1978).

3.5 Kurtaran's Separation Result

In this section, we focus on the structural result proposed by Kurtaran *Kurtaran* (1979). We restrict to the two controller system ($K = 2$) and delay $n = 2$. For this case, we have

$$C_t = (Y_{1:t-2}^1, Y_{1:t-2}^2, U_{1:t-2}^1, U_{1:t-2}^2),$$
$$P_t^1 = (Y_t^1, Y_{t-1}^1, U_{t-1}^1), \quad P_t^2 = (Y_t^2, Y_{t-1}^2, U_{t-1}^2),$$

and

$$Z_{t+1} = (Y_{t-1}^1, Y_{t-1}^2, U_{t-1}^1, U_{t-1}^2).$$

Kurtaran's structural result for this case states that without loss of optimality we can restrict attention to control strategies of the form:

$$U_t^k = g_t^k(P_t^k, \Phi_t), \quad k = 1, 2, \tag{3.70}$$

where

$$\Phi_t \coloneqq \mathbb{P}^{\boldsymbol{g}}\{X_{t-2}, U_{t-1}^1, U_{t-1}^2 | C_t\}.$$

Kurtaran *Kurtaran* (1979) proved this result for the terminal time-step T and simply stated that the result for $t = 1, \ldots, T-1$ can be established by the dynamic programming argument given in *Kurtaran* (1976). We believe that this is not the case.

In the dynamic programming argument in *Kurtaran* (1976), a critical step is the update of the information state Φ_t, which is given by (*Kurtaran*, 1976, Eq (30)). For the result presented in *Kurtaran* (1979), the corresponding equation is

$$\Phi_{t+1} = F_t(\Phi_t, Y^1_{t-1}, Y^2_{t-1}, U^1_{t-1}, U^2_{t-1}). \tag{3.71}$$

We believe that such an update equation cannot be established.

To see the difficulty in establishing (3.71), lets follow an argument similar to the proof of (*Kurtaran*, 1976, Eq (30)) given in (*Kurtaran*, 1976, Appendix B). For a fixed strategy $\boldsymbol{g}$, and a realization c_{t+1} of C_{t+1}, the realization φ_{t+1} of Φ_{t+1} is given by

$$\begin{aligned}
\varphi_{t+1} &= \mathbb{P}(x_{t-1}, u^1_t, u^2_t | c_{t+1}) \\
&= \mathbb{P}(x_{t-1}, u^1_t, u^2_t | c_t, y^1_{t-1}, y^2_{t-1}, u^1_{t-1}, u^2_{t-1}) \\
&= \frac{\mathbb{P}(x_{t-1}, u^1_t, u^2_t, y^1_{t-1}, y^2_{t-1}, u^1_{t-1}, u^2_{t-1} | c_t)}{\displaystyle\sum_{(x', a^1, a^2) \in \mathcal{X} \times \mathcal{U}^1 \times \mathcal{U}^2} \mathbb{P}(X_{t-1} = x', U^1_t = a^1, U^2_t = a^2, y^1_{t-1}, y^2_{t-1}, u^1_{t-1}, u^2_{t-1} \mid c_t)}
\end{aligned} \tag{3.72}$$

The numerator can be expressed as:

$$\begin{aligned}
&\mathbb{P}(x_{t-1}, u^1_t, u^2_t, y^1_{t-1}, y^2_{t-1}, u^1_{t-1}, u^2_{t-1} | c_t) \\
&\quad = \sum_{(x_{t-2}, y^1_t, y^2_t) \in \mathcal{X} \cdot \mathcal{Y}^1 \cdot \mathcal{Y}^2} \mathbb{P}(x_{t-1}, u^1_t, u^2_t, y^1_{t-1}, y^2_{t-1}, u^1_{t-1}, u^2_{t-1}, x_{t-2}, y^1_t, y^2_t | c_t)
\end{aligned}$$

$$
\begin{aligned}
&= \sum_{(x_{t-2}, y_t^1, y_t^2) \in \mathcal{X} \cdot \mathcal{Y}^1 \cdot \mathcal{Y}^2} \mathbb{1}_{g_t^1(c_t, u_{t-1}^1, y_{t-1}^1, y_t^1)}[u_t^1] \cdot \mathbb{1}_{g_t^2(c_t, u_{t-1}^2, y_{t-1}^2, y_t^2)}[u_t^2] \\
&\qquad \cdot \mathbb{P}(y_t^1 | x_{t-1}) \cdot \mathbb{P}(y_t^2 | x_{t-1}) \cdot \mathbb{P}(x_{t-1} | x_{t-2}, u_{t-1}^1, u_{t-1}^2) \\
&\qquad \cdot \mathbb{1}_{g_{t-1}^1(c_{t-1}, u_{t-2}^1, y_{t-2}^1, y_{t-1}^1)}[u_{t-1}^1] \cdot \mathbb{1}_{g_t^2(c_{t-1}, u_{t-2}^2, y_{t-2}^2, y_{t-1}^2)}[u_{t-2}^2] \\
&\qquad \cdot \mathbb{P}(y_{t-1}^1 | x_{t-2}) \cdot \mathbb{P}(y_{t-1}^2 | x_{t-2}) \cdot \mathbb{P}(x_{t-2} | c_t)
\end{aligned} \tag{3.73}
$$

If, in addition to φ_t, y_{t-1}^1, y_{t-1}^2, u_{t-1}^1, and u_{t-1}^2, each term of (3.73) depended only on terms that are being summed over (x_{t-2}, y_t^1, y_t^2), then (3.73) would prove (3.71). However, this is not the case: the first two terms also depend on c_t. Therefore, the above calculation shows that φ_{t+1} is a function of $\varphi_t, Y_{t-1}^1, Y_{t-1}^2, U_{t-1}^1, U_{t-1}^2$ *and* c_t. This dependence on c_t is not an artifact of the order in which we decided to use the chain rule in (3.73) (we choose the natural sequential order in the system). No matter how we try to write φ_{t+1} in terms of φ_t, there will be a dependence on c_t.

The above argument shows that it is not possible to establish (3.71). Consequently, the dynamic programming argument presented in *Kurtaran* (1976) breaks down when working with the information state of *Kurtaran* (1979), and, hence, the proof in *Kurtaran* (1979) is incomplete. So far, we have not been able to correct the proof or find a counterexample to it.

3.6 Conclusion

We studied the stochastic control problem with n-step delay sharing information structure and established two structural results for it. Both the results characterize optimal control laws with time-invariant domains. Our second result also establishes a partial separation result, that is, it shows that the information state at time t is separated from choice of laws before time $t - n + 1$. Both the results agree with Witsenhausen's conjecture for $n = 1$. To derive our structural results, we formulated an alternative problem from the point of a coordinator of the system that knows the common information among the controllers. In the subsequent chapters, we will extend this idea of formulating an alternative problem from the point of view of a coordinator which has access to common information for general sequential decision making problems.

CHAPTER IV

Common Information and the General Sequential Decision-Making Problems

4.1 General Sequential Decision-Making Problems

In this Chapter, we consider a general model of sequential decision-making problems with finite number of decisions. We borrow the general model of a sequential decision making problem from *Witsenhausen* (1973). We impose the assumption of finiteness of the underlying probability space and finiteness of all observation and decision spaces. Other than the assumptions of finiteness, the following model is general enough to capture any sequential decision-making problem. As pointed out in *Witsenhausen* (1973), *any sequential stochastic control problem can be reduced to our model by elimination of intermediate random variables.*

We define the concepts of common information and private information for the general problem. We then investigate the general sequential problem from the perspective of a coordinator who knows the common information and who has to provide prescriptions to each decision-maker on how to use their private information. We show that the coordinator's perspective allows us to formulate an equivalent decision-making problem where the coordinator is the only decision maker (acting at multiple times). The coordinator's problem is a sequential problem with classical information structure for which we can identify an information state and a dynamic program to find the optimal prescriptions that the coordinator prescribes to each decision-maker. The coordinator's dynamic program implies a sequential decomposition for the general sequential decision-making problem.

The result of this chapter show that the common information methodology used in Chapters II,III can be extended to **any sequential decision-making problem** with finite spaces of observation and decisions and finite number of decisions. To the best of our knowledge, the only other result that provides a sequential decomposition for the general model is that by *Witsenhausen* (1973). Our result, however, is the first to use the concept of common information. Thus, our sequential decomposition is different from Witsehausen's. Our sequential decomposition coincides with Witsenhausen's only in problems where common information is absent.

Further, our sequential decomposition specializes to the classical dynamic program if the decision-making problem has a classical information structure. Thus, our common information methodology shows that the classical dynamic programming results of classical information structures and Witsenhausen's sequential decomposition for non-classical information structures both follow from the **same conceptual framework of common information** in sequential decision-making problems.

The rest of this chapter is organized as follows. We present a general model of sequential decision making problems in Section 4.1.1. We define private and common information for this model in Section 4.1.2 and present a sequential decomposition based on common information in Section 4.1.3. We consider two special cases of sequential problems in Section 4.2 and conclude in Section 4.3.

Remark IV.1. When dealing with collections of random variables, we will at times treat the collection as a random vector of appropriate dimension. At other times, it will be convenient to think of different collections of random variables as sets on which one can define the usual set operations. For example consider random vectors

$A = (A_1, A_2, A_3)$ and $\tilde{A} = (A_1, A_2)$. Then, treating A and $\tilde{A}$ as sets would allow us to write $A \cap \tilde{A} = \{A_1, A_2\}$.

4.1.1 The Model

The model we study consists of the following components:

Components of a general sequential decision making problem:

1. A finite probability space $(\Omega, \mathcal{B}, \mathbb{P})$.

2. Primitive Random Variables: A collection $Q = \{Q^1, Q^2, \ldots, Q^N\}$ defined on the above probability space (where N is a finite positive number). We will treat Q as a random vector that takes values in a finite space $\mathcal{Q}$ endowed with the power-set sigma algebra $2^{\mathcal{Q}}$.

3. A finite number T of decision-makers (DMs) that make decisions in a pre-determined sequence. Since the order of decisions is fixed, we may think of each decision being made at the tick of a discrete time clock that runs from the initial time $t = 1$ to the final time $t = T$. Any correspondence with physical time is irrelevant for the decision problem.

4. Decision Spaces: For $t = 1, 2, \ldots, T$, $(\mathcal{U}_t, 2^{\mathcal{U}_t})$ are finite measurable spaces.

5. Observation Spaces and Observation Maps: For $t = 1, 2, \ldots, T$, N_t is the finite number of observations available for making the t^{th} decision. We have the following finite observation space

$$\mathcal{I}_t = \mathcal{Y}_t^1 \times \mathcal{Y}_t^2 \times \ldots \times \mathcal{Y}_t^{N_t}$$

equipped with the power-set sigma algebra.

Further, we define the collection of observation maps

$$H_t = (h_t^1, h_t^2, \ldots, h_t^{N_t}),$$

where for each $j = 1, 2, \ldots, N_t$, h_t^j is a measurable map from $(\mathcal{Q} \times \mathcal{U}_1 \times \mathcal{U}_2 \times \ldots \times \mathcal{U}_{t-1}, 2^{\mathcal{Q}} \times 2^{\mathcal{U}_1} \times \ldots \times 2^{\mathcal{U}_{t-1}})$ to $(\mathcal{Y}_t^j, 2^{\mathcal{Y}_t^j})$.

6. Reward: R is a real-valued function defined on the measurable space $(\mathcal{Q} \times \mathcal{U}_1 \times \mathcal{U}_2 \times \ldots \times \mathcal{U}_T, 2^{\mathcal{Q}} \times 2^{\mathcal{U}_1} \times \ldots \times 2^{\mathcal{U}_T})$.

7. Decision Strategy Space: The decision strategy space $\mathcal{G}_t$ is the set of all measurable mappings from $(\mathcal{I}_t, 2^{\mathcal{I}_t})$ to $(\mathcal{U}_t, 2^{\mathcal{U}_t})$. The space of strategy profiles is $\mathcal{G} := \mathcal{G}_1 \times \ldots \times \mathcal{G}_T$.

Description of the Decision-Making Problem:

The above components of the sequential decision making problem can be interpreted as follows. The primitive random variables $Q = \{Q^1, Q^2, \ldots, Q^N\}$ are variables chosen by nature over which the designer has no control. In other words, the statistics and the realizations of these random variables are independent of the choice of decision strategy profile. The information available to the first decision-maker is a random vector I_1 that takes values in the observation space $\mathcal{I}_1$. Further, $I_1 = (Y_1^1, Y_1^2, \ldots, Y_1^{N_1})$, where for each $j = 1, \ldots, N_1$, Y_1^j is related to the primitive random variables via an observation map h_1^j. That is,

$$Y_1^j = h_1^j(Q) \tag{4.1}$$

The first decision maker uses a decision strategy $g_1 \in \mathcal{G}_1$ to map its information to its decision. Thus,

$$U_1 = g_1(I_1) \tag{4.2}$$

Subsequently, the information available to the second decision-maker is a random vector I_2 that takes values in the observation space $\mathcal{I}_2$. Further, $I_2 = (Y_2^1, Y_2^2, \ldots, Y_2^{N_2})$, where for each $j = 1, \ldots, N_2$, Y_2^j is related to the primitive random variables and U_1 via an observation map h_2^j. That is,

$$Y_2^j = h_2^j(Q, U_1) \tag{4.3}$$

The second decision maker uses a decision strategy $g_2 \in \mathcal{G}_2$ to map its information to its decision. Thus,

$$U_2 = g_2(I_2) \tag{4.4}$$

Proceeding sequentially, the t^{th} decision-maker's information is a random vector $I_t = (Y_t^1, Y_t^2, \ldots, Y_t^{N_t})$ that takes values in the observation space $\mathcal{I}_t$ and is given by the observation maps

$$Y_t^j = h_t^j(Q, U_1, U_2, \ldots, U_{t-1}), \tag{4.5}$$

for $j = 1, 2, \ldots, N_t$. The t^{th} decision maker uses a decision strategy $g_t \in \mathcal{G}_t$ to map its information to its decision. Thus,

$$U_t = g_t(I_t) \tag{4.6}$$

The collection $\boldsymbol{g} = (g_1, g_2, \ldots, g_T)$ is referred to as a *decision strategy profile*, while the function g_t is referred to as a *decision strategy/decision rule* at time t. Given a choice of decision strategy, the decisions $U_1, U_2, \ldots U_T$, the observation vectors $I_1, I_2, \ldots I_T$ as well as the reward $R(Q, U_1, U_2, \ldots, U_T)$ are well-defined random variables.

The value of a decision strategy profile $\boldsymbol{g}$ is defined as

$$\mathcal{J}(\boldsymbol{g}) = \mathbb{E}^{\boldsymbol{g}}[R(Q, U_1, U_2, \ldots, U_T)] \tag{4.7}$$

We can now formulate the following problem:

Problem 3. *For the general model of sequential decision making problems described above, choose a decision strategy profile $\boldsymbol{g} \in \mathcal{G}$ in order to maximize the value $\mathcal{J}(\boldsymbol{g})$ given by equation (4.7).*

4.1.2 Private and Common Information

For each decision-maker, we have a collection of observation maps $H_t = (h_t^1, h_t^2, \ldots, h_t^{N_t})$, where for each $j = 1, 2, \ldots, N_t$, h_t^j is a measurable map from $(\mathcal{Q} \times \mathcal{U}_1 \times \mathcal{U}_2 \times \ldots \times \mathcal{U}_{t-1}, 2^{\mathcal{Q}} \times 2^{\mathcal{U}_1} \times \ldots \times 2^{\mathcal{U}_{t-1}})$ to $(\mathcal{Y}_t^j, 2^{\mathcal{Y}_t^j})$. These observation maps represent information about primitive random variables and preceding decisions that is available to t^{th} decision-maker.

Definition IV.2 (Equivalent Observation Maps)**.** Consider the decision makers at time t and t', with $t < t'$. We will say that two observation maps h_t^j and $h_{t'}^k$ are equivalent if $\mathcal{Y}_t^j = \mathcal{Y}_{t'}^k$ and for any $q \in \mathcal{Q}$, and $(u_1, u_2, \ldots, u_{t'-1}) \in \mathcal{U}_1 \times \mathcal{U}_2 \ldots \times \mathcal{U}_{t'-1}$,

$$h_t^j(q, u_1, u_2, \ldots, u_{t-1}) = h_{t'}^k(q, u_1, u_2, \ldots, u_{t'-1})$$

Because equivalent observation maps provide the same information, we will treat them as identical observation maps. With the above equivalence between observation maps of different decision-makers, we will now define the common observation maps at each time.

Definition IV.3 (Common Observation Maps)**.** At each time t, we define H_t^{common} to be the collection of all observation maps that are available to all decision makers that act at or after time t. That is,

$$H_t^{common} := \cap_{k \geq t} H_k, \tag{4.8}$$

where we use the equivalence of observation maps defined in Definition IV.2 to identify identical observation maps in the right hand side of equation 4.8. We refer to H_t^{common} as the collection of common observation maps.

Definition IV.4 (Private Observation Maps)**.** At each time t, we define $H_t^{private}$ to be the collection of all observation maps that is available to the t^{th} decision maker but not to all decision makers that act after the t^{th} decision-maker. That is,

$$H_t^{private} := H_t \setminus H_t^{common} \tag{4.9}$$

We refer to $H_t^{private}$ as the collection of private observation maps.

Definition IV.5 (Common Information)**.** When a decision strategy profile has been chosen, the common observation maps will result in observations that are available to the t^{th} decision maker as well as all decision makers that act after it. We define the collection of these commonly known observations as the **common information at the t^{th} decision, and we denote it by C_t.** In terms of random variables that are observable to different decision makers, we have

$$C_t = \cap_{k \geq t} I_k, \tag{4.10}$$

where I_k is interpreted as a collection of random variables. We denote by $\mathcal{C}_t$ the set of all possible values of the random vector C_t.

Definition IV.6 (Private Information)**.** When a decision strategy profile has been chosen, the private observation maps will result in observations that are available to the t^{th} decision maker but not to all decision makers that act after it. We define the collection of these observations as the **private information at the t^{th} decision, and we denote it by P_t.** In terms of random variables that are observable to different decision makers, we have

$$P_t = I_t \setminus C_t, \tag{4.11}$$

where I_t, C_t is interpreted as collections of random variables. We denote by $\mathcal{P}_t$ the set of all possible values of the random vector P_t.

Remark IV.7. By definition, we have that $C_1 \subseteq C_2 \subseteq \ldots \subseteq C_T$. We can define $Z_t = C_t \setminus C_{t-1}$ as the increase in common information from time $t-1$ to t.

Remark IV.8. From the definitions of common and private information, any decision rule of the form $U_t = g_t(I_t)$ can be written as $U_t = g_t(P_t, C_t)$.

4.1.3 A Sequential Decomposition

In this section, we provide a sequential decomposition for the problem of obtaining an optimal decision strategy profile for Problem 3. We follow the same philosophy as in the delayed sharing problem. We proceed as follows:

1. First, we formulate a centralized stochastic control problem from the point of view of a coordinator who observes the common information C_t at each time t, but does not observe the private information P_t. The coordinator's problem is to provide prescriptions to each decision-maker on how to map its private information to its decision.

2. Next, we argue that for any strategy for the coordinator's problem there is an equivalent decision strategy profile in the original problem that achieves the same expected reward and vice versa. Hence, the two problems are equivalent.

3. Finally, we transform the coordinator's problem into a MDP (Markov decision process), and obtain a sequential decomposition for the coordinator's problem.

This sequential decomposition is also a sequential decomposition for the original problem due to the equivalence between the two problems.

Below, we elaborate on each of these stages.

Stage 1

We consider the following modified problem. In the model described in Section 4.1.1, we introduce a coordinator that knows the common information C_t at each time t. At time t, the coordinator decides the *partial decision rule*

$$\gamma_t : \mathcal{P}_t \mapsto \mathcal{U}_t$$

In case $\mathcal{P}_t$ is empty, γ_t is simply an element of the set $\mathcal{U}_t$.

The partial decision rule is to be interpreted as a prescription from the coordinator to the t^{th} decision maker informing it on how to use its private information to make its decision. In the case where there is no private information (that is, $I_t = C_t$), the prescription from the coordinator will simply be the decision for the t^{th} decision maker.

The choice of the partial decision rule at time t is based on the realization of the common information and the partial decision rules selected before time t. The coordinator then informs all decision makers that act at or after time t of the partial decision rule it selected at time t. The decision maker acting at time t then uses its assigned partial decision rule to generate a control action as follows.

$$U_t = \gamma_t(P_t). \tag{4.12}$$

In the above formulation, the only decision maker is the coordinator: the individual decision makers simply carry out the necessary evaluations prescribed by (4.12). At time t, the coordinator knows the common information C_t and all past partial decision rules $\gamma_{1:t-1}$. The coordinator uses a decision rule ψ_t to map this information to its decision, that is,

$$\gamma_t = \psi_t(C_t, \gamma_{1:t-1}). \tag{4.13}$$

The choice of $\boldsymbol{\psi} = \{\psi_t; t = 1, \dots, T\}$ is called a *coordination strategy.* Ψ denotes the class of all possible coordination strategies. The performance of a coordinating strategy is given by the expected total reward under that strategy, that is,

$$\hat{\mathcal{J}}(\boldsymbol{\psi}) = \mathbb{E}^{\boldsymbol{\psi}}\{R(Q, U_1, U_2, \dots, U_T)\} \tag{4.14}$$

where the expectation is with respect to the joint measure on all the system variables induced by the choice of $\boldsymbol{\psi}$. The coordinator has to solve the following optimization problem.

Problem 4 (The Coordinator's Optimization Problem)**.** *For the coordinator's problem described above, choose a coordination strategy $\boldsymbol{\psi}$ from Ψ in order to maximize the value $\mathcal{J}(\boldsymbol{\psi})$ given by equation* (4.14)*.*

Stage 2

We now show that the Problem 4 is equivalent to Problem 3. Specifically, we will show that any decision strategy profile $\boldsymbol{g}$ for Problem 3 can be implemented by the coordinator in Problem 4 with the same value of the problem objective. Conversely, any coordination strategy $\boldsymbol{\psi}$ in Problem 4 can be implemented in Problem 3 with the same value of the objective.

Any decision strategy profile $\boldsymbol{g}$ for Problem 3 can be implemented by the coordinator in Problem 4 as follows. Let c_t be the realization of common information at time t. The coordinator selects partial decision rule $\tilde{\gamma}_t$ using the common information

c_t as follows.

$$\tilde{\gamma}_t(\cdot) = g_t(\cdot, c_t) =: \psi_t(c_t). \tag{4.15}$$

Then, for any given realization of the primitive random variables, the choice of $\boldsymbol{\psi} = (\psi_1, \psi_2, \dots, \psi_T)$ according to (4.15) implies that the realization of the the observations $\{I_t; t = 1, \dots, T\}$, and the control actions $\{U_t; t = 1, \dots, T\}$ are identical in Problems 3 and 4. Thus, any strategy profile $\boldsymbol{g}$ for Problem 3 can be implemented by the coordinator in Problem 4 by using a coordination strategy given by (4.15) and the total expected reward under $\boldsymbol{g}$ in Problem 3 is same as the total expected cost under the coordination strategy given by (4.15) in Problem 4.

By a similar argument, any coordination strategy $\boldsymbol{\psi}$ for Problem 4 can be implemented by the decision makers in Problem 3 as follows. At time 1, decision maker 1 knows c_1; so, it can compute $\tilde{\gamma}_1 = \psi_1(c_1)$ and chooses action $u_1 = \tilde{\gamma}_1(p_1)$. Thus,

$$g_1(p_1, c_1) = \psi_1(c_1)(p_1). \tag{4.16a}$$

At time 2, decision maker 2 knows c_2 and $\tilde{\gamma}_1$, since $\tilde{\gamma}_1$ was chosen based on c_1 which is contained in c_2. Therefore, decision maker 2 can compute $\tilde{\gamma}_2 = \psi_2(c_2, \tilde{\gamma}_1)$. Then decision maker 2 chooses action $u_2 = \tilde{\gamma}_2(p_2)$. Thus,

$$g_2(p_2, c_2) = \psi_2(c_2, \tilde{\gamma}_1)(p_2). \tag{4.16b}$$

Proceeding this way, at time t decision maker t knows c_t and $\tilde{\gamma}_{1:t-1}$, so it can compute $\tilde{\gamma}_t = \psi_t(c_t, \tilde{\gamma}_{1:t-1})$. Then, decision maker t chooses action $u_t = \tilde{\gamma}_t(p_t)$. Thus,

$$g_t(p_t, c_t) = \psi_t(c_t, \tilde{\gamma}_{1:t-1})(p_t). \tag{4.16c}$$

Then, for any given realization of the primitive random variables, the choice of $\boldsymbol{g}$ according to (4.16) implies that the realization of the the observations $\{I_t; t =$

$1, \dots, T\}$, and the control actions $\{U_t; t = 1, \dots, T\}$ are identical in Problems 4 and 3. Thus, any design $\boldsymbol{\psi}$ for Problem 4 can be implemented in Problem 3 with the same expected reward.

Since Problems 3 and 4 are equivalent, we derive a sequential decomposition for the latter problem.

Stage 3

We now consider the coordinator's problem formulated in Problem 4. We first define information states for the coordinator.

Definition IV.9 (Information States)**.** For a coordination strategy $\boldsymbol{\psi}$, define *information states* Π_t, $t = 1, 2, \dots, T$ as

$$\Pi_t(q, u_1, u_2, \dots, u_{t-1}) := \mathbb{P}^{\boldsymbol{\psi}}(Q = q, U_1 = u_1, U_2 = u_2, \dots, U_{t-1} = u_{t-1} | C_t, \gamma_{1:t-1}). \tag{4.17}$$

We also define

$$\Pi_{T+1}(q, u_1, u_2, \dots, u_T) := \mathbb{P}^{\boldsymbol{\psi}}(Q = q, U_1 = u_1, U_2 = u_2, \dots, U_T = u_T | C_T, \gamma_{1:T}). \tag{4.18}$$

The following result shows how the coordinator's information state evolves depending on its observations and its decisions.

Proposition IV.10. *For $t = 1, \dots, T-1$, there exists functions F_t (which do not depend on the coordinator's strategy) such that*

$$\Pi_{t+1} = F_{t+1}(\Pi_t, \gamma_t, Z_{t+1}), \tag{4.19}$$

where $Z_{t+1} = C_{t+1} \setminus C_t$. Further, there exists functions F_{T+1} (which does not depend *on the coordinator's strategy) such that*

$$\Pi_{T+1} = F_{T+1}(\Pi_T, \gamma_T). \tag{4.20}$$

Proof. See Appendix C. $\square$

At $t = 1$, $\Pi_1(q) := \mathbb{P}^{\boldsymbol{\psi}}(Q = q|C_1)$ Since $C_1 \subseteq I_1$ which only depends on the primitive random variables, it follows that the above conditional probability is independent of the coordinator's strategy $\boldsymbol{\psi}$. Thus, Π_1 does not depend on the choice of coordinator's strategy ψ. Proposition IV.10 shows that at $t = 2, \dots, T$, Π_t depends on the strategy $\boldsymbol{\psi}$ only through the choices of $\gamma_{1:t-1}$.

Proposition IV.11. *For the coordinator, the process Π_t, $t = 1, 2, \dots, T$ is a controlled Markov chain with γ_t as the control action at time t, i.e.,*

$$\mathbb{P}^{\boldsymbol{\psi}}(\Pi_{t+1}|C_t, \Pi_{1:t}, \gamma_{1:t}) = \mathbb{P}(\Pi_{t+1}|\Pi_{1:t}, \gamma_{1:t}) \tag{4.21}$$

$$= \mathbb{P}(\Pi_{t+1}|\Pi_t, \gamma_t), \tag{4.22}$$

where the transition probabilities on the right hand side of (4.22) *do not depend on the coordination strategy $\boldsymbol{\psi}$. Furthermore, there exists a deterministic real-valued function $\tilde{R}$ defined on the space of PMFs on $\mathcal{Q} \times \mathcal{U}_1 \times \dots \times \mathcal{U}_T$ such that*

$$\mathbb{E}\{R(Q, U_1, U_2, \dots, U_T)\} = \mathbb{E}\{\tilde{R}(\Pi_{T+1})\}. \tag{4.23}$$

Proof. See Appendix C. $\square$

The controlled Markov property of the process $\{\Pi_t, t = 1, \dots, T\}$ immediately gives rise to the following result.

Theorem IV.12. *In Problem 4, without loss of optimality we can restrict attention* *to coordination strategies of the form*

$$\gamma_t = \psi_t(\Pi_t), \quad t = 1, \dots, T. \tag{4.24}$$

Further, we can write a dynamic program for the coordinator as follows: For any PMF π *on the* $\mathcal{Q} \times \mathcal{U}_1 \times \ldots \times \mathcal{U}_{T-1}$*, define*

$$\begin{aligned} J_T(\pi) &= \sup_{\tilde{\gamma}_T} \mathbb{E}\{\tilde{R}(\Pi_{T+1}) | \Pi_T = \pi, \gamma_T = \tilde{\gamma}_T\} \\ &= \sup_{\tilde{\gamma}_T} \mathbb{E}\{\tilde{R}(F_{T+1}(\pi, \tilde{\gamma}_T)) | \Pi_T = \pi, \gamma_T = \tilde{\gamma}_T\} \\ &= \sup_{\tilde{\gamma}_T} \tilde{R}(F_{T+1}(\pi, \tilde{\gamma}_T)) \end{aligned} \tag{4.25}$$

For $t = 1, \dots, T-1$*, and for any PMF* π *on the* $\mathcal{Q} \times \mathcal{U}_1 \times \ldots \times \mathcal{U}_{t-1}$*, define*

$$\begin{aligned} J_t(\pi) &= \sup_{\tilde{\gamma}_t} \mathbb{E}\{J_{t+1}(\Pi_{t+1}) | \Pi_t = \pi, \gamma_t = \tilde{\gamma}_t\} \\ &= \sup_{\tilde{\gamma}_t} \mathbb{E}\{J_{t+1}(F_{t+1}(\pi, \tilde{\gamma}_t, Z_{t+1})) | \Pi_t = \pi, \gamma_t = \tilde{\gamma}_t\} \end{aligned} \tag{4.26}$$

The arg sup γ_t^* *in the RHS of* (4.26) *is the optimal action for the coordinator at time* t *when* $\Pi_t = \pi$*. Thus,*

$$\gamma_t^* = \psi_t^*(\pi)$$

Proof. From Proposition IV.11, we conclude that the optimization problem for the coordinator is to control the evolution of the controlled Markov process $\{\Pi_t,\ t = 1, 2, \dots, T\}$ by selecting the partial functions $\{\gamma_t, t = 1, 2, \dots, T\}$ in order to maximize $\mathbb{E}\{\tilde{R}(\Pi_{T+1})\}$.

Thus, the coordinator's optimization problem can be viewed as a Markov decision problem with Π_t as the state of the Markov process. The structural result and the dynamic program follow from standard results in Markov decision theory [*Kumar and Varaiya* (1986a)]. □

Theorem IV.12 provides a sequential decomposition to find the optimal coordination strategy in Problem 4. Because of the equivalence between Problems 4 and 3, optimal decision strategies in Problem 3 can be constructed from an optimal coordination strategy. Thus, we have the following result:

Theorem IV.13. *In Problem 3, without loss of optimality we can restrict attention to decision strategies of the form*

$$U_t = g_t(P_t, \Pi_t), \quad t = 1, \dots, T, \tag{4.27}$$

where

$$\Pi_t(q, u_1, \dots, u_{t-1}) := \mathbb{P}^{g_{1:t-1}}(Q = q, U_1 = u_1, \dots, U_{t-1} = u_{t-1} | C_t).$$

Further, the optimal strategy at time t ,g_t^* , is related to the arg sup γ_t^* in the RHS of $J_t(\pi)$ in Theorem IV.12 (equation (4.26)*) as follows:*

$$g_t^*(\cdot, \pi) = \gamma_t^*$$

4.2 Two Important Special Cases

The previous section establishes the role of common information in finding sequential decompositions for general sequential decision making problems. The fictitious coordinator approach allows us to sequentially find prescriptions for decision makers that map their private information to their decision. In the following two sections, we will specialize our results to two specific cases. The two cases present two extremes of information structures: in the first case, all information of the t^{th} decision maker is private and not shared with any future decision maker while in the second case all information of the t^{th} decision maker is shared with all future decision makers. The second case is also commonly known as the case of perfect recall in stochastic control problems.

4.2.1 No Common Information

In this section, we specialize the results of Section 4.1 to the case when there is no common information at any time $t < T$. That is, we have

$$C_t = \cap_{k \geq t} I_k = \emptyset,$$

for $t < T$. This has the following implications for $t < T$:

1. $P_t = I_t \setminus C_t = I_t$

2. $Z_{t+1} = C_{t+1} \setminus C_t = \emptyset$.

3. The coordinator's partial decision rule maps private information to decision. Since, in this case, all information is private, we have that $\gamma_t : \mathcal{I}_t \mapsto \mathcal{U}_t$. In other words, the coordinator's *partial* decision rule is in fact the *actual* decision rule for the t^{th} decision-maker. Because of this, we will replace γ_t by our usual notation for decision rule g_t.

4. In absence of any common information, the coordinator selects g_t as

$$g_t = \psi_t(g_{1:t-1})$$

   Thus, under any given coordination strategy ψ_t, the sequence of decision rules g_t, $1 \leq t \leq T$ is a deterministic sequence.

5. In absence of common information, Proposition IV.10 implies that the coordi-

nator's information states

$$\Pi_t = \mathbb{P}(Q = q, U_1 = u_1, U_2 = u_2, \ldots, U_{t-1} = u_{t-1} | g_{1:t-1}),$$

evolve deterministically as: $\Pi_{t+1} = F_{t+1}(\Pi_t, g_t)$.

Note that at the final time, $C_T = \cap_{k \geq T} I_k = I_T$, thus $P_T = \emptyset$ and the prescription from the coordinator is the decision for the T^{th} decision maker.

The result of Theorem IV.12 gives us the following sequential decomposition for this case

Theorem IV.14. *In Problem 3 with no common information at any time, we can restrict attention to coordination strategies of the form*

$$g_t = \psi_t(\Pi_t), \quad t = 1, \ldots, T-1, \tag{4.28}$$

and

$$U_T = \psi_T(\Pi_T). \tag{4.29}$$

Further, we can write a dynamic program for the coordinator as follows: For any PMF π on the $\mathcal{Q} \times \mathcal{U}_1 \times \ldots \times \mathcal{U}_{T-1}$, define

$$J_T(\pi) = \sup_{u_T} \tilde{R}(F_{T+1}(\pi, u_T)) \tag{4.30}$$

For $t = 1, \ldots, T-1$, and for any PMF π on the $\mathcal{Q} \times \mathcal{U}_1 \times \ldots \times \mathcal{U}_{t-1}$, define

$$\begin{aligned} J_t(\pi) &= \sup_{g_t} \mathbb{E}\{\tilde{J}_{t+1}(F_{t+1}(\pi, g_t)) | \Pi_t = \pi, g_t\} \\ &= \sup_{g_t} J_{t+1}(F_{t+1}(\pi, g_t)) \end{aligned} \tag{4.31}$$

The arg sup g_t^* in the RHS of $J_t(\pi)$ is the optimal decision rule at time t when $\Pi_t = \pi$.

Thus,

$$g_t^* = \psi_t^*(\pi)$$

The dynamic program in Theorem IV.14 provides a sequential decomposition for the problem of finding an optimal decision strategies in the sequential decision making problem with no common information at any time. In absence of common information, the coordinator's problem is to choose decision strategies $g_1, g_2, \ldots, g_T$. The dynamic program of Theorem IV.14 provides a backward inductive method of finding an optimal strategy where at each step the coordinator selects an optimal decision rule g_t^* given that the future strategies are chosen optimally. In its essence, this sequential decomposition is identical to the designer's approach first described in *Witsenhausen* (1973). Thus, in absence of common information, the coordinator's approach described in this chapter is identical to the designer's approach suggested in *Witsenhausen* (1973).

In all sequential problems, the designer's approach will result in a deterministic dynamic program where the optimization at each step is over the space of decision rules. In the next section, we consider a case where the coordinator's approach based on common information results in a dynamic program that is stochastic and in which the optimization at each step is over the space of decisions.

4.2.2 No Private Information: The Case of Perfect Recall

In this section, we specialize the results of Section 4.1 to the case when there is no private information at any time t. This happens if we have for all $t = 1, ..T$,

$$C_t = \cap_{k \geq t} I_k = I_t \tag{4.32}$$

which implies that

$$P_t = I_t \setminus C_t = \emptyset \tag{4.33}$$

Remark IV.15 (Perfect Recall). A direct implication of equation (4.32) is that $I_t \subset I_k$, $\forall k > t$. That is, all decision makers that act after time t "remember" the information available to the t^{th} decision maker. Since this holds for all t, we have that the k^{th} decision maker has all the information of *all preceding decision makers*. That is,

$$I_1 \subset I_2 \subset \cdots \subset I_k \subset \cdots \subset I_T \tag{4.34}$$

This is commonly referred to as the condition of perfect recall in sequential decision making problems. For decision problems with perfect recall, classical dynamic program provide a sequential decomposition where at each time step the optimization is over the space of decisions [*Bertsekas* (1976)].

We can now conclude the following from equations (4.32) and (4.33):

1. As mentioned earlier, the coordinator's decision is a prescription to the t^{th} decision maker informing it on how to use its private information to make its decision. In the case where there is no private information, the prescription from the coordinator will simply be the decision for the t^{th} decision maker. In other words, the coordinator's *partial* decision rule is in fact the *actual decision* for the t^{th} decision-maker. Because of this, we will replace γ_t by our usual notation for decision U_t.

2. The coordinator selects U_t as

$$U_t = \psi_t(C_t, U_{1:t-1}) = \psi_t(I_t, U_{1:t-1})$$

3. Further, the coordinator's information states are

$$\begin{aligned}\Pi_t &= \mathbb{P}(Q = q, U_1 = u_1, U_2 = u_2, \ldots, U_{t-1} = u_{t-1} | C_t, U_{1:t-1}) \\ &= \mathbb{P}(Q = q, U_1 = u_1, U_2 = u_2, \ldots, U_{t-1} = u_{t-1} | I_t, U_{1:t-1}) \end{aligned} \tag{4.35}$$

4. Replacing γ_t by U_t in propositions IV.10 and IV.11 give us the following results: For $t = 1, \ldots, T-1$, there exists functions F_t (which do not depend on the coordinator's strategy) such that

$$\Pi_{t+1} = F_{t+1}(\Pi_t, U_t, Z_{t+1}), \tag{4.36}$$

where $Z_{t+1} = C_{t+1} \setminus C_t$. Also, there exists functions F_{T+1} (which does not depend on the coordinator's strategy) such that

$$\Pi_{T+1} = F_{T+1}(\Pi_T, U_T). \tag{4.37}$$

Further, Π_t, $t = 1, 2, \ldots, T+1$ is a controlled Markov chain with U_t as the control actions.

With the above implications, the result of Theorem IV.12 can be specialized to give the following sequential decomposition for this case.

Theorem IV.16. *In Problem 3 with no private information at any time, we can restrict attention to decision strategies of the form*

$$U_t = \psi_t(\Pi_t), \quad t = 1, \ldots, T. \tag{4.38}$$

Further, we can write a dynamic program for the coordinator as follows: For any PMF π on the $\mathcal{Q} \times \mathcal{U}_1 \times \ldots \times \mathcal{U}_{T-1}$, define

$$J_T(\pi) = \sup_{u_T} \tilde{R}(F_{T+1}(\pi, u_T)) \tag{4.39}$$

For $t = 1, \ldots, T-1$, and for any PMF π on the $\mathcal{Q} \times \mathcal{U}_1 \times \ldots \times \mathcal{U}_{t-1}$, define

$$\begin{aligned} J_t(\pi) &= \sup_{u_t} \mathbb{E}\{J_{t+1}(\Pi_{t+1})|\Pi_t = \pi, U_t = u_t\} \\ &= \sup_{u_t} \mathbb{E}\{J_{t+1}(F_{t+1}(\pi, u_t, Z_{t+1}))|\Pi_t = \pi, U_t = u_t\} \end{aligned} \tag{4.40}$$

The arg sup u_t^* in the RHS of $J_t(\pi)$ is the optimal decision at time t when $\Pi_t = \pi$. Thus,

$$u_t^* = \psi_t^*(\pi)$$

The dynamic program of Theorem IV.16 is identical to the classical dynamic program of sequential decision problems of classical information structure. At the final time, the decision maker selects a decision that is optimal with respect to its belief on the primitive random variables and past actions. (In this case, the past actions are known exactly.) Then, proceeding backwards, at time t, the decision maker selects a decision optimal to its current belief on primitive random variables and past actions assuming that the future decisions will be chosen optimally given its current decision. Thus, when all information is common, the coordinator's approach described in this chapter is identical to the classical backward inductive dynamic program.

4.3 Conclusion

In this Chapter, we presented a general model of sequential decision making problem and derived a sequential decomposition of the problem of finding optimal strategies based on the common information methodology. This result shows the universal applicability of the common information idea to sequential decision making problems. We also showed that our sequential decomposition unifies the separate results of dynamic programming in classical information structures and the designer based sequential decompositions of *Witsenhausen* (1973) for non-classical information structures. We showed that both these results are special cases of our common information based sequential decomposition.

CHAPTER V

Common Information and the Notion of State

The result of Chapter IV provides a sequential decomposition for any sequential problem. The information state for this sequential decomposition is a probability distribution on **all primitive random variables and preceding decisions conditioned on the common information**. In terms of the coordinator's perspective that we used to obtain that result, we require that at any time t, the coordinator forms a belief on all relevant system variables that have been realized so far. The question we want to address in this chapter is the following: When can we simplify the information state for sequential decomposition? In other words, are there cases when the coordinator does not need to form a belief on all system variables?

Our approach for simplifying the coordinator's information state is to use the concept of state. In broad terms, a state is a summary of past data that is sufficient for an input-output description of the system from the coordinator's point of view. What is the past data? At time t, clearly the decisions made before time t are a part of past data. But what about the primitive random variables? In our general model of a sequential problem, we did not assume any temporal order in the realization of the primitive random variables. In order to develop the concept of state, we will now impose a temporal order on the primitive random variables. That is, we will classify the primitive random variables into sub-groups each of which is realized at a different time instant. With this modification, we can exactly identify the primitive random variables that are part of the past data at time t - these are simply the variables realized before the decision is made at time t.

In general, there is no algorithmic way of identifying a state sufficient for input-output map from the coordinator's perspective. Instead, we will present sufficient conditions for a function of past data to be a state sufficient for the coordinator's input-output map. We will show that when these conditions are met, we can simplify the information state for the coordinator. Instead of a posterior on all past data, the coordinator would only need a posterior on the state sufficient for input-output map as its information state.

Organization: The rest of this chapter is organized as follows. In Section 5.1, we present a modified model of sequential decision making problems with a temporal order on the primitive random variables. In Section 5.2, we present sufficient conditions for a function of past data to be a state sufficient for coordinator's input-output map and provide a simplified sequential decomposition when such a state can be found. We finally consider the special cases of partially observable Markov decision problems and a two agent team problem in Section 5.3.

5.1 The Model

We modify the model in Section 4.1.1 of Chapter Chapter IV to introduce a temporal order in the realization of primitive random variables. The model we now study consists of the following components:

Components of a Sequential Decision Making Problem:

1. A finite probability space $(\Omega, \mathcal{B}, \mathbb{P})$.

2. Primitive Random Variables: A collection $Q = \{Q_1, Q_2, \ldots, Q_T\}$ of random vectors defined on the above probability space. We assume that the primitive

random variables are selected by nature sequentially. The random vector Q_t is realized at time t just before the t^{th} decision is made. We will assume Q_t takes values in a finite space $\mathcal{Q}_t$ endowed with the power-set sigma algebra $2^{\mathcal{Q}_t}$.

3. A finite number T of decision-makers (DMs) that make decisions in a pre-determined sequence.

4. Decision Spaces: For $t = 1, 2, \ldots, T$, $(\mathcal{U}_t, 2^{\mathcal{U}_t})$ are finite measurable spaces.

5. Observation Spaces and Observation Maps: For $t = 1, 2, \ldots, T$, N_t is the finite number of observations available for making the t^{th} decision. We have the following finite observation space

$$\mathcal{I}_t = \mathcal{Y}_t^1 \times \mathcal{Y}_t^2 \times \ldots \times \mathcal{Y}_t^{N_t}$$

equipped with the power-set sigma algebra.

Further, we define the collection of observation maps

$$H_t = (h_t^1, h_t^2, \ldots, h_t^{N_t}),$$

where for each $j = 1, 2, \ldots, N_t$, h_t^j is a measurable map from $(\mathcal{Q}_1 \times \mathcal{Q}_2 \ldots \times \mathcal{Q}_t \times \mathcal{U}_1 \times \mathcal{U}_2 \times \ldots \times \mathcal{U}_{t-1}, 2^{\mathcal{Q}_1} \times \ldots \times 2^{\mathcal{Q}_t} \times 2^{\mathcal{U}_1} \times \ldots \times 2^{\mathcal{U}_{t-1}})$ to $(\mathcal{Y}_t^j, 2^{\mathcal{Y}_t^j})$. Note that the domain of h_t^j does not include $\mathcal{Q}_{t+1}, \mathcal{Q}_{t+2}, \ldots, \mathcal{Q}_T$.

6. Reward: Let R be a real-valued function defined on the measurable space $(\mathcal{Q}_1 \times \mathcal{Q}_2 \times \ldots \times \mathcal{Q}_T \times \mathcal{U}_1 \times \mathcal{U}_2 \times \ldots \times \mathcal{U}_T, 2^{\mathcal{Q}} \times 2^{\mathcal{U}_1} \times \ldots \times 2^{\mathcal{U}_T})$.

7. Decision Strategy Space: Let $\mathcal{G}_t$ be the set of all measurable mappings from $(\mathcal{I}_t, 2^{\mathcal{I}_t})$ to $(\mathcal{U}_t, 2^{\mathcal{U}_t})$ and $\mathcal{G} = \mathcal{G}_1 \times \ldots \times \mathcal{G}_T$.

Description of the Decision-Making Problem:

The above components of the sequential decision making problem can be interpreted as follows. The information available to the first decision-maker is a random vector I_1 that takes values in the observation space $\mathcal{I}_1$. Further, $I_1 = (Y_1^1, Y_1^2, \ldots, Y_1^{N_1})$, where for each $j = 1, \ldots, N_1$, Y_1^j is related to the primitive random vector Q_1 via an observation map h_1^j. That is,

$$Y_1^j = h_1^j(Q_1) \tag{5.1}$$

The first decision maker uses a decision rule $g_1 \in \mathcal{G}_1$ to map its information to its decision. Thus,

$$U_1 = g_1(I_1) \tag{5.2}$$

Subsequently, the information available to the second decision-maker is a random vector I_2 that takes values in the observation space $\mathcal{I}_2$. Further, $I_2 = (Y_2^1, Y_2^2, \ldots, Y_1^{N_2})$, where for each $j = 1, \ldots, N_2$, Y_2^j is related to the primitive random vectors Q_1, Q_2 and U_1 via an observation map h_2^j. That is,

$$Y_2^j = h_2^j(Q_1, Q_2, U_1) \tag{5.3}$$

The second decision maker uses a decision rule $g_2 \in \mathcal{G}_2$ to map its information to its decision. Thus,

$$U_2 = g_2(I_2) \tag{5.4}$$

Proceeding sequentially, the t^{th} decision-maker's information is a random vector I_t that takes values in the observation space $\mathcal{I}_t$ and is given by the observation maps

$$Y_t^j = h_t^j(Q_1, Q_2, \ldots, Q_t, U_1, U_2, \ldots, U_{t-1}), \tag{5.5}$$

for $j = 1, 2, \ldots, N_t$. Note that the observations of the t^{th} decision maker depend only on the primitive random vectors chosen before the t^{th} decision, i.e., $Q_1, Q_2, \ldots, Q_t$. The t^{th} decision maker uses a decision rule $g_t \in \mathcal{G}_t$ to map its information to its decision. Thus,

$$U_t = g_t(I_t) \tag{5.6}$$

The collection $\boldsymbol{g} = (g_1, g_2, \ldots, g_T)$ is referred to as a *decision strategy profile*, while the function g_t is referred to as a *decision rule/decision strategy* at time t. Given a choice of decision strategy profile, the decisions $U_1, U_2, \ldots U_T$, the observation vectors $I_1, I_2, \ldots I_T$ as well as the reward $R(Q_{1:T}, U_{1:T})$ are well-defined random variables.

The value of a decision strategy $\boldsymbol{g}$ is defined as

$$\mathcal{J}(\boldsymbol{g}) = \mathbb{E}^{\boldsymbol{g}}[R(Q_1, Q_2, \ldots, Q_T, U_1, U_2, \ldots, U_T)] \tag{5.7}$$

We can now formulate the following problem:

Problem 5. *For the model of sequential decision making problems described above, choose a decision strategy profile $\boldsymbol{g} \in \mathcal{G}$ in order to maximize the value $\mathcal{J}(\boldsymbol{g})$ given by equation (5.7).*

5.2 The Coordinator's Input-Output State

Clearly, the model described above is a special case of the sequential decision-making problems of Chapter IV. Therefore, the information state established for the coordinator in Theorem IV.12 applies here as well. Our goal in this section is to find conditions under which the coordinator's information state can be simplified. To do so, our approach is to check if we can find a "state description" of the system when viewed from the coordinator's perspective. The state we want to find should be a summary of past data that is sufficient for input-output map as viewed from the coordinator.

We do not have an algorithmic way of identifying a state sufficient for input-output map from the coordinator's perspective. Instead, we will present sufficient conditions for a function of past data to be a state sufficient for coordinator's input-output map. Thus, in specific instances of the model described above, we will have to guess a potential state for the coordinator and then use the conditions presented below to verify if our guess is indeed a true state sufficient for coordinator's input-output map.

We now present sufficient conditions for a function of past data to be a state sufficient for the coordinator's input-output map. (Recall that we denote by C_t the common information at time t, by P_t the private information at time t and $Z_{t+1} = C_{t+1}/C_t$.)

Definition V.1 (State Sufficient for the Coordinator's Input-Output Map)**.** For each $t = 1, 2, \ldots, T$, let $L_t = d_t(Q_{1:t}, U_{1:t-1})$ be a function of the past primitive random variables and decisions and let $S_t = (L_t, P_t)$. We assume L_t takes values in the set $\mathcal{L}_t$ and S_t takes values in the set $\mathcal{S}_t = \mathcal{L}_t \times \mathcal{P}_t$. Then, S_t is a state sufficient for coordinator's input-output map if the following two conditions hold:

1. **Condition 1:** Under any coordination strategy $\boldsymbol{\psi}$,

$$\mathbb{P}^{\boldsymbol{\psi}}(S_{t+1}, Z_{t+1}|Q_{1:t}, U_{1:t}) = \mathbb{P}(S_{t+1}, Z_{t+1}|S_t, U_t), \tag{5.8}$$

   where the right hand side of equation (5.8) does not depend on the coordination strategy $\boldsymbol{\psi}$.

2. **Condition 2:** The reward function can be expressed as an accumulated sum of the form:

$$R(Q_{1:T}, U_{1:T}) = \sum_{t=1}^{T} R_t(S_t, U_t)$$

5.2.1 A Sequential Decomposition

In this section, we provide a sequential decomposition for the coordinator when a state sufficient for the coordinator's input output map (that is, a state satisfying

Conditions 1 and 2) can be found. The main idea here is that instead of a posterior probability distribution on all primitive random variables and preceding decisions, the coordinator can use its posterior belief on the state sufficient for the input-output map as its information state.

Definition V.2 (Information States)**.** For a coordination strategy $\boldsymbol{\psi}$, define *information states* Π_t, $t = 1, 2, \dots, T$ as

$$\Pi_t(l, p) := \mathbb{P}^{\boldsymbol{\psi}}(L_t = l, P_t = p | C_t, \gamma_{1:t-1}). \tag{5.9}$$

Recall, from Chapter IV, that γ_t is the partial decision rule prescribed by the coordinator to the decision-maker at time t.

The following result shows how the coordinator's information state evolves depending on its observations and its decisions.

Proposition V.3. *For $t = 1, \dots, T-1$, there exists functions F_t (which do not depend on the coordinator's strategy) such that*

$$\Pi_{t+1} = F_{t+1}(\Pi_t, \gamma_t, Z_{t+1}), \tag{5.10}$$

Proof. See Appendix D. □

At $t = 1$, $\Pi_1(l, p) := \mathbb{P}^{\boldsymbol{\psi}}(L_1 = l, P_1 = p | C_1)$. Since L_1, P_1 and C_1 only depend on the primitive random variables, it follows that the above conditional probability is independent of the coordinator's strategy $\boldsymbol{\psi}$. Thus, Π_1 does not depend on the choice of coordinator's strategy $\boldsymbol{\psi}$. Proposition V.3 shows that at $t = 2, \dots, T$, Π_t depends on the strategy $\boldsymbol{\psi}$ only through the choices of $\gamma_{1:t-1}$.

Proposition V.4. *The process Π_t, $t = 1, 2, \dots, T$ is a controlled Markov chain with*

γ_t as the control action at time t, i.e.,

$$\begin{aligned}\mathbb{P}^{\boldsymbol{\psi}}(\Pi_{t+1}|C_t, \Pi_{1:t}, \gamma_{1:t}) &= \mathbb{P}(\Pi_{t+1}|\Pi_{1:t}, \gamma_{1:t}) \\ &= \mathbb{P}(\Pi_{t+1}|\Pi_t, \gamma_t), \end{aligned} \tag{5.11}$$

where the transition probabilities on the right hand side of (5.11) *do not depend on the coordination strategy $\boldsymbol{\psi}$. Furthermore, there exist deterministic functions $\tilde{R}_t$ such that*

$$\mathbb{E}\{R_t(S_t, U_t))\} = \mathbb{E}\{\tilde{R}_t(\Pi_t, \gamma_t)\}. \tag{5.12}$$

Proof. See Appendix D. □

The controlled Markov property of the process $\{\Pi_t, t = 1, \ldots, T\}$ immediately gives rise to the following result.

Theorem V.5. *In Problem 5, when there exists a state process S_t as defined in Definition V.1 and satisfying Conditions 1 and 2, we can restrict attention to coordination strategies of the form*

$$\gamma_t = \psi_t(\Pi_t), \quad t = 1, \ldots, T. \tag{5.13}$$

without loss of optimality. Further, we can write a dynamic program for the coordinator as follows: For any PMF π on the $\mathcal{L}_t \times \mathcal{P}_t$, define

$$\begin{aligned}J_T(\pi) &= \sup_{\tilde{\gamma}_T} \mathbb{E}\{\tilde{R}_T(\Pi_T, \gamma_T)|\Pi_T = \pi, \gamma_T = \tilde{\gamma}_T\} \\ &= \sup_{\tilde{\gamma}_T} \tilde{R}_T(\pi, \tilde{\gamma}_T) \end{aligned} \tag{5.14}$$

For $t = 1, \ldots, T-1$, and for any PMF π on the $\mathcal{L}_t \times \mathcal{P}_t$, define

$$\begin{aligned} J_t(\pi) &= \sup_{\tilde{\gamma}_t} \mathbb{E}\{\tilde{R}_t(\Pi_t, \gamma_t) + J_{t+1}(\Pi_{t+1}) | \Pi_t = \pi, \gamma_t = \tilde{\gamma}_t\} \\ &= \sup_{\tilde{\gamma}_t} \mathbb{E}\{\tilde{R}_t(\pi, \tilde{\gamma}_t) + J_{t+1}(F_{t+1}(\pi, \tilde{\gamma}_t, Z_{t+1})) | \Pi_t = \pi, \gamma_t = \tilde{\gamma}_t\} \end{aligned} \tag{5.15}$$

The arg sup γ_t^* in the RHS of $J_t(\pi)$ is the optimal action for the coordinator at time t then $\Pi_t = \pi$. Thus,

$$\gamma_t^* = \psi_t^*(\pi)$$

Proof. From Proposition V.4, we conclude that the optimization problem for the coordinator is to control the evolution of the controlled Markov process $\{\Pi_t,\ t = 1, 2, \ldots, T\}$ by selecting the partial functions $\{\gamma_t, t = 1, 2, \ldots, T\}$ in order to maximize $\mathbb{E}\{\sum_{t=1}^{T} \tilde{R}_t(\Pi_t, \gamma_t)\}$.

Thus, the coordinator's optimization problem can be viewed as a Markov decision problem with Π_t as the state of the Markov process. The structural result and the dynamic program follow from standard results in Markov decision theory [*Kumar and Varaiya* (1986a)]. □

The solution of the coordinator's problem given by Theorem V.5 can be adapted to the original problem Problem 5 as follows:

Theorem V.6. *Consider Problem 5 described in Section 5.1. Suppose there exist a state process S_t, $t = 1, 2, \ldots, T$ as defined in Definition V.1 and satisfying Conditions 1 and 2. Then, without loss of optimality we can restrict attention to decision strategies of the form*

$$U_t = g_t(P_t, \Pi_t), \quad t = 1, \ldots, T, \tag{5.16}$$

where

$$\Pi_t(l, p) := \mathbb{P}^{g_{1:t-1}}(L_t = l, P_t = p | C_t).$$

Further, the optimal strategy is related to the arg inf γ_t^* in the RHS of $J_t(\pi)$ in Theorem V.5 as follows:

$$g_t^*(\cdot, \pi) = \gamma_t^*$$

5.3 Examples

In this Section, we use the result of Theorem V.6 to two special models of sequential decision making problems. We first consider the well-studied partially observable Markov decision problem (POMDP) and show that the result of Theorem V.6 specialize to well-known structural and dynamic programming results for this problem. We then consider a model of two-agent teams similar to the one studied in *Mahajan* (2008) and show that the result of Theorem V.6 are analogous to the sequential decomposition presented in *Mahajan* (2008). As in Chapter IV, our common information methodology shows that the classical results of centralized models like POMDPs and the sequential decomposition for non-classical team problems both follow from the **same conceptual framework of common information and an input-output state description** for the coordinator who only knows the common information.

5.3.1 Partially Observable Markov Decision Problems

5.3.1.1 The Model

Consider a system consisting of a plant and a controller. At time $t = 1$, the plant is in the initial state given by the random variable X_1. With time, the plant evolves according to

$$X_{t+1} = f_t(X_t, U_t, V_t) \tag{5.17}$$

where V_t is a random variable taking values in a finite set $\mathcal{V}$. $\{V_t; t = 1, \dots, T-1\}$ is a sequence of independent random variables that are also independent of X_1.

The controller's observations are generated according to

$$Y_t = h_t(X_t, W_t) \tag{5.18}$$

where W_t are random variables taking values in a finite set $\mathcal{W}_t$. $\{W_t; t = 1, \dots, T;\}$ are independent random variables that are also independent of X_1 and $\{V_t; t = 1, \dots, T\}$. The controller must select its action U_t as a function of the form:

$$U_t = g_t(Y_{1:t}, U_{1:t-1}) \tag{5.19}$$

The controller's objective is to choose $\boldsymbol{g} := (g_1, g_2, \dots, g_T)$ to maximize

$$\mathbb{E}^{\boldsymbol{g}}\Big[\sum_{t=1}^{T} R_t(X_t, U_t)\Big] \tag{5.20}$$

5.3.1.2 The Input-Output State for the Coordinator

The above decision-making problem can be easily described in terms of the model of Section 5.1. The primitive random variables are given as $Q = \{Q_1, Q_2, \dots, Q_T\}$ where $Q_1 = (X_1, W_1)$ and $Q_t = (V_{t-1}, W_t)$, for $1 < t \leq T$. The information available for making the t^{th} decision is $I_t = (Y_{1:t}, U_{1:t-1})$. It is easy to see that in this case, we have

$$C_t = \cap_{k \geq t} I_k = I_t \tag{5.21}$$

and

$$P_t = I_t \setminus C_t = \emptyset$$

and

$$Z_{t+1} = C_{t+1} \setminus C_t = (Y_{t+1}, U_t)$$

As mentioned earlier, the coordinator's decision is a prescription to the t^{th} decision maker informing it on how to use its private information to make its decision. Since in this case there is no private information, the prescription from the coordinator will simply be the decision for the t^{th} decision maker. Because of this, we will replace γ_t by our usual notation for decision U_t. The coordinator selects its prescription U_t as a function of the common information at time t and the past prescriptions. Thus,

$$U_t = \psi_t(C_t, U_{1:t-1}) = \psi_t(I_t, U_{1:t-1}) = \psi_t(I_t)$$

A Candidate for Input-Output State

As a summary of the past data, we define $L_t = X_t$. Then, $S_t = (L_t, P_t) = L_t = X_t$ (since P_t is empty). With the above definition of S_t, we can now verify if Conditions 1 and 2 are satisfied.

1. **Condition 1:**

$$\begin{aligned}
&\mathbb{P}^{\psi}(S_{t+1}, Z_{t+1}|q_{1:t}, u_{1:t}) \\
&= \mathbb{P}^{\psi}(X_{t+1}, Y_{t+1}, U_t|x_1, w_{1:t}, v_{1:t-1}, u_{1:t}) \\
&= \mathbb{1}_{\{U_t=u_t\}}\mathbb{P}(Y_{t+1}|X_{t+1})\mathbb{P}^{\psi}(X_{t+1}|x_1, w_{1:t}, v_{1:t-1}, u_{1:t}, x_t)
\end{aligned} \tag{5.22}$$

where we added x_t in the conditioning in (5.22) since it is a function of the other terms already fixed in the conditioning. Because of the nature of plant dynamics and independence of primitive random variables, we can write (5.22) as

$$\begin{aligned}
&\mathbb{1}_{\{U_t=u_t\}}\mathbb{P}(Y_{t+1}|X_{t+1})\mathbb{P}(X_{t+1}|x_t, u_t) \\
&= \mathbb{P}(U_t, Y_{t+1}, X_{t+1}|x_t, u_t) \\
&= \mathbb{P}(Z_{t+1}, S_{t+1}|s_t, u_t)
\end{aligned} \tag{5.23}$$

2. **Condition 2:** The reward function trivially satisfies Condition 2.

Since S_t satisfies Conditions 1 and 2, we can define the coordinator's information states as

$$\begin{aligned}\Pi_t &= \mathbb{P}(S_t | C_t, \gamma_{1:t-1}) \\ &= \mathbb{P}(X_t | Y_{1:t}, U_{1:t-1}) \end{aligned} \tag{5.24}$$

Replacing γ_t by U_t and Z_{t+1} by Y_{t+1}, U_t in propositions V.3 and V.4 gives us the following results:
For $t = 1, \dots, T-1$, there exists functions F_t (which do not depend on the coordinator's strategy) such that

$$\Pi_{t+1} = F_{t+1}(\Pi_t, U_t, Y_{t+1}), \tag{5.25}$$

since $Y_{t+1} = Z_{t+1}$; and

$$\mathbb{E}\{R_t(S_t, U_t))\} = \mathbb{E}\{\tilde{R}_t(\Pi_t, U_t)\}$$

Further, Π_t, $t = 1, 2, \dots, T$ is a controlled Markov chain with U_t as the control actions.

With the above results, Theorem V.5 can be specialized to give the following sequential decomposition for this case.

Theorem V.7. *In the partially observable Markov decision problem described in Section 5.3.1.1, we can restrict attention to decision strategies of the form*

$$U_t = \psi_t(\Pi_t), \quad t = 1, \dots, T. \tag{5.26}$$

Further, we can write a dynamic program for the coordinator as follows: For any

PMF π *on the* $\mathcal{S}_T$*, define*

$$J_T(\pi) = \sup_{u_T} \tilde{R}_T(\pi, u_T) \tag{5.27}$$

For $t = 1, \ldots, T-1$*, and for any PMF* π *on the* $\mathcal{S}_t$*, define*

$$J_t(\pi) = \sup_{u_t} \mathbb{E}\{\tilde{R}_t(\pi, u_t) + J_{t+1}(F_{t+1}(\pi, u_t, Z_{t+1}))|\Pi_t = \pi, U_t = u_t\} \tag{5.28}$$

The arg sup u_t^* *in the RHS of* $J_t(\pi)$ *is the optimal decision at time* t *when* $\Pi_t = \pi$*. Thus,*

$$u_t^* = \psi_t^*(\pi)$$

The above result is identical to the classical structural and dynamic programming result for POMDPs [*Kumar and Varaiya* (1986a)].

5.3.2 A General Two Agent Team Problem

We next consider a general model of two-agent team similar to the model described in *Mahajan* (2008). The model consists of a plant and two agents. In the original model, at each time, Agent 1 takes an action followed by Agent 2. For the ease of notation in our setup, we will divide each time instant into two sub-time instants such that Agent 1 acts on odd times and Agent 2 acts on even times. The total time horizon is thus effectively $2T$. Also, unlike the original model of *Mahajan* (2008), we will assume memory update rules of each agent (described further below) are fixed.

5.3.2.1 The Model

At time $t = 1$, the plant is in the initial state given by the random variable X_1. With time, the plant evolves according to

$$X_{2k} = X_{2k-1}, \tag{5.29a}$$

for $k = 1, 2, \ldots, T$ and

$$X_{2k+1} = f_{2k}(X_{2k}, U^1_{2k-1}, U^2_{2k}, V_{2k}), \tag{5.29b}$$

for $k = 1, 2, \ldots, T-1$, where V_{2k} is a random variable taking values in a finite set $\mathcal{V}$. $\{V_{2k}; k = 1, \ldots, T-1\}$ is a sequence of independent random variables that are also independent of X_1. At each odd time $t = 2k-1$, Agent 1 makes an observation according to

$$Y^1_{2k-1} = h^1_{2k-1}(X_{2k-1}, W_{2k-1}) \tag{5.30}$$

At each even time $t = 2k$, Agent 2 makes an observation according to

$$Y^2_{2k} = h^2_{2k}(X_{2k}, U^1_{2k-1}, W_{2k}) \tag{5.31}$$

Each agent has a memory that is initialized to 0 ($M^1_1 = M^2_2 = 0$) and updated as follows:

$$M^1_{2k+1} = d^1_{2k-1}(M^1_{2k-1}, Y^1_{2k+1}, U^1_{2k+1}) \tag{5.32a}$$

$$M^2_{2k+2} = d^2_{2k}(M^2_{2k}, Y^2_{2k+2}, U^2_{2k+2}) \tag{5.32b}$$

We assume that the memory update rules are fixed. The decisions are chosen as:

$$U^1_{2k+1} = g^1_{2k+1}(M^1_{2k-1}, Y^1_{2k+1}) \tag{5.33a}$$

$$U^2_{2k} = g^2_{2k}(M^2_{2k-2}, Y^2_{2k}) \tag{5.33b}$$

The agents' objective is choose $\boldsymbol{g}^1 := (g_1^1, g_3^1, \ldots, g_{2T-1}^1)$ and $\boldsymbol{g}^2 := (g_2^2, g_4^2, \ldots, g_{2T}^2)$ to maximize

$$\mathbb{E}^{\boldsymbol{g}^1,\boldsymbol{g}^2}[\sum_{k=1}^{T} R_k(X_{2k}, U_{2k-1}^1, U_{2k}^2)] \tag{5.34}$$

5.3.2.2 The Input-Output State for the Coordinator

The above decision-making problem can be described in terms of the model of Section 5.1. The primitive random variables are given as $Q = \{Q_1, Q_2, \ldots, Q_{2T}\}$ where $Q_1 = (X_1, W_1)$ and $Q_{2k} = W_{2k}$ and $Q_{2k+1} = (V_{2k}, W_{2k+1})$. The information available for making the t^{th} decision is

$$I_t = (Y_t^1, M_{t-2}^1), \text{ if } t \text{ is odd}$$

and

$$I_t = (Y_t^2, M_{t-2}^2), \text{ if } t \text{ is even}$$

. It is easy to see that in this case, we have

$$C_t = \cap_{k \geq t} I_k = \emptyset \tag{5.35}$$

and

$$P_t = I_t \setminus C_t = I_t$$

and

$$Z_{t+1} = C_{t+1} \setminus C_t = \emptyset$$

Since in this case all information is private, the prescription from the coordinator will be the decision rule for the t^{th} decision maker. The coordinator selects its prescription g_t as a function of the common information at time t and the past prescriptions. Thus,

$$g_t = \psi_t(C_t, g_{1:t-1}) = \psi_t(g_{1:t-1})$$

A Candidate for Input-Output State

For $k = 1, 2, \ldots, T$, we define $L_{2k+1} = (X_{2k+1}, M^2_{2k})$ and $L_{2k} = (X_{2k}, M^1_{2k-1}, U^1_{2k-1})$. Then, we have

$$S_{2k+1} = (L_{2k+1}, P_{2k+1}) = (X_{2k+1}, M^2_{2k}, Y^1_{2k+1}, M^1_{2k-1})$$

and

$$S_{2k} = (L_{2k}, P_{2k}) = (X_{2k}, M^1_{2k-1}, U^1_{2k-1}, Y^2_{2k}, M^2_{2k-2})$$

We can now verify if S_t, $t = 1, 2, ..., 2T$, satisfies Conditions 1 and 2:

1. **Condition 1:** We first focus on $t = 2k$,

$$\begin{aligned}
&\mathbb{P}^{\psi}(S_{2k+1}, Z_{2k+1} | q_{1:2k}, u^1_1, \ldots, u^1_{2k-1}, u^2_2, \ldots, u^2_{2k}) \\
&= \mathbb{P}^{\psi}(X_{2k+1}, M^2_{2k}, Y^1_{2k+1}, M^1_{2k-1} | x_1, w_{1:2k}, v_2, v_4, \ldots, v_{2k-2}, u^1_1, \ldots, u^1_{2k-1}, u^2_2, \ldots, u^2_{2k}) \\
&= \mathbb{P}^{\psi}\left(X_{2k+1}, M^2_{2k}, Y^1_{2k+1}, M^1_{2k-1} \middle| \begin{array}{l} x_{1:2k}, w_{1:2k}, v_2, v_4, \ldots, v_{2k}, \\ u^1_1, \ldots, u^1_{2k-1}, u^2_2, \ldots, u^2_{2k}, m^1_{2k-1}, m^2_{2k-2}, y^2_{2k} \end{array} \right)
\end{aligned} \tag{5.36}$$

where we added $x_{2:2k}, y^2_{2k}, m^1_{2k-1}, m^2_{2k-2}$ in the conditioning in (5.36) since they are function of the primitive random variables and decisions included in the conditioning terms. Equation (5.36) can then be written as

$$\mathbb{P}(Y^1_{2k+1} | X_{2k+1}) \mathbb{P}(X_{2k+1} | x_{2k}, u^1_{2k-1}, u^2_{2k}) \mathbb{1}_{\{M^1_{2k-1} = m^1_{2k-1}\}} \mathbb{1}_{\{M^2_{2k} = d^2_{2k}(m^2_{2k-2}, y^2_{2k}, u^2_{2k})\}} \tag{5.37}$$

$$= \mathbb{P}(S_{2k+1} | s_{2k}, u^2_{2k}) \tag{5.38}$$

where (5.38) follows from (5.36) and (5.37) since of all the conditioning terms in (5.36), its simplification in (5.37) only uses those included in s_{2k}, u^2_{2k}.

Also, for $t = 2k-1$,

$$
\begin{aligned}
&\mathbb{P}^{\psi}(S_{2k}, Z_{2k} | q_{1:2k-1}, u^1_1, \ldots, u^1_{2k-1}, u^2_2, \ldots, u^2_{2k-2}) \\
&= \mathbb{P}^{\psi}\left(X_{2k}, M^1_{2k-1}, U^1_{2k-1} Y^2_{2k}, M^2_{2k-2} \middle| \begin{array}{l} x_1, w_{1:2k-1}, v_2, v_4, \ldots, v_{2k-2}, \\ u^1_1, \ldots, u^1_{2k-1}, u^2_2, \ldots, u^2_{2k-2} \end{array} \right) \\
&= \mathbb{P}^{\psi}\left(X_{2k}, M^1_{2k-1}, U^1_{2k-1} Y^2_{2k}, M^2_{2k-2} \middle| \begin{array}{l} x_{1:2k-1}, w_{1:2k-1}, v_2, v_4, \ldots, v_{2k-2}, \\ u^1_1, \ldots, u^1_{2k-1}, u^2_2, \ldots, u^2_{2k-2}, m^1_{2k-1}, m^2_{2k-2}, y^1_{2k-1} \end{array} \right)
\end{aligned}
\tag{5.39}
$$

where we added $x_{2:2k-1}, m^1_{2k-3}, m^2_{2k-2}, y^1_{2k-1}$ in the conditioning in (5.39) since they are function of the primitive random variables and decisions included in the conditioning terms. Equation (5.39) can then be written as

$$
\mathbb{P}(Y^1_{2k} | X_{2k}, u^1_{2k-1}) \mathbb{P}(X_{2k} | x_{2k-1}) \mathbb{1}_{\{M^1_{2k-1} = d^1_{2k-1}(m^1_{2k-3}, y^1_{2k-1}, u^1_{2k-1})\}} \mathbb{1}_{\{M^2_{2k-2} = m^2_{2k-2}\}} \tag{5.40}
$$

$$
= \mathbb{P}(S_{2k} | s_{2k-1}, u^1_{2k-1}) \tag{5.41}
$$

where (5.41) follows from (5.39) and (5.40) since of all the conditioning terms in (5.39), its simplification in (5.40) only uses those included in s_{2k-1}, u^1_{2k-1}. Thus, condition 1 is satisfied at all times $t = 1, 2, \ldots, 2T$.

2. **Condition 2:** By a simple change of variables, the reward function can be written as

$$
\sum_{k=1}^{T} R_{2k}(S_{2k}, U^2_{2k})
$$

which is of the form required in Condition 2 with the reward at odd times being 0.

Since S_t satisfies Conditions 1 and 2, we can define the coordinator's information states as

$$\begin{aligned}\Pi_t &= \mathbb{P}(S_t|C_t, \gamma_{1:t-1}) \\ &= \mathbb{P}(S_t|g_{1:t-1})\end{aligned} \tag{5.42}$$

Replacing γ_t by g_t in propositions V.3 and V.4 gives us the following results:
For $t = 1, \ldots, 2T-1$, there exists functions F_t (which do not depend on the coordinator's strategy) such that

$$\Pi_{t+1} = F_{t+1}(\Pi_t, g_t) \tag{5.43}$$

and for $k = 1, 2, \ldots, T$

$$\mathbb{E}\{R_{2k}(S_{2k}, U^2_{2k})\} = \mathbb{E}\{\tilde{R}_{2k}(\Pi_{2k}, g^2_{2k})\}$$

Further, Π_t, $t = 1, 2, \ldots, T+1$ is a controlled Markov chain with g_t as the control actions.

With the above results, Theorem V.5 can be specialized to give the following sequential decomposition for this case

Theorem V.8. *In the two agent team problem described in Section 5.3.2.1, we can write a dynamic program for the coordinator as follows: For any PMF π on $\mathcal{S}_T$, define*

$$J_{2T}(\pi) = \sup_{g^2_{2T}} \tilde{R}_{2T}(\pi, g^2_{2T}) \tag{5.44}$$

For $k = 1, \ldots, T$, and for any PMF π on $\mathcal{S}_{2k-1}$, define

$$J_{2k-1}(\pi) = \sup_{g^1_{2k-1}} J_{2k}(F_{2k}(\pi, g^1_{2k-1})) \tag{5.45}$$

For $k = 1, \dots, T-1$, and for any PMF π on $\mathcal{S}_{2k}$, define

$$J_{2k}(\pi) = \sup_{g^2_{2k}} [\tilde{R}_{2k}(\pi, g^2_{2k}) + J_{2k+1}(F_{2k+1}(\pi, g^2_{2k}))] \tag{5.46}$$

The arg sup in the RHS of $J_t(\pi)$ is the optimal decision at time t when $\Pi_t = \pi$.

Theorem V.8 provides a sequential decomposition of a two-agent team problem that is similar to the sequential decomposition of *Mahajan* (2008).

5.4 Other Simplifications of the Coordinator's Problem

In the preceding sections, we identified conditions under which the coordinator can simplify its information state. We now focus on another aspect of simplifying the coordinator's dynamic program. Recall that at each time t, the coordinator selects a partial decision rule γ_t that maps private information at time t to decision at time t. The optimal choice of γ_t must be chosen from the set of all functions from the space of private information $\mathcal{P}_t$ to space of decisions $\mathcal{U}_t$. Let $\Gamma(\mathcal{P}_t, \mathcal{U}_t)$ be the set of all functions from $\mathcal{P}_t$ to $\mathcal{U}_t$. The size of the set $\Gamma(\mathcal{P}_t, \mathcal{U}_t)$ is a natural measure of the complexity of the coordinator's optimization problem at each step of its dynamic program. One way to simplify the coordinator's optimization is to characterize a set $A \subset \Gamma(\mathcal{P}_t, \mathcal{U}_t)$ for which we can guarantee that there always exists an optimal choice of coordinator's decision in the set A. We state, without proof, the following lemma that provides sufficient conditions for one such characterization.

Lemma V.9. *Consider a sequential decision making problem of the form described in Section 5.1. For $t = 1, 2, \dots, T$, let $P'_t \subset P_t$, such that the following conditions hold:*

1. *For any choice of decision strategies $g_{1:T-1}$ of the form $U_t = g_t(P_t, C_t)$, $1 \leq t \leq$*

$T-1$, and for all $u \in \mathcal{U}_T$

$$\mathbb{E}[R(Q_{1:T}, U_{1:T})|p_T, c_T, U_T = u] = \mathbb{E}[R(Q_{1:T}, U_{1:T})|p'_T, c_T, U_T = u] \tag{5.47}$$

where p_T, p'_T, c_T are any realizations of P_T, P'_T, C_T that occur with non-zero probability;

2. *For any choice of decision strategies $g_{1:t-1}$ of the form $U_k = g_k(P_k, C_k)$, $1 \leq k \leq t-1$, and $g_{t+1:T}$ of the form $U_k = g_k(P'_k, C_k)$, $t+1 \leq k \leq T$ and for all $u \in \mathcal{U}_T$*

$$\mathbb{E}[R(Q_{1:T}, U_{1:T})|p_t, c_t, U_t = u] = \mathbb{E}[R(Q_{1:T}, U_{1:T})|p'_t, c_t, U_t = u] \tag{5.48}$$

where p_t, p'_t, c_t are any realizations of P_t, P'_t, C_t that occur with non-zero probability.

Then, there exist optimal decision strategies of the form $U_t = g_t(P'_t, C_t)$, for all $t = 1, 2, \ldots, T$. In particular, the coordinator can select its partial decision rule from $\Gamma(\mathcal{P}'_t, \mathcal{U}_t) \subset \Gamma(\mathcal{P}_t, \mathcal{U}_t)$ without losing optimality.

5.5 Conclusions

In this Chapter, we used the concept of state to simplify the common information based sequential decomposition described in Chapter IV. In broad terms, a state is a summary of past data that is sufficient for an input-output description of the system from the point of view of a coordinator that knows the common information. We identified sufficient conditions for a function of past data to be a state sufficient for the coordinator's input output map. We obtained a simplified sequential decomposition for the coordinator when such a state exists. We re-derived sequential decompositions of POMDP and a general two-agent team problem by identifying a suitable state for the coordinator and using the common information based sequential decomposition.

CHAPTER VI

Conclusions and Reflections

In this thesis, we studied sequential decision making problems in cooperative systems where different agents with different information want to achieve a common objective. The sequential nature of the decision problem implies that all decisions can be arranged in a sequence such that the information available to make the t^{th} decision only depends on preceding decisions. Markov decision theory provides tools for addressing sequential decision making problems with classical information structures. In this thesis, we focused on decision making problems with non-classical information structures. We introduced a new approach for such decision making problems. This approach relies on the idea of common information among decision-makers. Intuitively, common information consists of past observations and decisions that are commonly known to the current and future decision makers. We showed that a common information based approach can allow us to discover new structural results of optimal decision strategies; and provide simpler sequential decomposition of the decision-making problems than earlier approaches. We first demonstrated this approach on two specific instances of sequential problems, namely, a real-time multi-terminal communication system and a decentralized control system with delayed sharing of information. We then showed that the common information methodology applies more generally to any sequential decision making problem. Moreover, we showed that our common information methodology unifies the separate sequential decomposition results available for classical and non-classical information structures. We also presented sufficient conditions for simplifying common information based sequential decompositions using the concept of state sufficient for the input output map of a coordinator that only knows the common information.

6.1 Common Knowledge and Common Information

The central question in decentralized decision making in cooperative systems is one of coordination. How can two (or more) agents with different information coordinate their decisions to achieve the best performance? For effective coordination, an agent needs to infer what other agents know about the system, what other agents know about what it knows about the system, what other agents know about what it knows about their knowledge about the system and so on. Common knowledge, as defined in *Aumann* (1976), describes the knowledge that all agents know, that all agents know that all agents know it and all agents know that all agents know that all agents know it and so on. The above self-referential verbal logic is expressed much easily mathematically: if two agents' knowledge is identified with sigma algebras $\mathcal{F}$ and $\mathcal{G}$, then the common knowledge is simply $\mathcal{F} \cap \mathcal{G}$. Now consider two agents that observe the following set of primitive random variables defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$:

Agent	Observations
1	X, Y
2	Y, Z

Then, one can identify Agent 1's knowledge with the sigma field $\sigma(X, Y) \subset \mathcal{F}$ and Agent 2's knowledge with the sigma field $\sigma(Y, Z) \subset \mathcal{F}$. Then, the common knowledge between the agents is $\sigma(X, Y) \cap \sigma(Y, Z)$. Note that the common information between the agents (as per our definition) is simply the observation Y. Clearly, $\sigma(Y) \subset$

$\sigma(X,Y) \cap \sigma(Y,Z)$. Thus, common information describes a part of common knowledge between the agents. Although, common information may capture only a part of the entire common knowledge, describing common information is much simpler than describing the entire common knowledge.

In the above example, all observations were primitive random variables. Hence, one could define agents' knowledge and the common knowledge without specifying any decision strategy. In more general problems, an agent's observation depends on other agents decisions. For example, the observation Z in the above example could be a function of primitive random variable and the decision U_1 of agent 1. In this case, Z is well-defined as a random variable *only if a decision strategy of agent 1 has been specified.* If g_1 is the strategy of agent 1, then one may define agent 2's knowledge as $\sigma^{g_1}(Y,Z) \subset \mathcal{F}$, where we include g^1 in the superscript to denote the dependence of this sub-sigma field on the choice of strategy g_1. Thus, the common knowledge between two agents depends, in general, on the choice of strategies. Note, however, we can define common information without specifying any decision strategies.

6.2 Common Information and Decision Strategies

The goal of any decision problem is to identify optimal decision strategies for all agents. Each decision strategy is a complete prescription of an agent's behavior for all possible realizations of its information. A common information methodology is essentially a way of dividing the problem of finding behavioral prescriptions into sub-problems. When all agents observe the realization of the common information, they only need prescriptions *for the commonly known realization of common information* and prescriptions for unobserved realizations of common information are inconsequential.

Consider again a sequential problem with two agents where agent 1 observes the primitive random variables X and Y and agent 2 observes the primitive random variables Y and Z. The two agents are required to make decisions $U_1 \in \mathcal{U}_1$ and $U_2 \in \mathcal{U}_2$ respectively in order to maximize the expected value of a reward $R(X, Y, Z, U_1, U_2)$. Assume that the primitive random variables X, Y, Z take values in finite sets $\mathcal{X}, \mathcal{Y}, \mathcal{Z}$ respectively. Then, in order to find the best strategy profile, one has to search from $N := |\mathcal{U}_1|^{|\mathcal{X}||\mathcal{Y}|} \times |\mathcal{U}_2|^{|\mathcal{Y}||\mathcal{Z}|}$ possible strategy profiles for this decision problem. Consider now a common information based approach. For each realization of common information Y, we only need to find how agents use their information *given the realization of Y*. Thus, to find the best prescription for the given realization of Y, we need to search from $n = |\mathcal{U}_1|^{|\mathcal{X}|} \times |\mathcal{U}_2|^{|\mathcal{Z}|}$ possible choices. Since, we need prescriptions for all realizations of Y, the complexity of this approach is roughly $|\mathcal{Y}| \cdot |\mathcal{U}_1|^{|\mathcal{X}|} \times |\mathcal{U}_2|^{|\mathcal{Z}|}$. This is a considerable improvement from the complexity of a brute force search over N choices of strategy profiles.

In the above simple example, the second agent's observation did not depend on the first agent's decision. This is not the case in more general forms of decision problems. In problems where second agent's observation depends on the decision of the first agent, there is a *communication aspect in the decision problem.* This is because the first agent may try to convey some information to the second agent by means of its decision. In order for agent 2 to interpret what information agent 1 is trying to convey, agent 2 needs to know *how agent 1 maps information to decision*, that is, agent 2 needs to know the decision strategy of agent 1. This communication aspect of decision problems has been identified as the reason that sequential decompositions based on designer's approach result in optimization over strategy space [*Mahajan* (2008)].

Consider again a two agent problem where agent 1 observes the primitive random variables X, Y. Agent 1 has to make a decision $U_1 = g_1(X, Y)$ according to a decision strategy g_1. Agent 2 observes the primitive random variable Y and the decision of agent 1, U_1. Agent 2 must make a decision U_2 in order to maximize the expected value of a reward $R(X, Y, U_2)$. Consider a sequential decomposition based on designer's approach. Proceeding backwards, the designer needs to start with specifying agent 2's strategy. What should agent 2 do if it observes a realization y, u_1 of its observations? In order to find agent 2's best decision, the designer must know how can agent 2 interpret u_1. That is, the designer needs to know what values of agent 1's observations X, Y could have led to the observed decision. In other words, the designer needs to know the decision strategy g_1. This is the main reason why the designer's sequential decomposition always involve optimization over strategies.

Consider now a common information based approach for the above problem. Given a realization y of the common information Y, agent 2 only needs to know what values of agent 1's observation X *along with the observed realization y of common information could have led to the observed decision u_1.* That is, agent 2 only needs to know the *partial decision strategy* $g(\cdot, y)$ to interpret agent 1's decision. This is one of the main reasons why common information based sequential decompositions involve optimization over partial decision strategies. Thus, the common information methodology simplifies the communication aspect of decision problems.

6.3 Future Directions

The common information methodology developed in this thesis relies only on the sequentiality of decision-making problems. This suggests that this approach can be used to address a wide class of decision-making problems- especially those with substantial common information. The delayed sharing information structure is an example of decentralized control problem where the system architecture (that is, the sharing of information over a communication medium with delay) introduces common information among decision-makers. Other kinds of information sharing mechanisms in decentralized control - namely, sharing of control actions, periodic sharing of information, sharing with random delay- also result in the presence of common information among controllers. Thus, we expect that the common information methodology should be able to provide structural results and sequential decompositions for decentralized control problems with the above mentioned information sharing mechanisms.

In Chapter V, (Section 5.4), we presented one way of simplifying the coordinator's sequential decomposition. Essentially, this approach relied on removing redundant private information from decision-makers' information. Lesser the private information, the simpler are the prescriptions that the coordinator has to find. Another way of reducing private information can be to make it common information. Can we *add redundant information* to a decision-maker's information so as to have more common information? Consider a simple two-stage decision problem as an example. DM1 observes a random variable X_1, takes an action U_1, DM 2 observes $X_2 = X_1 + U_1 + W_1$, where W_1 is a random variable independent of X_1. DM 2 takes an action U_2 and a reward $R(X_2, U_2)$ is received. Is there common information at time $t = 1$? No, since X_1 is unavailable to DM 2. However, it is possible to show that even if X_1 were available to DM 2, it will not alter the optimal decision strategy for DM 2. Thus, as far as DM 2 is concerned, X_1 is a redundant observation. Adding this redundant observation to DM 2, however, allows us to have X_1 as common information at time $t = 1$. This implies that the coordinator's prescription at time $t = 1$ can now simply be a decision (since there is no private information). Finding appropriate redundant information that may be added or removed from a decision-makers information in order to simplify the coordinator's problem seems to require guesswork and ingenuity, although the work in *Mahajan and Tatikonda* (2009) suggests an automated way may be possible.

In general, having more information as common information among decision-makers can provide improvements in performance. If a group of decision-makers have more information as common information, they will do no worse than in the case where they had less common information. Moreover, having more information as common implies lesser private information which results in a simpler problem for the coordinator. If the amount of information that should be common can be altered (say, by a system designer), how can we decide how much information should be common information? In other words, can we assess different information structures in terms of the cost of making more information common versus the benefit of enhanced performance and simpler solutions?

Finally, in this thesis we developed the common information methodology only for decision-making problems where all agents have the same objective. The cooperative nature of the problem is essential for this methodology. When the fictitious coordinator makes its prescriptions, the cooperative nature of the problem ensures that all agents have no reason to deviate from the prescribed behavior. This is clearly not the case in game-theoretic problems where different agents may have different objectives. Can common information methodology have any role to play in game-theoretic problems? We believe the answer is yes. Using the fact that all agents have some information in common may simplify the problem of finding equilibrium decision strategies. Instead of finding decision strategies that are in Nash equilibrium, we expect that one may be able to simplify the problem to finding partial decision strategies that are in Nash equilibrium for each realization of common information.

We hope the above questions will provide useful starting points in future efforts to extend the work presented in this thesis.

APPENDICES

APPENDIX A

Appendix for Multi-Terminal Communication System

Proof of Lemma II.4

For a realization $x^1_{1:t}$ of $X^1_{1:t}$, we have by definition,

$$
\begin{aligned}
b^1_t(a) =& P(A = a|x^1_{1:t}) \\
=& P(A = a, x^1_t|x^1_{1:t-1}) / \sum_{a' \in \mathcal{A}} P(A = a', x^1_t|x^1_{1:t-1})
\end{aligned}
\tag{A.1}
$$

where we used Bayes' rule in (A.1). The numerator in (A.1) can be written as,

$$
\begin{aligned}
& P(X^1_t = x^1_t|A = a, x^1_{1:t-1}).P(A = a|x^1_{1:t-1}) \\
=& P(X^1_t = x^1_t|A = a, x^1_{t-1}).b^1_{t-1}(a)
\end{aligned}
\tag{A.2}
$$

where we used the Markov nature of X^1_t when conditioned on A. Thus, for a fixed a, the numerator in (A.1) depends only on x^1_t, x^1_{t-1} and the previous belief b^1_{t-1}. Since the same factorization holds for each term in the denominator, we have that

$$b_t^1 = \alpha_t^1(b_{t-1}^1, X_t^1, X_{t-1}^1),$$

where $\alpha_t^1, t = 2, 3, ..., T$ are deterministic transformations.

Proof of Lemma II.5

By definition of μ_t^1, we have

$$\begin{aligned} \mu_t^1(m) &= P(M_{t-1}^1 = m | Z_{1:t-1}^1, l_{1:t-1}^1) \\ &= P(l_{t-1}^1(M_{t-2}^1, Y_{t-1}^1) = m | Z_{1:t-1}^1, l_{1:t-1}^1) \end{aligned} \tag{A.3}$$

With the memory update rules $l_{1:t-1}^1$ fixed, the probability in (A.3) can be evaluated from the conditional distribution $P(M_{t-2}^1, Y_{t-1}^1 | Z_{1:t-1}^1, l_{1:t-1}^1)$. For $m' \in \mathcal{M}^1$ and $y \in \mathcal{Y}^1$, this conditional distribution is given as

$$P(M_{t-2}^1 = m', Y_{t-1}^1 = y | Z_{1:t-1}^1, l_{1:t-1}^1) \tag{A.4}$$

$$=P(Y_{t-1}^1 = y | M_{t-2}^1 = m', Z_{1:t-1}^1, l_{1:t-1}^1) \times$$

$$P(M_{t-2}^1 = m' | Z_{1:t-1}^1, l_{1:t-1}^1)$$

$$=P(Y_{t-1}^1 = y | Z_{t-1}^1).P(M_{t-2}^1 = m' | Z_{1:t-2}^1, l_{1:t-2}^1) \tag{A.5}$$

$$=P(Y_{t-1}^1 = y | Z_{t-1}^1).\mu_{t-1}^1(m') \tag{A.6}$$

where we used the fact that the channel noise at time t (N_t^1) is independent of the past noise variables and the Markov source in (A.5). Thus, we only need Z_{t-1}^1 and μ_{t-1}^1 to form the joint belief in (A.4). Consequently, we can evaluate $\mu_t^1(m)$ just from Z_{t-1}^1 and μ_{t-1}^1. Thus,

$$\mu_t^1 = \beta_t^1(\mu_{t-1}^1, Z_{t-1}^1)$$

where $\beta_t^1, t = 2, 3, ..., T$ are deterministic transformations.

Proof Of Lemma II.13

For fixed $f_{1:T}^2$ and for a given realization of the received symbols $z_{1:t}^1, z_{1:t}^2$ and the partial encoding functions $\tilde{w}_{1:t}^1$, the receiver's belief on the state of the source at time t is given as:

$$\psi_t(x) := P(X_t = x | z_{1:t}^1, z_{1:t}^2, \tilde{w}_{1:t}^1, f_{1:t}^2) \tag{A.7}$$

where $x = (x^1, x^2, a)$. Using Bayes' rule, we have

$$\psi_t(x) = P(X_t = x, Z_{1:t}^2 = z_{1:t}^2 | z_{1:t}^1, \tilde{w}_{1:t}^1, f_{1:t}^2) / \sum_{x' \in \mathcal{X}} P(X_t = x', Z_{1:t}^2 = z_{1:t}^2 | z_{1:t}^1, \tilde{w}_{1:t}^1, f_{1:t}^2) \tag{A.8}$$

The numerator in right hand side of (A.8) can be written as

$$\begin{aligned} & P(Z_{1:t}^2 = z_{1:t}^2 | z_{1:t}^1, X_t = x, \tilde{w}_{1:t}^1, f_{1:t}^2) \\ & \times P(X_t = x | z_{1:t}^1, \tilde{w}_{1:t}^1, f_{1:t}^2) \\ =& P(Z_{1:t}^2 = z_{1:t}^2 | X_t^2 = x^2, A = a, f_{1:t}^2) \\ & \times P(X_t^2 = x^2 | A = a) \times P(X_t^1 = x_t^1, A = a | z_{1:t}^1, \tilde{w}_{1:t}^1) \end{aligned} \tag{A.9}$$

where we used conditional independence of $Z_{1:t}^2, X_t^2$ and $Z_{1:t}^1, X_t^1$ given A for the first term in (A.9) and the fact that $X_t = (X_t^1, X_t^2, A)$ in the second term of left hand side of (A.9).

Since the second encoder is fixed, the first term in the right hand side of (A.9) is a known statistic which depends on $z_{1:t}^2$. The second term is again a known source statistic. Consider the last term in (A.9). It can be expressed as follows:

$$
\begin{aligned}
&\sum_{b' \in \Delta(\mathcal{A})} P(X_t^1 = x_t^1, A = a, b_t^1 = b' | z_{1:t}^1, \tilde{w}_{1:t}^1) \\
&= \sum_{b' \in \Delta(\mathcal{A})} [P(A = a | b_t^1 = b', x_t^1, z_{1:t}^1, \tilde{w}_{1:t}^1) \\
&\quad \times P(X_t^1 = x_t^1, b_t^1 = b' | z_{1:t}^1, \tilde{w}_{1:t}^1)] && \text{(A.10)} \\
&= \sum_{b' \in \Delta(\mathcal{A})} b'(a) \times P(X_t^1 = x_t^1, b_t^1 = b' | z_{1:t}^1, \tilde{w}_{1:t}^1) \\
&= \sum_{b' \in \Delta(\mathcal{A})} b'(a) \times \tilde{\xi}_t^1(x_t^1, b') && \text{(A.11)}
\end{aligned}
$$

Similar representations also hold for each term in the denominator of (A.8). It follows then that with a fixed $f_{1:t}^2$, $\psi_t(x)$ depends only on the realization of second encoder's messages $Z_{1:t}^2$ and ξ_t^1. Thus, from (A.9) and (A.11), we conclude that ψ_t can be evaluated from ξ_t^1 and $Z_{1:t}^2$ by means of deterministic transformations. We will call this overall transformation as δ_t. Thus, we have

$$\psi_t = \delta_t(\xi_t^1, Z_{1:t}^2) \tag{A.12}$$

Since the estimate $\hat{X}_t$ is a function of ψ_t (cf. Theorem II.11), we conclude that

$$\hat{X}_t = \tau_t(\delta_t(\xi_t^1, Z_{1:t}^2))$$

Proof of Lemma II.14

1) Consider a realization $z_{1:t}^1$ and $\tilde{w}_{1:t}^1$.

By definition, the realization of ξ_{t-1}^1 is given as

$$\tilde{\xi}_t^1(x_t^1, \tilde{b}_t^1) = P(X_t^1 = x_t^1, b_t^1 = \tilde{b}_t^1 | z_{1:t}^1, \tilde{w}_{1:t}^1) \tag{A.13}$$

Using Bayes' rule, we have

$$\tilde{\xi}_t^1(x_t^1, \tilde{b}_t^1) = P(X_t^1 = x_t^1, b_t^1 = \tilde{b}_t^1, Z_t^1 = z_t^1 | z_{1:t-1}^1, \tilde{w}_{1:t}^1)$$
$$/ \sum_{x' \in \mathcal{X}^1} \sum_{b' \in \Delta(\mathcal{A})} P(X_t^1 = x', b_t^1 = b', Z_t^1 = z_t^1 | z_{1:t-1}^1, \tilde{w}_{1:t}^1) \tag{A.14}$$

We can write the numerator as:

$$P(Z_t^1 = z_t^1 | X_t^1 = x_t^1, b_t^1 = \tilde{b}_t^1, z_{1:t-1}^1, \tilde{w}_{1:t}^1)$$
$$\times P(X_t^1 = x_t^1, b_t^1 = \tilde{b}_t^1 | z_{1:t-1}^1, \tilde{w}_{1:t}^1)$$
$$= P(Z_t^1 = z_t^1 | X_t^1 = x_t^1, b_t^1 = \tilde{b}_t^1, \tilde{w}_t^1)$$
$$\times P(X_t^1 = x_t^1, b_t^1 = \tilde{b}_t^1 | z_{1:t-1}^1, \tilde{w}_{1:t}^1) \tag{A.15}$$

the first term in (A.15) is true since $z_t^1 = w_t^1(x_t^1, \tilde{b}_t^1)$. The second term in (A.15) can be further written as:

$$
\begin{aligned}
&\sum_{\substack{x''\in\mathcal{X}^1,\\ a\in\mathcal{A}}} \sum_{b'\in\Delta(\mathcal{A})} P(X_t^1 = x_t^1, b_t^1 = \tilde{b}_t^1, X_{t-1}^1 = x'' \\
&\qquad , A = a, b_{t-1}^1 = b' | z_{1:t-1}^1, \tilde{w}_{1:t}^1) \\
&= \sum_{\substack{x''\in\mathcal{X}^1,\\ a\in\mathcal{A}}} \sum_{b'\in\Delta(\mathcal{A})} [P(b_t^1 = \tilde{b}_t^1 | b_{t-1}^1 = b', X_t^1 = x_t^1, X_{t-1}^1 = x'') \\
&\qquad \times P(X_t^1 = x_t^1 | A = a, X_{t-1}^1 = x'') \\
&\qquad \times P(A = a | b_{t-1}^1 = b', X_{t-1}^1 = x'', \\
&\qquad z_{1:t-1}^1, \tilde{w}_{1:t-1}^1) \\
&\qquad \times P(X_{t-1}^1 = x'', b_{t-1}^1 = b' | z_{1:t-1}^1, \tilde{w}_{1:t-1}^1)] \qquad \text{(A.16)} \\
&= \sum_{\substack{x''\in\mathcal{X}^1,\\ a\in\mathcal{A}}} \sum_{b'\in\Delta(\mathcal{A})} [P(b_t^1 = \tilde{b}_t^1 | b_{t-1}^1 = b', X_t^1 = x_t^1, X_{t-1}^1 = x'') \\
&\qquad \times P(X_t^1 = x_t^1 | A = a, X_{t-1}^1 = x'') \\
&\qquad \times P(A = a | b_{t-1}^1 = b') \times \tilde{\xi}_{t-1}^1(x'', b')] \qquad \text{(A.17)}
\end{aligned}
$$

where we used Lemma II.4 and the Markov property of X_t^1 given A in (A.16). The first term in (A.17) is simply 1 or 0 since b_t^1 is a deterministic function of b_{t-1}^1, X_t^1 and X_{t-1}^1. The second term is a known source statistic and the third term is $b'(a)$. Similar expressions hold for the denominator in (A.14). Thus, from (A.14)-(A.17), we conclude that to evaluate $\xi_t^1(x_t^1, \tilde{b}_t^1)$ we only need Z_t^1, w_t^1 and ξ_{t-1}^1. This establishes equation (2.43).

2) With encoder 2's strategy fixed, the expected instantaneous cost from the

coordinator's perspective is given as:

$$
\begin{aligned}
&\mathbb{E}\{\rho_t(X_t, \hat{X}_t)|z^1_{1:t}, \tilde{w}^1_{1:t}\} \\
=&\mathbb{E}\{\rho_t(X^1_t, X^2_t, A, \tau_t(\delta_t(\xi^1_t, Z^2_{1:t})))|z^1_{1:t}, \tilde{w}^1_{1:t}, \tilde{\xi}^1_t\},
\end{aligned} \tag{A.18}
$$

since $\tilde{\xi}^1_t$ is a function of $z^1_{1:t}, \tilde{w}^1_{1:t}$, hence it can be included in the conditioning variables. Thus, the only random variables in the above expectation are X^1_t, X^2_t, A and $Z^2_{1:t}$. Therefore, the above expectation is a function of the following probability mass function:

$$
\begin{aligned}
&P(X^1_t = x^1_t, X^2_t = x^2_t, A = a, Z^2_{1:t} = z^2_{1:t}|z^1_{1:t}, \tilde{w}^1_{1:t}, \tilde{\xi}^1_t) \\
=&P(Z^2_{1:t} = z^2_{1:t}, X^2_t = x^2_t|A = a, X^1_t = x^1_t, z^1_{1:t}, \tilde{w}^1_{1:t}, \tilde{\xi}^1_t) \\
&\times P(X^1_t = x^1_t, A = a|z^1_{1:t}, \tilde{w}^1_{1:t}, \tilde{\xi}^1_t) \\
=&P(Z^2_{1:t} = z^2_{1:t}, X^2_t = x^2_t|A = a)\times \\
&\sum_{b' \in \Delta(\mathcal{A})} [P(X^1_t = x^1_t, A = a, b^1_t = b'|z^1_{1:t}, \tilde{w}^1_{1:t}, \tilde{\xi}^1_t)] \\
=&P(Z^2_{1:t} = z^2_{1:t}, X^2_t = x^2_t|A = a)\times \\
&\sum_{b' \in \Delta(\mathcal{A})} [P(A = a|b^1_t = b', z^1_{1:t}, \tilde{w}^1_{1:t}, \tilde{\xi}^1_t) \\
&\times P(X^1_t = x^1_t, b^1_t = b'|z^1_{1:t}, \tilde{w}^1_{1:t}, \tilde{\xi}^1_t)]
\end{aligned} \tag{A.19}
$$

$$
\begin{aligned}
=&P(Z^2_{1:t} = z^2_{1:t}, X^2_t = x^2_t|A = a) \times \sum_{b' \in \Delta(\mathcal{A})} [b'(a) \\
&\times P(X^1_t = x^1_t, b^1_t = b'|z^1_{1:t}, \tilde{w}^1_{1:t}, \tilde{\xi}^1_t)] \\
=&P(Z^2_{1:t} = z^2_{1:t}, X^2_t = x^2_t|A = a) \times \sum_{b' \in \Delta(\mathcal{A})} [b'(a) \\
&\times \tilde{\xi}^1_t(x^1_t, b')]
\end{aligned} \tag{A.20}
$$

where we used conditional independence of the encoder's observations and actions given A in (A.19). In (A.20), the first term is a fixed statistic when encoder 2's strategy is fixed and the second term depends only on $\tilde{\xi}_t^1$. Thus, the expectation in (A.18) can be evaluated using $\tilde{\xi}_t^1$. This establishes the second part of the Lemma (equation 2.44).

APPENDIX B

Appendix for Delayed Sharing Information Structures

Proof of Proposition III.4

Fix a coordinator strategy $\boldsymbol{\psi}$. Consider a realization c_{t+1} of the common information C_{t+1}. Let $(\tilde{\gamma}^1_{1:t}, \tilde{\gamma}^2_{1:t})$ be the corresponding realization of partial functions until time t. Assume that the realization $(c_{t+1}, \pi_{1:t}, \gamma^1_{1:t}, \tilde{\gamma}^2_{1:t})$ has non-zero probability. Then, the realization π_{t+1} of Π_{t+1} is given by

$$\pi_{t+1}(s_{t+1}) = \mathbb{P}^{\boldsymbol{\psi}}(S_{t+1} = s_{t+1} | c_{t+1}, \tilde{\gamma}^1_{1:t}, \tilde{\gamma}^2_{1:t}). \tag{B.1}$$

Using Proposition III.2 , this can be written as

$$\begin{aligned} \sum_{s_t, v_t, w^1_{t+1}, w^2_{t+1}} & \mathbb{1}_{s_{t+1}}(\hat{f}_{t+1}(s_t, v_t, w^1_{t+1}, w^2_{t+1}, \tilde{\gamma}^1_t, \tilde{\gamma}^2_t)) \\ & \cdot \mathbb{P}(V_t = v_t) \cdot \mathbb{P}(W^1_{t+1} = w^1_{t+1}) \\ & \cdot \mathbb{P}(W^2_{t+1} = w^2_{t+1}) \cdot \mathbb{P}^{\boldsymbol{\psi}}(S_t = s_t | c_{t+1}, \tilde{\gamma}^1_{1:t}, \tilde{\gamma}^2_{1:t}). \end{aligned} \tag{B.2}$$

Since $c_{t+1} = (c_t, z_{t+1})$, the last term of (B.2) can be written as

$$\mathbb{P}^{\boldsymbol{\psi}}(S_t = s_t | c_t, z_{t+1}, \tilde{\gamma}^1_{1:t}, \tilde{\gamma}^2_{1:t}) \\ = \frac{\mathbb{P}^{\boldsymbol{\psi}}(S_t = s_t, Z_{t+1} = z_{t+1} | c_t, \tilde{\gamma}^1_{1:t}, \tilde{\gamma}^2_{1:t})}{\sum_{s'} \mathbb{P}^{\boldsymbol{\psi}}(S_t = s', Z_{t+1} = z_{t+1} | c_t, \tilde{\gamma}^1_{1:t}, \tilde{\gamma}^2_{1:t})}. \quad \text{(B.3)}$$

Use (3.20) and the sequential order in which the system variables are generated to write

$$\mathbb{P}^{\boldsymbol{\psi}}(S_t = s_t, Z_{t+1} = z_{t+1} | c_t, \tilde{\gamma}^1_{1:t}, \tilde{\gamma}^2_{1:t}) \\ = \mathbb{1}_{\hat{h}_t(s_t)}(z_{t+1}) \cdot \mathbb{P}^{\boldsymbol{\psi}}(S_t = s_t | c_t, \tilde{\gamma}^1_{1:t-1}, \tilde{\gamma}^2_{1:t-1}) \quad \text{(B.4)} \\ = \mathbb{1}_{\hat{h}_t(s_t)}(z_{t+1}) \cdot \pi_t(s_t). \quad \text{(B.5)}$$

where $\tilde{\gamma}^1_t, \tilde{\gamma}^2_t$ are dropped from conditioning in (B.4) because for the given coordinator's strategy, they are functions of the rest of the terms in the conditioning. Substitute (B.5), (B.3), and (B.2) into (B.1), to get

$$\pi_{t+1}(s_{t+1}) = F_{t+1}(\pi_t, \tilde{\gamma}^1_t, \tilde{\gamma}^2_t, z_{t+1})(s_{t+1})$$

where $F_{t+1}(\cdot)$ is given by (B.1), (B.2), (B.3), and (B.5).

Proof of Proposition III.5

Fix a coordinator strategy $\boldsymbol{\psi}$. Consider a realization c_{t+1} of the common information C_{t+1}. Let $\pi_{1:t}$ be the corresponding realization of $\Pi_{1:t}$ and $(\tilde{\gamma}^1_{1:t}, \tilde{\gamma}^2_{1:t})$ the corresponding choice of partial functions until time t. Assume that the realization $(c_{t+1}, \pi_{1:t}, \gamma^1_{1:t}, \tilde{\gamma}^2_{1:t})$ has a non-zero probability. Then, for any Borel subset $A \subset \mathcal{P}\{\mathcal{S}\}$, where $\mathcal{P}\{\mathcal{S}\}$ is the space of probability mass functions over the finite set $\mathcal{S}$ (the space of realization of S_t), use Proposition III.4 to write

$$\begin{aligned}
&\mathbb{P}(\Pi_{t+1} \in A | c_t, \pi_{1:t}, \tilde{\gamma}^1_{1:t}, \tilde{\gamma}^2_{1:t}) \\
&\quad = \sum_{z_{t+1}} \mathbb{1}_A(F_{t+1}(\pi_t, \tilde{\gamma}^1_t, \tilde{\gamma}^2_t, z_{t+1})) \\
&\quad\quad \cdot \mathbb{P}(Z_{t+1} = z_{t+1} | c_t, \pi_{1:t}, \tilde{\gamma}^1_{1:t}, \tilde{\gamma}^2_{1:t})
\end{aligned} \tag{B.6}$$

Now, use (3.20), to obtain

$$\begin{aligned}
&\mathbb{P}(Z_{t+1} = z_{t+1} | c_t, \pi_{1:t}, \tilde{\gamma}^1_{1:t}, \tilde{\gamma}^2_{1:t}) \\
&\quad = \sum_{s_t} \mathbb{1}_{\hat{h}_t(s_t)}(z_{t+1}) \cdot \mathbb{P}(S_t = s_t | c_t, \pi_{1:t}, \tilde{\gamma}^1_{1:t}, \tilde{\gamma}^2_{1:t}) \\
&\quad = \sum_{s_t} \mathbb{1}_{\hat{h}_t(s_t)}(z_{t+1}) \cdot \pi_t(s_t)
\end{aligned} \tag{B.7}$$

where we used the fact that for any realization $(c_t, \pi_{1:t}, \tilde{\gamma}^1_{1:t}, \tilde{\gamma}^2_{1:t})$ of positive probability, the conditional probability $\mathbb{P}(S_t = s_t | c_t, \pi_{1:t}, \tilde{\gamma}^1_{1:t}, \tilde{\gamma}^2_{1:t})$ is same as $\pi_t(s_t)$. Substitute (B.7) back in (B.6), to get

$$\begin{aligned}
&\mathbb{P}(\Pi_{t+1} \in A | c_t, \pi_{1:t}, \tilde{\gamma}^1_{1:t}, \tilde{\gamma}^2_{1:t}) \\
&\quad = \sum_{z_{t+1}} \sum_{s_t} \mathbb{1}_A(F_{t+1}(\pi_t, \tilde{\gamma}^1_t, \tilde{\gamma}^2_t, z_{t+1})) \\
&\quad\quad \cdot \mathbb{1}_{\hat{h}_t(s_t)}(z_{t+1}) \cdot \pi_t(s_t) \\
&\quad = \mathbb{P}(\Pi_{t+1} \in A | \pi_t, \tilde{\gamma}^1_t, \tilde{\gamma}^2_t),
\end{aligned} \tag{B.8}$$

thereby proving (3.24).

Now, use Proposition III.2 to write,

$$
\begin{aligned}
&\mathbb{E}\{\hat{R}_t(S_t, \gamma_t^1, \gamma_t^2, S_{t+1}) | c_t, \pi_{1:t}, \tilde{\gamma}_{1:t}^1, \tilde{\gamma}_{1:t}^2\} \\
&\quad = \sum_{s_t, v_t, w_{t+1}^1, w_{t+1}^2} \hat{R}_t(s_t, \tilde{\gamma}_t^1, \tilde{\gamma}_t^2, \hat{f}_{t+1}(s_t, v_t, w_{t+1}^1, w_{t+1}^2, \tilde{\gamma}_t^1, \tilde{\gamma}_t^2)) \\
&\qquad \cdot \mathbb{P}(V_t = v_t) \cdot \mathbb{P}(W_{t+1}^1 = w_{t+1}^1) \cdot \mathbb{P}(W_{t+1}^2 = w_{t+1}^2) \\
&\qquad \cdot \mathbb{P}(S_t = s_t | c_t, \pi_{1:t}, \tilde{\gamma}_{1:t}^1, \tilde{\gamma}_{1:t}^2) \\
&\quad = \sum_{s_t, v_t, w_{t+1}^1, w_{t+1}^2} \hat{R}_t(s_t, \tilde{\gamma}_t^1, \tilde{\gamma}_t^2, \hat{f}_{t+1}(s_t, v_t, w_{t+1}^1, w_{t+1}^2, \tilde{\gamma}_t^1, \tilde{\gamma}_t^2)) \\
&\qquad \cdot \mathbb{P}(V_t = v_t) \cdot \mathbb{P}(W_{t+1}^1 = w_{t+1}^1) \\
&\qquad \cdot \mathbb{P}(W_{t+1}^2 = w_{t+1}^2) \cdot \pi_t(s_t) \\
&\quad =: \tilde{R}_t(\pi_t, \tilde{\gamma}_t^1, \tilde{\gamma}_t^2). \qquad \text{(B.9)}
\end{aligned}
$$

This proves (3.25).

Piecewise linearity and concavity of value function

Lemma B.1. *For any realization $\tilde{\gamma}_t^{1:2}$ of $\gamma_t^{1:2}$, the cost $\tilde{R}_t(\pi_t, \tilde{\gamma}_t^1, \tilde{\gamma}_t^2)$ is linear in π_t.*

Proof.

$$
\begin{aligned}
&\tilde{R}_t(\pi_t, \tilde{\gamma}_t^1, \tilde{\gamma}_t^2) \\
&\quad = \mathbb{E}\{\hat{R}_t(S_t, \gamma_t^1, \gamma_t^2, S_{t+1} | \Pi_t = \pi_t, \gamma_t^{1:2} = \tilde{\gamma}_t^{1:2}\} \\
&\quad = \sum \hat{R}_t(s_t, \tilde{\gamma}^1, \tilde{\gamma}^2, \hat{f}_{t+1}(s_t, v_t, w_{t+1}^1, w_{t+1}^2, \tilde{\gamma}^1, \tilde{\gamma}^2)) \\
&\qquad \cdot \mathbb{P}(V_t = v_t) \cdot \mathbb{P}(W_{t+1}^{1:2} = w_{t+1}^{1:2}) \cdot \pi_t(s_t)
\end{aligned}
$$

where the summation is over all realizations of $(s_t, v_t, w_{t+1}^{1:2})$. Hence $\tilde{R}_t(\pi_t, \tilde{\gamma}_t^1, \tilde{\gamma}_t^2)$ is linear in π_t. □

We prove the piecewise linearity and concavity of the value function by induction.

For $t = T$,

$$J_T(\pi) = \inf_{\tilde{\gamma}_t^{1:2}} \tilde{R}_T(\pi, \tilde{\gamma}_t^1, \tilde{\gamma}_t^2).$$

Lemma B.1 implies that $J_T(\pi)$ is the inifimum of finitely many linear functions of π. Thus, $J_T(\pi)$ is piecewise linear and concave in π. This forms the basis of induction. Now assume that $J_{t+1}(\pi)$ is piecewise linear and concave in π. Then, J_{t+1} can be written as the infimum of a finite family I of linear functions as

$$J_{t+1}(\pi) = \inf_{i \in I} \Big\{ \sum_{s \in \mathcal{S}} a_i(s) \cdot \pi(s) + b_i \Big\}, \tag{B.10}$$

where $a_i(s)$, b_i, $i \in I$, $s \in \mathcal{S}$ are real numbers. Using this, we will prove that the piecewise linearity and concavity of $J_t(\pi)$.

$$J_t(\pi) = \inf_{\tilde{\gamma}^1, \tilde{\gamma}^2} \Big[\tilde{R}_t(\pi, \tilde{\gamma}^1, \tilde{\gamma}^2) + \mathbb{E}\{J_{t+1}(\Pi_{t+1}) | \Pi_t = \pi, \gamma_t^{1:2} = \tilde{\gamma}^{1:2}\} \Big]. \tag{B.11}$$

For a particular choice of $\tilde{\gamma}^{1:2}$, we concentrate on the terms inside the square brackets. By Lemma B.1 the first term is linear in π. The second term can be written as

$$\begin{aligned}
&\mathbb{E}\{J_{t+1}(\Pi_{t+1}) | \Pi_t = \pi, \gamma_t^{1:2} = \tilde{\gamma}^{1:2}\} && \text{(B.12)}\\
&\quad = \sum_{z_{t+1}} J_{t+1}(F_{t+1}(\pi, \tilde{\gamma}^1, \tilde{\gamma}^2, z_{t+1}))\\
&\qquad \cdot \mathbb{P}(Z_{t+1} = z_{t+1} | \Pi_t = \pi, \gamma_t^{1:2} = \tilde{\gamma}^{1:2})\\
&\quad = \sum_{z_{t+1}} \Big[\inf_{i \in I} \Big\{ \sum_{s} a_i(s) \cdot (F_{t+1}(\pi, \tilde{\gamma}^1, \tilde{\gamma}^2, z_{t+1}))(s) + b_i \Big\}\\
&\qquad \cdot \mathbb{P}(Z_{t+1} = z_{t+1} | \Pi_t = \pi, \gamma_t^{1:2} = \tilde{\gamma}^{1:2}) \Big] && \text{(B.13)}
\end{aligned}$$

where the last expression follows from (B.10).

Note that

$$\mathbb{P}(Z_{t+1} = z_{t+1} | \Pi_t = \pi, \gamma_t^{1:2} = \tilde{\gamma}^{1:2}) = \sum_{s' \in \mathcal{S}} \mathbb{1}_{\hat{h}_t(s')}(z_{t+1}) \cdot \pi(s') \quad \text{(B.14)}$$

Focus on each term in the outer summation in (B.13). For each value of z_{t+1}, these terms can be written as:

$$\begin{aligned}
\inf_{i \in I} \Big\{ \sum_s a_i(s) \cdot (F_{t+1}(\pi, \tilde{\gamma}^1, \tilde{\gamma}^2, z_{t+1}))(s) \\
\cdot \sum_{s' \in \mathcal{S}} \mathbb{1}_{\hat{h}_t(s')}(z_{t+1}) \cdot \pi(s') \\
+ b_i \cdot \sum_{s' \in \mathcal{S}} \mathbb{1}_{\hat{h}_t(s')}(z_{t+1}) \cdot \pi(s')
\end{aligned} \quad \text{(B.15)}$$

The second summand is linear in π. Using the characterization of F_{t+1} from the proof of Proposition III.4 (Appendix B), we can write the first summand as

$$\begin{aligned}
a_i(s) \cdot \Big\{ \sum_{s_t, v_t, w_{t+1}^1, w_{t+1}^2} \mathbb{1}_s(\hat{f}_{t+1}(s_t, v_t, w_{t+1}^1, w_{t+1}^2, \tilde{\gamma}_t^1, \tilde{\gamma}_t^2)) \\
\cdot \mathbb{P}(V_t = v_t) \cdot \mathbb{P}(W_{t+1}^{1:2} = w_{t+1}^{1:2}) \\
\cdot \mathbb{1}_{\hat{h}(s_t)}(z_{t+1}) \pi(s_t) \Big\}
\end{aligned} \quad \text{(B.16)}$$

which is also linear in π. Substituting (B.15) and (B.16) in (B.13), we get that for a given choice of $\tilde{\gamma}^1, \tilde{\gamma}^2$, the second expectation in (B.11) is concave in π. Thus, the value function $J_t(\pi)$ is the minimum of finitely many functions each of which is linear in π. This implies that J_t is piecewise linear and concave in π. This completes the induction argument.

Proof of Proposition III.11

We prove the three parts separately.

Part 1)

We first prove that Θ_{t+1} is a function of Θ_t and Z_{t+1}. Recall that $Z_{t+1} = (Y^1_{t-n+1}, Y^2_{t-n+1}, U^1_{t-n+1}, U^2_{t-n+1})$ and $C_{t+1} = (C_t, Z_{t+1})$. Fix a coordination strategy $\boldsymbol{\psi}$ and consider a realization c_{t+1} of C_{t+1}. Then,

$$
\begin{aligned}
&\theta_{t+1}(x_{t-n+1}) \\
&\coloneqq \mathbb{P}(X_{t-n+1} = x_{t-n+1} | c_{t+1}) \\
&= \mathbb{P}(X_{t-n+1} = x_{t-n+1} | c_t, y^{1:2}_{t-n+1}, y^2_{t-n+1}, u^{1:2}_{t-n+1}) \\
&= \sum_{x \in \mathcal{X}} \mathbb{P}(X_{t-n+1} = x_{t-n+1} | X_{t-n} = x, u^{1:2}_{t-n+1}) \\
&\quad \cdot \mathbb{P}(X_{t-n} = x | c_t, y^{1:2}_{t-n+1}, u^{1:2}_{t-n+1}) \\
&= \sum_{x \in \mathcal{X}} \mathbb{P}(X_{t-n+1} = x_{t-n+1} | X_{t-n} = x, u^{1:2}_{t-n+1}) \\
&\quad \cdot \frac{\mathbb{P}(X_{t-n} = x, y^{1:2}_{t-n+1}, u^{1:2}_{t-n+1} | c_t)}{\sum_{x'} \mathbb{P}(X_{t-n} = x', y^{1:2}_{t-n+1}, u^{1:2}_{t-n+1} | c_t)}
\end{aligned} \tag{B.17}
$$

Consider the second term of (B.17), and note that under any coordination strategy $\boldsymbol{\psi}$, the variables $u^{1:2}_{t-n+1}$ are deterministic functions of $y^{1:2}_{t-n+1}$ and c_t (which is same as $y^{1:2}_{1:t-n}, u^{1:2}_{1:t-n}$). Therefore, the numerator of the second term of (B.17) can be written as

$$
\begin{aligned}
&\mathbb{P}^{\boldsymbol{\psi}}(x_{t-n}, y^{1:2}_{t-n+1}, u^{1:2}_{t-n+1} | c_t) \\
&= \mathbb{P}^{\boldsymbol{\psi}}(u^{1:2}_{t-n+1} | x_{t-n}, y^{1:2}_{t-n+1}, c_t) \\
&\quad \cdot \mathbb{P}^{\boldsymbol{\psi}}(y^{1:2}_{t-n+1} | x_{t-n}, c_t) \cdot \mathbb{P}^{\boldsymbol{\psi}}(x_{t-n} | c_t) \\
&= \mathbb{P}^{\boldsymbol{\psi}}(u^{1:2}_{t-n+1} | y^{1:2}_{t-n+1}, c_t) \\
&\quad \cdot \mathbb{P}(y^{1:2}_{t-n+1} | x_{t-n}) \cdot \theta_t(x_{t-n})
\end{aligned} \tag{B.18}
$$

Substitute (B.18) in (B.17) and cancel $\mathbb{P}^{\psi}(u^{1:2}_{t-n+1}|y^{1:2}_{t-n+1}, c_t)$ from the numerator and denominator. Thus, θ_{t+1} is a function of θ_t and z_{t+1}.

Next we prove that r^k_{t+1} is a function of r^k_t, Z_{t+1} and γ^k_t. Recall that

$$r^k_{t+1} \coloneqq (r^k_{m,(t+1)}, t-n+2 \leq m \leq t).$$

We prove the result by showing that each component $r^k_{m,(t+1)}$, $t-n+2 \leq m \leq t$ is a function of r^k_t, Z_{t+1} and γ^k_t.

1. For $m = t$, we have $r^k_{t,(t+1)} \coloneqq \gamma^k_t(\cdot, Y^k_{t-n+1})$. Since Y^k_{t-n+1} is a part of Z_{t+1}, $r^k_{t,(t+1)}$ is a function of γ^k_t and Z_{t+1}.

2. For $m = t-n+2, t-n+3, \ldots, t-1$,

$$\begin{aligned} r^k_{m,t+1}(\cdot) &\coloneqq \gamma^k_m(\cdot, Y^k_{m-n+1:t+1-n}, U^k_{m-n+1:t+1-n}) \\ &= \gamma^k_m(\cdot, Y^k_{t-n+1}, U^k_{t-n+1}, Y^k_{m-n+1:t-n}, U^k_{m-n+1:t-n}) \\ &= r^k_{m,t}(\cdot, Y^k_{t-n+1}, U^k_{t-n+1}) \end{aligned} \tag{B.19}$$

   Thus, for $m = t-n+2, t-n+3, \ldots, t-1$, $r^k_{m,t+1}$ is a function of $r^k_{m,t}$ and Z_{t+1}.

Part 2)

First, let us assume that the coordinator's belief Π_t defined in (3.22) is a function of (Θ_t, r^1_t, r^2_t), that is, there exist functions H_t, for $t = 1, 2, \ldots, T$, such that

$$\Pi_t = H_t(\Theta_t, r^1_t, r^2_t) \tag{B.20}$$

From (3.25) of Proposition III.5, we have that

$$
\begin{aligned}
&\mathbb{E}\{\hat{R}_t(S_t, \gamma_t^1, \gamma_t^2, S_{t+1}) | C_t, \Pi_{1:t}, \gamma_{1:t}^1, \gamma_{1:t}^2\} \\
&\quad = \tilde{R}_t(\Theta_t, r_t^1, r_t^2, \gamma_1^1, \gamma_t^2)
\end{aligned} \tag{B.21}
$$

where the last equation uses (B.20). Thus, to prove this part of the proposition, only need to prove (B.20). For that matter, we need the following lemma.

Lemma B.2. *$S_t \coloneqq (X_{t-1}, P_t^1, P_t^2)$ is a deterministic function of $(X_{t-n}, V_{t-n+1:t-1}, W_{t-n+1:t}^1, W_{t-n+1:t}^2, r_t^1, r_t^2)$, that is, there exists a fixed deterministic function D_t such that*

$$
\begin{aligned}
S_t &\coloneqq (X_{t-1}, P_t^1, P_t^2) \\
&= D_t(X_{t-n}, V_{t-n+1:t-1}, W_{t-n+1:t}^{1:2}, r_t^{1:2})
\end{aligned} \tag{B.22}
$$

Proof. We first prove a slightly weaker result: for $t-n+1 \le m \le t-1$, there exists a deterministic function $\hat{D}_{m,t}$ such that

$$
\begin{aligned}
&(X_{t-n+1:m}, Y_{t-n+1:m}^{1:2}, U_{t-n+1:m}^{1:2}) \\
&\qquad = \hat{D}_{m,t}(X_{t-n}, V_{t-n+1:m}, W_{t-n+1:m}^{1:2}, r_{t-n+1:m,t}^{1:2})
\end{aligned} \tag{B.23}
$$

using induction. First consider $m = t-n+1$. For this case, the LHS of (B.23) equals $(X_{t-n+1}, Y_{t-n+1}^{1:2}, U_{t-n+1}^{1:2})$. For $k = 1, 2$,

$$
\begin{aligned}
Y_{t-n+1}^k &= h_{t-n+1}^k(X_{t-n}, W_{t-n+1}^k) \\
U_{t-n+1}^k &= r_{t-n+1,t}^k(Y_{t-n+1}^k)
\end{aligned}
$$

Furthermore, by the system dynamics,

$$
X_{t-n+1} = f_t(X_{t-n}, U_{t-n+1}^{1:2}, V_{t-n+1})
$$

Thus $(X_{t-n+1}, Y^{1:2}_{t-n+1}, U^{1:2}_{t-n+1})$ is a deterministic function of $(X_{t-n}, W^{1:2}_{t-n+1}, V_{t-n+1}, r^{1:2}_{t-n+1,t})$. This proves (B.23) for $m = t-n+1$. Now assume that (B.23) is true for some m, $t-n+1 \leq m < t-1$. We show that this implies that (B.23) is also true for $m+1$. For $k = 1, 2$,

$$\begin{aligned} Y^k_{m+1} &= h^k_{m+1}(X_m, W^k_{m+1}) \\ U^k_{m+1} &= r^k_{m+1,t}(Y^k_{t-n+1:m+1}, U^k_{t-n+1:m}) \end{aligned}$$

Furthermore, by the system dynamics,

$$X_{m+1} = f_t(X_m, U^{1:2}_{m+1}, V_{m+1})$$

Thus, $(X_{m+1}, Y^{1:2}_{m+1}, U^{1:2}_{m+1})$ is a deterministic function of

$$(X_m, Y^{1:2}_{t-n+1:m}, U^{1:2}_{t-n+1:m}, W^{1:2}_{m+1}, V_{m+1}, r^{1:2}_{m+1,t})$$

Combining this with the induction hypothesis, we conclude that $(X_{t-n+1:m+1}, Y^{1:2}_{t-n+1:m+1}, U^{1:2}_{t-n+1:m+1})$ is a function of $(X_{t-n}, W^{1:2}_{t-n+1:m+1}, V_{t-n+1:m+1}, r^{1:2}_{t-n+1:m+1,t})$. Thus, by induction (B.23) is true for $t-n+1 \leq m \leq t-1$.

Now we use (B.23) to prove the lemma. For $k = 1, 2$

$$\begin{aligned} Y^k_t &= h^k_t(X_{t-1}, W^k_t) \\ r^k_t &= r^k_{t-n+1:t-1,t} \end{aligned}$$

Combining this with (B.23) for $m = t-1$ implies that there exists a deterministic function $\hat{D}_t$ such that

$$(X_{t-n+1:t-1}, Y^{1:2}_{t-n+1:t}, U^{1:2}_{t-n+1:t-1}) \quad = \quad \hat{D}_t(X_{t-n}, V_{t-n+1:t-1}, W^{1:2}_{t-n+1:t}, r^{1:2}_t)$$

This implies that there exists a function D_t such that Lemma B.2 is true. $\square$

Now consider

$$\begin{aligned}\Pi_t(s_t) &:= \mathbb{P}^{\psi}(S_t = s_t | C_t, \gamma^1_{1:t-1}, \gamma^2_{1:t-1}) \\ &= \sum \mathbb{1}_{s_t}\{D_t(x_{t-n}, v_{t-n+1:t-1}, w^{1:2}_{t-n+1:t}, \tilde{r}^{1:2}_t\} \\ &\quad \cdot \mathbb{P}(x_{t-n}, v_{t-n+1:t-1}, w^{1:2}_{t-n+1:t}, \tilde{r}^{1:2}_t | C_t, \gamma^{1:2}_{1:t-1})\end{aligned} \tag{B.24}$$

where the summation is over all choices of $(x_{t-n}, v_{t-n+1:t-1}, w^{1:2}_{t-n+1:t}, \tilde{r}^{1:2}_t)$. The vectors $r^{1:2}_t$ are completely determined by C_t and $\gamma^{1:2}_{1:t-1}$; the noise random variables $v_{t-n+1:t-1}, w^{1:2}_{t-n+1:t}$ are independent of the conditioning terms and X_{t-n}. Therefore, we can write (B.24) as

$$\begin{aligned}&\sum \mathbb{1}_{s_t}\{D_t(x_{t-n}, v_{t-n+1:t-1}, w^{1:2}_{t-n+1:t}, \tilde{r}^1_t, \tilde{r}^2_t)\} \\ &\quad \cdot \mathbb{P}(v_{t-n+1:t-1}, w^{1:2}_{t-n+1:t}) \cdot \mathbb{1}_{\tilde{r}^1_t, \tilde{r}^2_t}(r^1_t, r^2_t) \\ &\quad \cdot \mathbb{P}(x_{t-n} | C_t, \gamma^1_{1:t-1}, \gamma^2_{1:t-1})\end{aligned} \tag{B.25}$$

In the last term of (B.25), we dropped $\gamma^{1:2}_{1:t-1}$ from the conditioning terms because they are functions of C_t. The last term is therefore same as $\mathbb{P}(x_{t-n}|C_t) = \Theta_t$. Thus, Π_t is a function of Θ_t and r^1_t, r^2_t, thereby proving (B.20).

Part 3)

Consider the LHS of (3.45)

$$\begin{aligned}
&\mathbb{P}(\Theta_{t+1} = \theta_{t+1}, r^{1:2}_{t+1} = \tilde{r}^{1:2}_{t+1}, |c_t, \theta_{1:t}, \tilde{\gamma}^{1:2}_{1:t}, \tilde{r}^{1:2}_{1:t}) \\
&= \sum_{z_{t+1}} \mathbb{1}_{\theta_{t+1}}(Q_{t+1}(\theta_t, z_{t+1})) \cdot \mathbb{1}_{\tilde{r}^1_{t+1}}(Q^1_{t+1}(\tilde{r}^1_t, \tilde{\gamma}^1_t, z_{t+1})) \\
&\quad \cdot \mathbb{1}_{\tilde{r}^2_{t+1}}(Q^2_{t+1}(\tilde{r}^2_t, \tilde{\gamma}^2_t, z_{t+1})) \\
&\quad \cdot \mathbb{P}(Z_{t+1} = z_{t+1} | c_t, \tilde{\gamma}^{1:2}_{1:t}, \tilde{r}^1_{1:t}, \tilde{r}^2_{1:t})
\end{aligned} \tag{B.26}$$

The last term of (B.26) can be written as

$$\begin{aligned}
&\mathbb{P}(Z_{t+1} = z_{t+1} | c_t, \tilde{\gamma}^{1:2}_{1:t}, \tilde{r}^1_{1:t}, \tilde{r}^2_{1:t}) \\
&= \sum_{s_t} \mathbb{1}_{\hat{h}_t(s_t)}(z_{t+1}) \cdot \mathbb{P}(S_t = s_t | c_t, \tilde{\gamma}^{1:2}_{1:t}, \tilde{r}^1_{1:t}, \tilde{r}^2_{1:t}) \\
&= \sum_{s_t} \mathbb{1}_{\hat{h}_t(s_t)}(z_{t+1}) \cdot \mathbb{P}(S_t = s_t | c_t) \\
&= \sum_{s_t} \mathbb{1}_{\hat{h}_t(s_t)}(z_{t+1}) \cdot \pi_t(s_t) \\
&= \sum_{s_t} \mathbb{1}_{\hat{h}_t(s_t)}(z_{t+1}) \cdot H_t(\theta_t, \tilde{r}^1_t, \tilde{r}^2_t)(s_t)
\end{aligned} \tag{B.27}$$

Substituting (B.27) back in (B.26), we get

$$\begin{aligned}
&\mathbb{P}(\Theta_{t+1} = \theta_{t+1}, r^{1:2}_{t+1} = \tilde{r}^{1:2}_{t+1} | c_t, \theta_{1:t}, \tilde{\gamma}^{1:2}_{1:t}, \tilde{r}^{1:2}_{1:t}) \\
&= \sum_{z_{t+1}, s_t} \mathbb{1}_{\theta_{t+1}}(Q_{t+1}(\theta_t, z_{t+1})) \\
&\quad \cdot \mathbb{1}_{\tilde{r}^1_{t+1}}(Q^1_{t+1}(\tilde{r}^1_t, \tilde{\gamma}^1_t, z_{t+1})) \\
&\quad \cdot \mathbb{1}_{\tilde{r}^2_{t+1}}(Q^2_{t+1}(\tilde{r}^2_t, \tilde{\gamma}^2_t, z_{t+1})) \\
&\quad \cdot \mathbb{1}_{\hat{h}_t(s_t)}(z_{t+1}) \cdot H_t(\theta_t, \tilde{r}^1_t, \tilde{r}^2_t)(s_t) \\
&= \mathbb{P}(\Theta_{t+1} = \theta_{t+1}, r^{1:2}_{t+1} = \tilde{r}^{1:2}_{t+1} | \theta_t, \tilde{r}^{1:2}_t, \tilde{\gamma}^{1:2}_t)
\end{aligned} \tag{B.28}$$

thereby proving (3.45).

APPENDIX C

Appendix for General Sequential Problem

Proof of Proposition IV.10

Fix a coordinator strategy $\boldsymbol{\psi}$. Let $1 \le t \le T-1$. Consider a realization c_{t+1} of the common information C_{t+1}. Let $(\tilde{\gamma}_{1:t})$ be the corresponding realization of partial functions until time t. Assume that the realization $(c_{t+1}, \pi_{1:t}, \tilde{\gamma}_{1:t})$ has non-zero probability. Then, the realization π_{t+1} of Π_{t+1} is given by

$$\begin{aligned}
&\pi_{t+1}(q, u_1, u_2, \dots, u_t) \\
&= \mathbb{P}^{\boldsymbol{\psi}}(Q = q, U_1 = u_1, U_2 = u_2, \dots, U_t = u_t | c_{t+1}, \tilde{\gamma}_{1:t}) \\
&= \mathbb{P}^{\boldsymbol{\psi}}(Q = q, U_1 = u_1, U_2 = u_2, \dots, U_t = u_t | c_t, z_{t+1}, \tilde{\gamma}_{1:t}) && \text{(C.1)} \\
&= \frac{\mathbb{P}^{\boldsymbol{\psi}}(Q = q, U_1 = u_1, U_2 = u_2, \dots, U_t = u_t, Z_{t+1} = z_{t+1} | c_t, \tilde{\gamma}_{1:t})}{\sum_{q', u'_1, .., u'_t} \mathbb{P}^{\boldsymbol{\psi}}(Q = q', U_1 = u'_1, \dots, U_t = u'_t, Z_{t+1} = z_{t+1} | c_t, \tilde{\gamma}_{1:t})} && \text{(C.2)}
\end{aligned}$$

where we used the fact that c_{t+1} consists of c_t and z_{t+1} in (C.1) and Bayes' rule in (C.2). Focusing on the numerator in (C.2) gives

$$
\begin{aligned}
&\mathbb{P}^{\psi}(Q = q, U_1 = u_1, U_2 = u_2, \dots, U_t = u_t, Z_{t+1} = z_{t+1}|c_t, \tilde{\gamma}_{1:t}) \\
&= \mathbb{P}(Z_{t+1} = z_{t+1}|Q = q, U_1 = u_1, U_2 = u_2, \dots, U_t = u_t) \\
&\cdot \mathbb{P}^{\psi}(Q = q, U_1 = u_1, U_2 = u_2, \dots, U_t = u_t|c_t, \tilde{\gamma}_{1:t}),
\end{aligned}
\tag{C.3}
$$

where we used the fact that the random vector Z_{t+1} (being a part of I_{t+1}) is a known measurable function of the primitive random variables and the decisions $U_{1:t}$. The above expression can be further written as

$$
\begin{aligned}
&\mathbb{P}(Z_{t+1} = z_{t+1}|Q = q, U_1 = u_1, U_2 = u_2, \dots, U_t = u_t) \\
&\cdot \mathbb{P}(U_t = u_t|Q = q, U_1 = u_1, U_2 = u_2, \dots, U_{t-1} = u_{t-1}, c_t, \tilde{\gamma}_{1:t}) \\
&\cdot \mathbb{P}^{\psi}(Q = q, U_1 = u_1, U_2 = u_2, \dots, U_{t-1} = u_{t-1}|c_t, \tilde{\gamma}_{1:t}) \\
&= \mathbb{P}(Z_{t+1} = z_{t+1}|Q = q, U_1 = u_1, U_2 = u_2, \dots, U_t = u_t) \\
&\cdot \mathbb{P}(U_t = u_t|Q = q, U_1 = u_1, U_2 = u_2, \dots, U_{t-1} = u_{t-1}, \tilde{\gamma}_t) \\
&\cdot \mathbb{P}^{\psi}(Q = q, U_1 = u_1, U_2 = u_2, \dots, U_{t-1} = u_{t-1}|c_t, \tilde{\gamma}_{1:t-1})
\end{aligned}
\tag{C.4}
$$

where we used the fact that $U_t = \tilde{\gamma}_t(P_t)$ and P_t is a known measurable function of the primitive random variables and the decisions $U_{1:t-1}$ to simplify the second term in (C.4). Also, we dropped $\tilde{\gamma}_t$ from the conditioning terms in the third term of (C.4) since it is a fixed function of the remaining terms in the conditioning under the given coordination strategy. Recognizing the last term in (C.4) to be $\pi_t(q, u_1, \dots, u_t)$, we get the following expression for the numerator of (C.2)

$$
\begin{aligned}
&\mathbb{P}(Z_{t+1} = z_{t+1} | Q = q, U_1 = u_1, U_2 = u_2, \ldots, U_t = u_t) \\
&\cdot \mathbb{P}(U_t = u_t | Q = q, U_1 = u_1, U_2 = u_2, \ldots, U_{t-1} = u_{t-1}, \tilde{\gamma}_t) \\
&\cdot \pi_t(q, u_1, \ldots, u_t) \qquad \text{(C.5)}
\end{aligned}
$$

Similar expressions hold for the denominator terms in (C.2). Thus, we get that

$$
\pi_{t+1} = F_{t+1}(\pi_t, \tilde{\gamma}_t, z_{t+1})
$$

where $F_{t+1}(\cdot)$ is given by (C.2), (C.3), and (C.5).

Finally, for π_{T+1} we have

$$
\begin{aligned}
&\pi_{T+1}(q, u_1, u_2, \ldots, u_T) \\
&= \mathbb{P}^{\psi}(Q = q, U_1 = u_1, U_2 = u_2, \ldots, U_T = u_T | c_T, \tilde{\gamma}_{1:T}) \\
&= \mathbb{P}(U_T = u_T | Q = q, U_1 = u_1, U_2 = u_2, \ldots, U_{T-1} = u_{T-1}, c_T, \tilde{\gamma}_{1:T}) \\
&\cdot \mathbb{P}^{\psi}(Q = q, U_1 = u_1, U_2 = u_2, \ldots, U_{T-1} = u_{T-1} | c_T, \tilde{\gamma}_{1:T}) \\
&= \mathbb{P}(U_T = u_T | Q = q, U_1 = u_1, U_2 = u_2, \ldots, U_{T-1} = u_{T-1}, \tilde{\gamma}_T) \\
&\cdot \mathbb{P}^{\psi}(Q = q, U_1 = u_1, U_2 = u_2, \ldots, U_{T-1} = u_{T-1} | c_T, \tilde{\gamma}_{1:T-1}) \qquad \text{(C.6)}
\end{aligned}
$$

where we used the fact that $U_T = \tilde{\gamma}_T(P_T)$ and P_T is a known measurable function of the primitive random variables and the decisions $U_{1:T-1}$ to simplify the second term in (C.6). Also, we dropped $\tilde{\gamma}_T$ from the conditioning terms in the third term of (C.6) since it is a fixed function of the remaining terms in the conditioning under the given coordination strategy. Recognizing the last term in (C.6) to be $\pi_T(q, u_1, \ldots, u_{T-1})$, we get

$$
\begin{aligned}
&\pi_{T+1}(q, u_1, u_2, \dots, u_T) \\
&\mathbb{P}(U_T = u_T | Q = q, U_1 = u_1, U_2 = u_2, \dots, U_{T-1} = u_{T-1}, \tilde{\gamma}_T) \cdot \pi_T(q, u_1, \dots, u_{T-1})
\end{aligned}
\tag{C.7}
$$

Thus,

$$\pi_{T+1} = F_{T+1}(\pi_T, \tilde{\gamma}_T)$$

Proof of Proposition IV.11

Fix a coordinator strategy $\boldsymbol{\psi}$. Let $1 \leq t \leq T-1$. Consider a realization c_{t+1} of the common information C_{t+1}. Let $\tilde{\gamma}_{1:t}$ be the corresponding realization of partial functions until time t. Assume that the realization $(c_{t+1}, \pi_{1:t}, \tilde{\gamma}_{1:t})$ has non-zero probability. Consider,

$$
\begin{aligned}
&\mathbb{P}^{\boldsymbol{\psi}}(\Pi_{t+1} = \pi_{t+1} | c_t, \pi_{1:t}, \tilde{\gamma}_{1:t}) \\
&= \mathbb{P}^{\boldsymbol{\psi}}(F_{t+1}(\pi_t, \tilde{\gamma}_t, Z_{t+1}) = \pi_{t+1} | c_t, \pi_{1:t}, \tilde{\gamma}_{1:t}) \\
&= \sum_{z_{t+1}} \mathbb{1}_{\pi_{t+1}}(F_{t+1}(\pi_t, \tilde{\gamma}_t, z_{t+1})) \mathbb{P}^{\boldsymbol{\psi}}(Z_{t+1} = z_{t+1} | c_t, \pi_{1:t}, \tilde{\gamma}_{1:t})
\end{aligned}
\tag{C.8}
$$

Focusing on the conditional probability of Z_{t+1} in the above summation, we have

$$
\begin{aligned}
&\mathbb{P}^{\boldsymbol{\psi}}(Z_{t+1} = z_{t+1} | c_t, \pi_{1:t}, \tilde{\gamma}_{1:t}) \\
&= \sum_{q, u_1, .., u_t} \mathbb{P}^{\boldsymbol{\psi}}(Q = q, U_1 = u_1, U_2 = u_2, \dots, U_t = u_t, Z_{t+1} = z_{t+1} | c_t, \pi_{1:t}, \tilde{\gamma}_{1:t}) \\
&= \sum_{q, u_1, .., u_t} \mathbb{P}(Z_{t+1} = z_{t+1} | Q = q, U_1 = u_1, U_2 = u_2, \dots, U_t = u_t) \\
&\cdot \mathbb{P}^{\boldsymbol{\psi}}(Q = q, U_1 = u_1, U_2 = u_2, \dots, U_t = u_t | c_t, \pi_{1:t}, \tilde{\gamma}_{1:t}),
\end{aligned}
\tag{C.9}
$$

where we used the fact that the random vector Z_{t+1} (being a part of I_{t+1}) is a known measurable function of the primitive random variables and the decisions $U_{1:t}$. Each term in the above summation can be further written as

$$
\begin{aligned}
&\mathbb{P}(Z_{t+1} = z_{t+1} | Q = q, U_1 = u_1, U_2 = u_2, \ldots, U_t = u_t) \\
&\cdot \mathbb{P}(U_t = u_t | Q = q, U_1 = u_1, U_2 = u_2, \ldots, U_{t-1} = u_{t-1}, \tilde{\gamma}_t) \\
&\cdot \mathbb{P}^{\psi}(Q = q, U_1 = u_1, U_2 = u_2, \ldots, U_{t-1} = u_{t-1} | c_t, \tilde{\gamma}_{1:t-1})
\end{aligned}
\tag{C.10}
$$

where we used the fact that $U_t = \tilde{\gamma}_t(P_t)$ and P_t is a known measurable function of the primitive random variables and the decisions $U_{1:t-1}$ to simplify the second term in (C.10). Also, we dropped $\tilde{\gamma}_t$ from the conditioning terms in the third term of (C.10) since it is a fixed function of the remaining terms in the conditioning under the given coordination strategy. Recognizing the last term in (C.10) to be $\pi_t(q, u_1, \ldots, u_{t-1})$ gives the following simplification of (C.10)

$$
\begin{aligned}
&\mathbb{P}(Z_{t+1} = z_{t+1} | Q = q, U_1 = u_1, U_2 = u_2, \ldots, U_t = u_t) \\
&\cdot \mathbb{P}(U_t = u_t | Q = q, U_1 = u_1, U_2 = u_2, \ldots, U_{t-1} = u_{t-1}, \tilde{\gamma}_t) \cdot \pi_t(q, u_1, \ldots, u_{t-1})
\end{aligned}
\tag{C.11}
$$

Using expressions (C.9), (C.10) and (C.11) in (C.8), we get

$$
\begin{aligned}
&\mathbb{P}^{\psi}(\Pi_{t+1} = \pi_{t+1} | c_t, \pi_{1:t}, \tilde{\gamma}_{1:t}) \\
&= \sum_{z_{t+1}} \mathbb{1}_{\pi_{t+1}}(F_{t+1}(\pi_t, \tilde{\gamma}_t, z_{t+1})) \\
&\cdot \sum_{q, u_1, .., u_t} \mathbb{P}(Z_{t+1} = z_{t+1} | Q = q, U_1 = u_1, U_2 = u_2, \ldots, U_t = u_t) \\
&\cdot \mathbb{P}(U_t = u_t | Q = q, U_1 = u_1, U_2 = u_2, \ldots, U_{t-1} = u_{t-1}, \tilde{\gamma}_t) \cdot \pi_t(q, u_1, \ldots, u_{t-1})
\end{aligned}
\tag{C.12}
$$

Among all the conditioning terms in the left hand side of equation (C.12), the right hand side of equation (C.12) depends only on the realization π_t and $\tilde{\gamma}_t$. Also, it should be noted that none of the terms on the right hand side of (C.12) depend on $\boldsymbol{\psi}$. Thus, we can conclude that

$$\begin{aligned}
&\mathbb{P}^{\boldsymbol{\psi}}(\Pi_{t+1} = \pi_{t+1} | c_t, \pi_{1:t}, \tilde{\gamma}_{1:t}) \\
&= \mathbb{P}(\Pi_{t+1} = \pi_{t+1} | \pi_t, \tilde{\gamma}_t).
\end{aligned} \tag{C.13}$$

Similar arguments establish the Markovian property at time $T+1$.

Finally, we note that the expected reward can be written as

$$\mathbb{E}\{R(Q, U_1, U_2, \ldots, U_T)\} = \mathbb{E}\big[\mathbb{E}\{R(Q, U_1, U_2, \ldots, U_T) | C_T, \gamma_{1:T}\}\big] \tag{C.14}$$

The inner expectation in (C.14) can be written as:

$$\begin{aligned}
&\sum_{q, u_1, ..u_T} R(q, u_1, u_2, .., u_T) \mathbb{P}(Q = q, U_1 = u_1, U_2 = u_2, \ldots, U_T = u_T | C_T, \gamma_{1:T}) \\
&= \sum_{q, u_1, ..u_T} R(q, u_1, u_2, .., u_T) \Pi_{T+1}(q, u_1, u_2, \ldots, u_T) \\
&=: \tilde{R}(\Pi_{T+1})
\end{aligned} \tag{C.15}$$

Thus,

$$\mathbb{E}\{R(Q, U_1, U_2, \ldots, U_T)\} = \mathbb{E}\{\tilde{R}(\Pi_{T+1})\}. \tag{C.16}$$

APPENDIX D

Appendix for Sequential Problems with State

Proof of Proposition V.3

Fix a coordinator strategy $\boldsymbol{\psi}$. Let $1 \leq t \leq T-1$. Consider a realization c_{t+1} of the common information C_{t+1}. Let $\tilde{\gamma}_{1:t}$ be the corresponding realization of partial functions until time t. Assume that the realization $(c_{t+1}, \pi_{1:t}, \tilde{\gamma}_{1:t})$ has non-zero probability. Then, the realization π_{t+1} of Π_{t+1} is given by

$$
\begin{aligned}
&\pi_{t+1}(l,p) \\
&= \mathbb{P}^{\boldsymbol{\psi}}(L_{t+1} = l, P_{t+1} = p | c_{t+1}, \tilde{\gamma}_{1:t}) \\
&= \mathbb{P}^{\boldsymbol{\psi}}(L_{t+1} = l, P_{t+1} = p | c_t, z_{t+1}, \tilde{\gamma}_{1:t}) && \text{(D.1)} \\
&= \frac{\mathbb{P}^{\boldsymbol{\psi}}(L_{t+1} = l, P_{t+1} = p, Z_{t+1} = z_{t+1} | c_t, \tilde{\gamma}_{1:t})}{\sum_{l',p'} \mathbb{P}^{\boldsymbol{\psi}}(L_{t+1} = l', P_{t+1} = p', Z_{t+1} = z_{t+1} | c_t, \tilde{\gamma}_{1:t})} && \text{(D.2)}
\end{aligned}
$$

where we used the fact that c_{t+1} consists of c_t and z_{t+1} in (D.1) and Bayes' rule in (D.2). Focusing on the numerator in (D.2) gives

$$
\begin{aligned}
&\mathbb{P}^{\psi}(L_{t+1} = l, P_{t+1} = p, Z_{t+1} = z_{t+1} | c_t, \tilde{\gamma}_{1:t}) \\
&= \sum_{l',p',u} \mathbb{P}^{\psi}(L_{t+1} = l, P_{t+1} = p, Z_{t+1} = z_{t+1}, L_t = l', P_t = p', U_t = u | c_t, \tilde{\gamma}_{1:t}) \\
&= \sum_{l',p',u} \mathbb{P}^{\psi}(L_{t+1} = l, P_{t+1} = p, Z_{t+1} = z_{t+1} | L_t = l', P_t = p', U_t = u, c_t, \tilde{\gamma}_{1:t}) \\
&\times \mathbb{P}^{\psi}(L_t = l', P_t = p', U_t = U_t | c_t, \tilde{\gamma}_{1:t}) \\
&= \sum_{l',p',u} \mathbb{P}^{\psi}(L_{t+1} = l, P_{t+1} = p, Z_{t+1} = z_{t+1} | L_t = l', P_t = p', U_t = u) \\
&\times \mathbb{P}^{\psi}(L_t = l', P_t = p', U_t = U_t | c_t, \tilde{\gamma}_{1:t}) \qquad \text{(D.3)}
\end{aligned}
$$

where we used Condition 1 of the definition of state sufficient for input output map in the first term of equation (D.3). Since the conditioning terms include L_t, P_t, U_t, Condition 1 allows us to remove the remaining terms that are functions of $Q_{1:t}, U_{1:t}$. The above expression can be further written as

$$
\begin{aligned}
&\sum_{l',p',u} \mathbb{P}^{\psi}(L_{t+1} = l, P_{t+1} = p, Z_{t+1} = z_{t+1} | L_t = l', P_t = p', U_t = u) \\
&\times \mathbb{1}_u\{\tilde{\gamma}_t(p')\} \mathbb{P}^{\psi}(L_t = l', P_t = p' | c_t, \tilde{\gamma}_{1:t}) \qquad \text{(D.4)} \\
&= \sum_{l',p',u} \mathbb{P}^{\psi}(L_{t+1} = l, P_{t+1} = p, Z_{t+1} = z_{t+1} | L_t = l', P_t = p', U_t = u) \\
&\times \mathbb{1}_u\{\tilde{\gamma}_t(p')\} \mathbb{P}^{\psi}(L_t = l', P_t = p' | c_t, \tilde{\gamma}_{1:t-1}) \qquad \text{(D.5)}
\end{aligned}
$$

where we used the fact that $U_t = \tilde{\gamma}_t(P_t)$ in (D.4). Also, we dropped $\tilde{\gamma}_t$ from the conditioning terms in the third term of (D.5) since it is a fixed function of the remaining terms in the conditioning under the given coordination strategy. Recognizing the last term in (D.5) to be $\pi_t(l', p')$, we get the following expression for numerator in (D.2):

$$\sum_{l',p',u} \mathbb{P}^{\boldsymbol{\psi}}(L_{t+1} = l, P_{t+1} = p, Z_{t+1} = z_{t+1} | L_t = l', P_t = p', U_t = u) \\ \times \mathbb{1}_u\{\tilde{\gamma}_t(p')\}\pi(l'p') \tag{D.6}$$

Similar expressions hold for the denominator terms in (D.2). Thus, we get that

$$\pi_{t+1} = F_{t+1}(\pi_t, \tilde{\gamma}_t, z_{t+1})$$

where $F_{t+1}(\cdot)$ is given by (D.2), (D.3), and (D.6).

Proof of Proposition V.4

Fix a coordinator strategy $\boldsymbol{\psi}$. Let $1 \leq t \leq T-1$. Consider a realization c_{t+1} of the common information C_{t+1}. Let $\tilde{\gamma}_{1:t}$ be the corresponding realization of partial functions until time t. Assume that the realization $(c_{t+1}, \pi_{1:t}, \tilde{\gamma}_{1:t})$ has non-zero probability. Consider,

$$\begin{aligned} &\mathbb{P}^{\boldsymbol{\psi}}(\Pi_{t+1} = \pi_{t+1} | c_t, \pi_{1:t}, \tilde{\gamma}_{1:t}) \\ &= \mathbb{P}^{\boldsymbol{\psi}}(F_{t+1}(\pi_t, \tilde{\gamma}_t, Z_{t+1}) = \pi_{t+1} | c_t, \pi_{1:t}, \tilde{\gamma}_{1:t}) \\ &= \sum_{z_{t+1}} \mathbb{1}_{\pi_{t+1}}(F_{t+1}(\pi_t, \tilde{\gamma}_t, z_{t+1}))\mathbb{P}^{\boldsymbol{\psi}}(Z_{t+1} = z_{t+1} | c_t, \pi_{1:t}, \tilde{\gamma}_{1:t}) \end{aligned} \tag{D.7}$$

Focusing on the conditional probability of Z_{t+1} in the above summation, we have

$$\begin{aligned}
&\mathbb{P}^{\boldsymbol{\psi}}(Z_{t+1} = z_{t+1}|c_t, \pi_{1:t}, \tilde{\gamma}_{1:t}) \\
&= \sum_{l_t, p_t} \mathbb{P}^{\boldsymbol{\psi}}(Z_{t+1} = z_{t+1}|L_t = l_t, P_t = p_t, c_t, \pi_{1:t}, \tilde{\gamma}_{1:t}) \\
&\times \mathbb{P}^{\boldsymbol{\psi}}(L_t = l_t, P_t = p_t|c_t, \pi_{1:t}, \tilde{\gamma}_{1:t}) \\
&= \sum_{l_t, p_t} \mathbb{P}^{\boldsymbol{\psi}}(Z_{t+1} = z_{t+1}|L_t = l_t, P_t = p_t, c_t, \pi_{1:t}, \tilde{\gamma}_{1:t}) \\
&\times \pi_t(l_t, p_t) \qquad \text{(D.8)}
\end{aligned}$$

where we used the definition of π_t in (D.8). Since $U_t = \gamma_t(P_t)$, we can add U_t in the list of conditioning terms in (D.8) to get

$$\sum_{l_t, p_t} \mathbb{P}^{\boldsymbol{\psi}}(Z_{t+1} = z_{t+1}|L_t = l_t, P_t = p_t, c_t, \pi_{1:t}, \tilde{\gamma}_{1:t}, U_t = \tilde{\gamma}_t(p_t))\pi_t(l_t, p_t) \qquad \text{(D.9)}$$

A direct consequence of Condition 1 is that

$$\mathbb{P}^{\boldsymbol{\psi}}(Z_{t+1}|Q_{1:t}, U_{1:t}) = \mathbb{P}(Z_{t+1}|S_t, U_t) \qquad \text{(D.10)}$$

Using equation (D.10) in (D.9), we get

$$\sum_{l_t, p_t} \mathbb{P}(Z_{t+1} = z_{t+1}|L_t = l_t, P_t = p_t, U_t = \tilde{\gamma}_t(p_t))\pi_t(l_t, p_t) \qquad \text{(D.11)}$$

Since the conditioning terms in (D.9) include L_t, P_t, U_t, (D.10) allowed us to remove the remaining terms that are functions of $Q_{1:t}, U_{1:t}$. Using (D.11) in (D.7), we get

$$
\begin{aligned}
&\mathbb{P}^{\boldsymbol{\psi}}(\Pi_{t+1} = \pi_{t+1} | c_t, \pi_{1:t}, \tilde{\gamma}_{1:t}) \\
&= \sum_{z_{t+1}} \mathbb{1}_{\pi_{t+1}}(F_{t+1}(\pi_t, \tilde{\gamma}_t, z_{t+1})) \\
&\cdot \sum_{l_t, p_t} \mathbb{P}(Z_{t+1} = z_{t+1} | L_t = l_t, P_t = p_t, U_t = \tilde{\gamma}_t(p_t))\pi_t(l_t, p_t) \qquad \text{(D.12)}
\end{aligned}
$$

Among all the conditioning terms in the left hand side of equation (D.12), the right hand side of equation (D.12) depends only on the realization π_t and $\tilde{\gamma}_t$. Also, it should be noted that none of the terms on the right hand side of (D.12) depend on $\boldsymbol{\psi}$. Thus, we conclude that

$$
\begin{aligned}
&\mathbb{P}^{\boldsymbol{\psi}}(\Pi_{t+1} = \pi_{t+1} | c_t, \pi_{1:t}, \tilde{\gamma}_{1:t}) \\
&= \mathbb{P}(\Pi_{t+1} = \pi_{t+1} | \pi_t, \tilde{\gamma}_t). \qquad \text{(D.13)}
\end{aligned}
$$

Finally, we note that the expected reward at time t can be written as

$$
\begin{aligned}
&\mathbb{E}\{R_t(S_t, U_t)\} = \mathbb{E}\{R_t(L_t, P_t, U_t)\} = \mathbb{E}\{R_t(L_t, P_t, \gamma_t(P_t))\} \qquad \text{(D.14)} \\
&= \mathbb{E}[\mathbb{E}\{R_t(L_t, P_t, \gamma_t(P_t)) | C_t, \gamma_{1:t}\}] \qquad \text{(D.15)}
\end{aligned}
$$

The inner expectation in (D.15) can be written as:

$$
\begin{aligned}
&\sum_{l,p} R_t(l, p, \gamma_t(p))\mathbb{P}(L_t = l, P_t = p | C_t, \gamma_{1:t}) \\
&= \sum_{l,p} R_t(l, p, \gamma_t(p))\Pi_t(l, p) \\
&=: \tilde{R}_t(\Pi_t, \gamma_t) \qquad \text{(D.16)}
\end{aligned}
$$

Thus,

$$\mathbb{E}\{R_t(S_t, U_t))\} = \mathbb{E}\{\tilde{R}_t(\Pi_t, \gamma_t)\}. \tag{D.17}$$

BIBLIOGRAPHY

BIBLIOGRAPHY

A. Kh. Al Jabri, and S. Al-Issa (1997), Zero-error codes for correlated information sources, in *Proceedings of Cryptography*, pp. 17–22, Cirencester, UK.

Ahlswede, R., and T. S. Han (1983), On source coding with side information via a multiple- access channel and related problems in multi-user information theory, *IEEE Trans. Inform. Theory*, *29*, 396–412.

Aicardi, M., F. Davoli, and R. Minciardi (1987), Decentralized optimal control of Markov chains with a common past information set, *32*(11), 1028–1031.

Aumann, R. J. (1976), Agreeing to disagree, *Annals of Statistics*, *4*(6), 1236–39.

Berger, T., and S.Tung (1975), Encoding of correlated analog sources, *Proc. IEEEUSSR Joint Workshop on Information Theory*, pp. 7–10.

Berger, T., Z. Zhang, and H. Viswanathan (1996), The CEO problem, *IEEE Trans. Inform. Theory*, *42*, 887–902.

Bertsekas, D. P. (1976), *Dynamic Programming and Stochastic Control*, Academic Press.

Borkar, V. S., S. K. Mitter, and S. Tatikonda (2001), Optimal sequential vector quantization of markov sources, *SIAM Journal of Control Optim.*, *40*(1), 135–148.

Chen, J., X. Zhang, T. Berger, and S. B. Wicker (2004), An upper bound on the sum-rate distortion function and its corresponding rate allocation schemes for the ceo problem, *IEEE Journal on Selected Areas in Communications*, *22*(6).

Cover, T. M., A. El Gamal, and M. Salehi (1980), Multiple-access channel with arbitrarily correlated sources, *IEEE Trans. Inform. Theory*, pp. 648–659.

Csiszar, I., and J. Korner (1981), *Information Theory: Coding Theorems for Discrete Memoryless Systems*, Academic Press, London.

DeGroot, M. H. (1970), *Optimal Statistical Decisions*, New York: McGraw-Hill.

Draper, S. C., and G. W. Wornell (2002), Successively structured ceo problems, in *in Proc. ISIT 2002*, p. 65, Lausanne, Switzerland.

Flynn, T. J., and R. M. Gray (1987), Encoding of correlated observations, *IEEE Transactions on Information theory*, *IT-33*(6).

Gallager, R. G. (1968), *Information Theory and Reliable Communication*, New York: Wiley.

Gelfand, S. I., and M. S. Pinsker (1979), Coding of sources on the basis of observations with incomplete information, *Probl. Inform. Transmission*, *15*(2), 115–125.

Han, T. S., and K. Kobayashi (1987), A dichotomy of functions F(X,Y) of correlated sources (X,Y), *IEEE Trans. Inform. Theory*, *33*, 69–76.

Ho, Y. C., and K. C. Chu (1972), Team decision theory and information structures in optimal control problems-part 1, *IEEE Transaction on Automatic Control*, *17*(1), 15–22.

Ho, Y. C., M. Kastner, and E. Wong (1978), Teams, signaling and information theory, *IEEE Transaction on Automatic Control*, *23*(2), 305–312.

Korner, J., and K. Marton (1979), How to encode the modulo-two sum of binary sources, *IEEE Trans. Inform. Theory*, *25*, 219–221.

Krithivasan, D., and S. S. Pradhan (2007), Lattices for distributed source coding: Jointly gaussian sources and reconstruction of a linear function, *IEEE Trans. Inform. Theory*, submitted.

Kumar, P. R., and P. Varaiya (1986a), *Stochastic Systems: Estimation Identification and Adaptive Control*, Prentice Hall.

Kumar, P. R., and P. Varaiya (1986b), *Stochastic Systems: Estimation Identification and Adaptive Control*, Prentice Hall, Englewood Cliffs, NJ.

Kurtaran, B. (1976), Decentralized stochastic control with delayed sharing information pattern, *21*, 576–581.

Kurtaran, B. (1979), Corrections and extensions to "decentralized stochastic control with delayed sharing information pattern", *24*(4), 656–657.

Mahajan, A. (2008), Sequential decomposition of sequential dynamic teams: applications to real-time communication and networked control systems, Ph.D. thesis, University of Michigan.

Mahajan, A., and S. Tatikonda (2009), Sequential team form and its simplification using graphical models, in *Proceedings of 47th Annual Allerton Conference*, pp. 330–339.

Mahajan, A., and D. Teneketzis (2008), On the design of globally optimal communication strategies for real-time communication systems with noisy feedback, *IEEE J. Selected Areas in Communication*, *26*(4), 580–595.

Marschak, J., and R. Radner (1972), *Economic Theory of Teams*, Yale University Press, New Haven.

Nayyar, A., and D. Teneketzis (2008), On jointly optimal real-time encoding and decoding strategies in multiterminal communication systems, in *Proceedings of 47th IEEE Conference of Decision and Control*.

Oohama, Y. (1997), Gaussian multiterminal source coding, *IEEE Trans. Inform. Theory*, *43*, 1912–1923.

Oohama, Y. (1998), The rate-distortion function for the quadratic gaussian ceo problem, *IEEE Trans. Inform. Theory*, *44*, 1057–1070.

Pradhan, S. S., and K. Ramchandran (1999), Distributed source coding using syndromes (discus) design and construction, in *Proc.Data Compression Conf. (DCC)*, pp. 158–167, Snowbird, UT.

Radner, R. (1962), Team decision problems, *The Annals of Math. Statistics*, *33*(3), 857–881.

Slepian, D., and J. K. Wolf (1973a), Noiseless coding of correlated information sources, *IEEE Trans. Inform. Theory*, *IT-19*, 471–480.

Slepian, D., and J. K. Wolf (1973b), A coding theorem for multiple access channels with correlated sources, *Bell System Tech. Journal*, *52*, 1037–1076.

Smallwood, R. D., and E. J. Sondik (1973), The optimal control of partially observable Markov processes over a finite horizon, *Operations Research*, *11*, 1071–1088.

Teneketzis, D. (2006), On the structure of real-time encoders and decoders in noisy communication, *IEEE Trans. Inf. Theory*, *52*(9), 5–11.

Varaiya, P., and J. Walrand (1978), On delayed sharing patterns, *23*(3), 443–445.

Viswanathan, H., and T. Berger (1997), The quadratic gaussian ceo problem, *IEE Trans. Inform. Theory*, *43*, 1549–1559.

Walrand, J., and P. Varaiya (1983a), Optimal causal coding-decoding problems, *IEEE Trans. Inf. theory*, *IT-29*(6), 814–820.

Walrand, J. C., and P. Varaiya (1983b), Causal coding and control of markov chains, *Systems and Control Letters 3*, pp. 189–192.

Witsenhausen, H. (1973), A standard form for sequential stochastic control, *Mathematical Systems Theory*, *7*(1), 5–11.

Witsenhausen, H. (1978), On the structure of real-time source coders, *Bell Sys. Tech. J.*, *58*, 1437–1451.

Witsenhausen, H. S. (1968), A counterexample in stochastic optimum control, *SIAM Journal of Control and Optimization*, *6*(1), 131–147.

Witsenhausen, H. S. (1971), Separation of estimation and control for discrete time systems, *59*(11), 1557–1566.

Witsenhausen, H. S. (1976), Some remarks on the concept of state, in *Directions in Large-Scale Systems*, edited by Y. C. Ho and S. K. Mitter, pp. 69–75, Plenum.

Wyner, A. (1974), Recent results in the shannon theory, *IEEE Transactions on Information Theory*, pp. 2–10.

Y. C. Ho (1980), Team decision theory and information structures, in *Proceedings of the IEEE*, vol. 68, pp. 644–654.

Yuksel, S., and T. Basar (2008), Causal coding of markov sources with continuous alphabets, in *Analysis and Design of Nonlinear Control Systems: In Honor of Alberto Isidori Astolfi, Alessandro; Marconi, Lorenzo (Eds.)*, Springer.

Zamir, R., and T. Berger (1999), Multiterminal source coding with high resolution, *IEEE Trans. Inform. Theory*, *45*, 106–117.

Zhang, H. (2009), Partially observable markov decision processes: A geometric technique and analysis, *Operations Research*.

Zhao, Q., and M. Effros (2001), Optimal code design for lossless and near lossless source coding in multiple access networks, in *Proc.Data Compression Conf. (DCC)*, pp. 263–272, Snowbird, UT.