Sequential decomposition of sequential dynamic teams: applications to real-time communication and networked control systems

by

Aditya Mahajan

A dissertation submitted in the partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Electrical Engineering : Systems)
in the University of Michigan
2008

Doctoral Committee:

Professor Demosthenis Teneketzis, Chair
Professor Jeffrey C. Lagarias
Associate Professor Achilleas Anastasopoulos
Associate Professor S. Sandeep Pradhan

To my parents

Acknowledgement

I entered the Ph.D. program five years ago certain that I wanted to do research in communication and equally certain that I did not want to do anything related to control. I finish my Ph.D. uncertain about the distinction between control and communication and wondering why they are taught separately. I would like to express my sincerely gratitude to my advisor, Professor Demos Teneketzis, for inculcating this unified viewpoint of systems. His experience and unique perspective on the history and philosophy of various areas taught me more than what I could have learnt on my own. I would always cherish the thoughtful and engrossing discussions that I had with him.

I am grateful to my thesis committee for their guidance and encouragement. I thank Professor Achilleas Anastasopoulos for showing exciting about my work when I was uncertain about its merit. He helped me understand how real-time communication compares with other results in coding theory and also provided calm and rational advice on numerous occasions. I thank Professor Sandeep Pradhan for helping me explain the practical significance of my work. His insistence on understanding the vague statements in early drafts of my papers resulted in a cleaner exposition of my results. I am grateful to Professor Jeffrey Lagarias for providing various suggestions on improving the presentation of my work. I am also thankful to my other instructors at Michigan, Professors Stéphane Lafortune, Mingyan Liu, Serap Savari, and Kim Winick for enriching my academic experience.

I would like to thank Shrutivandana Sharma, Ashutosh Nayyar and David Shumann for many useful discussions and helpful suggestion. I am also thankful to my friends who made my stay in Ann Arbor a memorable experience, in particular, to Dinesh Krithivasan, Ramji Venkataramanan, Shivakumar Natarajan, Raghu Kainkaryam, Naveen Gupta, Manoj Rajagopalan, and Rohan Rao. I would also to thank everyone in the quizclub and the volleyball group for all the fun.

Finally, I wish to express my sincerest gratitude to my family who have always provided unconditional love, support, and encouragement. My parents took deep interest in my education and taught me how to think independently. Unfortunately, my father passed away in October, 2005 when I was working on this research. His unselfishness, sincerity, honesty, and hard work will always be an inspiration to me. My mother's sacrifices, strength and conviction during these years were invaluable in helping me complete my thesis. I thank my bother, Ankur, for always managing to make me smile.

Table of Contents

List of Figures

List of Tables

List of Acronyms

AWGN	additive white Gaussian noise
CPU	central processing unit
DMC	discrete memoryless channel
LHS	left hand side
LQG	linear quadratic Gaussian
MANET	mobile ad-hoc networks
MDP	Markov decision process
NCS	networked control system
PMF	probability mass function
POMDP	partially observable Markov decision process
RHS	right hand side
UAV	unmanned aerial vehicle

Abstract

Optimal design of multi-agent sequential teams is investigated in this thesis. A systematic methodology is presented to convert the search for an optimal multi-stage design into a sequence of nested optimization problems, where at each step the best decision rule of a agent at a given time is search. This conversion is called *sequential decomposition* and it drastically simplifies the search of optimal solution for both finite and infinite horizon problems. The main idea is as follows. A state sufficient for input-output mapping of the system is identified. A joint probability measure on this state is an information state sufficient for performance evaluation. This information state evolves in time in a deterministic manner depending on the choice of decision rules of the agents. Thus, these information states are a controlled Markov process where the control actions are the decision rules of the agents. The optimal control of the time-evolution of these information states results in a sequential decomposition of the problem. Applications of this methodology to real-time communication and optimal feedback control over noisy communication channels is also investigated.

Chapter 1

Introduction

1.1 Motivation

Decentralized systems arise in a variety of branches of engineering. Examples include the Internet, telecommunication networks, sensor networks, surveillance networks, monitoring and diagnostic systems, MANET (mobile ad-hoc networks), cognitive radio, control of UAVs (unmanned aerial vehicles), robotics, multi-core CPUs, etc. Most of these applications are independent areas of research with dedicated conferences and journals. However, from an abstract level, these applications have similar salient features and similar design difficulties. We believe that if we can capture these salient feature in a simple model and understand how to resolve the conceptual difficulties for that model, then these insights would provide design guidelines for these applications. This is the main premise of this thesis. We study a "simple" model of a decentralized system, and show how results for that model can help in optimally designing real-time communication systems and networked control systems.

The salient features of decentralized systems are as follows. Decentralized systems consist of multiple components (or agents); each component has partial information about the state of the system but there is no centralization of information, i.e., no agent knows the information available to all other agents. In many decentralized systems, all components/agents have a common objective: optimize the performance with respect to a system-wide objective (e.g., probability of correct detection with minimum energy consumption in sensor, surveillance, and UAV networks, congestion avoidance in transportation and telecommunication networks,

throughput in MANETs and telecommunication networks, etc.). The agents can exchange information with one another and coordinate their activities to achieve their objective.

The decentralization of information makes the design of decentralized systems drastically different from the design of centralized systems. In centralized partially observed systems, control has a ***dual aspect***, or function: (i) ***control***—the control action can alter the future values of the state of the system; and (ii) ***estimation***—the control action can alter the future information available to the control agent and hence affect the knowledge that the control agent has about the state of the system. In decentralized systems control has an additional function: (iii) ***communication***—the control action of an agent can alter the future information available to *other agents*, and affect the knowledge that other agents have about the state of the system. Thus, in decentralized systems control has a ***triple aspect***,[1] or function: control, estimation, and communication.

The communication aspect of control in decentralized systems is not well understood. The decentralization of information and the noisy nature of communication makes efficient communication extremely difficult. For example, when an agent communicates a message how can it ascertain that the receiver, which has different information, will interpret the message in the same way as its intended meaning? Moreover, the presence of multiple agents affects the control and estimation aspects of control. An agent has to take other agents' control strategies into account while determining how its control action alters the future state of the system and its own future information. In order to understand the design of decentralized systems, we need to understand the aforementioned three aspects of control. This thesis is an attempt in that direction.

The rest of this chapter is organized as follows. We begin by a classification of multi-agent systems and explain the class of systems that we will study in this thesis. We then explain the main ideas for the optimal design of dynamic teams, in particular, the notion of sequential decomposition and its advantages. We then describe the organization and present the contributions of the thesis.

[1] The term "triple aspect of control" or "triple control" is due to Pravin Varaiya. Ho (1980) used the term "signalling" for the third aspect of control.

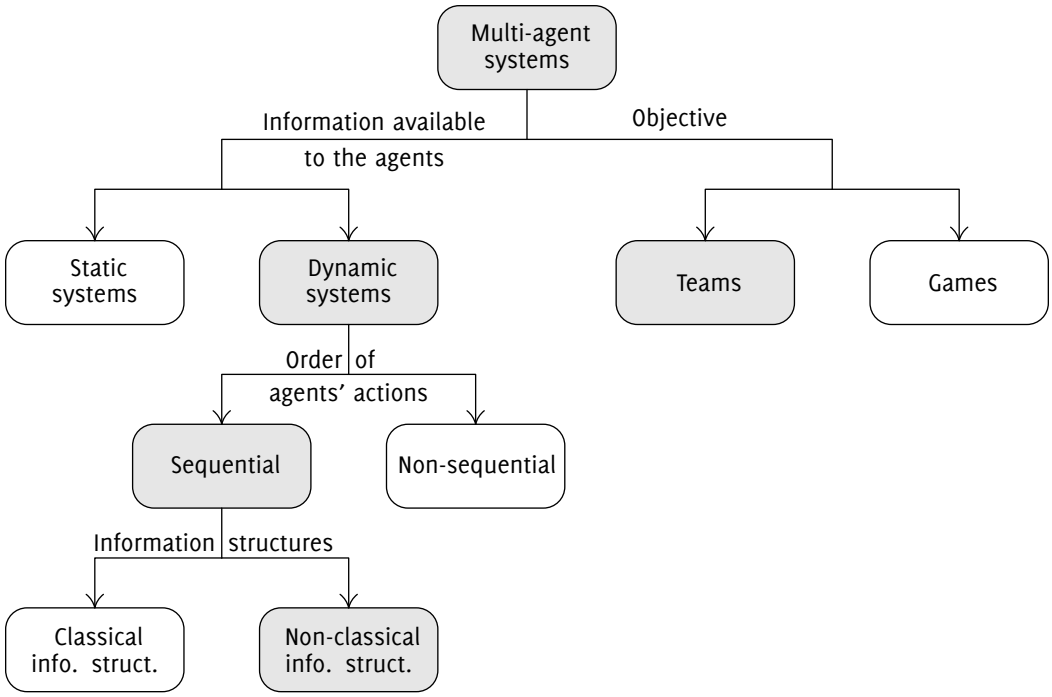


Figure 1.1: Classification of multi-agent systems. In this thesis we are interested in sequential dynamic teams with non-classical information structures.

1.2 Classification of multi-agent systems

Multi-agent systems can be classified either on the basis of the objective of the agents as teams and games, or on the basis of the information available to the agents as static and dynamic systems. Dynamic systems can be further classified as sequential and non-sequential. Sequential systems can be differentiated on the basis of their information structures. This classification systems is shown in Figure 1.1 and described in detail below.

Teams and games

Multi-agent systems can be classified as teams and games on the basis of the objective of the agents. In teams all agents have the same objective; in games, each agent has its own objective. Historically, games were first studied in the seminal work of von Neumann and Morgenstern (1944) and were later developed as a subfield of mathematical economics called game theory (Aumann and Hart, 1992, 1994, 2002). Teams were first studied in mathematical economics by Radner (1962) and Marschak and Radner (1972), and later in control systems by Witsenhausen (1971a, 1973), Ho et al. (1978), Ho (1980), and others.

Both teams and games have two solution philosophies: equilibrium solutions and optimal solutions. For each of these solution philosophies, the actual solution concept of teams and games are different due to the difference in the objectives of the agents.

A set of strategies of all agents is in equilibrium if no agent can improve its performance (which is same as the system performance in case of teams) by unilaterally changing its strategy. In teams, an equilibrium solution is called person by person optimal or member by member optimal solution. In games, there are various notions of equilibrium solutions, such as Nash equilibrium and its refinements, etc. For both teams and games, it is usually desirable to find all equilibrium designs.

A set of strategies of all agents is optimal if no other design gives a better performance to all agents. In teams, an optimal solution is also called globally optimal solution (to contrast it with person by person optimal solution which can be thought of as a locally optimal solution). In games, optimal solutions are generally called Pareto optimal in honor of Vilfredo Pareto, who used the concept in his studies of economic efficiency and income distribution. In teams it is usually sufficient to find one optimal design while in games it is desirable to find the Pareto frontier, which is the set of all Pareto optimal designs.

Static and dynamic systems

Multi-agent systems, both teams and games, can be classified on the basis of the information available to the agents into static and dynamic systems. The distinction is based on primitive random variables, which are random variables that represent the randomness generated by nature and are usually assumed to be mutually independent. If the observations of all agents depend only on primitive random variables then the system is called ***static***; if the observations of some agents also depend on the decision rule of any agent that acted in the past (including the agent whose observations we are interested in) then the system is called ***dynamic***. Static systems are also called ***single-stage*** systems, while dynamic systems are also called ***multi-stage***.

The search for an optimal design of static systems is called ***static optimization***. Examples include linear programming, non-linear programming, and convex optimization techniques. The search for an optimal design of dynamic systems is called

dynamic optimization. Examples include dynamic programming (backward induction) and forward induction.

Sequential and non-sequential systems

Multi-stage systems can be further classified into sequential and non-sequential systems. In ***sequential*** systems the order in which agents act does not depend on events in nature and the actions taken by the agents. Thus, the order of agents' actions can be fixed before the system starts operating and agents act in the same order along all behaviors (sample paths) of the systems. In ***non-sequential*** systems the order in which agents act depends on events in nature and actions taken by the agents. Thus, the order in which agents act cannot be determined before the system starts operating and the agents act in different order along different behaviors (sample paths) of the system.

The distinction between sequential and non-sequential systems has been explained in Witsenhausen (1971b, 1975). Optimal design of sequential systems was investigated in Witsenhausen (1973). Properties and design of non-sequential systems was investigated in Witsenhausen (1971b, 1975), Andersland (1991), Andersland and Teneketzis (1992, 1994), Teneketzis (1996) and Teneketzis and Andersland (2000).

Classical and non-classical information structures

Sequential systems can be further classified on the basis of their information structure (also called information pattern). Information structure is the set of data available to each agent to make a decision. If each agent knows everything that was known to all agents that acted before it, the system has a ***classical information structure***. If the information structure can be converted into a classical information structure by a change of variables, it is called ***quasi-classical***. An information structure that is neither classical nor quasi-classical is called a (strictly) ***non-classical*** information structure.

The importance of information structures was first highlighted in Witsenhausen (1971a). The role information structures in specific team problems was explored in Ho and Chu (1972), Chu (1972) Yoshikawa (1978), Ho (1980), and others.

To understand the simplification in determining optimal designs provided by a classical information structure, consider the simplest system with a non-classical

information structure— a one agent system that *does not* have perfect recall, i.e., the agent *does not* remember everything that it has seen in the past and everything that it has done in the past. In this case, at any time the agent can try to communicate to its future self through the state of the system. In a system with perfect recall, i.e., in a one agent system with perfect recall, the future self of the agent will know everything that the agent currently knows, so the agent has no information that it can communicate to its future self. Thus, in a system with classical information structure, control only has the dual function of control and estimation; the third function of communication is not needed. The absence of communication aspect drastically simplifies the search of optimal designs. We will explain this simplification later.

1.3 Scope of this thesis

We are interested in the optimal design of sequential dynamic teams. Witsenhausen (1988) showed that under a technical condition, which is almost always satisfied, a dynamic team can be converted to a static team. This conversion comes at the cost of expansion of the state space of system variables. Our belief is that one should try to exploit the dynamic nature of the problem rather than work around it. For that reason, we want to obtain a ***sequential decomposition*** of the optimal design of a dynamic (multi-stage) team. Sequential decomposition is a divide and conquer technique which decomposes the one shot optimization problem of choosing an optimal design into a sequence of nested optimization problems, each of which is drastically simpler to solve than the original problem.

Teams with a classical information structure are centralized problems; for such problems Markov decision theory (Kumar and Varaiya, 1986) provides a systematic methodology to obtain a sequential decomposition. Teams with non-classical information structure are strictly decentralized problems; for such finite-horizon problems the standard form (Witsenhausen, 1973) provides a sequential decomposition. There is no solution methodology for infinite-horizon decentralized team problems. This thesis provides a sequential decomposition for both finite and infinite horizon dynamic teams with non-classical information structures.

1.4 Main ideas for optimal design of dynamic teams

An optimal design of a finite-horizon dynamic team always exists when all system variables are finite valued; it can be found by a brute force search of all designs. The number of possible designs increase exponentially with the number of agents and the time horizon for which the system runs. So, the complexity of brute force search is exponential in the number of agents and the time horizon of the system. For this reason, we want to determine a *systematic* method to search for an optimal solution *efficiently*.

For infinite horizon performance criteria, the situation is different. Identifying an optimal (or near-optimal) design by brute force is not possible because there are countably infinite number of designs. Furthermore, it is not possible to implement a general infinite horizon design. So, we need to identify some qualitative properties of optimal designs that will enable us to search and implement optimal designs *compactly*.

A sequential team with a classical information structure is equivalent to either a MDP (Markov decision process), where the controller perfectly observes the state of the system, or a POMDP (partially observable Markov decision process), where the controller takes noisy observations of the state of the system. For both these cases, the results from Markov decision theory allow for efficient search and compact implementation of optimal designs. We explain why this is the case,[2] then contemplate on how these ideas can be extended to teams with non-classical information structure.

Consider a MDP with finite state space $\mathcal{X}$ and finite action space $\mathcal{U}$ which operates for a finite horizon T. In general, the control action at any time can depend on all the past observations and all the past control actions. So, at time t there are $|\mathcal{U}|^{|\mathcal{X}|^t \times |\mathcal{U}|^{t-1}}$ control laws; for the entire horizon there are approximately $|\mathcal{U}|^{(|\mathcal{X}|\times|\mathcal{U}|)^T}$ designs. Hence, a brute force search is doubly exponential in the size of the horizon. Markov decision theory provides two simplifications. The first is the structural results which state that we only need to look at controllers where the control action depends on the current state. Consequently, at time t, we only need to look at $|\mathcal{U}|^{|\mathcal{X}|}$ functions; for the entire horizon there are $|\mathcal{U}|^{T\times|\mathcal{X}|}$ designs. Hence, the structural

[2] We do a loose hand-waving complexity analysis here. See Blondel and Tsitsiklis (2000) for a more detailed survey of complexity results for MDP and POMDP.

results simplify the problem exponentially: a brute force search is now "only" exponential in T. The second simplification is the sequential decomposition provided by the dynamic programming equations which transform the one shot optimization problem into a sequence of T nested optimization problems. For each step of the dynamic problem, we need to need to find ***a value function***, which is a function with (finite) domain $\mathcal{X}$. For each value of $x \in \mathcal{X}$, evaluating the value function requires approximately $|\mathcal{U}|^2$ calculations. Hence, for the entire horizon we need to evaluate $T \times |\mathcal{X}| \times |\mathcal{U}|^2$ computations. Thus, the structural results and sequential decomposition drastically reduce the complexity of search for an optimal solution.

For infinite horizon MDP, Markov decision theory shows that we can restrict attention to control laws that do not change with time. This makes it easy to implement an optimal design since we only need to implement one control law. This also allows us to extend the sequential decomposition of the finite horizon to infinite horizon: for infinite horizon problems, optimal time-invariant control law is given by the fixed point of a functional equation. Finding fixed points of these functional equations can be converted into linear programs with $|\mathcal{X}|$ variables and $|\mathcal{X}|\times|\mathcal{U}|$ constraints (see Manne (1960) and d'Epenoux (1960)). Linear programs can be solved in polynomial time (Karmarkar, 1984), which means that infinite horizon MDPs with finite state and action spaces can be solved in polynomial time.

Next consider a finite horizon POMDP with finite state space $\mathcal{X}$, finite observation space $\mathcal{Y}$, and finite action space $\mathcal{U}$. In general, the control action at time t can depend on all the past observations and all the past control actions. So, at time t there are $|\mathcal{U}|^{|\mathcal{Y}|^t \times |\mathcal{U}|^{t-1}}$ control laws; for the entire horizon, there are approximately $|\mathcal{U}|^{(|\mathcal{Y}|\times|\mathcal{U}|)^T}$ designs. For POMDPs, Markov decision theory provides two simplifications. The first is the structural results which state that we only need to look at controllers where the control action depends on the controller's belief about the state of the system. However, these structural results do not immediately reduce the complexity of the search of an optimal design; the belief space is uncountable, so there are uncountable number of control laws that satisfy the structural properties. The second simplification—the sequential decomposition of dynamic programming equations—helps in reducing the complexity of the search of an optimal design. The nested optimality equations of dynamic programming reduce the one shot optimization problem into T nested optimization problems. At each step of the optimization problem we need to find a value function; for POMDPs, the value

function has an uncountable domain (the space of beliefs on the state of the system). In spite of this, these value functions can be computed exactly because they are piecewise linear and concave (Smallwood and Sondik, 1973) and as such can be represented compactly by a family of linear functions that form the upper envelope of the value function. In the worst case, we need to construct approximately $|\mathcal{U}|^{|\mathcal{Y}|^T}$ linear envelopes. For specific instances, the family of envelopes may increase polynomially with the time horizon (Littman, 1996).

For infinite horizon POMDP, Markov decision theory shows that we can restrict attention to control laws that do not change with time. An optimal time-invariant control law can be obtained by the fixed point of a functional equation. These functional equations cannot be solved exactly; however, they can be approximated efficiently by using randomized algorithms that discretize the belief space; Rust (1997) shows that the worst case complexity of solving discounted cost POMDPs with finite state and action spaces is polynomial in $|\mathcal{X}|$ and $|\mathcal{U}|$. There are other results which exploit the special structure of POMDPs arising in specific application domains to solve the finite and infinite horizon optimality equations more efficiently.

Thus, Markov decision theory provides an efficient method to search for optimal (or near optimal) solutions for MDPs and POMDPs, and a compact way to implement infinite horizon solutions. Furthermore, in many problems the optimality equations can be used to identify more refined qualitative properties of optimal control laws, which further simplify the search of an optimal solution. For example, for sequential hypothesis testing (Wald, 1947) the optimality equations can be used to prove that optimal decision rule is of a "threshold type"; for centralized LQG (linear quadratic Gaussian) problems the optimality equations can be used to prove that optimal control laws are affine. Such structural results significantly simplify the search of optimal designs.

So, when the sequential team has a classical information structure, both finite and infinite horizon systems can be optimally designed in an efficient manner using Markov decision theory. However, no such systematic methodology exists when sequential teams have non-classical information structure. In general, multi-agent teams are NEXP-complete, i.e., they *provably* do not admit a polynomial time solution (Bernstein et al., 2000). However, it is worthwhile to determine search methodologies which are better than brute force search and identify specific instances that

can be solved efficiently. In this thesis, we identify instances of multi-agent problem which can be reduced to POMDPs where the unobserved state is either finite or uncountable, and the action space is finite (but exponential in the size of the alphabets). This allows us to leverage the huge existing literature on numerically solving POMDPs to multi-agent teams.

We would like to exploit the sequentiality of the system to decompose the brute force one-shot optimization problem (where we evaluate the performance of a choice of design rules for all agents for the entire horizon in a single step) into a sequence of nested optimization problems (where at each step we evaluate the effect of a decision rule of a single agent at a single time step on the overall performance); that is, obtain a *sequential decomposition*. The number of nested optimization problems resulting from the sequential decomposition are linear in the number of agents and the time horizon. So, if each step of a sequential decomposition can be solved efficiently, the sequential decomposition provides a tractable method of optimally designing sequential multi-agent systems. We would also like to find qualitative properties of optimal decision rules to reduce the space over which we search at each step. We would then like to extend this sequential decomposition to infinite horizon problem, and hope that the search for an optimal infinite horizon design reduces to finding the fixed point of a functional equation; this would make the implementation of an optimal infinite horizon design easier and could also help in identifying approximation algorithms to search for near optimal designs.

For finite-horizon decentralized problems, the standard form (Witsenhausen, 1973) provides a sequential decomposition. However, this sequential decomposition is only applicable to finite-horizon problems and cannot be extended to infinite horizon problems even for time-homogeneous systems. We are interested in methodologies that extend to infinite horizon.

1.5 Organization of the thesis

Sequential decomposition of dynamic teams is difficult both from a conceptual and computational viewpoint. We focus on resolving the conceptual difficulties. For that matter, we consider the simplest multi-agent team—a two-agent team. We believe that once we make the conceptual jump from understanding the design of

one-agent centralized systems to understanding the design of two-agent decentralized systems, extending the same line of reasoning to general multi-agent systems will be significantly easier.

The research presented in this thesis was carried out in the following order. We first investigated sequential decomposition of globally optimal design of real-time communication over noisy forward channel. Teneketzis (2006) had derived qualitative properties of optimal encoders and decoders, and we used these qualitative properties as a starting point for determining globally optimal encoders and decoders. We then investigated networked control systems and real-time communication over noisy forward and backward channels. These applications are essentially two-agent teams. They are rich enough to capture the fundamental conceptual difficulties in optimally designing two-agent teams. By carefully analyzing the sequential decomposition of real-time communication and networked control systems, we were able to develop a framework for sequential decomposition of a general two-agent team.

In this thesis we do not present the research in the chronological order that it was carried out. Rather, we first present a solution methodology for the design of a general two-agent team problem (Chapter 2), and then explain how this methodology can be applied to real-time communication (Chapter 3) and networked control systems (Chapter 4). As a result of this organization, the fundamental ideas of sequential decomposition of two agent teams can be separated from application specific details. This organization also shows how the results of the general solution methodology of Chapter 2 could be applied to other applications. We take a critical look at the strength and weakness of our solution framework in Chapter 5 and conclude with some possible future directions.

1.6 Contribution of the thesis

The contributions of this thesis lie in both conceptual and technical aspects of the design of decentralized systems. This thesis provides a conceptual framework for the design of decentralized two-agent teams with strictly non-classical information structures. We use sequential decomposition as a solution concept for the design of dynamic teams. The notion of information state sufficient for performance analysis is the key concept to obtain a sequential decomposition. ***This thesis presents***

properties that such information states must satisfy. This is the first description of the properties of information states in the literature and ***the most important conceptual contribution of this thesis.*** There was no general methodology to identify information states appropriate for both finite and infinite horizon problems. As such, the only way to identify appropriate information states was to guess information states and check if they lead to a sequential decomposition. In light of the properties of information states presented in this thesis, we can guess information states and check if they satisfy the aforementioned properties. Checking whether our choice of information states satisfies a few properties is a huge simplification over directly checking if they lead to a sequential decomposition.

We also provide an intuitive explanation of our choice of information states that satisfy the aforementioned properties. We can look at the design of two-agent teams as controlled input-output systems from the point of view of the system designer. The designer provides the decision rules of each agent as control inputs and, in general, does not observe any output; hence, he has to optimally design a partially observed system. Therefore, the designer can use his belief on *the state sufficient for input-output mapping on the system* as an information state. This suggests that ***it may be possible to identify information states sufficient for performance analysis simply by identifying state sufficient for input-output mapping***, although we have not explored this angle in this thesis.

The technical contributions of this thesis are three-fold—in the contexts of two-agent teams, real-time communication, and networked control systems. In Chapter 2 we investigate sequential decomposition of finite and infinite horizon dynamic two-agent teams. In order to obtain a sequential decomposition of any dynamic optimization problem, we need to identify an *information state sufficient for performance evaluation*. We present properties that must be satisfied by information states sufficient for performance analysis, guess information states that satisfy these properties, and show that optimally controlling the time evolution of these information states leads to a sequential decomposition of finite horizon two-agent teams. When all system variables are finite valued, the optimality equations of the sequential decomposition can be viewed as the optimality equations of POMDPs (partially observable Markov decision processes) for which efficient computational algorithms exist.

Next we consider time-homogeneous infinite-horizon problems for two performance criteria: total expected discounted cost and average cost per unit time. For each of these performance criteria, we consider four variations, called variations v1, v2, v3, and v4, depending on the whether the agents have finite time-invariant state space or perfect recall. We show how to obtain a sequential decomposition of infinite horizon problems for three of these four variations; the sequential decomposition exploits the fact that there is no loss of optimality in restricting attention to time-invariant *meta-strategies* (explained in Chapter 2). For the total expected discounted cost criteria, we show that optimal strategies can be determined by the unique fixed point of a functional equation. For the average cost per unit time criteria, we show that, under a technical condition, near-optimal strategies can be determined by the fixed point of a functional equation. The fixed point functional equations for these three variations can be viewed as fixed point functional equations that arise in infinite horizon POMDPs. When all system variables are finite valued, the fixed point equations of variation v1 belong to a class for which efficient approximate computational algorithms exist; the fixed point equations of variations v2 and v3 belong to a class for which finding efficient approximate computational algorithms is an active area of research.

In variations v2 and v3 one agent has perfect recall (i.e., it remembers everything that it has seen and done in the past) and one agent has time-invariant memory. For these variations we derive qualitative/structural properties of optimal controllers and show that, without any loss of optimality, the agent with perfect recall can choose its control actions only based on its belief about the state of the plant and the state of the other agent; thus, the agent with perfect recall can restrict attention to control laws belonging to a time-invariant functional space. This restriction allows us to derive optimality equations for these variations of the infinite horizon problems.

In Chapter 3 we consider four models of point-to-point real-time communication, consisting of a Markov source, a real-time encoder, a real-time receiver, a forward channel, which is either noiseless or noisy, between the encoder and the receiver; some models also include a backward channel, which is either noiseless or noisy, between the receiver and the encoder. For each of these models, we consider both finite and infinite horizon problems; for infinite horizon problems, we consider three variations corresponding to variations v1, v2 and v3 of Chapter 2. We show

that these models are instances of two agent teams; thus, we can use the results of Chapter 2 to derive qualitative properties of optimal encoding and decoding strategies and to obtain a sequential decomposition of both finite and infinite horizon problems. Before this thesis, the qualitative/structural properties were known for variation v2 of three models and and the sequential decomposition was known for only the variation v4 for on model.

(variation v4 which we do not consider in this thesis) of one of the four models.

In this thesis, we provide a unified framework to study all models of point-to-point real-time communication. We show that the various structural properties of optimal real-time communication systems previously derived in the literature are special cases of the structural properties of optimal two-agent teams derived in Chapter 2. Furthermore, we provide a sequential decomposition for all models of real-time communication.

In Chapter 4 we consider the optimal design of a simple model for a NCS (networked control system) for both finite and infinite horizon problems. We assume that the NCS consists of a plant, a sensor, and a controller. There is a communication channel between the sensor and the controller, and another communication channel between the controller and the plant. Both communication channels are assumed to be noisy. This model captures the salient features of a general NCS, viz., non-linear plant dynamics, noisy communication channels, and resource and power limitations at the sensor. We show that this model is an instance of a two agent team; thus, we can use the results of Chapter 2 to derive qualitative properties of optimal sensors and controllers and to obtain a sequential decomposition of both finite and infinite horizon cases. Before this thesis, the optimal design of NCS was understood only for linear plant dynamics and noiseless communication channels. In this thesis we provide a framework to study NCS with non-linear plant dynamics and noisy communication channels.

Chapter 2

Optimal design of two-agent systems

In this chapter we consider the optimal design of a general two-agent team. We explain what an information state means, and then identify appropriate information states for two-agent teams. Controlling the time evolution of these information states in an optimal manner leads to a sequential decomposition of the optimization problem. We then consider time homogeneous models with infinite horizon cost criteria. We consider four variations of the infinite horizon problem; for three of the variations we show that optimal designs are given by fixed points of functional equations. We then present the intuition behind our choice of information states and conclude the chapter.

Notation

We use uppercase letters N, M, S, U, W, X, Y, and Z to represent random variables, corresponding lowercase letters n, m, etc. to represent their realizations, and corresponding calligraphic letters $\mathcal{N}$,$\mathcal{M}$, etc. to represent their alphabets. We use lowercase letters c, d, f, g, h, and l to represent functions, corresponding uppercase letters C, D, etc. to represent collection of functions, and corresponding script letters $\mathscr{C}$, $\mathscr{G}$, etc. to represent family of functions. We use Gothic letters $\mathfrak{F}$, $\mathfrak{J}$, etc. to represent σ-algebras.

For random variables and functions, x^t is a short hand for the sequence $x_1, \dots, x_t$, and x_a^b is a short hand for $x_a, \dots, x_b$. $\mathbb{E}\{\cdot\}$ denotes the expectation of a random variable, $\Pr(\cdot)$ denotes the probability of an event, $\mathbb{I}[\cdot]$ denotes the indicator function of a statement, and $\mathbb{P}\{\mathcal{X}\}$ denotes the space of all PMF (probability mass functions) on $\mathcal{X}$. In order to denote that the expectation of a random variable or the probability of an event depends on a function φ, we use $\mathbb{E}\{\cdot \,|\, \varphi\}$ and $\Pr(\cdot \,|\, \varphi)$, respectively. This

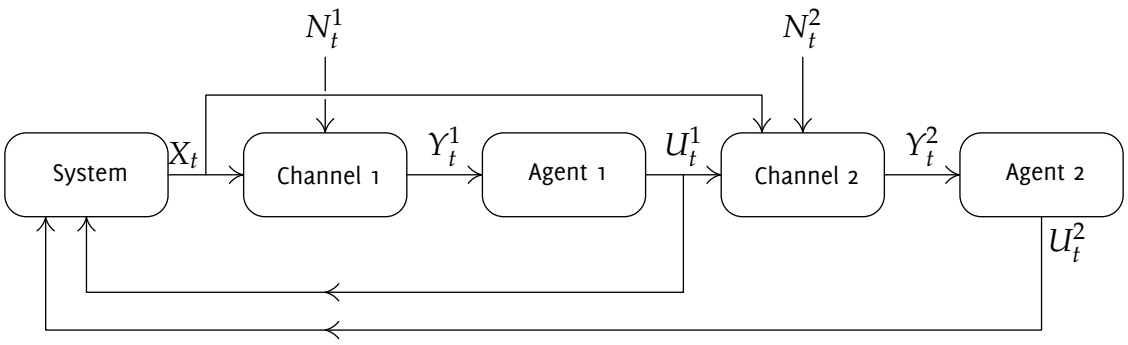


Figure 2.1: A general two-agent system

slightly unusual notation is chosen since we want to keep track of all functional dependencies and the conventional notation of $\mathbb{E}^{\varphi}\{\cdot\}$ and $\Pr^{\varphi}(\cdot)$ is too cumbersome to use.

2.1 A general finite-horizon two-agent problem

System model

Consider a two-agent system, shown in Figure 2.1, that operates in discrete time for a horizon T as follows:

$$X_{t+1} = f_t(X_t, U_t^1, U_t^2, W_t). \tag{2.1}$$

Here $X_t \in \mathcal{X}_t$ denotes the state of the system at time t, U_t^1 and U_t^2 denote the control actions of agents 1 and 2, respectively, at time t, and W_t denotes the process noise at time t. The function $f_t(\cdot)$ is the ***plant function***. The observations Y_t^1 and Y_t^2 of the agents are given by

$$Y_t^1 = h_t^1(X_t, N_t^1), \tag{2.2a}$$

$$Y_t^2 = h_t^2(X_t, U_t^1, N_t^2), \tag{2.2b}$$

where $N_t^k \in \mathcal{N}_t^k$ and h_t^k, $k = 1, 2$, denotes the channel noise and observation channel of agent k at time t, respectively. The control actions U_t^1 and U_t^2 are generated according to

$$U_t^1 = g_t^1(Y_t^1, S_{t-1}^1), \tag{2.3a}$$

$$U_t^2 = g_t^2(Y_t^2, S_{t-1}^2), \tag{2.3b}$$

where $S_t^k \in \mathcal{S}_t^k$, $k = 1, 2$, denote the state of agents 1 and 2. These states are updated according to

$$S_t^1 = l_t^1(Y_t^1, U_t^1, S_{t-1}^1), \tag{2.4a}$$

$$S_t^2 = l_t^2(Y_t^2, U_t^2, S_{t-1}^2). \tag{2.4b}$$

The functions $g_t^k(\cdot)$ and $l_t^k(\cdot)$ are the ***control law*** of agent k and ***state-update rule/ function*** of agent k at time t, respectively. We will use ***decision rule*** as a generic term to refer to either the control law or the state-update function. At each time an instantaneous cost of $\rho_t(X_t, U_t^1, U_t^2)$ is incurred. We assume that the function ρ_t is positive and uniformly bounded.

For ease of exposition, we will assume that all system variables are finite. Under some technical assumptions, the results presented in this chapter can be extend to the case where the system variables are continuous.

Agents' strategies and systems' design

The choice of $G^k := (g_1^k, \dots, g_T^k)$ is called a ***control strategy*** of agent k, $k = 1, 2$; the choice of $L^k := (l_1^k, \dots, l_T^k)$ is called a ***state-update strategy*** of agent k, $k = 1, 2$. The choice (G^1, L^1, G^2, L^2) of control and state-update strategies of both agents is called a ***design*** of the system.

Primitive random variables

All the randomness in the system is generated by the random variables $(X_1, W_1, \dots, W_T, N_1^1, \dots, N_T^1, N_1^2, \dots, N_T^2)$. These random variables are called *primitive random variables* and are assumed to be mutually independent and defined on a common probability space $(\Omega, \mathfrak{F}, P)$. We denote the PMF (probability mass function) of X_1 by P_{X_1}, the PMF of W_t by P_{W_t} and the PMF of N_t^k by $P_{N_t^k}$. Once a design is specified, all system variables are measurable with respect to the underlying probability space $(\Omega, \mathfrak{F}, P)$.

Performance measure and optimization

The performance of a design is quantified by the total expected cost under that design, i.e.,

$$\mathcal{J}_T(G^1, L^1, G^2, L^2) \coloneqq \mathbb{E}\left\{ \sum_{t=1}^{T} \rho_t(X_t, U_t^1, U_t^2) \,\middle|\, G^1, L^1, G^2, L^2 \right\} \tag{2.5}$$

where the expectation is taken with respect to the measure induced on $\{(X_t, U_t^1, U_t^2), t = 1, \dots, T\}$ by the joint measure on the primitive random variables and the choice of the design (G^1, L^1, G^2, L^2).

We are interested in the following optimization problem.

Problem 2.1. *Consider a two-agent sequential team where the plant function f_t, the observation functions h_t^1 and h_t^2, the cost function ρ_t, and the PMF of the primitive random variables are given for $t = 1, \dots, T$. Determine a design $(G^{1,*}, L^{1,*}, G^{2,*}, L^{2,*})$ that minimizes the total expected cost under that design, i.e.,*

$$\mathcal{J}_T(G^{1,*}, L^{1,*}, G^{2,*}, L^{2,*}) = \mathcal{J}_T^* \coloneqq \min_{\substack{G^1 \in \mathscr{G}^{1,T} \\ L^1 \in \mathscr{L}^{1,T} \\ G^2 \in \mathscr{G}^{2,T} \\ L^2 \in \mathscr{L}^{2,T}}} \mathcal{J}_T(G^1, L^1, G^2, L^2), \tag{2.6}$$

where for $k = 1, 2$, $\mathscr{G}^{k,T} \coloneqq \mathscr{G}_1^k \times \cdots \times \mathscr{G}_T^k$; $\mathscr{G}_t^k$ is the family of functions from $\mathcal{Y}_t^k \times \mathcal{S}_{t-1}^k$ to $\mathcal{U}_t^k$; $\mathscr{L}^{k,T} \coloneqq \mathscr{L}_1^k \times \cdots \times \mathscr{L}_T^k$; and $\mathscr{L}_t^k$ is the family of functions from $\mathcal{Y}_t^k \times \mathcal{U}_t^k \times \mathcal{S}_{t-1}^k$ to $\mathcal{S}_t^k$.

If part of the design is pre-specified (e.g., the state-update strategy is fixed), then we are only interested in determining the remaining components of the design optimally.

We assume that all decision rules are pure (i.e., not randomized). We could have generalized the model by allowing randomized decision rules; however, randomized decision rules would not lead to a lower cost (see Gihman and Skorohod (1979, Theorem 1.6)).

The sequential nature of the problem

Problem 2.1 is a multi-stage sequential team with non-classical information structure (see the classification in Section 1.2). To explicitly highlight the sequential nature of the problem, we need to refine the notion of time. We call each step of the evolution of the system a *stage*. For each stage we consider four time instances: ${}^1t$, ${}^2t$, ${}^3t$ and ${}^4t$. We assume that agent 1 generates a decision U_t^1 at time ${}^1t$ and updates its state S_t^1 at time ${}^2t$, agent 2 generates a decision U_t^2 at time ${}^3t$ and updates its state S_t^2 at time ${}^4t$. We will sometimes refer to the agent acting at time ${}^it$ as "agent ${}^it$". The

sequential ordering of the system variables is shown in Figure 2.2 (some of these variables will be defined later).

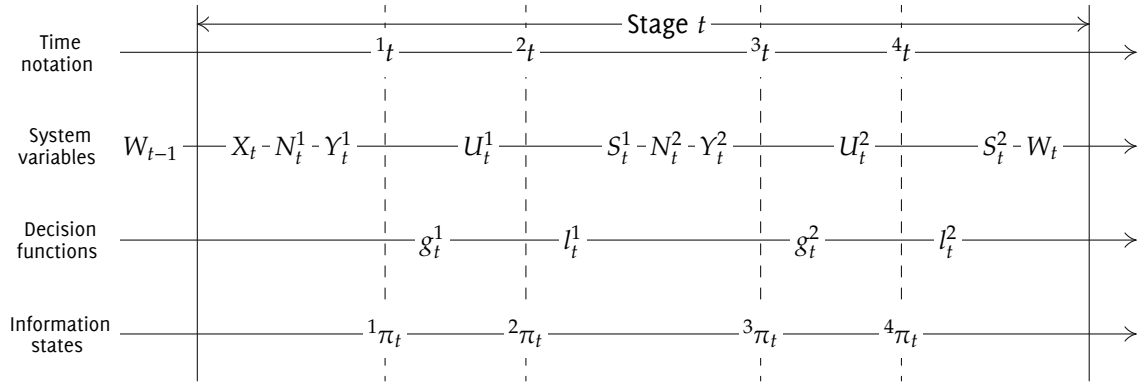


Figure 2.2: Sequential ordering of the system variables for the two-agent system

For the ease of notation, let ${}^i\varphi_t$ denote the function used by agent ${}^it$, and let ${}^i\varphi^{t-1}$ denote all the past functions until time ${}^it$, i.e.,

$$ {}^1\varphi_t = g_t^1, \qquad {}^1\varphi^{t-1} = (g^{1,t-1}, l^{1,t-1}, g^{2,t-1}, l^{2,t-1}); \tag{2.7a} $$

$$ {}^2\varphi_t = l_t^1, \qquad {}^2\varphi^{t-1} = (g^{1,t}, l^{1,t-1}, g^{2,t-1}, l^{2,t-1}); \tag{2.7b} $$

$$ {}^3\varphi_t = g_t^2, \qquad {}^3\varphi^{t-1} = (g^{1,t}, l^{1,t}, g^{2,t-1}, l^{2,t-1}); \tag{2.7c} $$

$$ {}^4\varphi_t = l_t^2, \qquad {}^4\varphi^{t-1} = (g^{1,t}, l^{1,t}, g^{2,t}, l^{2,t-1}). \tag{2.7d} $$

Data and Information Fields

Definition 2.1 (Data at agents). *Let ${}^iO_t^k$ denote the data available at agent k at time ${}^it$. Then,*

$$ {}^1O_t^1 := (Y_t^1, S_{t-1}^1), \qquad {}^2O_t^1 := (Y_t^1, U_t^1, S_{t-1}^1), \tag{2.8a} $$

$$ {}^3O_t^1 := S_t^1, \qquad {}^4O_t^1 := S_t^1, \tag{2.8b} $$

and

$$ {}^1O_t^2 := S_{t-1}^2, \qquad {}^2O_t^2 := S_{t-1}^2, \tag{2.8c} $$

$$ {}^3O_t^2 := (Y_t^2, S_{t-1}^2), \qquad {}^4O_t^2 := (Y_t^2, U_t^2, S_{t-1}^2). \tag{2.8d} $$

Let ${}^i\boldsymbol{O}_t^k$ denote the space of realizations of ${}^iO_t^k$.

For any choice of ${}^i\varphi^{t-1}$ of past decision rules, the data ${}^iO_t^k$ of agent k are measurable with respect to $\mathfrak{F}$. All the information (about the randomness of $\mathfrak{F}$) that agent k

can obtain from its data is called its ***information field*** ${}^i\mathfrak{I}_t^k$. This information field equals the smallest sub-field $\sigma({}^iO_t^k; {}^i\varphi^{t-1})$ of $\mathfrak{F}$ with respect to which ${}^iO_t^k$ is measurable.

Definition 2.2 (Information Fields). *Let ${}^i\mathfrak{I}_t^k$ denote the information field of agent k at time time ${}^it$. Then,*

$$ {}^1\mathfrak{I}_t^1 \coloneqq \sigma(Y_t^1, S_{t-1}^1; {}^1\varphi^{t-1}), \qquad {}^2\mathfrak{I}_t^1 \coloneqq \sigma(Y_t^1, U_t^1, S_{t-1}^1; {}^2\varphi^{t-1}), \tag{2.9a} $$

$$ {}^3\mathfrak{I}_t^1 \coloneqq \sigma(S_t^1; {}^3\varphi^{t-1}), \qquad {}^4\mathfrak{I}_t^1 \coloneqq \sigma(S_t^1; {}^4\varphi^{t-1}), \tag{2.9b} $$

and

$$ {}^1\mathfrak{I}_t^2 \coloneqq \sigma(S_{t-1}^2; {}^1\varphi^{t-1}), \qquad {}^2\mathfrak{I}_t^2 \coloneqq \sigma(S_{t-1}^2; {}^2\varphi^{t-1}), \tag{2.9c} $$

$$ {}^3\mathfrak{I}_t^2 \coloneqq \sigma(Y_t^2, S_{t-1}^2; {}^3\varphi^{t-1}), \qquad {}^4\mathfrak{I}_t^2 \coloneqq \sigma(Y_t^2, U_t^2, S_{t-1}^2; {}^4\varphi^{t-1}). \tag{2.9d} $$

Observe that agent 1 does not change its information at time ${}^2t$ and ${}^4t$, thus

$$ {}^3\mathfrak{I}_t^1 = {}^4\mathfrak{I}_t^1 \text{ and } {}^1\mathfrak{I}_{t+1}^1 = {}^2\mathfrak{I}_{t+1}^1. \tag{2.10a} $$

Similarly, for agent 2

$$ {}^1\mathfrak{I}_t^2 = {}^2\mathfrak{I}_t^2 \text{ and } {}^3\mathfrak{I}_t^2 = {}^4\mathfrak{I}_t^2. \tag{2.10b} $$

The information fields of agent 1 change with time as follows. While going from time ${}^4(t-1)$ to ${}^1t$, agent 1 observes new data, so ${}^4\mathfrak{I}_{t-1}^1 \subseteq {}^1\mathfrak{I}_t^1$. While going from time ${}^1t$ to ${}^2t$, agent 1 generates a control action; this control action does not contain any self information since randomized strategies are not allowed. So, ${}^1\mathfrak{I}_t^1 = {}^2\mathfrak{I}_t^1$. While going from ${}^2t$ to ${}^3t$, agent 1 either stores the current observation and control action or sheds information. In the first case ${}^2\mathfrak{I}_t^1 = {}^3\mathfrak{I}_t^1$; in the second ${}^2\mathfrak{I}_t^1 \supset {}^3\mathfrak{I}_t^1$. While going from ${}^3t$ to ${}^4t$, agent 1 neither observes any additional data, nor does it take any decision. So ${}^3\mathfrak{I}_t^1 = {}^4\mathfrak{I}_t^1$. Thus,

$$ \cdots {}^4\mathfrak{I}_{t-1}^1 \subseteq {}^1\mathfrak{I}_t^1 = {}^2\mathfrak{I}_t^1 \supseteq {}^3\mathfrak{I}_t^1 = {}^4\mathfrak{I}_t^1 \subseteq {}^1\mathfrak{I}_{t+1}^1 \cdots \tag{2.11a} $$

By a similar argument, the information fields of agent 2 change as follows:

$$ \cdots {}^4\mathfrak{I}_{t-1}^2 \supseteq {}^1\mathfrak{I}_t^2 = {}^2\mathfrak{I}_t^2 \subseteq {}^3\mathfrak{I}_t^2 = {}^4\mathfrak{I}_t^2 \supseteq {}^1\mathfrak{I}_{t+1}^2 \cdots \tag{2.11b} $$

Furthermore, due to noise in the observation channel

$$ \left({}^1\mathfrak{I}_t^1 = {}^2\mathfrak{I}_t^1\right) \nsubseteq \nsupseteq \left({}^1\mathfrak{I}_t^2 = {}^2\mathfrak{I}_t^2\right) \quad \text{and} \quad \left({}^3\mathfrak{I}_t^1 = {}^4\mathfrak{I}_t^1\right) \nsubseteq \nsupseteq \left({}^3\mathfrak{I}_t^2 = {}^4\mathfrak{I}_t^2\right). $$

Therefore, agent 1 does not know everything that is known to agent 2 and vice-versa; so, the system has non-classical information structure.

2.2 Global Optimization

As mentioned in the introduction, our goal is to obtain a sequential decomposition of a generic decentralized optimization problem. We proceed as follows. We first obtain a sequential decomposition of a finite horizon problem. We then investigate conditions under which this sequential decomposition can be extended to infinite horizon problems. In particular, we consider four variations of the two-agent problem that depend on whether an agent has perfect recall or time-invariant state. For three of these four cases, we show how to obtain a sequential decomposition that works for both finite and infinite horizon problems.

In order to obtain a sequential decomposition for the finite horizon problem, we need to identify "*information states sufficient for performance evaluation*". There is no known methodology for identifying appropriate information states that lead to a sequential decomposition of both finite and infinite horizon problems. We first explain the properties that appropriate information states must satisfy. We then use these properties as a guide to guess appropriate information states and show how they lead to a sequential decomposition of the two-agent team.

Information states

A critical step in obtaining a sequential decomposition for problems with non-classical information structures is identifying an information state sufficient for performance evaluation. An information state is a sufficient statistic that satisfies certain properties. Unfortunately, all definitions of information states in the literature, with the exception of Witsenhausen (1976), are in terms of their properties for systems with a classical information structure; the only explanation of the properties of information states for systems with a non-classical information structure is in Witsenhausen (1976):

> The (information) state should be a summary ('compression') of some data (the 'past') known to someone (an observer or a controller) and sufficient for some purposes (input-output map, optimization, dynamic programming).

In this section we define the properties that the information states sufficient for performance analysis should satisfy and explain what these properties mean in the context of the two-agent sequential team. Consider a two-agent team, where the agents sit together before the system starts operating and plan what decision rules they will use. Since both agents have the same objective, such a "pre-game" agreement is admissible. In order to obtain a sequential decomposition , the agents need to determine their decision rules in a backward manner. Thus at time t, the agents need to do the following:

1. Determine the information that would be common knowledge to them (in the sense of Aumann (1976)).
2. Using this common information and assuming that future decision rules have been optimally chosen, determine decision rules that are optimal at time t.

Hence, the information that is common knowledge at time t can be used to obtain a sequential decomposition. In general, the agents can chose *information states*, which are "smaller" than the *entire* common knowledge but still lead to a sequential decomposition. Such information states should be measurable with respect to the common knowledge and sufficient to determine optimal decision rule. More precisely, the should satisfy the following properties:

P1. ***Sufficient summary of past information***

The information state should be a representation of all the past information that is sufficient for future performance evaluation. This has the following interpretation.

The two-agent decentralized team is a controlled stochastic input-output system. The stochastic inputs are $\{X_1, W_t, N_t^k, k = 1, 2, t = 1, \dots, T\}$, the controlled inputs are $\{(g_t^k, l_t^k), k = 1, 2, t = 1, \dots, T\}$, and the outputs are $\{(U_t^1, U_t^2), t = 1, \dots, T\}$. The system designer has to choose a design (G^1, L^1, G^2, L^2). Suppose the system is at time ${}^1t$ (similar interpretations will hold for ${}^2t$, ${}^3t$, and ${}^4t$): nature has produced $(x_1, w^t, n^{1,t-1}, n^{2,t-1})$, the designer has chosen ${}^1\varphi^{t-1}$ (which equals $(g^{1,t-1}, l^{1,t-1}, g^{2,t-1}, l^{2,t-1})$), and the system has produced $(u^{1,t-1}, u^{2,t-1})$ and incurred a cost $\sum_{t'=1}^{t-1} \rho_{t'}(x_{t'}, u_{t'}^1, u_{t'}^2)$. The designer has to choose ${}^1\varphi_t^T$ (which equals $(g_t^{1,T}, l_t^{1,T}, g_t^{2,T}, l_t^{2,T})$) to minimize the expected future cost $\mathbb{E}\Big\{ \sum_{t'=t}^{T} \rho_{t'}(x_{t'}, U_{t'}^1, U_{t'}^2) \Big|\ {}^1\varphi^{t-1}, {}^1\varphi_t^T \Big\}$. Different choices of past decision rules are equivalent for the purpose of evaluating future performance

if any choice of future decision rules lead to the same expected future performance. In other words, two choices of past decision rules ${}^1\varphi^{t-1,(1)}$ and ${}^1\varphi^{t-1,(2)}$ are equivalent, denoted by ${}^1\varphi^{t-1,(1)} \sim {}^1\varphi^{t-1,(2)}$, if for any choice of future decision rules ${}^1\varphi_t^T = ({}^1g_t^T, {}^1l_t^T, {}^2g_t^T, {}^2l_t^T)$, we have

$$\mathbb{E}\left\{ \sum_{t'=t}^{T} \rho_{t'}(X_{t'}, U^1_{t'}, U^2_{t'}) \,\middle|\, {}^1\varphi^{t-1,(1)}, {}^1\varphi_t^T \right\} = \mathbb{E}\left\{ \sum_{t'=t}^{T} \rho_{t'}(X_{t'}, U^1_{t'}, U^2_{t'}) \,\middle|\, {}^1\varphi^{t-1,(2)}, {}^1\varphi_t^T \right\}$$

Recall that the designer has already chosen ${}^1\varphi^{t-1}$ and wants to choose ${}^1\varphi_t^T$ to minimize the expected future cost. If ${}^1\varphi^{t-1,(1)} \sim {}^1\varphi^{t-1,(2)}$ then the optimal future decision rules will be the same for both of them. So, to evaluate future performance and choose future decision rules, it is sufficient for the designer to keep track of the equivalence class of the past decision rules.

Let ${}^i\Phi^{t-1}$ denote the space of realization of all past decision rules, and let ${}^i\Pi_t$ be any arbitrary space. Suppose ${}^i\pi_t : {}^i\Phi^{t-1} \to {}^i\Pi_t$ is a function such that for any ${}^i\varphi^{t-1,(1)}, {}^i\varphi^{t-1,(2)} \in {}^i\Phi^{t-1}$, if ${}^i\pi_t({}^i\varphi^{t-1,(1)}) = {}^i\pi_t({}^i\varphi^{t-1,(2)})$ then ${}^i\varphi^{t-1,(1)} \sim {}^i\varphi^{t-1,(2)}$. Any such ${}^i\pi_t$ is a sufficient statistic for future performance evaluation.

P2. ***Common knowledge and sequential update***

All agents in the system should be able to solve the sequential decomposition of the problem. So, the information state cannot depend on data that is observed locally by one of the agents. In fact, the information state should be common knowledge between the two agents in the sense of Aumann (1976), and the agents should be able to keep track of how the information state evolves with time.

In centralized stochastic optimization (i.e., problems with classical information structure), the conditional probability of the state of the plant conditioned on the agent's data is an information state appropriate for performance evaluation. However, in decentralized stochastic optimization (i.e., problems with non-classical information structures) such conditional probabilities cannot be information states as they are not common knowledge: the data observed at each agent is not common knowledge, hence conditional probabilities based on this data is not common knowledge. The sufficient statistics ${}^i\pi_t$ of (P1) are derived from past decision rules, which are common knowledge. So, they can be evaluated by both agents.

Furthermore, in order for both agents to carry out the sequential decomposition, for any realization of current information state and any choice of current decision rules, both agents should be able to determine the next realization of information state. This means that if ${}^1\pi_t$, ${}^2\pi_t$, ${}^3\pi_t$ and ${}^4\pi_t$ are information states then: ${}^2\pi_t({}^2\varphi^{t-1})$ should be a function of ${}^1\pi_t({}^1\varphi^{t-1})$ and ${}^1\varphi_t$ (recall that ${}^2\varphi^{t-1} = ({}^1\varphi^{t-1}, {}^1\varphi_t)$); ${}^3\pi_t({}^3\varphi^{t-1})$ should be a function of ${}^2\pi_t({}^2\varphi^{t-1})$ and ${}^2\varphi_t$; ${}^4\pi_t({}^4\varphi^{t-1})$ should be a function on ${}^3\pi_t({}^3\varphi^{t-1})$ and ${}^3\varphi_t$; and ${}^1\pi_{t+1}({}^1\varphi^t)$ should be a function of ${}^4\pi_t({}^4\varphi^{t-1})$ and ${}^4\varphi_t$.

Any sequence of functions $\{{}^i\pi_t,\ i = 1,2,3,4,\ t = 1,\dots,T\}$ that have properties (P1) and (P2) is a valid choice of information states, and can be used to obtain a sequential decomposition for the finite horizon problem. We want to develop a methodology that can be extended to infinite horizon problem. For that matter, we require the following additional property.

P3. Time invariant domain

We want to identify functions ${}^i\pi_t : {}^i\Phi^{t-1} \to {}^i\Pi$ such that $\{{}^i\pi_t,\ i = 1,2,3,4,\ t = 1,\dots,T\}$ satisfy (P1) and (P2) and the sets ${}^1\Pi$, ${}^2\Pi$, ${}^3\Pi$ and ${}^4\Pi$ do not depend on the time horizon T.

An information state should provide a sufficient representation of past knowledge that is also efficient, both in calculating optimal decision rules and in their implementation. The smaller the set of all realizations of the information state, the more efficient it is to compute optimal decision rules. So, the following property is desirable.

P4. Minimality.

If more than one appropriate information state exist, working with the information state is computationally most efficient. However, we have not been able to establish a good way of comparing information states of infinite horizon problems. ***So, in the rest of the chapter, we will not consider minimality.***

In summary, for finite horizon problems we want to identify information states that satisfy properties (P1) and (P2); for infinite horizon problems, the information states should also satisfy property (P3). Now, we will restate the above properties more formally so that they can be verified more easily.

Property (P1) is formally equivalent to the following two statements:

S1. ***The information state is a summary of past information.***

Thus, ${}^1\pi_t$ should be a function of ${}^1\varphi^{t-1}$; ${}^2\pi_t$ should be a function of ${}^2\varphi^{t-1}$; ${}^3\pi_t$ should be a function of ${}^3\varphi^{t-1}$; and ${}^4\pi_t$ should be a function of ${}^4\varphi^{t-1}$.

S2. ***The information state absorbs the effect of past decisions on future performance.***

This means that

$$\begin{aligned}
&\mathbb{E}\left\{\sum_{t'=t}^{T} \rho_{t'}(X_{t'}, U^1_{t'}, U^2_{t'}) \,\middle|\, G^1, L^1, G^2, L^2\right\} \\
&\qquad = \mathbb{E}\left\{\sum_{t'=t}^{T} \rho_{t'}(X_{t'}, U^1_{t'}, U^2_{t'}) \,\middle|\, {}^1\pi_t, g^{1,T}_t, l^{1,T}_t, g^{2,T}_t, l^{2,T}_t\right\} \\
&\qquad = \mathbb{E}\left\{\sum_{t'=t}^{T} \rho_{t'}(X_{t'}, U^1_{t'}, U^2_{t'}) \,\middle|\, {}^2\pi_t, g^{1,T}_{t+1}, l^{1,T}_t, g^{2,T}_t, l^{2,T}_t\right\} \\
&\qquad = \mathbb{E}\left\{\sum_{t'=t}^{T} \rho_{t'}(X_{t'}, U^1_{t'}, U^2_{t'}) \,\middle|\, {}^3\pi_t, g^{1,T}_{t+1}, l^{1,T}_{t+1}, g^{2,T}_t, l^{2,T}_t\right\} \\
&\qquad = \mathbb{E}\left\{\sum_{t'=t}^{T} \rho_{t'}(X_{t'}, U^1_{t'}, U^2_{t'}) \,\middle|\, {}^4\pi_t, g^{1,T}_{t+1}, l^{1,T}_{t+1}, g^{2,T}_{t+1}, l^{2,T}_t\right\}
\end{aligned}$$

Property (P2) is equivalent to (S1) and the following statement.

S3. ***Both agents should be able to keep track of the information states.***

This means that ${}^2\pi_t$ should be determined from ${}^1\pi_t$ and ${}^1\varphi_t$ (i.e., ${}^1\pi_t$ and g^1_t); ${}^3\pi_t$ should be determined from ${}^2\pi_t$ and ${}^2\varphi_t$ (i.e., ${}^2\pi_t$ and l^1_t); ${}^4\pi_t$ should be determined from ${}^3\pi_t$ and ${}^3\varphi_t$ (i.e., ${}^3\pi_t$ and g^2_t); and ${}^1\pi_{t+1}$ should be determined from ${}^4\pi_t$ and ${}^4\varphi_t$ (i.e., ${}^4\pi_t$ and l^2_t).

Furthermore, statements (S1) and (S3) imply that statement (S2) is equivalent to the following:

S2′. ***The information states should be sufficient to evaluate the instantaneous cost.***

This means that

$$\begin{aligned}\mathbb{E}\left\{\rho_t(X_t, U_t^1, U_t^2) \,\middle|\, G^1, L^1, G^2, L^2\right\} &= \mathbb{E}\left\{\rho_t(X_t, U_t^1, U_t^2) \,\middle|\, {}^1\pi_t, g_t^1, l_t^1, g_t^2\right\} \\ &= \mathbb{E}\left\{\rho_t(X_t, U_t^1, U_t^2) \,\middle|\, {}^2\pi_t, l_t^1, g_t^2\right\} \\ &= \mathbb{E}\left\{\rho_t(X_t, U_t^1, U_t^2) \,\middle|\, {}^3\pi_t, g_t^2\right\} \\ &= \mathbb{E}\left\{\rho_t(X_t, U_t^1, U_t^2) \,\middle|\, {}^4\pi_t\right\}\end{aligned}$$

For finite horizon problems, *information states sufficient for performance analysis* must satisfy statements (s1), (s2) and (s3) or equivalently satisfy statements (s1), (s2′) and (s3). For infinite horizon problems, the information states should also satisfy property (p3) which is equivalent to the following:

s4 . The information states belong to time-invariant spaces

This means that there exist spaces ${}^1\Pi$, ${}^2\Pi$, ${}^3\Pi$ and ${}^4\Pi$ such that for all t, ${}^i\pi_t \in {}^i\Pi$, $i = 1, 2, 3, 4$.

As mentioned earlier, there is no general method of identifying appropriate information states for problems with a non-classical information structure. That is why we first guess information states that satisfy the above properties, and then show how to obtain a sequential decomposition using these information states.

Definition 2.3. *Define ${}^1\pi_t$, ${}^2\pi_t$, ${}^3\pi_t$ and ${}^4\pi_t$ as follows:*

$$ {}^1\pi_t \coloneqq \Pr\left(X_t, Y_t^1, S_{t-1}^1, S_{t-1}^2 \,\middle|\, {}^1\varphi^{t-1}\right), \tag{2.12a}$$

$$ {}^2\pi_t \coloneqq \Pr\left(X_t, Y_t^1, U_t^1, S_{t-1}^1, S_{t-1}^2 \,\middle|\, {}^2\varphi^{t-1}\right), \tag{2.12b}$$

$$ {}^3\pi_t \coloneqq \Pr\left(X_t, Y_t^2, U_t^1, S_t^1, S_{t-1}^2 \,\middle|\, {}^3\varphi^{t-1}\right), \tag{2.12c}$$

$$ {}^4\pi_t \coloneqq \Pr\left(X_t, Y_t^2, U_t^1, U_t^2, S_t^1, S_{t-1}^2 \,\middle|\, {}^4\varphi^{t-1}\right). \tag{2.12d}$$

Let ${}^i\Pi_t$, i = 1,2,3,4, denote the space of probability measures on $(\mathcal{X}_t \times \mathcal{Y}_t^1 \times \mathcal{S}_{t-1}^1 \times \mathcal{S}_{t-1}^2)$, $(\mathcal{X}_t \times \mathcal{Y}_t^1 \times \mathcal{U}_t^1 \times \mathcal{S}_{t-1}^1 \times \mathcal{S}_{t-1}^2)$, $(\mathcal{X}_t \times \mathcal{Y}_t^2 \times \mathcal{U}_t^1 \times \mathcal{S}_t^1 \times \mathcal{S}_{t-1}^2)$, and $(\mathcal{X}_t \times \mathcal{Y}_t^2 \times \mathcal{U}_t^1 \times \mathcal{U}_t^2 \times \mathcal{S}_t^1 \times \mathcal{S}_{t-1}^2)$, respectively. Then ${}^i\pi_t$ takes values in ${}^i\Pi_t$.

The above definitions are to be interpreted as follows. Let $(\Omega, \mathfrak{F}, P)$ denote the probability space on which all primitive random variables are defined. For any choice ${}^1\varphi^{t-1}$ of past decision rules at ${}^1t$, the state X_t of the system, the observation Y_t^1 of agent 1, and the states S_{t-1}^1, S_{t-1}^2 of both the agents are $\mathfrak{F}$–measurable. Thus,

for any choice of ${}^1\varphi^{t-1}$, the vector $(X_t, Y_t^1, S_{t-1}^1, S_{t-1}^2)$ is $\mathfrak{F}$–measurable. ${}^1\pi_t$ is the corresponding induced measure on $(\mathcal{X}_t \times \mathcal{Y}_t^1 \times \mathcal{S}_{t-1}^1 \times \mathcal{S}_{t-1}^2)$. Similar interpretations hold for ${}^2\pi_t$, ${}^3\pi_t$, and ${}^4\pi_t$.

The above defined probability measures are related as follows:

Lemma 2.1. *${}^1\pi_t$, ${}^2\pi_t$, ${}^3\pi_t$ and ${}^4\pi_t$ are information states for the control law and state-update rules at agents 1 and 2, respectively. Specifically,*

1. *There exist linear transformations ${}^1Q_t$, ${}^2Q_t$, ${}^3Q_t$ and ${}^4Q_t$ such that*

$$ {}^2\pi_t = {}^1Q_t(g_t^1)\,{}^1\pi_t, \tag{2.13a} $$

$$ {}^3\pi_t = {}^2Q_t(l_t^1)\,{}^2\pi_t, \tag{2.13b} $$

$$ {}^4\pi_t = {}^3Q_t(g_t^2)\,{}^3\pi_t, \tag{2.13c} $$

$$ {}^1\pi_{t+1} = {}^4Q_t(l_t^2)\,{}^4\pi_t. \tag{2.13d} $$

2. *The expected instantaneous cost can be expressed as*

$$ \mathbb{E}\left\{\rho_t(X_t, U_t^1, U_t^2) \,\middle|\, {}^4\varphi^{t-1}\right\} = \hat{\rho}_t({}^4\pi_t). \tag{2.14} $$

where $\hat{\rho}_t$ is a linear function of ${}^4\pi_t$ that depends on (G^1, L^1, G^2, L^2) only through ${}^4\pi_t$.

Proof. We will prove each part separately.

1. Consider any $x_t \in \mathcal{X}_t$, $y_t^1 \in \mathcal{Y}_t^1$, $u_t^1 = \mathcal{U}_t^1$, $s_{t-1}^1 \in \mathcal{S}_{t-1}^1$, $s_{t-1}^2 \in \mathcal{S}_{t-1}^2$, and ${}^2\varphi^{t-1} = ({}^1\varphi^{t-1}, g_t^1)$. A component of ${}^2\pi_t$ is given by

$$
\begin{aligned}
&{}^2\pi_t(x_t, y_t^1, u_t^1, s_{t-1}^1, s_{t-1}^2) \\
&\quad = \Pr\left(X_t = x_t, Y_t^1 = y_t^1, U_t^1 = u_t^1, S_{t-1}^1 = s_{t-1}^1, S_{t-1}^2 = s_{t-1}^2 \,\middle|\, {}^2\varphi^{t-1}\right) \\
&\quad = \Pr\left(U_t^1 = u_t^1 \,\middle|\, X_t = x_t, Y_t^1 = y_t^1, S_{t-1}^1 = s_{t-1}^1, S_{t-1}^2 = s_{t-1}^2; {}^1\varphi^{t-1}, g_t^1\right) \\
&\qquad \times \Pr\left(X_t = x_t, Y_t^1 = y_t^1, S_{t-1}^1 = s_{t-1}^1, S_{t-1}^2 = s_{t-1}^2 \,\middle|\, {}^1\varphi^{t-1}, g_t^1\right) \\
&\quad \overset{(a)}{=} \mathbb{I}\left[u_t^1 = g_t^1(y_t^1, s_{t-1}^1)\right] \Pr\left(X_t = x_t, Y_t^1 = y_t^1, S_{t-1}^1 = s_{t-1}^1, S_{t-1}^2 = s_{t-1}^2 \,\middle|\, {}^1\varphi^{t-1}\right) \\
&\quad = \mathbb{I}\left[u_t^1 = g_t^1(y_t^1, s_{t-1}^1)\right] {}^1\pi_t(x_t, y_t^1, s_{t-1}^1, s_{t-1}^2) \\
&\quad =: \left({}^1Q_t(g_t^1)\,{}^1\pi_t\right)(x_t, y_t^1, u_t^1, s_{t-1}^1, s_{t-1}^2)
\end{aligned} \tag{2.15}
$$

   where (a) follows from the sequential order in which the system variables are generated.

2. Consider any $x_t \in \mathcal{X}_t$, $y_t^2 \in \mathcal{Y}_t^2$, $u_t^1 \in \mathcal{U}_t^1$, $s_t^1 \in \mathcal{S}_t^1$, $s_{t-1}^2 \in \mathcal{S}_{t-1}^2$, and ${}^3\varphi^{t-1} = ({}^2\varphi^{t-1}, l_t^1)$. A component of ${}^3\pi_t$ is given by

$$
\begin{aligned}
&{}^3\pi_t(x_t, y_t^2, u_t^1, s_t^1, s_{t-1}^2) \\
&\quad = \Pr\left(X_t = x_t, Y_t^2 = y_t^2, U_t^1 = u_t^1, S_t^1 = s_t^1, S_{t-1}^2 = s_{t-1}^2 \,\middle|\, {}^3\varphi^{t-1}\right) \\
&\quad = \sum_{\substack{y_t^1 \in \mathcal{Y}_t^1 \\ s_{t-1}^1 \in \mathcal{S}_{t-1}^1}} \Pr\left(Y_t^2 = y_t^2 \,\middle|\, X_t = x_t, Y_t^1 = y_t^1, U_t^1 = u_t^1, S_t^1 = s_t^1, S_{t-1}^1 = s_{t-1}^1, S_{t-1}^2 = s_{t-1}^2; {}^3\varphi^{t-1}\right) \\
&\qquad\qquad \times \Pr\left(S_t^1 = s_t^1 \,\middle|\, X_t = x_t, Y_t^1 = y_t^1, U_t^1 = u_t^1, S_{t-1}^1 = s_{t-1}^1, S_{t-1}^2 = s_{t-1}^2; {}^2\varphi^{t-1}, l_t^1\right) \\
&\qquad\qquad \times \Pr\left(X_t = x_t, Y_t^1 = y_t^1, U_t^1 = u_t^1, S_{t-1}^1 = s_{t-1}^1, S_{t-1}^2 = s_{t-1}^2 \,\middle|\, {}^2\varphi^{t-1}, l_t^1\right) \\
&\quad \overset{(b)}{=} \sum_{\substack{y_t^1 \in \mathcal{Y}_t^1 \\ s_{t-1}^1 \in \mathcal{S}_{t-1}^1}} P_{N_t^2}\left(n_t^2 \in \mathcal{N}_t^2 : y_t^2 = h_t^2(x_t, u_t^1, n_t^2)\right) \mathbb{I}\left[s_t^1 = l_t^1(y_t^1, u_t^1, s_{t-1}^1)\right] \\
&\qquad\qquad \times \Pr\left(X_t = x_t, Y_t^1 = y_t^1, U_t^1 = u_t^1, S_{t-1}^1 = s_{t-1}^1, S_{t-1}^2 = s_{t-1}^2 \,\middle|\, {}^2\varphi^{t-1}\right) \\
&\quad = \sum_{\substack{y_t^1 \in \mathcal{Y}_t^1 \\ s_{t-1}^1 \in \mathcal{S}_{t-1}^1}} P_{N_t^2}\left(n_t^2 \in \mathcal{N}_t^2 : y_t^2 = h_t^2(x_t, u_t^1, n_t^2)\right) \mathbb{I}\left[s_t^1 = l_t^1(y_t^1, u_t^1, s_{t-1}^1)\right] \\
&\qquad\qquad \times {}^2\pi_t(x_t, y_t^1, u_t^1, s_{t-1}^1, s_{t-1}^2) \\
&\quad =: \left({}^2Q_t(l_t^1)\, {}^2\pi_t\right)(x_t, y_t^2, u_t^1, s_t^1, s_{t-1}^2)
\end{aligned}
\tag{2.16}
$$

where (b) follows from the sequential order in which the system variables are generated.

3. Consider any $x_t \in \mathcal{X}_t$, $y_t^2 \in \mathcal{Y}_t^2$, $u_t^1 \in \mathcal{U}_t^1$, $u_t^2 \in \mathcal{U}_t^2$, $s_t^1 \in \mathcal{S}_t^1$, $s_{t-1}^2 \in \mathcal{S}_{t-1}^2$, and ${}^4\varphi^{t-1} = ({}^3\varphi^{t-1}, g_t^2)$. A component of ${}^4\pi_t$ is given by,

$$
\begin{aligned}
&{}^4\pi_t(x_t, y_t^2, u_t^1, u_t^2, s_t^1, s_{t-1}^2) \\
&\quad = \Pr\left(X_t = x_t, Y_t^2 = y_t^2, U_t^1 = u_t^1, U_t^2 = u_t^2, S_t^1 = s_t^1, S_{t-1}^2 = s_{t-1}^2 \,\middle|\, {}^4\varphi^{t-1}\right) \\
&\quad = \Pr\left(U_t^2 = u_t^2 \,\middle|\, X_t = x_t, Y_t^2 = y_t^2, U_t^1 = u_t^1, S_t^1 = s_t^1, S_{t-1}^2 = s_{t-1}^2; {}^3\varphi^{t-1}, g_t^2\right) \\
&\qquad \times \Pr\left(X_t = x_t, Y_t^2 = y_t^2, U_t^1 = u_t^1, S_t^1 = s_t^1, S_{t-1}^2 = s_{t-1}^2 \,\middle|\, {}^3\varphi^{t-1}, g_t^2\right) \\
&\quad \overset{(c)}{=} \mathbb{I}\left[u_t^2 = g_t^2(y_t^2, s_{t-1}^2)\right] \Pr\left(X_t = x_t, Y_t^2 = y_t^2, U_t^1 = u_t^1, S_t^1 = s_t^1, S_{t-1}^2 = s_{t-1}^2 \,\middle|\, {}^3\varphi^{t-1}\right) \\
&\quad = \mathbb{I}\left[u_t^2 = g_t^2(y_t^2, s_{t-1}^2)\right] {}^3\pi_t(x_t, y_t^2, u_t^1, s_t^1, s_{t-1}^2) \\
&\quad =: \left({}^3Q_t(g_t^2)\, {}^3\pi_t\right)(x_t, y_t^2, u_t^1, u_t^2, s_t^1, s_{t-1}^2)
\end{aligned}
\tag{2.17}
$$

where (c) follows from the sequential order in which the system variables are generated.

4. Consider any $x_{t+1} \in \mathcal{X}_{t+1}$, $y^1_{t+1} \in \mathcal{Y}^1_{t+1}$, $s^1_t \in \mathcal{S}^1_t$, $s^2_t \in \mathcal{S}^2_t$, and ${}^1\varphi^t = ({}^4\varphi^{t-1}, l^2_t)$. Consider a component of ${}^1\pi_{t+1}$,

$$
\begin{aligned}
&{}^1\pi_{t+1}(x_{t+1}, y^1_{t+1}, s^1_t, s^2_t) \\
&\quad = \Pr\left(X_{t+1} = x_{t+1}, Y^1_{t+1} = y^1_{t+1}, S^1_t = s^1_t, S^2_t = s^2_t \,\middle|\, {}^1\varphi^t\right) \\
&\quad = \sum_{\substack{x_t \in \mathcal{X}_t,\, y^2_t \in \mathcal{Y}^2_t \\ u^1_t \in \mathcal{U}^1_t,\; u^2_t \in \mathcal{U}^2_t \\ s^2_{t-1} \in \mathcal{S}^2_{t-1}}} \Pr\Big(Y^1_{t+1} = y^1_{t+1} \;\Big|\; X_{t+1} = x_{t+1}, X_t = x_t, Y^2_t = y^2_t, U^1_t = u^1_t, \\
&\qquad\qquad\qquad\qquad U^2_t = u^2_t, S^1_t = s^1_t, S^2_{t-1} = s^2_{t-1}, S^2_t = s^2_t; {}^1\varphi^t\Big) \\
&\qquad\qquad \times \Pr\Big(X_{t+1} = x_{t+1} \;\Big|\; X_t = x_t, Y^2_t = y^2_t, U^1_t = u^1_t, \\
&\qquad\qquad\qquad\qquad U^2_t = u^2_t, S^1_t = s^1_t, S^2_{t-1} = s^2_{t-1}, S^2_t = s^2_t; {}^1\varphi^t\Big) \\
&\qquad\qquad \times \Pr\Big(S^2_t = s^2_t \;\Big|\; X_t = x_t, Y^2_t = y^2_t, U^1_t = u^1_t, \\
&\qquad\qquad\qquad\qquad U^2_t = u^2_t, S^1_t = s^1_t, S^2_{t-1} = s^2_{t-1}; {}^4\varphi^{t-1}, l^2_t\Big) \\
&\qquad\qquad \times \Pr\left(X_t = x_t, Y^2_t = y^2_t, U^1_t = u^1_t, U^2_t = u^2_t, S^1_t = s^1_t, S^2_{t-1} = s^2_{t-1} \,\middle|\, {}^4\varphi^{t-1}, l^2_t\right) \\
&\quad \overset{(d)}{=} \sum_{\substack{x_t \in \mathcal{X}_t,\, y^2_t \in \mathcal{Y}^2_t \\ u^1_t \in \mathcal{U}^1_t,\; u^2_t \in \mathcal{U}^2_t \\ s^2_{t-1} \in \mathcal{S}^2_{t-1}}} P_{N^1_{t+1}}\Big(n^1_{t+1} \in \mathcal{N}^1_{t+1} : y^1_{t+1} = h^1_{t+1}(x_{t+1}, n^1_{t+1})\Big) \\
&\qquad\qquad \times P_{W_t}\Big(w_t \in \mathcal{W}_t : x_{t+1} = f_t(x_t, u^1_t, u^2_t, w_t)\Big) \\
&\qquad\qquad \times \mathbb{I}\left[s^2_t = l^2_t(y^2_t, u^2_t, s^2_{t-1})\right] \\
&\qquad\qquad \times \Pr\left(X_t = x_t, Y^2_t = y^2_t, U^1_t = u^1_t, U^2_t = u^2_t, S^1_t = s^1_t, S^2_{t-1} = s^2_{t-1} \,\middle|\, {}^4\varphi^{t-1}\right) \\
&\quad = \sum_{\substack{x_t \in \mathcal{X}_t,\, y^2_t \in \mathcal{Y}^2_t \\ u^1_t \in \mathcal{U}^1_t,\; u^2_t \in \mathcal{U}^2_t \\ s^2_{t-1} \in \mathcal{S}^2_{t-1}}} P_{N^1_{t+1}}\Big(n^1_{t+1} \in \mathcal{N}^1_{t+1} : y^1_{t+1} = h^1_{t+1}(x_{t+1}, n^1_{t+1})\Big) \\
&\qquad\qquad \times P_{W_t}\Big(w_t \in \mathcal{W}_t : x_{t+1} = f_t(x_t, u^1_t, u^2_t, w_t)\Big) \\
&\qquad\qquad \times \mathbb{I}\left[s^2_t = l^2_t(y^2_t, u^2_t, s^2_{t-1})\right] {}^4\pi_t(x_t, y^2_t, u^1_t, u^2_t, s^1_t, s^2_{t-1}) \\
&\quad =: \left({}^4Q_t(l^2_t)\, {}^4\pi_t\right)(x_{t+1}, y^1_{t+1}, s^1_t, s^2_t)
\end{aligned}
\tag{2.18}
$$

where (d) follows from the sequential order in which the system variables are generated.

5. Consider the expected instantaneous cost

$$\begin{aligned}
&\mathbb{E}\left\{\rho_t(X_t, U_t^1, U_t^2) \,\middle|\, {}^4\varphi^{t-1}\right\} \\
&\quad = \sum_{(x_t \in \mathcal{X}_t,\, u_t^1 \in \mathcal{U}_t^1,\, u_t^2 \in \mathcal{U}_t^2)} \rho_t(x_t, u_t^1, u_t^2) \Pr\left(X_t = x_t, U_t^1 = u_t^1, U_t^2 = u_t^2 \,\middle|\, {}^4\varphi^{t-1}\right) \\
&\quad = \sum_{(x_t \in \mathcal{X}_t,\, u_t^1 \in \mathcal{U}_t^1,\, u_t^2 \in \mathcal{U}_t^2)} \rho_t(x_t, u_t^1, u_t^2) \\
&\qquad \times \sum_{(y_t^2 \in \mathcal{Y}_t^2,\, s_t^1 \in \mathcal{S}_t^1,\, s_{t-1}^2 \in \mathcal{S}_{t-1}^2)} \Pr\left(X_t = x_t, Y_t^2 = y_t^2, S_t^1 = s_t^1, S_{t-1}^2 = s_{t-1}^2, U_t^1 = u_t^1, U_t^2 = u_t^2 \,\middle|\, {}^4\varphi^{t-1}\right) \\
&\quad = \sum_{(x_t \in \mathcal{X}_t,\, u_t^1 \in \mathcal{U}_t^1,\, u_t^2 \in \mathcal{U}_t^2)} \rho_t(x_t, u_t^1, u_t^2) \quad \times \sum_{(y_t^2 \in \mathcal{Y}_t^2,\, s_t^1 \in \mathcal{S}_t^1,\, s_{t-1}^2 \in \mathcal{S}_{t-1}^2)} {}^4\pi_t(x_t, y_t^2, u_t^1, u_t^2, s_t^1, s_{t-1}^2) \\
&\quad =: \hat{\rho}_t({}^4\pi_t)
\end{aligned} \tag{2.19}$$

Equations (2.15)–(2.18) imply that the transformations ${}^1Q_t$, ${}^2Q_t$, ${}^3Q_t$ and ${}^4Q_t$ are linear in the sense that if ${}^i\pi_t^{(1)}, {}^i\pi_t^{(2)} \in {}^i\Pi_t$, ${}^i\varphi_t \in {}^i\Phi_t$, and $\lambda \in [0,1]$ then

$${}^iQ_t({}^i\varphi_t)\Big(\lambda\, {}^i\pi_t^{(1)} + (1-\lambda)\, {}^i\pi_t^{(2)}\Big) = \lambda\, {}^iQ_t({}^i\varphi_t)\, {}^i\pi_t^{(1)} + (1-\lambda)\, {}^iQ_t({}^i\varphi_t)\, {}^i\pi_t^{(2)}.$$

Further (2.19) implies that $\hat{\rho}$ is linear in ${}^4\pi$, i.e.,

$$\hat{\rho}_t\Big(\lambda\, {}^4\pi_t^{(1)} + (1-\lambda)\, {}^4\pi_t^{(2)}\Big) = \lambda \hat{\rho}_t({}^4\pi_t^{(1)}) + (1-\lambda)\hat{\rho}_t({}^4\pi_t^{(2)}) \qquad \square$$

The information states ${}^i\pi_t$, $i = 1,2,3,4$, satisfy (s1) by definition. Part 1 of Lemma 2.1 shows that they satisfy (s2); part 2 shows that they satisfy (s3). Thus, using these information states we should be able to obtain a sequential decomposition of Problem 2.1 However, (s4) is not satisfied in general. So, this sequential decomposition will not always extend to infinite horizon problems.

In order to obtain a sequential decomposition of Problem 2.1 we consider the problem of optimally controlling the time evolution of the information states ${}^1\pi_t$, ${}^2\pi_t$, ${}^3\pi_t$ and ${}^4\pi_t$, $t = 1, \dots, T$. We then consider four time-homogeneous variations of the system of Problem 2.1, and show how to extend the sequential decomposition to infinite horizon problems for three of these four variations.

An equivalent optimization problem

Consider a centralized deterministic optimization problem with state space alternating between ${}^1\Pi_t$, ${}^2\Pi_t$, ${}^3\Pi_t$ and ${}^4\Pi_t$ and action space alternating between $\mathscr{G}_t^1$, $\mathscr{L}_t^1$, $\mathscr{G}_t^2$, and $\mathscr{L}_t^2$. The system dynamics are given by (2.13) and at each stage t the decision rules g_t^1, l_t^1, g_t^2, and l_t^2 are determined according to ***meta-functions*** or ***meta-rules*** ${}^1\Delta_t$, ${}^2\Delta_t$, ${}^3\Delta_t$ and ${}^4\Delta_t$, where ${}^1\Delta_t$ is a function from ${}^1\Pi_t$ to $\mathscr{G}_t^1$, ${}^2\Delta_t$ is a function from ${}^2\Pi_t$ to $\mathscr{L}_t^1$, ${}^3\Delta_t$ is a function from ${}^3\Pi_t$ to $\mathscr{G}_t^2$, and ${}^4\Delta_t$ is a function from ${}^4\Pi_t$ to $\mathscr{L}_t^2$. Thus the system equations (2.13) can be written as

$$g_t^1 = {}^1\Delta_t({}^1\pi_t), \qquad {}^2\pi_t = {}^1Q_t(g_t^1)\,{}^1\pi_t, \tag{2.20a}$$

$$l_t^1 = {}^2\Delta_t({}^2\pi_t), \qquad {}^3\pi_t = {}^2Q_t(l_t^1)\,{}^2\pi_t, \tag{2.20b}$$

$$g_t^2 = {}^3\Delta_t({}^3\pi_t), \qquad {}^3\pi_t = {}^3Q_t(g_t^2)\,{}^1\pi_t, \tag{2.20c}$$

$$l_t^2 = {}^4\Delta_t({}^4\pi_t), \qquad {}^1\pi_{t+1} = {}^4Q_t(l_t^2)\,{}^4\pi_t. \tag{2.20d}$$

The initial state ${}^1\pi_1 = \Pr(X_1, Y_1)$ is given (in terms of P_{X_1} and $P_{N_1^1}$). An instantaneous cost $\hat{\rho}_t({}^4\pi_t)$ is incurred at each stage. The choice $({}^1\Delta_1, {}^2\Delta_1, {}^3\Delta_1, {}^4\Delta_t, \dots, {}^1\Delta_T, {}^2\Delta_T, {}^3\Delta_T, {}^4\Delta_T)$ is called a ***meta-strategy*** or a ***meta-design*** and denoted by Δ^T. The performance of a meta–strategy is given by the total cost incurred by that meta-strategy, i.e.,

$$\mathcal{J}_T(\Delta^T)({}^1\pi_1) = \sum_{t=1}^{T} \hat{\rho}_t({}^4\pi_t). \tag{2.21}$$

Now consider the following optimization problem:

Problem 2.2. *Consider the dynamic system* (2.20) *with known transformations* ${}^1Q_t$, ${}^2Q_t$, ${}^3Q_t$ *and* ${}^4Q_t$ *given by* (2.15)–(2.18). *The initial state* ${}^1\pi_1$ *is given. Determine a meta–strategy* Δ^T *to minimize the total cost given by* (2.21).

Given any meta-strategy Δ^T, the time evolution of ${}^i\pi_t$ is deterministic; ${}^i\pi_t$ and the corresponding ${}^i\varphi_t$ can be determined from (2.20). Thus, for any given initial state ${}^1\pi_1$ and any choice of meta-strategy, we can determine a corresponding design. Further, we can rewrite the performance criterion of (2.5) as

$$
\begin{aligned}
\mathcal{J}_T(G^1, L^1, G^2, L^2) &= \mathbb{E}\left\{ \sum_{t=1}^{T} \rho_t(X_t, U_t^1, U_t^2) \,\middle|\, G^1, L^1, G^2, L^2 \right\} \\
&\overset{(a)}{=} \sum_{t=1}^{T} \mathbb{E}\left\{ \rho_t(X_t, U_t^1, U_t^2) \,\middle|\, {}^4\varphi^{t-1} \right\} \\
&\overset{(b)}{=} \sum_{t=1}^{T} \hat{\rho}_t({}^4\pi_t) \\
&=: \mathcal{J}_T(\Delta^T)({}^1\pi_1) \qquad (2.22)
\end{aligned}
$$

where (a) follows from the sequential ordering of system variables and (b) follows from Lemma 2.1.

For any meta-strategy Δ^T and any initial state ${}^1\pi_1$, the system equations (2.20) determine a strategy (G^1, L^1, G^2, L^2). We call (G^1, L^1, G^2, L^2) the strategy corresponding to Δ^T and ${}^1\pi_1$. For any strategy (G^1, L^1, G^2, L^2) and any initial state ${}^1\pi_1$, there exist many meta-strategies Δ^T such that (G^1, L^1, G^2, L^2) is the strategy corresponding to Δ^T and ${}^1\pi_1$. One such meta-strategy is

$$
g_t^1 = {}^1\Delta_t({}^1\pi_t), \quad l_t^1 = {}^2\Delta_t({}^2\pi_t), \quad g_t^2 = {}^3\Delta_t({}^3\pi_t), \quad l_t^2 = {}^4\Delta_t({}^4\pi_t)
$$

for all ${}^i\pi_t \in {}^i\Pi_t$, $i = 1, 2, 3, 4$, $t = 1, \dots, T$. Equation (2.22) implies that for a given ${}^1\pi_1$, the performance of any meta-strategy Δ^T in Problem 2.2 is same as the performance of the corresponding strategy (G^1, L^1, G^2, L^2) in Problem 2.1. Therefore, the optimal performance of Problems 2.1 and 2.2 are same; furthermore, the design $(G^{1,*}, L^{1,*}, G^{2,*}, L^{2,*})$ corresponding to any optimal meta-strategy $\Delta^{*,T}$ for Problem 2.2 is optimal for Problem 2.1.

The global optimization algorithm

Problem 2.2 can be formulated as a classical centralized optimization problem by considering the information state ${}^i\pi_t$ as the "controlled state" at time ${}^it$, the decision rule ${}^i\varphi_t$ (which is equal to g_t^1, l_t^1, g_t^2, or l_t^2 depending on ${}^it$) as the "control action" (or decision) at time ${}^it$, and the meta-function ${}^i\Delta_t$ as the "control law" at time ${}^it$. Hence, an optimal meta-strategy for Problem 2.2 is given by the optimal "control strategy" of the centralized optimization problem and can be determined as follows:

Theorem 2.1 (Global optimization algorithm). *An optimal meta-strategy $\Delta^{*,T}$ for Problem 2.2 (and consequently an optimal design for Problem 2.1) can be determined by the solution of the following nested optimality equations. For all ${}^i\pi_t \in {}^i\Pi_t$, $i = 1, 2, 3, 4$, $t = 1, \dots, T$, define*

$$ {}^1V_{T+1}({}^1\pi_{T+1}) = 0, \tag{2.23a} $$

and for $t = 1, \dots, T$

$$ {}^1V_t({}^1\pi_t) = \inf_{g_t^1 \in \mathscr{G}_t^1} {}^2V_t\Big({}^1Q_t(g_t^1)\, {}^1\pi_t \Big), \tag{2.23b} $$

$$ {}^2V_t({}^2\pi_t) = \inf_{l_t^1 \in \mathscr{L}_t^1} {}^3V_t\Big({}^2Q_t(l_t^1)\, {}^2\pi_t \Big), \tag{2.23c} $$

$$ {}^3V_t({}^3\pi_t) = \inf_{g_t^2 \in \mathscr{G}_t^2} {}^4V_t\Big({}^3Q_t(g_t^2)\, {}^3\pi_t \Big), \tag{2.23d} $$

$$ {}^4V_t({}^4\pi_t) = \hat{\rho}_t({}^4\pi_t) + \inf_{l_t^2 \in \mathscr{L}_t^2} {}^1V_{t+1}\Big({}^4Q_t(l_t^2)\, {}^4\pi_t \Big). \tag{2.23e} $$

The functions ${}^iV_t$ are called ***value functions****; they represent the minimum expected future cost that the system in state ${}^i\pi$ will incur from time ${}^it$ onwards. These value functions can be determined sequentially by moving backwards in time. The optimal performance of Problem 2.2 (and Problem 2.1) is given by*

$$ \mathcal{J}_T^* = {}^1V_1({}^1\pi_1). \tag{2.24} $$

For any ${}^it$ and ${}^i\pi$, the $\arg\min$ *(or* $\arg\inf$*) in the RHS of ${}^iV_t({}^i\pi)$ equals to the optimal value of the meta-function ${}^i\Delta_t({}^i\pi_t)$. Thus, solving for the value functions for all values of the information state also determines an optimal meta-strategy $\Delta^{*,T}$ for Problem 2.2. Relations* (2.20) *can be used to determine optimal design for Problem 2.1.*

Proof. This is a standard result for a deterministic optimization problem, see Dynkin and Yushkevich (1975, Chapter 6). □

We can combine the four step T-stage sequential decomposition of (2.23) into a one-step T-stage sequential decomposition as follows

$$ {}^1V_{T+1}({}^1\pi_{T+1}) = 0, \tag{2.25a} $$

and for $t = 1, \dots, T$

$$
{}^1V_t({}^1\pi_t) = \inf_{\substack{g_t^1 \in \mathscr{G}_t^1 \\ l_t^1 \in \mathscr{L}_t^1 \\ g_t^2 \in \mathscr{G}_t^2 \\ l_t^2 \in \mathscr{L}_t^2}} \Bigg[\hat{\rho}_t\Big(\big({}^3Q_t(g_t^2) \circ {}^2Q_t(l_t^1) \circ {}^1Q_t(g_t^1) \big) {}^1\pi_t \Big) + {}^1V_{t+1}\Big(\big({}^4Q_t(l_t^2) \circ {}^3Q_t(g_t^2) \circ {}^2Q_t(l_t^1) \circ {}^1Q_t(g_t^1) \big) {}^1\pi_t \Big) \Bigg]. \tag{2.25b}
$$

The above decomposition is equivalent to a deterministic dynamic program in function space. Theorem 2.1 presents a finer decomposition which corresponds to the refinement of time presented in Figure 2.2; the decomposition given by (2.23) has a smaller search space than the decomposition given in (2.25).

We also have the following result

Theorem 2.2 (Concavity of Value Functions). *The value functions ${}^iV_t$, $i = 1, \dots, 4$, $t = 1, \dots, T$, given by* (2.23) *are concave in the corresponding ${}^i\pi$.*

Proof. Recall that ${}^1Q$, ${}^2Q$, ${}^3Q$ and ${}^4Q$ are linear in ${}^i\pi$ and $\hat{\rho}_t(\cdot)$ is linear in ${}^4\pi$. The result of the theorem follows from the fact that concavity is maintained under composition of a concave function with a linear transformation, summation of concave functions, and point–wise minimum/infimum of concave functions. We will proceed by backward induction. Observe that ${}^1V_{T+1}$ is a concave function of ${}^1\pi$. Assume that ${}^1V_{t+1}$ is a concave function of ${}^1\pi$. We will show that ${}^iV_t$, $i = 1, \dots, T$ are concave functions of ${}^i\pi$. Define

$$
{}^4W_t({}^4\pi_t, l_t^2) \coloneqq \hat{\rho}_t({}^4\pi_t) + {}^1V_{t+1}({}^4Q(l_t^2)\, {}^4\pi_t). \tag{2.26}
$$

${}^4W$ is a composition of a sum of two functions: one is linear in ${}^4\pi_t$ and the other is a composition of a concave function with a linear transformation. Hence ${}^4W$ is concave in ${}^4\pi_t$. Now,

$$
{}^4V_t({}^4\pi_t) = \min_{l_t^2 \in \mathscr{L}_t^2} {}^4W({}^4\pi_t, l_t^2). \tag{2.27}
$$

Since ${}^4W_t$ is concave in ${}^4\pi$, and ${}^4V_t$ is the point–wise minimum of ${}^4W_t$, ${}^4V_t$ is concave in ${}^4\pi$. Similar argument extends to the other three cases. Hence ${}^iV_t$ is concave in ${}^i\pi$, $i = 1, \dots, 4$. Thus, by induction ${}^iV_t$, $i = 1, \dots, 4$, $t = 1, \dots, T$ is concave in ${}^i\pi$. □

Summary of the result

We have shown that by an appropriate choice of information states, the general two-agent team problem (Problem 2.1) can be transformed into a deterministic perfectly

observed MDP (Problem 2.2) with state spaces ${}^{i}\Pi$ which are continuous and action spaces $\mathscr{G}_t^1$, $\mathscr{L}_t^1$, $\mathscr{G}_t^2$, and $\mathscr{L}_t^2$ which are function spaces. The information state is perfectly observed by both the agents since they know each other's decision rules. The search for an optimal design proceeds in two steps: the backward step and the forward step. In the backward step, both agents agree upon a rule to break ties and simultaneously determine meta-functions for each information state at each time by proceeding backwards in time. In the forward step, they start with the commonly known initial value of information state and use the result of the backward step to determine the decision rules for both of them for all times by moving forwards in time. This is similar to a standard deterministic MDP (e.g., travelling salesman problem), where the agent can simply store the control actions rather than the control laws.

Furthermore, when all the system variables are finite the nested optimality equations of (2.23) are similar to the nested optimality equations of POMDPs ??: the information state ${}^{1}\pi$, ${}^{2}\pi$, ${}^{3}\pi$ and ${}^{4}\pi$ are probability measures on finite spaces and the action spaces $\mathscr{G}_t^1$, $\mathscr{L}_t^1$, $\mathscr{G}_t^2$, and $\mathscr{L}_t^2$ are finite. So, the computational methods of solving POMDPs are also applicable to two-agent teams. However, there is a big difference between the sequential decomposition of POMDPs and two-agent teams. In POMDPs each step of the sequential decomposition ?? is a parameter optimization problem (we have to choose the best control action U_t) while in two-agent teams each step of the sequential decomposition (2.23) is a functional optimization problem (we have to choose the best decision function g_t^1, l_t^1, g_t^2, or l_t^2). This difference makes it harder to solve the sequential decomposition equations of two-agent teams than the sequential decomposition equations of POMDPs. We will talk about computational complexity in detail in later sections.

2.3 An example—real-time communication

Consider a real-time communication system that consists of a binary Markov source that must be transmitted over a binary discrete memoryless channel in real-time. Suppose that the system operates for three steps. The source output is denoted by $\{X_1, X_2, X_3\}$ and it is assumed that

$$P_{X_1} = \begin{bmatrix} 0.4 & 0.6 \end{bmatrix}, \qquad P_{X_2|X_1} = P_{X_3|X_2} = \begin{bmatrix} 1.0 & 0.0 \\ 0.1 & 0.9 \end{bmatrix}.$$

An encoder encodes the source output in real-time to generate binary encoded symbols $\{Z_1, Z_2, Z_3\}$ as follows:

$$Z_1 = c_1(X_1), \quad Z_2 = c_2(X_1, X_2), \quad Z_3 = c_3(X_1, X_2, X_3).$$

The functions $c_1(\cdot)$, $c_2(\cdot)$, and $c_3(\cdot)$ are called ***encoding functions.*** The encoded symbols $\{Z_1, Z_2, Z_3\}$ are transmitted over a memoryless Z-channel and generate binary source outputs $\{Y_1, Y_2, Y_3\}$ as follows:

$$Y_t = Z_t \cdot N_t, \quad t = 1, 2, 3,$$

where $\{N_1, N_2, N_3\}$ denotes the channel noise and are binary i.i.d. random variables with the distribution

$$P_{N_1} = P_{N_2} = P_{N_3} = \begin{bmatrix} 0.1 & 0.9 \end{bmatrix}.$$

A decoder observes $\{Y_1, Y_2, Y_3\}$ and generates estimates $\left\{\hat{X}_1, \hat{X}_2, \hat{X}_3\right\}$ of the source in real-time as follows:

$$\hat{X}_1 = g_1(Y_1), \quad \hat{X}_2 = g_2(Y_1, Y_2), \quad \hat{X}_3 = g_3(Y_1, Y_2, Y_3).$$

The functions $g_1(\cdot)$, $g_2(\cdot)$, and $g_3(\cdot)$ are called ***decoding functions.*** The objective is to determine encoding and decoding functions $(c_1, c_2, c_3, g_1, g_2, g_3)$ to minimize the probability of error

$$P_e = \mathbb{E}\left\{\rho(X_1, \hat{X}_1) + \rho(X_2, \hat{X}_2) + \rho(X_3, \hat{X}_3) \,\middle|\, c_1, c_2, c_3, g_1, g_2, g_3\right\}$$

where

$$\rho(X, \hat{X}) = \begin{cases} 0, & \text{if } X = \hat{X}, \\ 1, & \text{if } X \neq \hat{X}, \end{cases}$$

On solving the nested optimality equations of Theorem 2.1, we find that an optimal solution is (the optimal solution was obtained numerically, it is simply presented in analytic form here).

$$c_1(X_1) = 1 - X_1, \quad c_2(X_1, X_2) = 1 - X_1 \cdot X_2,$$

$$c_3(X_1, X_2, X_3) = 1 - X_1 \cdot X_2 \cdot X_3$$

and

$$g_1(Y_1) = 1 - Y_1, g_2(Y_1, Y_2) = (1 - Y_1) \cdot (1 - Y_2),$$
$$g_3(Y_1, Y_2, Y_3) = (1 - Y_1) \cdot (1 - Y_2) \cdot (1 - Y_3)$$

The probability of error under this scheme is 141/2500.

2.4 The time homogeneous system—the four variations

The information states of Definition 2.3 do not belong to time-invariant spaces. Consequently, the sequential decomposition of Theorem 2.1 cannot be extended to infinite horizons in general. In the remainder of this chapter, we identify instances of Problem 2.1 where the sequential decomposition can be extended to infinite horizon problems.

We restrict attention to time-homogeneous systems, i.e., systems where (i) the spaces $\mathcal{X}_t$, $\mathcal{Y}_t^1$, $\mathcal{Y}_t^2$, $\mathcal{U}_t^1$, $\mathcal{U}_t^2$, $\mathcal{W}_t$, $\mathcal{N}_t^1$, and $\mathcal{N}_t^2$ do not depend on t; (ii) the noise statistics P_{W_t}, $P_{N_t^1}$ and $P_{N_t^2}$ do not depend on t; (iii) the plant function $f_t(\cdot)$, the observation functions $h_t^1(\cdot)$ and $h_t^2(\cdot)$, and the cost function $\rho_t(\cdot)$ do not depend on time t; and (iv) each agent has either time-invariant state (i.e., $\mathcal{S}_t^k$ does not depend on t), or has perfect recall. Based on the state of each agent, there are four variations of a time-homogeneous two-agent system:

v1. both agents have time-invariant state space;
v2. agent 1 has perfect recall, agent 2 has time-invariant state space;
v3. agent 1 has time-invariant state space, agent 2 has perfect recall;
v4. both agents have perfect recall.

We will consider the first three of these four variation in the sequel.

2.5 Time-homogeneous system—Variation v1

Consider a time-homogeneous variation of the system of Problem 2.1 with both agents having time-invariant states denoted by $\mathcal{S}^1$ and $\mathcal{S}^2$, respectively. We first show how this simplifies the sequential decomposition of the finite horizon problem derived in Section 2.2. We then will consider two instances of the corresponding infinite horizon problem, and derive optimality equations whose solutions determine a globally optimal design for these instances.

The finite horizon problem

For a time-homogeneous variation of Problem 2.1 with time-invariant state at both agents, we can simplify some of the optimality equations derived in the previous section. The spaces $\mathscr{G}_t^1$, $\mathscr{L}_t^1$, $\mathscr{G}_t^2$, $\mathscr{L}_t^2$, ${}^1\Pi_t$, ${}^2\Pi_t$, ${}^3\Pi_t$ and ${}^4\Pi_t$ do not depend on time; so, we can drop the subscript t and simply denote them by $\mathscr{G}^1$, $\mathscr{L}^1$, $\mathscr{G}^2$, $\mathscr{L}^2$, ${}^1\Pi$, ${}^2\Pi$, ${}^3\Pi$ and ${}^4\Pi$, respectively. Furthermore, the transformations ${}^1Q_t$, ${}^2Q_t$, ${}^3Q_t$ and ${}^4Q_t$ and the function $\hat{\rho}_t$ of Lemma 2.1 do not depend on t; so, we can denote them by ${}^1Q$, ${}^2Q$, ${}^3Q$, ${}^4Q$ and $\hat{\rho}$, respectively. Thus, Problem 2.2 reduces to a time-homogeneous problem—the state space, the action space, the system update equations, and the instantaneous cost do not depend on t. Hence, we can simplify Theorem 2.1 as follows.

Corollary 2.1. *If the system of Problem 2.1 is time-homogeneous, the nested optimality equations* (2.23) *can be written as*

$$ {}^1V_{T+1}({}^1\pi) = 0, \tag{2.28a} $$

and for $t = 1, \dots, T$

$$ {}^1V_t({}^1\pi) = \inf_{g_t^1 \in \mathscr{G}^1} {}^2V_t\Big({}^1Q(g_t^1)\,{}^1\pi\Big), \tag{2.28b} $$

$$ {}^2V_t({}^2\pi) = \inf_{l_t^1 \in \mathscr{L}^1} {}^3V_t\Big({}^2Q(l_t^1)\,{}^2\pi\Big), \tag{2.28c} $$

$$ {}^3V_t({}^3\pi) = \inf_{g_t^2 \in \mathscr{G}^2} {}^4V_t\Big({}^3Q(g_t^2)\,{}^3\pi\Big), \tag{2.28d} $$

$$ {}^4V_t({}^4\pi) = \hat{\rho}({}^4\pi) + \inf_{l_t^2 \in \mathscr{L}^2} {}^1V_{t+1}\Big({}^4Q(l_t^2)\,{}^4\pi\Big). \tag{2.28e} $$

Infinite horizon problems

For the time-homogeneous model of the two-agent team of Section 2.1 we can formulate infinite horizon ($T \to \infty$) optimization problems using two criteria: the expected discounted cost and the average cost per unit time. Let (G^1, L^1, G^2, L^2), where $G^1 \coloneqq (g_1^1, g_2^1, \dots)$, $L^1 \coloneqq (l_1^1, l_2^1, \dots)$, $G^2 \coloneqq (g_1^2, g_2^2, \dots)$, $L^2 \coloneqq (l_1^2, l_2^2, \dots)$ denote an infinite horizon design. The two performance criteria that we consider are:

11. The expected discounted cost criterion

Under this criterion the performance of a design is given by

$$\mathcal{J}^{\beta}(G^1, L^1, G^2, L^2) := \mathbb{E}\left\{ \sum_{t=1}^{\infty} \beta^{t-1} \rho(X_t, U_t^1, U_t^2) \,\middle|\, G^1, L^1, G^2, L^2 \right\} \tag{2.29}$$

where $0 < \beta < 1$ is called the discount factor.

12. ***The average cost per unit time criterion***

Under this criterion the performance of a design is given by

$$\bar{\mathcal{J}}(G^1, L^1, G^2, L^2) := \limsup_{T \to \infty} \frac{1}{T} \mathbb{E}\left\{ \sum_{t=1}^{T} \rho(X_t, U_t^1, U_t^2) \,\middle|\, G^1, L^1, G^2, L^2 \right\}. \tag{2.30}$$

We take the lim sup rather than the lim as for some designs (G^1, L^1, G^2, L^2) the limit may not exist.

Ideally, while implementing a solution for infinite horizon problems, we would like to use time-invariant designs. This motivates the following definitions.

Definition 2.4 (Stationary design). *A infinite horizon design (G^1, L^1, G^2, L^2), where $G^1 := (g_1^1, g_2^1, \ldots)$, $L^1 := (l_1^1, l_2^1, \ldots)$, $G^2 := (g_1^2, g_2^2, \ldots)$, $L^2 := (l_1^2, l_2^2, \ldots)$, is called stationary (or time-invariant) if $g_1^1 = g_2^1 = \cdots := g^1$, $l_1^1 = l_2^1 = \cdots := l^1$, $g_1^2 = g_2^2 = \cdots := g^2$, and $l_1^2 = l_2^2 = \cdots := l^2$. Such a stationary design is equivalently denoted by $(g^{1,\infty}, l^{1,\infty}, g^{2,\infty}, l^{2,\infty})$.*

Definition 2.5 (Stationary meta-strategy). *A meta-strategy $\tilde{\Delta}^{\infty} = (\tilde{\Delta}_1, \tilde{\Delta}_2, \cdots)$, where $\tilde{\Delta}_t = ({}^1\Delta_t, {}^2\Delta_t, {}^3\Delta_t, {}^4\Delta_t)$, is called stationary (or time-invariant) if $\tilde{\Delta}_1 = \tilde{\Delta}_2 = \cdots := \tilde{\Delta}$.*

In time-homogeneous infinite-horizon stochastic optimization problems with classical information structures, there is no loss in optimality in restricting attention to stationary strategies, see Kumar and Varaiya (1986). This result drastically simplifies the search for an optimal solution. In general, restricting attention to stationary strategies is not optimal for problems with non-classical information structures, see Sandell (1974). However, there is no loss of optimality in restricting attention to stationary meta-strategies: for the expected discounted cost criterion there exist stationary meta-strategies that are optimal; for the average cost per unit time criterion, under a technical condition, there exist stationary meta-strategies that are arbitrarily close to optimal. ***However, the optimal design corresponding to the optimal stationary meta-strategy is, in general, time-varying.***

The expected discounted cost problem

Consider a time-homogeneous infinite-horizon problem for the system of Problem 2.1 with the expected discounted cost criterion of (2.29). Consider ${}^1\pi_t$, ${}^2\pi_t$, ${}^3\pi_t$

and ${}^4\pi_t$ as in Definition 2.3: they satisfy the properties of Lemma 2.1; further, since the system is time-invariant, the transformations ${}^1Q$, ${}^2Q$, ${}^3Q$ and ${}^4Q$ and the expected instantaneous cost $\hat{\rho}$ do not depend on t. Let $\gamma_t \coloneqq (g_t^1, l_t^1, g_t^2, l_t^2)$ denote the decision rules at time t and Γ denote the space $\mathscr{G}^1 \times \mathscr{L}^1 \times \mathscr{G}^2 \times \mathscr{L}^2$. We can combine (2.20) as

$$ {}^1\pi_{t+1} = \tilde{Q}(\gamma_t)\, {}^1\pi_t, \qquad \gamma_t = \tilde{\Delta}_t({}^1\pi_t) \tag{2.31} $$

where

$$ \tilde{Q}(\gamma_t) \coloneqq {}^4Q(l_t^2) \circ {}^3Q(g_t^2) \circ {}^2Q(l_t^1) \circ {}^1Q(g_t^1), $$

$$ \tilde{\Delta}_t({}^1\pi_t) \coloneqq \Big({}^1\Delta_t({}^1\pi_t), {}^2\Delta_t({}^2\pi_t), {}^3\Delta_t({}^3\pi_t), {}^4\Delta_t({}^4\pi_t) \Big) $$

and ${}^2\pi_t$, ${}^3\pi_t$, and ${}^4\pi_t$ are related to ${}^1\pi_t$ by the time-invariant analogue of (2.13). Further, the instantaneous cost at time t can be written as

$$ \tilde{\rho}({}^1\pi_t, {}^1\gamma_t) = \hat{\rho}\Big({}^3Q(g_t^2) \circ {}^2Q(l_t^1) \circ {}^1Q(g_t^1)\, {}^1\pi_t \Big) $$

Hence, the time-homogeneous infinite horizon problem for the system of Problem 2.1 with the expected discounted cost criterion of (2.29) is equivalent to the following deterministic optimization problem.

Problem 2.3. *Consider a deterministic system with state space ${}^1\Pi$ and action space Γ. The system dynamics are given by*

$$ {}^1\pi_{t+1} = \tilde{Q}(\gamma_t)\, {}^1\pi_t, \qquad \gamma_t = \tilde{\Delta}_t({}^1\pi_t) \tag{2.32} $$

where $\tilde{Q}$ is a known transformation and $\tilde{\Delta} : \Pi \to \Gamma$ is a meta-function that is to be determined. At each time an instantaneous cost $\tilde{\rho}({}^1\pi_t, \gamma_t)$ is incurred. The initial state ${}^1\pi_1$ is known. The objective is to determine a meta-strategy $\tilde{\Delta}^\infty = (\tilde{\Delta}_1, \tilde{\Delta}_2, \dots)$ so as to minimize the discounted infinite horizon total cost given by

$$ \mathcal{J}^\beta(\tilde{\Delta}^\infty) = \sum_{t=1}^{\infty} \beta^{t-1} \tilde{\rho}({}^1\pi_t, \gamma_t) \tag{2.33} $$

Problem 2.3 is a standard deterministic time-invariant infinite horizon problem with total discounted cost criterion. Since we have assumed $\rho(\cdot)$ to be uniformly bounded, $\hat{\rho}$ and $\tilde{\rho}$ are also uniformly bounded, therefore an optimal meta-strategy is guaranteed to exist and we have the following result:

Theorem 2.3. *For Problem 2.3 and consequently for the infinite horizon expected discounted cost problem for the system of Problem 2.1 with the performance criterion given by* (2.29)*, there is no loss of optimality in restricting attention to stationary meta-strategies. Specifically there exists a stationary meta-strategy* $\tilde{\Delta}^{*,\infty} := (\tilde{\Delta}^*, \tilde{\Delta}^*, \ldots)$*, and a corresponding infinite horizon design* $(G^{1,*}, L^{1,*}, G^{2,*}, L^{2,*})$*, where* $G^{1,*} := (g_1^{1,*}, g_2^{1,*}, \ldots)$*,* $L^{1,*} := (l_1^{1,*}, l_2^{1,*}, \ldots)$*,* $G^{2,*} := (g_1^{2,*}, g_2^{2,*}, \ldots)$*,* $L^{2,*} := (l_1^{2,*}, l_2^{2,*}, \ldots)$*, such that*

$$\mathcal{J}^{\beta}(\tilde{\Delta}^{*,\infty}) = V({}^1\pi_1), \tag{2.34}$$

where V *is the unique uniformly bounded fixed point of*

$$V({}^1\pi) = \min_{\gamma \in \Gamma} \left\{ \tilde{\rho}({}^1\pi, \gamma) + \beta V\Big(\tilde{Q}(\gamma)({}^1\pi)\Big) \right\}, \tag{2.35}$$

and $\tilde{\Delta}^*$ *satisfies*

$$V({}^1\pi) = \tilde{\rho}\Big({}^1\pi, \tilde{\Delta}^*({}^1\pi)\Big) + \beta V\Big(\tilde{Q}\Big(\tilde{\Delta}^*({}^1\pi)\Big)({}^1\pi)\Big). \tag{2.36}$$

Optimal decision rules $(g_t^{1,*}, l_t^{1,*}, g_t^{2,*}, l_t^{2,*})$ *at time t are given by*

$$(g_t^{1,*}, l_t^{1,*}, g_t^{2,*}, l_t^{2,*}) =: \gamma_t^* = \tilde{\Delta}^*({}^1\pi_t). \tag{2.37}$$

Proof. This is a standard result, see Dynkin and Yushkevich (1975, Chapter 6). □

Observe that the fixed point equation (2.35) can be decomposed into its "natural" sequential form as

$${}^1V({}^1\pi) = \inf_{g^1 \in \mathscr{G}^1} {}^2V({}^1Q(g^1)\,{}^1\pi) \tag{2.38a}$$

$${}^2V({}^2\pi) = \min_{l^1 \in \mathscr{L}^1} {}^3V({}^2Q(l^1)\,{}^2\pi) \tag{2.38b}$$

$${}^3V({}^3\pi) = \inf_{g^2 \in \mathscr{G}^2} {}^4V({}^3Q(g^2)\,{}^3\pi) \tag{2.38c}$$

$${}^4V({}^4\pi) = \tilde{\rho}({}^4\pi) + \min_{l^2 \in \mathscr{L}^2} {}^1V({}^4Q(l^2)\,{}^4\pi) \tag{2.38d}$$

These equations are the infinite horizon analogues of (2.23).

The average cost per unit time problem

Consider the time-homogeneous infinite horizon problem for the system of Problem 2.1 with the average cost per unit time criterion of (2.30). Using arguments

similar to those of the first paragraph of the previous section, this problem is equivalent to the following deterministic problem:

Problem 2.4. *Consider a deterministic system with state space* ${}^1\Pi$ *and action space* Γ. *The system dynamics are given by*

$$ {}^1\pi_{t+1} = \tilde{Q}(\gamma)\,{}^1\pi_t, \qquad \gamma_t = \tilde{\Delta}_t({}^1\pi_t) \tag{2.39} $$

where $\tilde{Q}$ *is a known transformation and* $\tilde{\Delta} : \Pi \to \Gamma$ *is a meta-function. At each time an instantaneous cost* $\tilde{\rho}({}^1\pi_t, \gamma_t)$ *is incurred. The initial state* ${}^1\pi_1$ *is known. The objective is to determine a meta-strategy* $\tilde{\Delta}^\infty = (\tilde{\Delta}_1, \tilde{\Delta}_2, \ldots)$ *so as to minimize the average cost per unit time over an infinite horizon, given by*

$$ \bar{\mathcal{J}}(\tilde{\Delta}^\infty) = \limsup_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} \tilde{\rho}({}^1\pi_t, \gamma_t). \tag{2.40} $$

Problem 2.4 *cannot* be solved by taking the limit $\beta \to 1$ in the result of Theorem 2.3. Such a result is valid only if the problem has finite state and action space, see Whittle (1983, Theorem 31.5.2), which is not the case here. See Arapostathis et al. (1993) for a survey of various results connecting the expected discounted cost problem with the average cost per unit time problem.

For Problem 2.4 an optimal meta-strategy may not exist. However, under suitable conditions, we can guarantee the existence of meta-strategies that are arbitrarily close to optimal. Specifically, we have the following result:

Theorem 2.4. *For Problem 2.4 and correspondingly for the infinite horizon average cost per unit time problem with the performance criterion given by* (2.30)*, assume that:*

A1. *For any* $\epsilon > 0$ *there exist bounded measurable functions* $v(\cdot)$ *and* $r(\cdot)$ *and meta-function* $\tilde{\Delta}^* : \Pi \to \Gamma$ *such that for all* ${}^1\pi$,

$$ v({}^1\pi) = \min_{\gamma\in\Gamma} v\Big(\tilde{Q}(\gamma)\,{}^1\pi\Big) = v\Big(\tilde{Q}\Big(\tilde{\Delta}^*({}^1\pi)\Big)\,{}^1\pi\Big), \tag{2.41} $$

and

$$ \tilde{\rho}({}^1\pi, \tilde{\Delta}^*({}^1\pi)) + r\Big(\tilde{Q}\Big(\tilde{\Delta}^*({}^1\pi)\Big)\,{}^1\pi\Big) \le v({}^1\pi) + r({}^1\pi) \le \min_{\gamma\in\Gamma} \Big\{\tilde{\rho}({}^1\pi, \gamma) + r(\tilde{Q}(\gamma)\,{}^1\pi)\Big\} + \epsilon. \tag{2.42} $$

Then for any horizon T and any meta-strategy $\tilde{\Delta}^T := (\tilde{\Delta}_1, \dots, \tilde{\Delta}_T)$*, the stationary meta-strategy* $\tilde{\Delta}^{*,T} := (\tilde{\Delta}^*, \dots, \tilde{\Delta}^*)$ *(T-times) satisfies*

$$\mathcal{J}_T(\tilde{\Delta}^{*,T}) \le r({}^1\pi_1) + Tv({}^1\pi_1) \le \mathcal{J}_T(\tilde{\Delta}^T) + T\epsilon \tag{2.43}$$

Further, the stationary meta-strategy $\tilde{\Delta}^{*,\infty} := (\tilde{\Delta}^*, \tilde{\Delta}^*, \dots)$ *is* ϵ*-optimal (i.e.,* ϵ *close to optimal). That is, for any infinite horizon meta-strategy* $\tilde{\Delta}^\infty := (\tilde{\Delta}_1, \tilde{\Delta}_2, \dots)$ *we have*

$$\overline{\mathcal{J}}(\tilde{\Delta}^{*,\infty}) \le v({}^1\pi_1) \le \underline{\mathcal{J}}(\tilde{\Delta}^\infty) + \epsilon \tag{2.44}$$

where

$$\overline{\mathcal{J}}(\tilde{\Delta}^{*,\infty}) := \limsup_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} \tilde{\rho}\Big({}^1\pi_t, \tilde{\Delta}^*({}^1\pi_t)\Big) \tag{2.45}$$

with ${}^1\pi_{t+1} = \tilde{Q}\Big(\tilde{\Delta}^*({}^1\pi_t)\, {}^1\pi_t\Big)$ *and*

$$\underline{\mathcal{J}}(\tilde{\Delta}^\infty) := \liminf_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} \tilde{\rho}\Big({}^1\pi_t, \tilde{\Delta}_t({}^1\pi_t)\Big) \tag{2.46}$$

with ${}^1\pi_{t+1} = \tilde{Q}\Big(\tilde{\Delta}_t({}^1\pi_t)\, {}^1\pi_t\Big)$*.* ϵ*-optimal decision rules* $(g_t^{1,*}, l_t^{1,*}, g_t^{2,*}, l_t^{2,*})$ *at time t are given by*

$$(g_t^{1,*}, l_t^{1,*}, g_t^{2,*}, l_t^{2,*}) =: \gamma_t^* = \tilde{\Delta}^*({}^1\pi_t). \tag{2.47}$$

Proof. This is a standard result, see Dynkin and Yushkevich (1975, Chapter 7). □

Conditions that guarantee that assumption (A1) of Theorem 2.4 is satisfied are fairly technical and do not provide much insight into the properties of the plant, the observation channels, and the cost functions that will guarantee the existence of such policies. The interested reader may look at Dynkin and Yushkevich (1975, Chapter 7, §10). It may be possible to extend the sufficiency conditions of Sennott (1999, 1997a, 1997b) to uncountable action spaces.

Significance of the results of variation v1

The sequential decomposition presented in this section provides an efficient way to search for an optimal design. For infinite horizon problem, it also provides an economical way to implement an optimal solution.

Consider the finite horizon problem. Suppose agent 1 and 2 are identical, i.e., $\mathcal{Y}^1 = \mathcal{Y}^2 = \mathcal{Y}$, $\mathcal{U}^1 = \mathcal{U}^2 = \mathcal{U}$, and $\mathcal{S}^1 = \mathcal{S}^2 = \mathcal{S}$. Then, searching for an optimal

design by brute force requires evaluation of $|\mathcal{U}|^{2\times T\times|\mathcal{Y}|\times|\mathcal{S}|}\times|\mathcal{S}|^{2\times T\times|\mathcal{Y}|\times|\mathcal{U}|\times|\mathcal{S}|}$ designs; this is exponential in the time horizon for which the system runs. As mentioned in the discussion of Section 2.2, the optimality equations are similar to POMDP; the information state is a belief over a $(|\mathcal{X}|\times|\mathcal{Y}|\times|\mathcal{U}|\times|\mathcal{S}|^2)$–dimensional vector (the actual size varies with time, so we chose a representative value), an action space of size $|\mathcal{U}|^{|\mathcal{Y}|\times|\mathcal{S}|}$ or $|\mathcal{S}|^{|\mathcal{Y}|\times|\mathcal{U}|\times|\mathcal{S}|}$ and observation space that has only one element (no observation). A technique similar to Smallwood and Sondik (1973), which represents the value functions by a family of linear functions that form their upper envelope, can be used to solve the finite horizon optimality equations exactly. In the worst case, the family of linear functions forming the upper envelopes is as large as all possible designs. However, we hope that for specific problems a lot of these functions can be pruned at each step, and, as is the case in POMDPs, the total number of envelopes may increase in a sub-exponential manner.

Consider the infinite horizon problem. As mentioned Section 2.2, we can view the decentralized two-agent control problem as an equivalent deterministic optimization problem by considering the information state as the "controlled state", the decision rules as the "control action" and the meta-function as the "control law" at each time. In classical infinite-horizon deterministic optimization problems, there is no loss of optimality in restricting attention to stationary design/strategy; by analogy, in the infinite-horizon decentralized two-agent problem, there is no loss of optimality in restricting attention to stationary meta-strategies. In classical infinite-horizon deterministic optimization problems, stationary actions are not optimal in general; by analogy, in infinite-horizon decentralized two-agent problem, stationary control and state-update strategies are not optimal in general. In the absence of a systematic framework, the task of finding and implementing an optimal infinite-horizon strategy is infeasible. The methodology of this section provides one systematic framework: ***obtain and implement time-varying optimal infinite-horizon control and state-update strategies by obtaining and implementing stationary infinite-horizon meta-strategies***. The off-line search simplifies to finding the fixed point of a functional equation. As is the case in POMDPs, we can find an approximate fixed point using randomized algorithms whose complexity is polynomial in the size of the alphabets (Rust, 1997). Once an optimal stationary meta-

strategy is obtained, both agents can store it, and use it to obtain the current optimal decision rules by keeping track of the current information state. This greatly simplifies the on-line implementation of a time-varying optimal design.

2.6 Time-homogeneous system—Variation v2

Consider a time-homogeneous variation of the model of Problem 2.1 where agent 1 has perfect recall and agent 2 has time-invariant and finite state. Since agent 1 has perfect recall the state-update function of agent 1 is fixed, and we only need to determine the control strategy of agent 1 and control state-update strategies for agent 2.

Data and information fields of agent 1

Agent 1 has perfect recall, hence its state is given by

$$S_t^1 = (Y^{1,t}, U^{1,t}).$$

Thus, the data at agent 1 can be written as (cf. (2.8))

$$^1O_t^1 := (Y^{1,t}, U^{1,t-1}), \tag{2.48a}$$

$$^2O_t^1 = {}^3O_t^1 = {}^4O_t^1 := (Y^{1,t}, U^{1,t}) \tag{2.48b}$$

Further, the information fields of agent 1 are given by

$$^1\mathfrak{I}_t^1 := \sigma(Y^{1,t}, U^{1,t-1}; {}^1\varphi^{t-1}), \qquad ^2\mathfrak{I}_t^1 := \sigma(Y^{1,t}, U^{1,t}; {}^2\varphi^{t-1}), \tag{2.49a}$$

$$^3\mathfrak{I}_t^1 := \sigma(Y^{1,t}, U^{1,t}; {}^3\varphi^{t-1}), \qquad ^4\mathfrak{I}_t^1 := \sigma(Y^{1,t}, U^{1,t}; {}^4\varphi^{t-1}), \tag{2.49b}$$

Agent 1 has perfect recall, so it does not shed information while going from time ^2t to ^3t. Thus, the time-evolution of the information fields of agent 1, which is given in general by (2.11a), can be written more precisely as

$$\cdots {}^4\mathfrak{I}_{t-1}^1 \subseteq {}^1\mathfrak{I}_t^1 = {}^2\mathfrak{I}_t^1 = {}^3\mathfrak{I}_t^1 = {}^4\mathfrak{I}_t^1 \subseteq {}^1\mathfrak{I}_{t+1}^1 \cdots \tag{2.50a}$$

Therefore, ***the information fields at agent 1 are a filtration***. Agent 2 does not have perfect recall; it sheds information while going from $^4(t-1)$ to ^1t. Therefore, the time-evolution of the information fields of agent 1, which is given in general by (2.11b), can be written more precisely as

$$\cdots {}^4\mathfrak{I}^2_{t-1} \supset {}^1\mathfrak{I}^2_t = {}^2\mathfrak{I}^2_t \subseteq {}^3\mathfrak{I}^2_t = {}^4\mathfrak{I}^2_t \supset {}^1\mathfrak{I}^2_{t+1} \cdots \tag{2.50b}$$

Agent 1's belief and their evolution

Agent 1 does not "know" what is "known" to agent 2. We can characterize what agent 1 "thinks" about nature and agent 2 by using agent 1's belief on the state of the plant and the data at agent 2 based on the information field of agent 1. Since the state of the plant and the data at agent 2 are time-invariant, agent 1's belief is time-invariant (Agent 2's data, and therefore agent 1's belief change between refinements of time at each stage, but not across stages). These beliefs are defined as follows:

Definition 2.6 (Agent 1's belief). *Let ${}^iB^1_t$ denote agent 1's belief on the state of the plant and the data at agent 2, i.e.,*

$${}^iB^1_t = \Pr\left(X_t, {}^iO^2_t \,\middle|\, {}^i\mathfrak{I}^1_t\right). \tag{2.51}$$

Let ${}^i\mathcal{B}^1 := \mathbb{P}\left\{\mathcal{X} \times {}^i\mathcal{O}^2\right\}$ denote the space of realizations of ${}^iB^1_t$.

Agent 1's belief can be written more elaborately as follows

$${}^1B^1_t := \Pr\left(X_t, S^2_{t-1} \,\middle|\, Y^{1,t}, U^{1,t-1}; {}^1\varphi^{t-1}\right) \tag{2.52a}$$

$${}^2B^1_t := \Pr\left(X_t, S^2_{t-1} \,\middle|\, Y^{1,t}, U^{1,t}; {}^2\varphi^{t-1}\right) \tag{2.52b}$$

$${}^3B^1_t := \Pr\left(X_t, Y^2_t, S^2_{t-1} \,\middle|\, Y^{1,t}, U^{1,t}; {}^3\varphi^{t-1}\right) \tag{2.52c}$$

$${}^4B^1_t := \Pr\left(X_t, Y^2_t, U^2_t, S^2_{t-1} \,\middle|\, Y^{1,t}, U^{1,t}; {}^4\varphi^{t-1}\right) \tag{2.52d}$$

The sequential ordering of these beliefs are shown in Figure 2.2. For any particular realization $(y^{1,t}, u^{1,t-1})$ of $(Y^{1,t}, U^{1,t-1})$ and any arbitrary (but fixed) choice of ${}^1\varphi^{t-1}$, the realization ${}^1b^1_t$ of ${}^1B^1_t$ is a PMF of $(\mathcal{X}\times\mathcal{S}^2)$. If $(Y^{1,t}, U^{1,t-1})$ is a random vector, then ${}^1B^1_t$ is a random vector belonging to $\mathbb{P}\left\{\mathcal{X} \times \mathcal{S}^2\right\}$. Similar interpretation hold for ${}^2B^1_t$, ${}^3B^1_t$, ${}^4B^1_t$.

Since the information fields of agent 1 are a filtration (see (2.50a)), the beliefs evolve in a state-like manner as follows:

Lemma 2.2 (Evolution of agent 1's beliefs). *For each stage t, there exists functions ${}^1F^1$, ${}^2F^1$, ${}^3F^1$ and ${}^4F^1$ such that*

$$^2B_t^1 = {}^1F^1({}^1B_t^1, U_t^1), \tag{2.53a}$$

$$^3B_t^1 = {}^2F^1({}^2B_t^1, U_t^1), \tag{2.53b}$$

$$^4B_t^1 = {}^3F^1({}^3B_t^1, g_t^2), \tag{2.53c}$$

$$^1B_{t+1}^1 = {}^4F^1({}^4B_t^1, l_t^2, Y_{t+1}^1, U_t^1). \tag{2.53d}$$

Proof. We will prove each part separately.

1. Consider $^1o_t^1 = (y^{1,t}, u^{1,t-1}) \in (\mathcal{Y}^{1,t} \times \mathcal{U}^{1,t-1})$, $u_t^1 \in \mathcal{U}^1$, $x_t \in \mathcal{X}$, $s_{t-1}^2 \in \mathcal{S}^2$, and $^2\varphi^{t-1} = ({}^1\varphi^{t-1}, g_t^1)$. Then, a component (x_t, s_{t-1}^2) of a realization $^2b_t^1$ of $^2B_t^1$ is given by

$$\begin{aligned} {}^2b_t^1(x_t, s_{t-1}^2) &= \Pr\left(X_t = x_t, S_{t-1}^2 = s_{t-1}^2 \,\middle|\, {}^1O_t^1 = {}^1o_t^1, U_t^1 = u_t^1; {}^2\varphi^{t-1}\right) \\ &= \frac{\Pr\left(X_t = x_t, S_{t-1}^2 = s_{t-1}^2, U_t^1 = u_t^1 \,\middle|\, {}^1O_t^1 = {}^1o_t^1; {}^2\varphi^{t-1}\right)}{\Pr\left(U_t^1 = u_t^1 \,\middle|\, {}^1O_t^1 = {}^1o_t^1; {}^2\varphi^{t-1}\right)} \end{aligned} \tag{2.54}$$

Now,

$$\begin{aligned} &\Pr\left(X_t = x_t, S_{t-1}^2 = s_{t-1}^2, U_t^1 = u_t^1 \,\middle|\, {}^1O_t^1 = {}^1o_t^1; {}^2\varphi^{t-1}\right) \\ &\quad = \Pr\left(U_t^1 = u_t^1 \,\middle|\, X_t = x_t, S_{t-1}^2 = s_{t-1}^2, {}^1O_t^1 = {}^1o_t^1; {}^1\varphi^{t-1}, g_t^1\right) \\ &\qquad \times \Pr\left(X_t = x_t, S_{t-1}^2 = s_{t-1}^2 \,\middle|\, {}^1O_t^1 = {}^1o_t^1; {}^1\varphi^{t-1}, g_t^1\right) \\ &\quad \overset{(a)}{=} \mathbb{I}\left[u_t^1 = g_t^1({}^1o_t^1)\right] \Pr\left(X_t = x_t, S_{t-1}^2 = s_{t-1}^2 \,\middle|\, {}^1O_t^1 = {}^1o_t^1; {}^1\varphi^{t-1}\right) \\ &\quad = \mathbb{I}\left[u_t^1 = g_t^1({}^1o_t^1)\right] {}^1b_t^1(x_t, s_{t-1}^2) \end{aligned} \tag{2.55}$$

where (a) follows from the sequential order in which the system variables are generated. Observe that $\Pr\left(U_t^1 = u_t^1 \,\middle|\, {}^1O_t^1 = {}^1o_t^1; {}^2\varphi^{t-1}\right)$ is the marginal of the LHS of (2.55). Combining (2.54) and (2.55) results in

$$^2b_t^1(x_t, s_{t-1}^2) =: {}^1F^1({}^1b_t^1, u_t^1)(x_t, s_{t-1}^2) \tag{2.56}$$

where $^1F^1$ is defined by (2.54) and (2.55).

2. Consider $(y^{1,t}, u^{1,t}) \in (\mathcal{Y}^{1,t} \times \mathcal{U}^{1,t})$, $x_t \in \mathcal{X}$, $y_t^2 \in \mathcal{Y}^2$, $s_{t-1}^2 \in \mathcal{S}^2$, and ${}^3\varphi^{t-1} = ({}^2\varphi^{t-1}, l_t^1)$. Then a component (x_t, y_t^1, s_{t-1}^2) of a realization ${}^1b_t^2$ of ${}^3B_t^1$ is given by

$$\begin{aligned}
{}^3b_t^1(x_t, y_t^2, s_{t-1}^2) &= \Pr\left(X_t = x_t, Y_t^2 = y_t^2, S_{t-1}^2 = s_{t-1}^2 \,\middle|\, Y^{1,t} = y^{1,t}, U^{1,t} = u^{1,t}; {}^3\varphi^{t-1}\right) \\
&= \Pr\left(Y_t^2 = y_t^2 \,\middle|\, X_t = x_t, S_{t-1}^2 = s_{t-1}^2, Y^{1,t} = y^{1,t}, U^{1,t} = u^{1,t}; {}^2\varphi^{t-1}, l_t^1\right) \\
&\quad \times \Pr\left(X_t = x_t, S_{t-1}^2 = s_{t-1}^2 \,\middle|\, Y^{1,t} = y^{1,t}, U^{1,t} = u^{1,t}; {}^2\varphi^{t-1}, l_t^1\right) \\
&\overset{(b)}{=} P_{N^2}\left(n_t^2 \in \mathcal{N}^2 : y_t^2 = h^2(x_t, u_t^1, n_t^2)\right) \\
&\quad \times \Pr\left(X_t = x_t, S_{t-1}^2 = s_{t-1}^2 \,\middle|\, Y^{1,t} = y^{1,t}, U^{1,t} = u^{1,t}; {}^2\varphi^{t-1}\right) \\
&= P_{N^2}\left(n_t^2 \in \mathcal{N}^2 : y_t^2 = h^2(x_t, u_t^1, n_t^2)\right) {}^2b_t^1(x_t, s_{t-1}^2) \\
&=: {}^3F^1({}^2b_t^1, u_t^1)(x_t, y_t^2, s_{t-1}^2)
\end{aligned} \tag{2.57}$$

where (b) follows from the sequential order in which the system variables are generated.

3. Consider $(y^{1,t}, u^{1,t}) \in (\mathcal{Y}^{1,t} \times \mathcal{U}^{1,t})$, $x_t \in \mathcal{X}$, $y_t^2 \in \mathcal{Y}^2$, $u_t^2 \in \mathcal{U}^2$, $s_{t-1}^2 \in \mathcal{S}^2$, and ${}^4\varphi^{t-1} = ({}^3\varphi^{t-1}, g_t^2)$. Then a component $(x_t, y_t^2, u_t^2, s_{t-1}^2)$ of a realization ${}^4b_t^1$ of ${}^4B_t^1$ is given by

$$\begin{aligned}
&{}^4b_t^1(x_t, y_t^2, u_t^2, s_{t-1}^2) \\
&\quad = \Pr\left(X_t = x_t, Y_t^2 = y_t^2, U_t^2 = u_t^2, S_{t-1}^2 = s_{t-1}^2 \,\middle|\, Y^{1,t} = y^{1,t}, U^{1,t} = u^{1,t}; {}^4\varphi^{t-1}\right) \\
&\quad = \Pr\left(U_t^2 = u_t^2 \,\middle|\, X_t = x_t, Y_t^2 = y_t^2, S_{t-1}^2 = s_{t-1}^2 Y^{1,t} = y^{1,t}, U^{1,t} = u^{1,t}; {}^3\varphi^{t-1}, g_t^2\right) \\
&\qquad \times \Pr\left(X_t = x_t, Y_t^2 = y_t^2, S_{t-1}^2 = s_{t-1}^2 \,\middle|\, Y^{1,t} = y^{1,t}, U^{1,t} = u^{1,t}; {}^3\varphi^{t-1}, g_t^2\right) \\
&\quad \overset{(c)}{=} \mathbb{I}\left[u_t^2 = g_t^2(y_t^2, s_{t-1}^2)\right] \\
&\qquad \times \Pr\left(X_t = x_t, Y_t^2 = y_t^2, S_{t-1}^2 = s_{t-1}^2 \,\middle|\, Y^{1,t} = y^{1,t}, U^{1,t} = u^{1,t}; {}^3\varphi^{t-1}\right) \\
&\quad = \mathbb{I}\left[u_t^2 = g_t^2(y_t^2, s_{t-1}^2)\right] {}^3b_t^1(x_t, y_t^2, s_{t-1}^2) \\
&\quad =: {}^3F^1({}^3b_t^1, g_t^2)(x_t, y_t^2, s_{t-1}^2)
\end{aligned} \tag{2.58}$$

where (c) follows from the sequential order in which the system variables are generated.

4. Consider ${}^4o_t^1 = (y^{1,t}, u^{1,t}) \in (\mathcal{Y}^{1,t} \times \mathcal{U}^{1,t})$, $y_{t+1}^1 \in \mathcal{Y}^1$, $x_{t+1} \in \mathcal{X}$, $s_t^2 \in \mathcal{S}^2$, and ${}^1\varphi^t = ({}^4\varphi^{t-1}, l_t^2)$. Then a component (x_{t+1}, s_t^2) of a realization ${}^1b_{t+1}^1$ of ${}^1B_{t+1}^1$ is given by

$$\begin{aligned} {}^1b_{t+1}^1(x_{t+1}, s_t^2) &= \Pr\left(X_{t+1} = x_{t+1}, S_t^2 = s_t^2 \,\middle|\, Y_{t+1}^1 = y_{t+1}^1, {}^4O_t^1 = {}^4o_t^1; {}^1\varphi^t\right) \\ &= \frac{\Pr\left(X_{t+1} = x_{t+1}, S_t^2 = s_t^2, Y_{t+1}^1 = y_{t+1}^1 \,\middle|\, {}^4O_t^1 = {}^4o_t^1; {}^1\varphi^t\right)}{\Pr\left(Y_{t+1}^1 = y_{t+1}^1 \,\middle|\, {}^4O_t^1 = {}^4o_t^1; {}^1\varphi^t\right)} \end{aligned} \tag{2.59}$$

Now,

$$\begin{aligned} &\Pr\left(X_{t+1} = x_{t+1}, S_t^2 = s_t^2, Y_{t+1}^1 = y_{t+1}^1 \,\middle|\, {}^4O_t^1 = {}^4o_t^1; {}^1\varphi^t\right) \\ &\quad = \Pr\left(Y_{t+1}^1 = y_{t+1}^1 \,\middle|\, X_{t+1} = x_{t+1}, S_t^2 = s_t^2, {}^4O_t^1 = {}^4o_t^1; {}^1\varphi^t\right) \\ &\qquad \times \Pr\left(X_{t+1} = x_{t+1}, S_t^2 = s_t^2 \,\middle|\, {}^4O_t^1 = {}^4o_t^1; {}^1\varphi^t\right) \\ &\quad = P_{N^1}\left(n_t^1 \in \mathcal{N}^1 : y_{t+1}^1 = h_t^1(x_{t+1}, n_{t+1}^1)\right) \\ &\qquad \times \Pr\left(X_{t+1} = x_{t+1}, S_t^2 = s_t^2 \,\middle|\, {}^4O_t^1 = {}^4o_t^1; {}^1\varphi^t\right) \end{aligned} \tag{2.60}$$

Further,

$$\begin{aligned} &\Pr\left(X_{t+1} = x_{t+1}, S_t^2 = s_t^2 \,\middle|\, {}^4O_t^1 = {}^4o_t^1; {}^1\varphi^t\right) \\ &\quad = \sum_{\substack{x_t \in \mathcal{X}, y_t^2 \in \mathcal{Y}^2 \\ u_t^2 \in \mathcal{U}^2, s_{t-1}^2 \in \mathcal{S}^2}} \Pr\Big(X_{t+1} = x_{t+1}, X_t = x_t, Y_t^2 = y_t^2, U_t^2 = u_t^2, \\ &\qquad\qquad S_{t-1}^2 = s_{t-1}^2, S_t^2 = s_t^2 \;\Big|\; {}^4O_t^1 = {}^4o_t^1; {}^1\varphi^t\Big) \\ &\quad = \sum_{\substack{x_t \in \mathcal{X}, y_t^2 \in \mathcal{Y}^2 \\ u_t^2 \in \mathcal{U}^2, s_{t-1}^2 \in \mathcal{S}^2}} \Pr\Big(X_{t+1} = x_{t+1} \;\Big|\; X_t = x_t, Y_t^2 = y_t^2, U_t^2 = u_t^2, \\ &\qquad\qquad S_{t-1}^2 = s_{t-1}^2, S_t^2 = s_t^2, {}^4O_t^1 = {}^4o_t^1; {}^1\varphi^t\Big) \\ &\qquad \times \Pr\Big(S_t^2 = s_t^2 \;\Big|\; X_t = x_t, Y_t^2 = y_t^2, U_t^2 = u_t^2, \\ &\qquad\qquad S_{t-1}^2 = s_{t-1}^2, {}^4O_t^1 = {}^4o_t^1; {}^4\varphi^{t-1}, l_t^2\Big) \\ &\qquad \times \Pr\left(X_t = x_t, Y_t^2 = y_t^2, U_t^2 = u_t^2, S_{t-1}^2 = s_{t-1}^2 \,\middle|\, {}^4O_t^1 = {}^4o_t^1; {}^4\varphi^{t-1}, l_t^2\right) \\ &\quad \overset{(d)}{=} \sum_{\substack{x_t \in \mathcal{X}, y_t^2 \in \mathcal{Y}^2 \\ u_t^2 \in \mathcal{U}^2, s_{t-1}^2 \in \mathcal{S}^2}} P_W\left(w_t \in \mathcal{W} : x_{t+1} = f(x_t, u_t^1, u_t^2, w_t)\right) \mathbb{I}\left[s_t^2 = l_t^2(y_t^2, u_t^2, s_{t-1}^1)\right] \\ &\qquad \times \Pr\left(X_t = x_t, Y_t^2 = y_t^2, U_t^2 = u_t^2, S_{t-1}^2 = s_{t-1}^2 \,\middle|\, {}^4O_t^1 = {}^4o_t^1; {}^4\varphi^{t-1}\right) \end{aligned}$$

$$= \sum_{\substack{x_t \in \mathcal{X}, y_t^2 \in \mathcal{Y}^2 \\ u_t^2 \in \mathcal{U}^2, s_{t-1}^2 \in \mathcal{S}^2}} P_W\Big(w_t \in \mathcal{W} : x_{t+1} = f(x_t, u_t^1, u_t^2, w_t)\Big) \mathbb{I}\left[s_t^2 = l_t^2(y_t^2, u_t^2, s_{t-1}^1)\right] \times {}^4b_t^1(x_t, y_t^2, u_t^2, s_{t-1}^2). \tag{2.61}$$

where (d) follows from the sequential order in which the system variables are generated. Combining (2.59)–(2.61) we get

$${}^1b_{t+1}^1(x_{t+1}, s_t^2) =: {}^4F^1({}^4b_t^1, l_t^2, y_{t+1}^1, u_t^1)(x_{t+1}, s_t^2) \tag{2.62}$$

where ${}^4F^1$ is given by (2.59)–(2.61). □

Structural properties

In this section, we provide structural/qualitative properties of optimal control laws of agent 1 that are true for every arbitrary but fixed control and state-update strategies of agent 2. These properties are subsequently used to convert the model of variation v2 into a model similar to that of variation v1.

Theorem 2.5 (Structure of optimal control laws of agent 1). *Consider variation v2 of the model of Problem 2.1. For any arbitrary but fixed control and state-update strategies of agent 2, there is no loss of optimality in restricting attention to control laws at agent 1 of the form*

$$U_t^1 = \hat{g}_t^1({}^1B_t^1), \quad t = 2, \dots, T. \tag{2.63}$$

Proof. We will look at the system from agent 1's point of view. The plant and agent 2 are fixed, and agent 1 has perfect recall. So, for fixed control and state-update strategies at agent 2, the design of agent 1 is a centralized optimization problem. The structural results follow from the standard result for POMDPs (partially observable Markov decision processes). In order to prove this explicitly, we need to show that the process $\{{}^1B_t^1, t = 1, \dots, T\}$ is controlled Markov process with control action U_t^1, and the expected instantaneous cost at time t can be written as a function of ${}^1B_t^1$ and U_t^1.

For any ${}^1b_{t+1}^1 \in {}^1\mathcal{B}^1$, any realization $({}^1b^{1,t}, u^{1,t})$ of $({}^1B^{1,t}, U^{1,t})$, and any choice of ${}^1\varphi^t$, we have

$$
\begin{aligned}
&\Pr\left({}^1B^1_{t+1} = {}^1b^1_{t+1} \,\middle|\, {}^1B^{1,t} = {}^1b^{1,t}, U^{1,t} = u^{1,t}; {}^1\varphi^t\right) \\
&\quad = \sum_{\substack{x^1_{t+1}\in\mathcal{X}^1_{t+1} \\ y^1_{t+1}\in\mathcal{Y}^1_{t+1}}} \Pr\left({}^1B^1_{t+1} = {}^1b^1_{t+1}, X_{t+1} = x_{t+1}, Y^1_{t+1} = y^1_{t+1} \,\middle|\, {}^1B^{1,t} = {}^1b^{1,t}, U^{1,t} = u^{1,t}; {}^1\varphi^t\right) \\
&\quad = \sum_{\substack{x^1_{t+1}\in\mathcal{X}^1_{t+1} \\ y^1_{t+1}\in\mathcal{Y}^1_{t+1}}} \Pr\left(Y^1_{t+1} = y^1_{t+1} \,\middle|\, {}^1B^{1,t+1} = {}^1b^{1,t+1}, X_{t+1} = x_{t+1}, U^{1,t} = u^{1,t}; {}^1\varphi^t\right) \\
&\qquad\qquad \times \Pr\left(X_{t+1} = x_{t+1} \,\middle|\, {}^1B^{1,t+1} = {}^1b^{1,t+1}, U^{1,t} = u^{1,t}; {}^1\varphi^t\right) \\
&\qquad\qquad \times \Pr\left({}^1B^1_{t+1} = {}^1b^1_{t+1} \,\middle|\, {}^1B^{1,t} = {}^1b^{1,t}, Y^1_{t+1} = y^1_{t+1}, U^{1,t} = u^{1,t}; {}^1\varphi^t\right) \\
&\quad \overset{(a)}{=} \sum_{\substack{x^1_{t+1}\in\mathcal{X}^1_{t+1} \\ y^1_{t+1}\in\mathcal{Y}^1_{t+1}}} P_{N^1}\left(n^1_{t+1} \in \mathcal{N}^1 : y^1_{t+1} = h^1(x_{t+1}, n^1_{t+1})\right) {}^1b^1_{t+1}(x_{t+1}) \\
&\qquad\qquad \times \mathbb{I}\left[{}^1b^1_{t+1} = {}^4F^1\Big({}^3F^1({}^2F^1({}^1F^1({}^1b^1_t, u^1_t), u^1_t), g^2_t), l^2_t, y^1_{t+1}, u^1_t\Big)\right] \\
&\quad = \Pr\left({}^1B^1_{t+1} = {}^1b^1_{t+1} \,\middle|\, {}^1B^1_t = {}^1b^1_t, U^1_t = u^1_t; g^2_t, l^2_t\right), \qquad (2.64)
\end{aligned}
$$

where ${}^1b^1_{t+1}(x_{t+1})$ denotes the marginal of ${}^1b^1_{t+1}(x_{t+1}, s^1_t)$ and (a) follows from Lemma 2.2. Further, the expected conditional instantaneous cost can be written as

$$
\begin{aligned}
&\mathbb{E}\left\{\rho(X_t, U^1_t, U^2_t) \,\middle|\, {}^4B^{1,t} = {}^4b^{1,t}, U^{1,t} = u^{1,t}; {}^4\varphi^{t-1}\right\} \\
&\quad = \sum_{\substack{x_t\in\mathcal{X} \\ u^2_t\in\mathcal{U}^2}} \rho(x_t, u^1_t, u^2_t) \Pr\left(X_t = x_t, U^2_t = u^2_t \,\middle|\, {}^4B^{1,t} = {}^4b^{1,t}, U^{1,t} = u^{1,t}; {}^4\varphi^{t-1}\right) \\
&\quad \overset{(b)}{=} \sum_{\substack{x_t\in\mathcal{X} \\ u^2_t\in\mathcal{U}^2}} \rho(x_t, u^1_t, u^2_t)\, {}^4b^1_t(x_t, u^2_t) \\
&\quad = \sum_{\substack{x_t\in\mathcal{X} \\ u^2_t\in\mathcal{U}^2}} \rho(x_t, u^1_t, u^2_t)\Big({}^3F^1\Big({}^2F^1\Big({}^1F^1({}^1b^1_t, u^1_t), u^1_t\Big), g^2_t\Big)\Big)(x_t, u^2_t) \\
&\quad =: \bar{\rho}({}^1b^1_t, u^1_t, g^2_t),
\end{aligned}
$$

where ${}^4b^1_t(x_t, u^2_t)$ is the marginal of ${}^4b^1_t(x_t, y^2_t, u^2_t, s^2_{t-1})$ in (b). Therefore, the total expected distortion can be written as

$$
\begin{aligned}
&\mathbb{E}\left\{\sum_{t=1}^{T}\rho(X_t,U_t^1,U_t^2)\,\middle|\,G^1,L^1,G^2,L^2\right\}\\
&\quad=\mathbb{E}\left\{\sum_{t=1}^{T}\mathbb{E}\left\{\rho(X_t,U_t^1,U_t^2)\,\middle|\,{}^4B^{1,t},U^{1,t};{}^4\varphi^{t-1}\right\}\,\middle|\,G^1,L^1,G^2,L^2\right\}\\
&\quad=\mathbb{E}\left\{\sum_{t=1}^{T}\bar{\rho}({}^1B_t^1,U_t^1,g_t^2)\,\middle|\,G^1,L^1,G^2,L^2\right\} \qquad (2.65)
\end{aligned}
$$

Thus, for a fixed $G^2 \coloneqq (g_1^2,\dots,g_T^2)$ and $L^2 \coloneqq (l_1^2,\dots,l_T^2)$, $\{{}^1B_t^1,\ t=1,\dots,T\}$ is a controlled Markov process with control action U_t^1, and the objective is to minimize a total expected cost where the instantaneous cost is a function of ${}^1B_t^1$ and U_t^1. From Markov decision theory Kumar and Varaiya (1986, Chapter 6) we know that there is no loss of optimality in restricting attention to control laws of the form (2.63).

□

Information states

Theorem 2.5 implies that we only need to consider control laws of the form (2.63) at agent 1. Control laws of the form (2.63) have the same structure as control laws of the form (2.3). Furthermore, agent 1's beliefs take values in a time-invariant space. So, if we think of agent 1's belief as its state, variation v2 is almost the same as variation v1; the only difference lies in the time instances at which agent 1's state is updated. In variation v1 the agent 1's state is updated only at ${}^2t$, while in variation v2 the agent 1's belief is updated at each time ${}^1t$, ${}^2t$, ${}^3t$ and ${}^4t$. This difference does not affect the solution methodology. We can still define information states in the same way as in Definition 2.3 and use them to obtain a sequential decomposition.

Definition 2.7. *Define ${}^1\pi_t$, ${}^2\pi_t$, ${}^3\pi_t$ and ${}^4\pi_t$ as follows:*

$$
{}^1\pi_t \coloneqq \Pr\left(X_t, Y_t^1, {}^1B_t^1, S_{t-1}^2 \,\middle|\, {}^1\varphi^{t-1}\right) \qquad (2.66a)
$$

$$
{}^2\pi_t \coloneqq \Pr\left(X_t, U_t^1, {}^2B_t^1, S_{t-1}^2 \,\middle|\, {}^2\varphi^{t-1}\right) \qquad (2.66b)
$$

$$
{}^3\pi_t \coloneqq \Pr\left(X_t, Y_t^2, U_t^1, {}^3B_t^1, S_{t-1}^2 \,\middle|\, {}^3\varphi^{t-1}\right) \qquad (2.66c)
$$

$$
{}^4\pi_t \coloneqq \Pr\left(X_t, Y_t^2, U_t^1, U_t^2, {}^4B_t^1, S_{t-1}^2 \,\middle|\, {}^4\varphi^{t-1}\right) \qquad (2.66d)
$$

These information states can be interpreted in a similar manner to information states of Definition 2.3. Observe that in Definition 2.3 ${}^2\pi_t$ includes a measure on

Y_t^1, while in Definition 2.7 it does not. In Definition 2.3 a measure on Y_t^1 is needed to generate a measure on S_t^1 at time ${}^3t$; in Definition 2.7 the information from Y_t^1 is absorbed in ${}^2b_t^1$ and the measure on Y_t^1 is not needed at ${}^3t$.

The time-evolution of the above defined information states is similar to that of Lemma 2.1; the time-evolution is time invariant and the evolution at ${}^2t$ is simpler as l_t^1 is fixed.

Lemma 2.3. *${}^1\pi_t$, ${}^2\pi_t$, ${}^3\pi_t$ and ${}^4\pi_t$ are information states for the control law and state-update rules at agents 1 and 2, respectively. Specifically,*

1. *There exist linear transformations ${}^1Q$, ${}^2Q$, ${}^3Q$ and ${}^4Q$ such that*

$$ {}^2\pi_t = {}^1Q(\hat{g}_t^1)\, {}^1\pi_t, \tag{2.67a} $$

$$ {}^3\pi_t = {}^2Q\, {}^2\pi_t, \tag{2.67b} $$

$$ {}^4\pi_t = {}^3Q(g_t^2)\, {}^3\pi_t, \tag{2.67c} $$

$$ {}^1\pi_{t+1} = {}^4Q(l_t^2)\, {}^4\pi_t. \tag{2.67d} $$

2. *The expected instantaneous cost can be expressed as*

$$ \mathbb{E}\left\{\rho(X_t, U_t^1, U_t^2) \,\middle|\, {}^4\varphi^{t-1}\right\} = \hat{\rho}({}^4\pi_t). \tag{2.68} $$

Proof. We will prove each part separately. The proof follows the same outline as the proof of Lemma 2.1.

1. Consider any $x_t \in \mathcal{X}$, $u_t^1 \in \mathcal{U}^1$, ${}^2b_t^1 \in {}^2\mathcal{B}^1$, $s_{t-1}^2 \in \mathcal{S}^2$, and ${}^2\varphi^{t-1} = ({}^1\varphi^{t-1}, \hat{g}_t^1)$ where $\hat{g}_t^1$ is of the form (2.63). A component of ${}^2\pi_t$ is given by

$$
\begin{aligned}
&{}^2\pi_t(x_t, u_t^1, {}^2b_t^1, s_{t-1}^2) \\
&\quad = \Pr\left(X_t = x_t, U_t^1 = u_t^1, {}^2B_t^1 = {}^2b_t^1, S_{t-1}^2 = s_{t-1}^2 \,\middle|\, {}^2\varphi^{t-1}\right) \\
&\quad = \int_{{}^1\mathcal{B}^1} \sum_{y_t \in \mathcal{Y}} \Pr\Big(X_t = x_t, Y_t^1 = y_t^1, U_t^1 = u_t^1, {}^2B_t^1 = {}^2b_t^1, \\
&\qquad\qquad {}^1B_t^1 = {}^1b_t^1, S_{t-1}^2 = s_{t-1}^2 \;\Big|\; {}^2\varphi^{t-1}\Big)\, \mathrm{d}\, {}^1b_t^1 \\
&\quad = \int_{{}^1\mathcal{B}^1} \sum_{y_t \in \mathcal{Y}} \Pr\left({}^2B_t^1 = {}^2b_t^1 \,\middle|\, X_t = x_t, Y_t^1 = y_t^1, U_t^1 = u_t^1, {}^1B_t^1 = {}^1b_t^1, S_{t-1}^2 = s_{t-1}^2; {}^2\varphi^{t-1}\right) \\
&\qquad\quad \times \Pr\left(U_t^1 = u_t^1 \,\middle|\, X_t = x_t, Y_t^1 = y_t^1, {}^1B_t^1 = {}^1b_t^1, S_{t-1}^2 = s_{t-1}^2; {}^1\varphi^{t-1}, \hat{g}_t^1\right) \\
&\qquad\quad \times \Pr\left(X_t = x_t, Y_t^1 = y_t^1, {}^1B_t^1 = {}^1b_t^1, S_{t-1}^2 = s_{t-1}^2 \,\middle|\, {}^1\varphi^{t-1}, \hat{g}_t^1\right)\, \mathrm{d}\, {}^1b_t^1
\end{aligned}
$$

$$
\begin{aligned}
&\overset{(a)}{=} \int_{{}^1\mathcal{B}^1} \sum_{y_t \in \mathcal{Y}} \mathbb{I}\left[{}^2b_t^1 = {}^2F^1({}^1b_t^1, u_t^1)\right] \mathbb{I}\left[u_t^1 = \hat{g}_t^1({}^1b_t^1)\right] \\
&\qquad \times \Pr\left(X_t = x_t, Y_t^1 = y_t^1, {}^1B_t^1 = {}^1b_t^1, S_{t-1}^2 = s_{t-1}^2 \,\middle|\, {}^1\varphi^{t-1}\right) \mathrm{d}\, {}^1b_t^1 \\
&= \int_{{}^1\mathcal{B}^1} \sum_{y_t \in \mathcal{Y}} \mathbb{I}\left[{}^2b_t^1 = {}^2F^1({}^1b_t^1, u_t^1)\right] \mathbb{I}\left[u_t^1 = \hat{g}_t^1({}^1b_t^1)\right] {}^1\pi_t(x_t, y_t^1, {}^1b_t^1, s_{t-1}^2) \,\mathrm{d}\, {}^1b_t^1 \\
&=: \left({}^1Q(\hat{g}_t^1)\, {}^1\pi_t\right)(x_t, u_t^1, {}^1b_t^1, s_{t-1}^2)
\end{aligned} \tag{2.69}
$$

where (a) follows from Lemma 2.2 and the sequential order in which the system variables are generated.

2. Consider any $x_t \in \mathcal{X}$, $y_t^2 \in \mathcal{Y}^2$, $u_t^1 \in \mathcal{U}^1$, ${}^3b_t^1 \in {}^3\mathcal{B}^1$, $s_{t-1}^2 \in \mathcal{S}^2$, and ${}^3\varphi^{t-1} = ({}^2\varphi^{t-1}, l_t^1)$. A component of ${}^3\pi_t$ is given by

$$
\begin{aligned}
&{}^3\pi_t(x_t, y_t^2, u_t^1, {}^3b_t^1, s_{t-1}^2) \\
&\quad = \Pr\left(X_t = x_t, Y_t^2 = y_t^2, U_t^1 = u_t^1, {}^3B_t^1 = {}^3b_t^1, S_{t-1}^2 = s_{t-1}^2 \,\middle|\, {}^3\varphi^{t-1}\right) \\
&\quad = \int_{{}^2\mathcal{B}^1} \Pr\Big(X_t = x_t, Y_t^2 = y_t^2, U_t^1 = u_t^1, \\
&\qquad\qquad {}^2B_t^1 = {}^2b_t^1, {}^3B_t^1 = {}^3b_t^1, S_{t-1}^2 = s_{t-1}^2 | \, {}^3\varphi^{t-1}\Big) \,\mathrm{d}\, {}^2b_t^1 \\
&\quad = \int_{{}^2\mathcal{B}^1} \Pr\Big(Y_t^2 = y_t^2 \,\Big|\, X_t = x_t, U_t^1 = u_t^1, \\
&\qquad\qquad {}^2B_t^1 = {}^2b_t^1, {}^3B_t^1 = {}^3b_t^1, S_{t-1}^2 = s_{t-1}^2; {}^3\varphi^{t-1}\Big) \,\mathrm{d}\, {}^2b_t^1 \\
&\qquad \times \Pr\left({}^3B_t^1 = {}^3b_t^1 \,\middle|\, X_t = x_t, U_t^1 = u_t^1, {}^2B_t^1 = {}^2b_t^1, S_{t-1}^2 = s_{t-1}^2; {}^2\varphi^{t-1}, l_t^1\right) \\
&\qquad \times \Pr\left(X_t = x_t, U_t^1 = u_t^1, {}^2B_t^1 = {}^2b_t^1, S_{t-1}^2 = s_{t-1}^2 \,\middle|\, {}^2\varphi^{t-1}, l_t^1\right) \mathrm{d}\, {}^2b_t^1 \\
&\quad \overset{(b)}{=} \int_{{}^2\mathcal{B}^1} P_{N^2}\left(n_t^2 \in \mathcal{N}^2 : y_t^2 = h^2(x_t, u_t^1, n_t^2)\right) \mathbb{I}\left[{}^3b_t^1 = {}^2F^1({}^2b_t^1, u_t^1)\right] \\
&\qquad\qquad \times \Pr\left(X_t = x_t, U_t^1 = u_t^1, {}^2B_t^1 = {}^2b_t^1, S_{t-1}^2 = s_{t-1}^2 \,\middle|\, {}^2\varphi^{t-1}\right) \mathrm{d}\, {}^2b_t^1 \\
&\quad = \int_{{}^2\mathcal{B}^1} P_{N^2}\left(n_t^2 \in \mathcal{N}^2 : y_t^2 = h^2(x_t, u_t^1, n_t^2)\right) \mathbb{I}\left[{}^3b_t^1 = {}^2F^1({}^2b_t^1, u_t^1)\right] \\
&\qquad\qquad \times {}^2\pi_t(x_t, u_t^1, {}^2b_t^1, s_{t-1}^2) \,\mathrm{d}\, {}^2b_t^1 \\
&\quad =: ({}^2Q\, {}^2\pi_t)(x_t, y_t^2, u_t^1, {}^3b_t^1, s_{t-1}^2)
\end{aligned} \tag{2.70}
$$

where (b) follows from Lemma 2.2 and the sequential order in which the system variables are generated.

3. Consider any $x_t \in \mathcal{X}_t$, $y_t^2 \in \mathcal{Y}^2$, $u_t^1 \in \mathcal{U}^1$, $u_t^2 \in \mathcal{U}^2$, ${}^4b_t^1 \in {}^4\mathcal{B}^1$, $s_{t-1}^2 \in \mathcal{S}^2$, and ${}^4\varphi^{t-1} = ({}^3\varphi^{t-1}, g_t^2)$. A component of ${}^4\pi_t$ is given by

$$
\begin{aligned}
&{}^4\pi_t(x_t, y_t^2, u_t^1, u_t^2, {}^4b_t^1, s_{t-1}^2) \\
&\quad = \Pr\left(X_t = x_t, Y_t^2 = y_t^2, U_t^1 = u_t^1, U_t^2 = u_t^2, {}^4B_t^1 = {}^4b_t^1, S_{t-1}^2 = s_{t-1}^2 \,\middle|\, {}^4\varphi^{t-1}\right) \\
&\quad = \int_{{}^3\mathcal{B}^1} \Pr\Big(X_t = x_t, Y_t^2 = y_t^2, U_t^1 = u_t^1, U_t^2 = u_t^2, \\
&\qquad\qquad {}^3B_t^1 = {}^3b_t^1, {}^4B_t^1 = {}^4b_t^1, S_{t-1}^2 = s_{t-1}^2 \,\Big|\, {}^4\varphi^{t-1}\Big) \, \mathrm{d}\, {}^3b_t^1 \\
&\quad = \int_{{}^3\mathcal{B}^1} \Pr\Big(U_t^2 = u_t^2 \,\Big|\, X_t = x_t, Y_t^2 = y_t^2, U_t^1 = u_t^1, \\
&\qquad\qquad {}^3B_t^1 = {}^3b_t^1, {}^4B_t^1 = {}^4b_t^1, S_{t-1}^2 = s_{t-1}^2; {}^4\varphi^{t-1}\Big) \, \mathrm{d}\, {}^3b_t^1 \\
&\qquad \times \Pr\left({}^4B_t^1 = {}^4b_t^1 \,\middle|\, X_t = x_t, Y_t^2 = y_t^2, U_t^1 = u_t^1, {}^3B_t^1 = {}^3b_t^1, S_{t-1}^2 = s_{t-1}^2; {}^4\varphi^{t-1}\right) \\
&\qquad \times \Pr\left(X_t = x_t, Y_t^2 = y_t^2, U_t^1 = u_t^1, {}^3B_t^1 = {}^3b_t^1, S_{t-1}^2 = s_{t-1}^2 \,\middle|\, {}^3\varphi^{t-1}, g_t^2\right) \mathrm{d}\, {}^3b_t^1 \\
&\quad \overset{(c)}{=} \int_{{}^3\mathcal{B}^1} \mathbb{I}\left[u_t^2 = g_t^2(y_t^2, s_{t-1}^2)\right] \mathbb{I}\left[{}^4b_t^1 = {}^3F^1({}^3b_t^1, g_t^2)\right] \\
&\qquad\qquad \times \Pr\left(X_t = x_t, Y_t^2 = y_t^2, U_t^1 = u_t^1, {}^3B_t^1 = {}^3b_t^1, S_{t-1}^2 = s_{t-1}^2 \,\middle|\, {}^3\varphi^{t-1}\right) \mathrm{d}\, {}^3b_t^1 \\
&\quad = \int_{{}^3\mathcal{B}^1} \mathbb{I}\left[u_t^2 = g_t^2(y_t^2, s_{t-1}^2)\right] \mathbb{I}\left[{}^4b_t^1 = {}^3F^1({}^3b_t^1, g_t^2)\right] {}^3\pi_t(x_t, y_t^2, u_t^1, {}^3b_t^1, s_{t-1}^2) \\
&\quad =: \left({}^3Q(g_t^2)\, {}^3\pi_t\right)(x_t, y_t^2, u_t^1, u_t^2, {}^4b_t^1, s_{t-1}^2)
\end{aligned} \tag{2.71}
$$

where (c) follows from Lemma 2.2 and the sequential order in which the system variables are generated.

4. Consider any $x_{t+1} \in \mathcal{X}$, $y_{t+1}^1 \in \mathcal{Y}^1$, ${}^1b_{t+1}^1 \in {}^1\mathcal{B}^1$, $s_t^2 \in \mathcal{S}^2$, and ${}^1\varphi^1 = ({}^4\varphi^{t-1}, l_t^2)$. Consider a component of ${}^1\pi_{t+1}$,

$$
\begin{aligned}
&{}^1\pi_{t+1}(x_{t+1}, y_{t+1}^1, {}^1b_{t+1}^1, s_t^2) \\
&\quad = \Pr\left(X_{t+1} = x_{t+1}, Y_{t+1}^1 = y_{t+1}^1, {}^1B_{t+1}^1 = {}^1b_{t+1}^1, S_t^2 = s_t^2 \,\middle|\, {}^1\varphi^{t-1}\right) \\
&\quad = \int_{{}^4\mathcal{B}^1} \sum_{\substack{x_t \in \mathcal{X}, y_t^2 \in \mathcal{Y}^2, \\ u_t^1 \in \mathcal{U}^1, u_t^2 \in \mathcal{U}^2 \\ s_{t-1}^2 \in \mathcal{S}_{t-1}^2}} \Pr\Big({}^1B_{t+1}^1 = {}^1b_{t+1}^1 \;\Big|\; X_{t+1} = x_{t+1}, X_t = x_t, \\
&\qquad\qquad Y_{t+1}^1 = y_{t+1}^1, Y_t^2 = y_t^2, U_t^1 = u_t^1, U_t^2 = u_t^2, \\
&\qquad\qquad {}^4B_t^1 = {}^4b_t^1, S_{t-1}^2 = s_{t-1}^2, S_t^2 = s_t^2; {}^4\varphi^{t-1}, l_t^2\Big) \\
&\qquad \times \Pr\Big(Y_{t+1}^1 = y_{t+1}^1 \;\Big|\; X_{t+1} = x_{t+1}, X_t = x_t, Y_t^2 = y_t^2, U_t^1 = u_t^1, \\
&\qquad\qquad U_t^2 = u_t^2, {}^4B_t^1 = {}^4b_t^1, S_{t-1}^2 = s_{t-1}^2, S_t^2 = s^2; {}^4\varphi^{t-1}, l_t^2\Big)
\end{aligned}
$$

$$
\begin{aligned}
&\times \Pr\Big(X_{t+1} = x_{t+1} \;\Big|\; X_t = x_t, Y_t^2 = y_t^2, U_t^1 = u_t^1, \\
&\qquad\qquad U_t^2 = u_t^2, {}^4B_t^1 = {}^4b_t^1, S_{t-1}^2 = s_{t-1}^2, S_t^2 = s_t^2; {}^4\varphi^{t-1}, I_t^2\Big) \\
&\times \Pr\Big(X_t = x_t, Y_t^2 = y_t^2, U_t^1 = u_t^1, \\
&\qquad U_t^2 = u_t^2, {}^4B_t^1 = {}^4b_t^1, S_{t-1}^2 = s_{t-1}^2 \;\Big|\; {}^4\varphi^{t-1}\Big) \, \mathrm{d}\, {}^4b_t^1 \\
\overset{(d)}{=} \int_{{}^4\mathcal{B}^1} \sum_{\substack{x_t \in \mathcal{X}, y_t^2 \in \mathcal{Y}^2, \\ u_t^1 \in \mathcal{U}^1, u_t^2 \in \mathcal{U}^2 \\ s_{t-1}^2 \in \mathcal{S}_{t-1}^2}} & \mathbb{I}\Big[{}^1b_{t+1}^1 = {}^4F^1({}^4b_t^1, I_t^2, y_{t+1}^1, u_t^1)\Big] \\
& \times P_{N^1}\Big(n_t^1 \in \mathcal{N}^1 : y_{t+1}^1 = h^1(x_{t+1}, n_{t+1}^1)\Big) \\
& \times P_W\Big(w_t \in \mathcal{W} : x_{t+1} = f_t(x_t, u_t^1, u_t^2, w_t)\Big) \\
& \times \Pr\Big(X_t = x_t, Y_t^2 = y_t^2, U_t^1 = u_t^1, \\
& \qquad U_t^2 = u_t^2, {}^4B_t^1 = {}^4b_t^1, S_{t-1}^2 = s_{t-1}^2 \;\Big|\; {}^4\varphi^{t-1}\Big) \, \mathrm{d}\, {}^4b_t^1 \\
= \int_{{}^4\mathcal{B}^1} \sum_{\substack{x_t \in \mathcal{X}, y_t^2 \in \mathcal{Y}^2, \\ u_t^1 \in \mathcal{U}^1, u_t^2 \in \mathcal{U}^2 \\ s_{t-1}^2 \in \mathcal{S}_{t-1}^2}} & \mathbb{I}\Big[{}^1b_{t+1}^1 = {}^4F^1({}^4b_t^1, I_t^2, y_{t+1}^1, u_t^1)\Big] \\
& \times P_{N^1}\Big(n_t^1 \in \mathcal{N}^1 : y_{t+1}^1 = h^1(x_{t+1}, n_{t+1}^1)\Big) \\
& \times P_W\Big(w_t \in \mathcal{W} : x_{t+1} = f_t(x_t, u_t^1, u_t^2, w_t)\Big) \\
& \times {}^4\pi_t(x_t, y_t^2, u_t^1, u_t^2, {}^4b_t^1, s_{t-1}^2) \, \mathrm{d}\, {}^4b_t^1 \\
=: \Big({}^4Q(I_t^2)\, {}^4\pi_t\Big)(x_{t+1}, y_{t+1}^1, {}^1b_{t+1}^1, s_t^2) & \qquad (2.72)
\end{aligned}
$$

where (d) follows from Lemma 2.2 and the sequential order in which the system variables are generated.

5. The expected instantaneous cost can be expressed as

$$
\begin{aligned}
&\mathbb{E}\Big\{\rho(X_t, U_t^1, U_t^2) \,\Big|\, {}^4\varphi^{t-1}\Big\} \\
&\quad = \sum_{(x_t \in \mathcal{X}, u_t^1 \in \mathcal{U}^1, u_t^2 \in \mathcal{U}^2)} \rho(x_t, u_t^1, u_t^2) \Pr\Big(X_t = x_t, U_t^1 = u_t^1, U_t^2 = u_t^2 \,\Big|\, {}^4\varphi^{t-1}\Big) \\
&\quad = \sum_{(x_t \in \mathcal{X}, u_t^1 \in \mathcal{U}^1, u_t^2 \in \mathcal{U}^2)} \rho(x_t, u_t^1, u_t^2) \times \int_{{}^4\mathcal{B}^1} \sum_{(y_t^2 \in \mathcal{Y}^2, s_{t-1}^2 \in \mathcal{S}^2)} {}^4\pi_t(x_t, y_t^2, u_t^1, u_t^2, {}^4b_t^1, s_{t-1}^2) \\
&\quad =: \hat{\rho}_t({}^4\pi_t) \qquad (2.73)
\end{aligned}
$$

□

Global optimization

The results for global optimization for finite and infinite horizon problems for variation v1 only depend on: (i) the information states ${}^1\pi_t$, ${}^2\pi_t$, ${}^3\pi_t$ and ${}^4\pi_t$ belonging to time-invariant spaces ${}^1\Pi$, ${}^2\Pi$, ${}^3\Pi$ and ${}^4\Pi$; and (ii) satisfying Lemma 2.1 where the transformations ${}^1Q$, ${}^2Q$, ${}^3Q$ and ${}^4Q$ and the function $\hat{\rho}$ do not depend on t. As shown above, in the case of variation v2, the information states specified by Definition 2.7 belong to time-invariant spaces and satisfy Lemma 2.1 in a time-invariant manner. Thus, the results of variation v1 are also applicable to variation v2 with information states specified by Definition 2.7. Hence, for the finite horizon problem, we can simplify Theorem 2.1 as follows.

Corollary 2.2. *For the time-homogeneous variation v2 of the model of Problem 2.1, the nested optimality equations* (2.23) *can be written as*

$$ {}^1V_{T+1}({}^1\pi) = 0, \tag{2.74a} $$

and for $t = 1, \dots, T$

$$ {}^1V_t({}^1\pi) = \inf_{\hat{g}_t^1 \in \hat{\mathscr{G}}^1} {}^2V_t\Big({}^1Q(\hat{g}_t^1)\, {}^1\pi \Big), \tag{2.74b} $$

$$ {}^2V_t({}^2\pi) = {}^3V_t\Big({}^2Q\, {}^2\pi \Big), \tag{2.74c} $$

$$ {}^3V_t({}^3\pi) = \inf_{g_t^2 \in \mathscr{G}^2} {}^4V_t\Big({}^3Q(g_t^2)\, {}^3\pi \Big), \tag{2.74d} $$

$$ {}^4V_t({}^4\pi) = \hat{\rho}({}^4\pi) + \inf_{l_t^2 \in \mathscr{L}^2} {}^1V_{t+1}\Big({}^4Q(l_t^2)\, {}^4\pi \Big). \tag{2.74e} $$

For the infinite horizon problems, the fixed point equations remain the same as those of Theorems 2.3 and 2.4 with $\gamma_t := (\hat{g}_t^1, g_t^2, l_t^2)$, $\Gamma := \hat{\mathscr{G}}^1 \times \mathscr{G}^2 \times \mathscr{L}^2$,

$$ \tilde{Q}(\gamma_t) := {}^4Q(l_t^2) \circ {}^3Q(g_t^2) \circ {}^2Q \circ {}^1Q(g_t^1) $$

and, instead of (2.37) and (2.47), optimal decision rules $(g_t^{1,*}, g_t^{2,*}, l_t^{2,*})$ are given by

$$ (g_t^{1,*}, g_t^{2,*}, l_t^{2,*}) =: \gamma_t^* = \tilde{\Delta}^*({}^1\pi_t). \tag{2.75} $$

Significance of the results of variation v2

The sequential decomposition presented in this section suggests an approach that can potentially lead to efficient algorithms for the search of globally optimal designs

for finite and infinite horizon problems. The structural results of (2.63) provide a compact representation of optimal control laws of agent 1; instead of implementing a control law with a time-varying domain of the form

$$U_t^1 = g_t^1(Y_1^1, \dots, Y_t^1)$$

the controller only has to use a control law of the form (2.63), which has a time-invariant domain.

The sequential decomposition of (2.74) is equivalent to the sequential decomposition of a POMDP where the (unobserved) state and the action spaces are continuous. Numerical methods for POMDPs of this form is an active area of research. For finite horizon problems, we are not aware of any good computational techniques; For infinite horizon problems, some preliminary results (Thrun, 2000 and Porta et al., 2006) exist.

2.7 Time-homogeneous system—Variation v3

Consider a time-homogeneous variation of the model of Problem 2.1 where agent 1 has time-invariant and finite state and agent 2 has perfect recall. This is similar to variation v2, with the structure of the states of agents 1 and 2 reversed. This variation can be solved in exactly the same manner as variation v2. We look at the system from agent 2's point-of-view and show that the control law at agent 2 can just be a function of its belief. This means that we can define information states in the same manner as in variation v2 and show that these information states satisfy Lemma 2.1 in a time-invariant manner. Hence, the results for global optimization for the finite and infinite horizon problems of variation v1 can also be used for variation v3. Details are provided below.

Data and information fields of agent 2

Agent 2 has perfect recall, hence its state is given by

$$S_t^2 = (Y^{2,t}, U^{2,t}).$$

Thus, the data at agent 2 can be written as (cf. (2.8))

$$^1O_t^2 = {}^2O_t^2 \coloneqq (Y^{2,t-1}, U^{2,t-1}), \tag{2.76a}$$

$$^3O_t^2 \coloneqq (Y^{2,t}, U^{2,t-1}) \tag{2.76b}$$

$$^4O_t^2 \coloneqq (Y^{2,t}, U^{2,t}) \tag{2.76c}$$

Agent 1 does not have perfect recall; it sheds information while going from time ^2t to ^3t. Thus, the time-evolution of the information fields of agent 1, which is given in general by (2.11a), can be written more precisely as

$$\cdots {}^4\mathfrak{I}_{t-1}^1 \subseteq {}^1\mathfrak{I}_t^1 = {}^2\mathfrak{I}_t^1 \supset {}^3\mathfrak{I}_t^1 = {}^4\mathfrak{I}_t^1 \subseteq {}^1\mathfrak{I}_{t+1}^1 \cdots \tag{2.77a}$$

Agent 2 has perfect recall, so it does not shed information while going from $^4(t-1)$ to ^1t. Therefore, the time-evolution of the information fields of agent 1, which is given in general by (2.11b), can be written more precisely as

$$\cdots {}^4\mathfrak{I}_{t-1}^2 = {}^1\mathfrak{I}_t^2 = {}^2\mathfrak{I}_t^2 \subseteq {}^3\mathfrak{I}_t^2 = {}^4\mathfrak{I}_t^2 = {}^1\mathfrak{I}_{t+1}^2 \cdots \tag{2.77b}$$

Therefore, ***the information fields at agent 2 are a filtration***.

Agent 2's belief and their evolution

We define agent 2's beliefs as follows:

Definition 2.8 (Agent 2's belief). *Let $^iB_t^2$ denote agent 2's belief on the state of the plant and the data at agent 1, i.e.,*

$$^iB_t^2 = \Pr\left(X_t, {}^iO_t^1 \,\middle|\, {}^i\mathfrak{I}_t^2\right). \tag{2.78}$$

Let $^i\mathcal{B}^2 \coloneqq \mathbb{P}\left\{\mathcal{X} \times {}^i\mathcal{O}^1\right\}$ denote the space of realizations of $^iB_t^2$.

Agent 2's belief can be written more elaborately as follows

$$^1B_t^2 \coloneqq \Pr\left(X_t, Y_t^1, S_{t-1}^1 \,\middle|\, Y^{2,t-1}, U^{2,t-1}; {}^1\varphi^{t-1}\right) \tag{2.79a}$$

$$^2B_t^2 \coloneqq \Pr\left(X_t, Y_t^1, U_t^1, S_{t-1}^1 \,\middle|\, Y^{2,t-1}, U^{2,t-1}; {}^2\varphi^{t-1}\right) \tag{2.79b}$$

$$^3B_t^2 \coloneqq \Pr\left(X_t, U_t^1, S_t^1 \,\middle|\, Y^{2,t}, U^{2,t-1}; {}^3\varphi^{t-1}\right) \tag{2.79c}$$

$$^4B_t^2 \coloneqq \Pr\left(X_t, U_t^1, S_t^1 \,\middle|\, Y^{2,t}, U^{2,t}; {}^4\varphi^{t-1}\right) \tag{2.79d}$$

The sequential ordering of these beliefs are shown in Figure 2.2. These beliefs should be interpreted in the same manner as the beliefs of agent 1 in variation v2.

Since the information fields of agent 2 are a filtration (see (2.77b)), the beliefs evolve in a state-like manner as follows:

Lemma 2.4 (Evolution of agent 2's beliefs). *For each stage t, there exists functions ${}^1F^2$, ${}^2F^2$, ${}^3F^2$ and ${}^4F^2$ such that*

$$ {}^2B_t^2 = {}^1F^2({}^1B_t^2, g_t^1), \tag{2.80a} $$

$$ {}^3B_t^2 = {}^2F^2({}^2B_t^2, Y_t^2, l_t^1), \tag{2.80b} $$

$$ {}^4B_t^2 = {}^3F^2({}^3B_t^2, U_t^2), \tag{2.80c} $$

$$ {}^1B_{t+1}^2 = {}^4F^2({}^4B_t^2, U_t^2), \tag{2.80d} $$

The proof is similar to that of Lemma 2.2.

Structural properties

The structural/qualitative properties of optimal control laws of agent 2 for variation v3 are similar to the structural properties of optimal control laws for agent 1 in variation v2.

Theorem 2.6 (Structure of optimal control laws of agent 2). *Consider variation v3 of the model of Problem 2.1. For any arbitrary but fixed control and state-update strategies of agent 1, there is no loss of optimality in restricting attention to control laws of the form*

$$ U_t^2 = \hat{g}_t^2({}^3B_t^2), \quad t = 2, \dots, T \tag{2.81} $$

for agent 2.

Proof. We look at the system from agent 2's point of view. The plant and agent 1 are fixed, and agent 2 has perfect recall. So, for fixed control and state-update strategies at agent 1, determining the optimal design of agent 2 is a centralized optimization problem. The structural form of optimal designs for agent 2 follows from the standard result for POMDP (Kumar and Varaiya, 1986, Chapter 6). The details are along the same lines as the proof of Theorem 2.5. □

Information states

As in variation v2, we can use the structural result of Theorem 2.6 as a guide to define information states for variation v3.

Definition 2.9. *Define* ${}^1\pi_t$, ${}^2\pi_t$, ${}^3\pi_t$ *and* ${}^4\pi_t$ *as follows:*

$$ {}^1\pi_t := \Pr\left(X_t, Y_t^1, S_{t-1}^1, {}^1B_t^2 \,\middle|\, {}^1\varphi^{t-1}\right), \tag{2.82a} $$

$$ {}^2\pi_t := \Pr\left(X_t, Y_t^1, U_t^1, S_{t-1}^1, {}^2B_t^2 \,\middle|\, {}^2\varphi^{t-1}\right), \tag{2.82b} $$

$$ {}^3\pi_t := \Pr\left(X_t, Y_t^2, U_t^1, S_t^1, {}^3B_t^2 \,\middle|\, {}^3\varphi^{t-1}\right), \tag{2.82c} $$

$$ {}^4\pi_t := \Pr\left(X_t, Y_t^2, U_t^1, U_t^2, S_t^1, {}^4B_t^2 \,\middle|\, {}^4\varphi^{t-1}\right). \tag{2.82d} $$

These information states can be interpreted in the same way as the information states of Definition 2.3. Observe that in Definition 2.3 ${}^4\pi_t$ includes a measure on Y_t^2, while in Definition 2.9 it does not. In Definition 2.3 a measure on Y_t^2 is needed to generate a measure on S_t^2 at time ${}^4t$; in Definition 2.9 the information from Y_t^2 is absorbed in ${}^4b_t^1$ and the measure on Y_t^2 is not needed at ${}^4t$.

The time-evolution of the above defined information states is similar to that of Lemma 2.1; the time-evolution is time invariant, and the evolution at ${}^4t$ is simpler as l_t^2 is fixed.

Lemma 2.5. ${}^1\pi_t$, ${}^2\pi_t$, ${}^3\pi_t$ *and* ${}^4\pi_t$ *are information states for the control law and state-update rules at agents 1 and 2, respectively. Specifically,*

1. *There exist linear transformations* ${}^1Q$, ${}^2Q$, ${}^3Q$ *and* ${}^4Q$ *such that*

$$ {}^2\pi_t = {}^1Q(g_t^1)\,{}^1\pi_t, \tag{2.83a} $$

$$ {}^3\pi_t = {}^2Q(l_t^1)\,{}^2\pi_t, \tag{2.83b} $$

$$ {}^4\pi_t = {}^3Q(g_t^2)\,{}^3\pi_t, \tag{2.83c} $$

$$ {}^1\pi_{t+1} = {}^4Q\,{}^4\pi_t. \tag{2.83d} $$

2. *The expected instantaneous cost can be expressed as*

$$ \mathbb{E}\left\{\rho(X_t, U_t^1, U_t^2) \,\middle|\, {}^4\varphi^{t-1}\right\} = \hat{\rho}({}^4\pi_t). \tag{2.84} $$

The proof proceeds along the same lines as the proof in case of variation v2.

Global optimization

Following the arguments for variation v2, we can use the results for global optimization for the finite and infinite horizon problems for variation v1 to variation v3 with the information states specified by Definition 2.9. Hence, for the finite horizon problem, we can simplify Theorem 2.1 as follows.

Corollary 2.3. *For the time-homogeneous variation v3 of the model of Problem 2.1, the nested optimality equations* (2.23) *can be written as*

$$ {}^1V_{T+1}({}^1\pi) = 0, \tag{2.85a} $$

and for $t = 1, \dots, T$

$$ {}^1V_t({}^1\pi) = \inf_{g_t^1 \in \mathscr{G}^1} {}^2V_t\Big({}^1Q(g_t^1)\, {}^1\pi \Big), \tag{2.85b} $$

$$ {}^2V_t({}^2\pi) = \inf_{l_t^1 \in \mathscr{L}^1} {}^3V_t\Big({}^2Q(l_t^1)\, {}^2\pi \Big), \tag{2.85c} $$

$$ {}^3V_t({}^3\pi) = \inf_{g_t^2 \in \mathscr{G}^2} {}^4V_t\Big({}^3Q(g_t^2)\, {}^3\pi \Big), \tag{2.85d} $$

$$ {}^4V_t({}^4\pi) = \hat{\rho}({}^4\pi) + {}^1V_{t+1}\Big({}^4Q\, {}^4\pi \Big). \tag{2.85e} $$

For the infinite horizon problems, the fixed point equations remain the same as those of Theorems 2.3 and 2.4 with $\gamma_t := (\hat{g}_t^1, l_t^1, g_t^2)$, $\Gamma := \hat{\mathscr{G}}^1 \times \mathscr{L}^1 \times \mathscr{G}^2$,

$$ \tilde{Q}(\gamma_t) := {}^4Q \circ {}^3Q(g_t^2) \circ {}^2Q(l_t^1) \circ {}^1Q(g_t^1) $$

and, instead of (2.37) and (2.47), optimal decision rules $(g_t^{1,*}, l^{1,*}, g_t^{2,*})$ are given by

$$ (g_t^{1,*}, l^{1,*}, g_t^{2,*}) =: \gamma_t^* = \tilde{\Delta}^*({}^1\pi_t). \tag{2.86} $$

Significance of the results of variation v3

The results for variation v3 are similar to those for variation v2; the significance of and limitations of the results for both cases are the same.

2.8 Intuition behind the choice of information state

The most critical part of the sequential decomposition is identifying information states sufficient for performance evaluation. In this chapter we explained the properties that such information states need to satisfy, identified information states that satisfy these properties, and showed that the optimal control of the evolution of these information states leads to a sequential decomposition of the optimization problem formulated in Problem 2.1. At first glance the choice of the information state presented in this chapter seems ad-hoc. The reader may be left wondering

why the probability measures presented in Sections 2.2, 2.5–2.7 were chosen as information states. This choice was guided by some intuition and a lot of trial and error. However, on hindsight this choice of information state seems obvious.

We can view the two-agent team from the designer's point: the designer knows the system model and the statistics of the primitive random variables but does not know the observations of any agent. He is concerned with determining optimal decision rules for both agents *before the system starts operating*. From the designer's point of view, the optimization problem is centralized. The designer can look at the system as a stochastic input-output system. The stochastic inputs are the primitive random variables, the controlled inputs are the decision rules, and the output is the instantaneous cost. The input-output relation can be described consistently by the tuple $(X_t, S^1_{t-1}, S^2_{t-1})$, which represents the state of the plant, the stage of agent 1 and the state of agent 2. Thus, this tuple represents a state sufficient for the input-output mapping of the system. However, this state cannot be used for optimization because the designer does not observe this state. So, the optimization problem at the designer is conceptually equivalent to a POMDP. Hence, the designer can obtain a sequential decomposition by forming a belief on the state (sufficient for input-output mapping) of the system based on all the past information available to him (i.e., all the past decision rules, since the designer does not observe anything). This "belief" can be described by

$$\Pr\left(X_t, S^1_{t-1}, S^2_{t-1} \,\middle|\, {}^1\varphi^{t-1}\right)$$

which is the "conditional probability density" of the "state" conditioned on the all the past observations and "control actions" of the designer. Technically ${}^1\pi_t$ is not a conditional probability measure, rather it is a unconditional probability measure; but, this fact is a technicality which does not affect the solution methodology.

In Definition 2.3, information state at time ${}^1t$ is defined as

$${}^1\pi_t = \Pr\left(X_t, Y^1_t, S^1_{t-1}, S^2_{t-1} \,\middle|\, {}^1\varphi^{t-1}\right)$$

which can be simplified to

$$\begin{aligned}
&= \Pr\left(Y^1_t \,\middle|\, X_t, S^1_{t-1}, S^2_{t-1}; {}^1\varphi^{t-1}\right) \Pr\left(X_t, S^1_{t-1}, S^2_{t-1} \,\middle|\, {}^1\varphi^{t-1}\right) \\
&= P_{N^1_t}\left(N^1_t \in \mathcal{N}^1_t : Y^1_t = h^1_t(X_t, N^1_t)\right) \Pr\left(X_t, S^1_{t-1}, S^2_{t-1} \,\middle|\, {}^1\varphi^{t-1}\right) \\
&= \Pr\left(Y^1_t \,\middle|\, X_t\right) \Pr\left(X_t, S^1_{t-1}, S^2_{t-1} \,\middle|\, {}^1\varphi^{t-1}\right)
\end{aligned}$$

$\Pr\left(Y_t^1 \,\middle|\, X_t\right)$ depends on the statistics of the observation channel of agent 1 and does not depend on the decision rules. Thus, the information state defined in Definition 2.3 is essentially equivalent to the belief $\Pr(X_t, S_{t-1}^1, S_{t-1}^2 \mid {}^1\varphi^{t-1})$ of the designer on the state $(X_t, S_{t-1}^1, S^{t-1})$ which is sufficient for input output mapping of the system. Similar considerations motivate the choice of information states at time ${}^2t$, ${}^3t$, and ${}^4t$.

2.9 Conclusion

In this chapter, we considered a general model for a two-agent team and showed how to obtain a sequential decomposition for both finite and infinite horizon problems.

First, we considered a general finite-horizon model for a two-agent team. We formulated it as a decentralized optimization problem. We presented general properties that information states sufficient for performance evaluation should satisfy. We identified information states that satisfy these properties, and obtained a sequential decomposition of the finite-horizon problem using these information states.

Next, we restricted attention to time-homogeneous systems and considered three variations of infinite horizon problems: in variation v1 both agents have finite memory; in variations v2 and v3 one agent has finite memory, the other has perfect recall. For variation v1 we showed how to extend the sequential decomposition of finite horizon problems to two infinite horizon cases: total discounted cost criterion, and average cost per unit time criterion. For both these criteria, in general stationary designs are not optimal. However, for the total discounted cost there is no loss of optimality in restricting attention to stationary meta-designs; for the average cost per unit time criterion, if a technical condition holds, there is no loss of optimality in restricting attention to stationary meta-designs. For both cases, we derived functional equations whose fixed points determine optimal meta-designs. For variations v2 and v3, we showed that the agent with perfect recall can make optimal decisions based on its belief on the state of the plant and the state of the other agent. This structural result converts the model of variations v2 and v3 into a model similar to that of variation v1; a slight modification of the information states leads to a sequential decomposition for the infinite–horizon problems.

In the next two chapters we apply the results of this chapter to specific applications. We consider real-time communication in Chapter 3 and networked control systems in Chapter 4. We show that simple models of real-time communication and networked control systems can be considered as two-agent teams; this allows us to use the results derived in this chapter to optimally design real-time communication and networked control systems.

Chapter 3

Real-time communication

3.1 Introduction

What is real-time communication

Consider a point-to-point communication system consisting of a first order Markov source, a causal encoder, and a causal decoder. At each time, the decoder needs to estimate the output of the source that was generated δ steps earlier. The quality of the estimate is measured by a given distortion function. The objective is to design the encoder and the decoder to minimize the total expected distortion over a finite horizon.

When delay δ is zero, the problem is called ***zero-delay*** communication; when delay δ is finite but non-zero, the problem is called ***finite-delay*** communication; ***real-time*** communication is a generic term for both zero- and finite-delay communication.

There are four kinds of real-time communication systems depending on the nature of the communication channel between the encoder and the decoder:

R1. When the encoder and the decoder are connected over a one-way noiseless communication link, the system is equivalent to a *real-time source coding system*.

R2. When the encoder and the decoder are connected over a one-way noisy communication link, the system is equivalent to a *real-time joint source-channel coding system*.

R3. When the encoder and the decoder are connected over a two-way communication link with a noisy forward channel and noiseless backward channel, the system is equivalent to a *real time joint source-channel coding system with noiseless feedback*.

R4. When the encoder and the decoder are connected over a two-way communication link with noisy forward and backward channels, the system is equivalent to a *real-time joint source-channel coding system with noisy feedback*.

Motivation

In many informationally decentralized systems, the nodes of the system need to communicate with one another to improve the system performance; this communication must take place within bounded delay. For example, in transportation networks, the sensors need to communicate their observations to a controller in a timely manner so that the controller can efficiently control the flow of traffic. Other examples include multimedia streaming over wired and wireless networks, distributed routing, decentralized resource allocation, information flow in sensor networks, and consensus in partially synchronous systems. The operation of all of the above described systems include a real-time communication component. So, in order to understand how to design such systems it is necessary to understand real-time communication of information.

Conceptual difficulties

The real-time constraint on information transmission makes the real-time communication problem drastically different from the classical information theoretic formulation (Shannon, 1948) which has no delay constraint. Information theory is an asymptotic theory; the fundamental concepts of information theory like source entropy and channel capacity are asymptotic concepts; the performance bounds of information theory are tight only for asymptotically large values of delay. Real-time communication is not asymptotic. Hence, the concepts and results from information theory are not appropriate for real-time communication. In particular, separate source and channel coding is not optimal and joint source-channel coding strategies must be considered.

Real-time communication can be considered as a multi-stage sequential team with two agents—the encoder and the decoder. Due to the noise in the communication channel, the encoder does not know the information available at the decoder and vice-versa; thus, the two agents have different information. Due to this decentralization of information, solving the real-time communication problem as an optimization problem is outside the domain of Markov decision theory (Kumar and Varaiya, 1986) since Markov decision theory is only applicable to stochastic optimization problems with centralized information. However, the solution methodology developed in the previous chapter can be used to obtain a sequential decomposition for real-time communication problems.

Literature Overview

There are three approaches to real-time communication; each of them have received attention in the literature.

1. ***Performance bounds of finite delay or real−time communication systems***

   The first approach aims at identifying performance bounds of real-time communication systems. This approach is inspired by information theory. Various methods have been used to derive performance bounds of real-time communication systems including mathematical programming, forward flow of information, conditional mutual information, determination of non-anticipatory rate distortion function, randomizing over a family of encoders-decoders in Witsenhausen (1978), Teneketzis (1979), Munson (1981), Gorbunov and Pinsker (1973, 1974), Pinsker and Gorbunov (1987), Ho et al. (1978), Tatikonda (2000), Tatikonda and Mitter (2004a) and Merhav and Kontoyiannis (2003). However, these bounds are not tight for small values of delay.

   A weaker constraint of *causal source coding* was investigated in Lloyd (1977), Piret (1979), Neuhoff and Gilbert (1982) and Linder and Zamir (2001). The performance bounds of causal source coding are an upper bound on the performance bounds of real-time source coding.

2. ***Asymptotically efficient real–time encoding and decoding of individual sequences***

   The second approach aims at identifying asymptotically efficient real-time communication schemes when no statistical information is available about the source. This approach is inspired by universal source coding. For noiseless channels (i.e., for real-time source coding problem) asymptotically efficient communication strategies were derived in Linder and Lugosi (2001), Weissman and Merhav (2002) and Gyorgy et al. (2004); for noisy channels such strategies were derived in Matloub and Weissman (2006).

3. ***Optimal real–time encoding and decoding of Markov sources***

   The third approach aims at identifying qualitative properties of optimal real-time communication schemes when the source statistics are known; it is usually assumed that the source is first-order Markov; for higher-order Markov sources, the source is transformed into a first-order Markov source and then the qualitative properties of optimal real-time communication strategies for first-order Markov sources can be translated to higher-order sources. This approach also aims at identifying algorithms to efficiently search for optimal communication schemes. This approach is inspired by Markov decision theory. Qualitative properties of real-time decoders for noisy observations of a Markov source were considered in Drake (1962) and Devore (1974). Qualitative properties of optimal real–time encoders for transmitting Markov sources over a noiseless channel were derived in Witsenhausen (1979), Gaarder and Slepian (1979, 1982) and Borkar et al. (2001). Qualitative properties of optimal real-time encoders and decoders for transmitting Markov sources over noisy channels with noiseless feedback and a methodology for determining globally optimal encoding and decoding strategies were derived in Walrand and Varaiya (1983a), Lipster and Shiryayev (1977) and Basar and Bansal (1989, 1994). Qualitative properties of optimal real–time encoders and decoders for transmitting Markov sources in systems with noisy channels and no feedback was considered in Teneketzis (2006).

In this chapter we follow the philosophy of the third approach. Specifically, we seek to find techniques for efficient search of an optimal communication scheme. So far, most of the research along the lines of the third approach has focussed on systems with either a noiseless forward channel or with a noisy forward channel

and noiseless feedback. In both these cases, the encoder knows everything that is known to the receiver. We concentrate on the other two cases: a noisy forward channel with no feedback; and a noisy forward channel with noisy feedback. We model these real-time communication systems as two-agent sequential teams and use the methodology of Chapter 2 to obtain a sequential decomposition.

Outline of the approach

In this chapter we consider four models for real-time communication.

R1. Real-time communication over noiseless forward channel.

R2. Real-time communication over noisy forward channel

R3. Real-time communication over noisy forward channel noiseless backward channels.

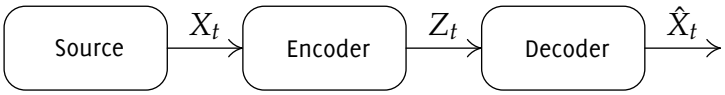
R4. Real-time communication over noisy forward and backward channels.

These models are shown in Figure 3.1. We will consider the simplest instance of these models.

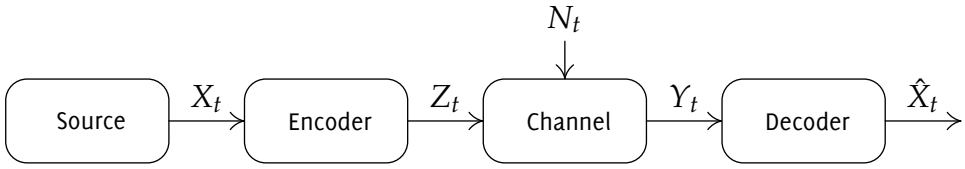
All the four models consist of a source, an encoder and a receiver. A communication channel, which is either noiseless or noisy, exists between the encoder and the receiver. For models R3 and R4 a communication channel, which is either noiseless or noisy, exists between the receiver and the encoder. For all four models, we assume that the source is first order Markov; the encoder and the receiver operate in real-time; the noisy communication channels are memoryless; and distortion is measured by a given metric that accepts zero delay.

These models can be generalized to more realistic models; the distortion metric may accept a fixed finite delay; the source may be higher-order Markov; the channels may have memory.

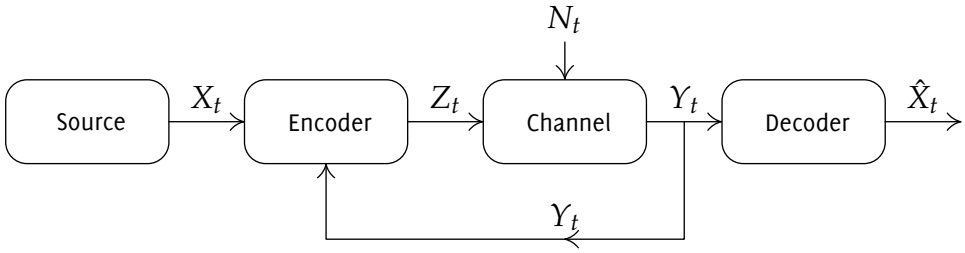
In model R2 if the forward channel is noiseless then the model reduces to model R1. In model R4 if the backward channel is noiseless then the model reduces to model R3. In this chapter, we show that models R2 and R4 are special cases of the two-agent team model considered in Chapter 2. This also implies that models R1 and R3 are special cases of the two-agent team model considered in Chapter 2. Thus, we can use the results of Chapter 2 to obtain a sequential decomposition of all four models of real-time communication considered in this chapter.



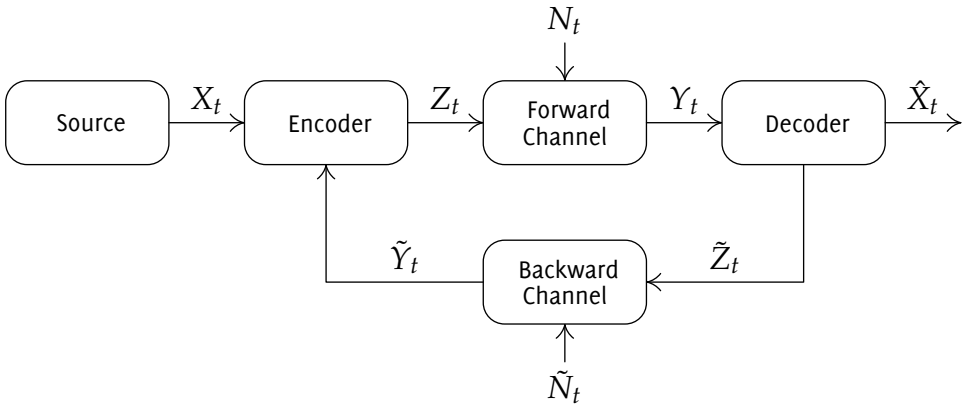
Model R1: Real-time communication over noiseless forward channel



Model R2: Real-time communication over noisy forward channel



Model R3: Real-time communication over noisy forward and noiseless backward channels



Model R4: Real-time communication over noisy forward and backward channels

Figure 3.1: Four models for point-to-point real-time communication systems

For all four models, we also consider four variations of infinite horizon problems along the lines of variations v1–v4 of Chapter 2 depending on whether the encoder and the receiver have finite memory or perfect recall. For variations v1, v2, and v3 of these models, we can search for optimal design of the encoder and the decoder using the results of Chapter 2.

The remainder of this chapter is organized as follows. In Section 3.2 we formally define model R2, show how it can be considered as a special instance of the two-

agent team model of Chapter 2, and show how to obtain a sequential decomposition of this model. In Section 3.3 we consider model R1 and show how the sequential decomposition of model R2 simplifies in this case. In Sections 3.4 and 3.5 we consider models R4 and R3, respectively. In Section 3.6 we compare the philosophy of our approach to real-time communication with the philosophy of information theory and coding theory. We conclude in Section 3.7.

3.2 Model R2: real-time communication over noisy channels

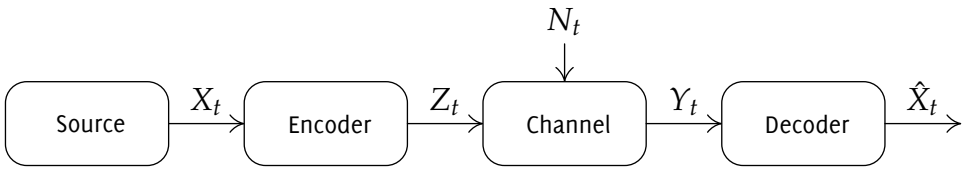


Figure 3.2: Real-time communication over noisy forward channel

Problem formulation

Consider the system of model R2 shown in Figure 3.2. The source is first-order Markov; it produces a random sequence $\{X_t, t = 1, \dots, T\}$. For simplicity of exposition we assume that X_t takes values in a finite alphabet $\mathcal{X}$. Let P_{X_1} denote the PMF (probability mass function) of the first source output X_1, and $P_{X_{t+1}|X_t}$ denote the transition probability at time t.

At each stage t, the encoder generates an encoded symbol Z_t taking values in a finite alphabet $\mathcal{Z}$ as follows:

$$Z_t = c_t(X_t, S_{t-1}), \tag{3.1}$$

where c_t is the ***encoding function*** and $S_{t-1} \in \mathcal{S}_{t-1}$ is the state or memory of the encoder. The size of the encoder's memory can increase with time; so, this model includes the case when the encoder has perfect recall. The encoder updates its memory according to

$$S_t = d_t(X_t, S_{t-1}), \tag{3.2}$$

where d_t is the ***encoder's memory-update rule***.

The encoded symbol Z_t is transmitted over a $|\mathcal{Z}|$-input $|\mathcal{Y}|$-output DMC (discrete memoryless channel) producing a channel output Y_t which belongs to a finite alphabet $\mathcal{Y}$. The channel can be described by

$$Y_t = h_t(Z_t, N_t), \tag{3.3}$$

where $h_t(\cdot)$ denotes the channel function at time t, and N_t, which belongs to $\mathcal{N}$, denotes the channel noise at time t. We assume that $\{N_t, t = 1, \dots, T\}$ is a sequence of independent random variables and denote the PMF (probability mass function) of N_t by P_{N_t}. We also assume that $\{N_t, t = 1, \dots, T\}$ is independent of the source output $\{X_t, t = 1, \dots, T\}$.

The receiver generates an estimate $\hat{X}_t$ of the source according to

$$\hat{X}_t = g_t(Y_t, M_{t-1}), \tag{3.4}$$

where $\hat{X}_t \in \hat{\mathcal{X}}$, g_t is the ***decoding function*** and $M_{t-1} \in \mathcal{M}_{t-1}$ is the state or memory of the receiver. The size of the receiver's memory can increase with time; so, this model includes the case when the receiver has perfect recall. The receiver updates its memory according to

$$M_t = l_t(Y_t, M_{t-1}), \tag{3.5}$$

where l_t is the ***receiver's memory-update rule***.

The performance of the system is determined by a sequence of distortion functions, $\rho_t : \mathcal{X} \times \hat{\mathcal{X}} \to [0, \rho_{\max}]$, where $\rho_{\max} < \infty$. The function $\rho_t(X_t, \hat{X}_t)$ measures the distortion at stage t.

The collection $C := (c_1, \dots, c_T)$ of encoding rules for the entire horizon is called an ***encoding strategy***; the collection $D := (d_1, \dots, d_T)$ of encoder's memory-update rules is called the ***encoder's memory-update strategy***. Similarly, $G := (g_1, \dots, g_T)$ is called a ***decoding strategy*** and $L := (l_1, \dots, l_T)$ is called the ***receiver's memory update strategy***. Further, the choice (C, D, G, L) of communication rules for the entire horizon is called a ***communication strategy*** or a ***design***. The performance of a communication strategy is quantified by the expected total distortion under that strategy and is given by

$$\mathcal{J}_T(C, D, G, L) := \mathbb{E}\left\{ \sum_{t=1}^{T} \rho_t(X_t, \hat{X}_t) \,\middle|\, C, D, G, L \right\}. \tag{3.6}$$

We are interested in the following optimization problem

Problem 3.1. *Assume that the encoder and the receiver know the time horizon T, the statistics of the source (i.e., the PMF of X_1 and the transition probabilities $P_{X_{t+1}|X_t}$), the channel function h_t, the statistics P_{N_t} of the noise, the distortion function $\rho_t(\cdot,\cdot)$, $t = 1,\dots,T$. Determine a communication strategy (C^*, D^*, G^*, L^*) that is optimal with respect to the performance criterion of* (3.6)*, i.e.,*

$$\mathcal{J}_T(C^*, D^*, G^*, L^*) = \mathcal{J}_T^* \coloneqq \min_{\substack{C\in\mathscr{C}^T\\ D\in\mathscr{D}^T\\ G\in\mathscr{G}^T\\ L\in\mathscr{L}^T}} \mathcal{J}_T(C, D, G, L), \tag{3.7}$$

where $\mathscr{C}^T \coloneqq \mathscr{C}_1 \times \cdots \times \mathscr{C}_T$; $\mathscr{C}_t$ is the family of functions from $\mathcal{X}_t \times \mathcal{S}_{t-1}$ to $\mathcal{Z}$; $\mathscr{D}^T \coloneqq \mathscr{D}_1 \times \cdots \times \mathscr{D}_T$; $\mathscr{D}_t$ is the family of functions from $\mathcal{X}_t \times \mathcal{S}_{t-1}$ to $\mathcal{S}_t$; $\mathscr{G}^T \coloneqq \mathscr{G}_1 \times \cdots \times \mathscr{G}_T$; $\mathscr{G}_t$ is the family of functions from $\mathcal{Y}_t \times \mathcal{M}_{t-1}$ to $\hat{\mathcal{X}}$; $\mathscr{L}^T \coloneqq \mathscr{L}_1 \times \cdots \times \mathscr{L}_T$; and $\mathscr{L}_t$ is the family of functions from $\mathcal{Y}_t \times \mathcal{M}_{t-1}$ to $\mathcal{M}_t$.

Reduction to the model of Chapter 2

Consider an instance of a two-agent team of Chapter 2 with the following restrictions:

1. The plant function f_t does not depend on control actions of the two agents, i.e.,

$$X_{t+1} = f_t(X_t, W_t). \tag{3.8a}$$

2. The observation channel of agent 1 is noiseless; the observation channel of agent 2 does not depend on the state of the plant, i.e.,

$$Y_t^1 = X_t, \qquad Y_t^2 = h_t^2(U_t^1, N_t^2). \tag{3.8b}$$

   and therefore the control laws can be written as

$$U_t^1 = g_t^1(X_t, S_{t-1}^1), \qquad U_t^2 = g_t^2(Y_t^2, S_{t-1}^2). \tag{3.8c}$$

3. The state-update functions of both agents do not depend on the control actions of the agents, i.e.,

$$S_t^1 = l_t^1(X_t, S_{t-1}^1), \qquad S_t^2 = l_t^2(Y_t^2, S_{t-1}^2). \tag{3.8d}$$

4. The instantaneous cost does not depend on the control action of agent 1, and is given by $\rho_t(X_t, U_t^2)$.

This instance of the two-agent problem of Section 2.1 is equivalent to model R2 of real-time communication over noisy channels. The relation between the variables of the two models is shown in Table 3.1.

Component	Variable	Two-agent team	Model R2
Plant	State	X_t	X_t
Agent 1	Observation	Y_t^1	X_t
	Control action	U_t^1	Z_t
	State	S_t^1	S_t
Agent 2	Observation	Y_t^2	Y_t
	Control action	U_t^2	$\hat{X}_t$
	State	S_t^2	M_t

Table 3.1: Model R2 as an instance of two-agent team. In model R2, the Markov source is the plant, the encoder is agent 1, and the receiver is agent 2.

Information states and global optimization

Substituting the above described reduction in Definition 2.3, we get that the information states for model R2 are given by

$$ {}^1\pi_t = \Pr\left(X_t, S_{t-1}, M_{t-1} \,\middle|\, {}^1\varphi^{t-1}\right) \tag{3.9a} $$

$$ {}^2\pi_t = \Pr\left(X_t, Z_t, S_{t-1}, M_{t-1} \,\middle|\, {}^2\varphi^{t-1}\right) \tag{3.9b} $$

$$ {}^3\pi_t = \Pr\left(X_t, Y_t, S_t, M_{t-1} \,\middle|\, {}^3\varphi^{t-1}\right) \tag{3.9c} $$

$$ {}^4\pi_t = \Pr\left(X_t, Y_t, \hat{X}_t, S_t, M_{t-1} \,\middle|\, {}^4\varphi^{t-1}\right) \tag{3.9d} $$

The time evolution of these information states is given by Lemma 2.1; the transformations ${}^1Q_t$, ${}^2Q_t$, ${}^3Q_t$ and ${}^4Q_t$ and the cost $\hat{\rho}_t$ in Lemma 2.1 can be simplified by incorporating the above described reduction in their definitions. Consequently, the sequential decomposition of Problem 3.1 is given by the nested optimality equations of Theorem 2.1.

Infinite horizon problems

We can consider the following three variations of the time-homogeneous infinite-horizon version of model R2 :

v1. Both the encoder and receiver have finite memory;
v2. The encoder has perfect recall while the receiver has finite memory;
v3. The encoder has finite memory while the receiver has perfect recall.

These are equivalent to the variations of the two-agent teams considered in Chapter 2. Thus, the fixed-point equations for the two infinite horizon criteria derived for variations v1, v2, and v3 in Chapter 2 are also applicable to the corresponding variations of the real-time communication problem of model R2.

Variation v2

For variation v2, the results of the finite and infinite horizon problems can be simplified further. Recall that in variation v2 agent 1 (the encoder) has perfect recall and agent 2 (the receive) has time-invariant state. We first derive qualitative properties by looking at the system from the point of view of the encoder. In model R2, the encoder observes the source output (the state of the plant). Therefore, the beliefs of agent 1 given by Definition 2.6 can be simplified to

$$ {}^{i}B^{1}_{t} = \Pr\Big(X_t, M_{t-1} \,\Big|\, {}^{i}\mathfrak{I}^{1}_{t}\Big) \equiv \Big(X_t, \Pr\Big(M_{t-1} \,\Big|\, {}^{i}\mathfrak{I}^{1}_{t}\Big)\Big). $$

Consequently, the structural results of variation v2 (Theorem 2.5) imply that there is no loss of optimality to restrict attention to encoders of the form

$$ Z_t = c_t\Big(X_t, \Pr\Big(M_{t-1} \,\Big|\, {}^{1}\mathfrak{I}^{1}_{t}\Big)\Big) \tag{3.10} $$

Hence, for variation v2, the information states simplify to

$$ {}^{1}\pi_t = \Pr\Big(X_t, {}^{1}\tilde{B}^{1}_{t}, M_{t-1} \,\Big|\, {}^{1}\varphi^{t-1}\Big) \tag{3.11a} $$

$$ {}^{2}\pi_t = \Pr\Big(X_t, Z_t, {}^{2}\tilde{B}^{1}_{t}, M_{t-1} \,\Big|\, {}^{2}\varphi^{t-1}\Big) \tag{3.11b} $$

$$ {}^{3}\pi_t = \Pr\Big(X_t, Y_t, {}^{3}\tilde{B}^{1}_{t}, M_{t-1} \,\Big|\, {}^{3}\varphi^{t-1}\Big) \tag{3.11c} $$

$$ {}^{4}\pi_t = \Pr\Big(X_t, Y_t, \hat{X}_t, {}^{4}\tilde{B}^{1}_{t}, M_{t-1} \,\Big|\, {}^{4}\varphi^{t-1}\Big) \tag{3.11d} $$

where ${}^{i}\tilde{B}^{1}_{t} := \Pr\left({}^{i}O^{2}_{t} \,\middle|\, {}^{i}\mathfrak{I}^{1}_{t}\right)$. This leads to the corresponding simplification of nested optimality equations of Theorem 2.1.

3.3 Model R1: real-time communication over noiseless channels

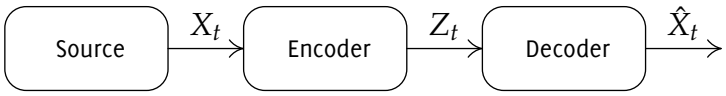


Figure 3.3: Real-time communication over noiseless forward channel

Finite horizon problem

Model R1, which is shown in Figure 3.3, is identical to model R2 except that $\mathcal{Y} = \mathcal{Z}$ and the forward channel h_t is noiseless, i.e., in model R1 the communication channel is given by

$$Y_t = h_t(Z_t, N_t) := Z_t \tag{3.12}$$

instead of (3.3).

For this model the information states of model R2, given by (3.9), simplify to

$${}^{1}\pi_t = \Pr\left(X_t, S_{t-1}, M_{t-1} \,\middle|\, {}^{1}\varphi^{t-1}\right) \tag{3.13a}$$

$${}^{2}\pi_t = \Pr\left(X_t, Z_t, S_{t-1}, M_{t-1} \,\middle|\, {}^{2}\varphi^{t-1}\right) \tag{3.13b}$$

$${}^{3}\pi_t = \Pr\left(X_t, Z_t, S_t, M_{t-1} \,\middle|\, {}^{3}\varphi^{t-1}\right) \tag{3.13c}$$

$${}^{4}\pi_t = \Pr\left(X_t, Z_t, \hat{X}_t, S_t, M_{t-1} \,\middle|\, {}^{4}\varphi^{t-1}\right) \tag{3.13d}$$

The time evolution of the information states (Lemma 2.1) simplify further due to the noiseless nature of h_t. Optimal encoding and decoding strategies can be determined by the nested optimality equations of Theorem 2.1.

Infinite horizon problem

For the infinite horizon problem we can consider variations v1, v2, and v3 as for model R2; optimal encoding and decoding strategies are given by the fixed point equations for variations v1, v2, and v3 derived in Chapter 2.

Variation v2

For variation v2, the results of the finite and infinite horizon can be simplified further. The encoder (agent 1) perfectly observes the Markov source (the plant) and the observations of the receiver (agent 2). Consequently, the belief of agent 1 (see Definition 2.6) simplify to

$$ {}^{i}B_t^1 = \Pr\left(X_t, M_{t-1} \,\middle|\, {}^{i}\mathfrak{I}_t^1\right) \equiv (X_t, M_{t-1}). $$

Therefore, the structural results of variation v2 (Theorem 2.5) imply that there is no loss of optimality to restrict attention

$$ Z_t = c_t(X_t, M_{t-1}) \tag{3.14} $$

Hence, for variation v2, the information states simplify to

$$ {}^{1}\pi_t = \Pr\left(X_t, M_{t-1} \,\middle|\, {}^{1}\varphi^{t-1}\right) \tag{3.15a} $$

$$ {}^{2}\pi_t = \Pr\left(X_t, Z_t, M_{t-1} \,\middle|\, {}^{2}\varphi^{t-1}\right) \tag{3.15b} $$

$$ {}^{3}\pi_t = \Pr\left(X_t, Z_t, M_{t-1} \,\middle|\, {}^{3}\varphi^{t-1}\right) = {}^{2}\pi_t \tag{3.15c} $$

$$ {}^{4}\pi_t = \Pr\left(X_t, Z_t, \hat{X}_t, M_{t-1} \,\middle|\, {}^{4}\varphi^{t-1}\right) \tag{3.15d} $$

This leads to the corresponding simplification of the nested optimality equations of Theorem 2.1.

3.4 Model R4: real-time communication over noisy channels with noisy feedback

Problem formulation

Consider the system of model R4 shown in Figure 3.4. The source is first-order Markov; it produces a random sequence $\{X_t,\, t = 1, \dots, T\}$. For simplicity of exposition we assume that X_t takes values in a finite alphabet $\mathcal{X}$. Let P_{X_1} denote the PMF (probability mass function) of the first output X_1, and $P_{X_{t+1}|X_t}$ denote the transition probability at time t.

At each stage t, the encoder generates an encoded symbol Z_t taking values in a finite alphabet $\mathcal{Z}$ as follows:

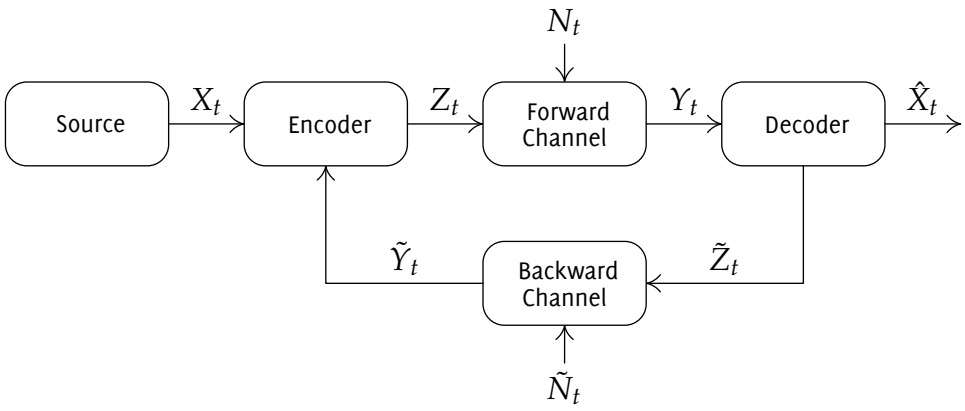


Figure 3.4: Real-time communication over noisy forward and backward channels

$$Z_t = c_t(X_t, \tilde{Y}_{t-1}, S_{t-1}), \tag{3.16}$$

where c_t is the ***encoding function***, $\tilde{Y}_{t-1} \in \tilde{\mathcal{Y}}$ is the output of the backward channel, and $S_{t-1} \in \mathcal{S}_{t-1}$ is the state or memory of the encoder. The size of the encoder's memory can increase with time; so, this model includes the case when the encoder has perfect recall. The encoder updates its memory according to

$$S_t = d_t(X_t, \tilde{Y}_{t-1}, S_{t-1}), \tag{3.17}$$

where d_t is the ***encoder's memory-update rule***.

The encoded symbol Z_t is transmitted over a $|\mathcal{Z}|$-input $|\mathcal{Y}|$-output DMC (discrete memoryless channel) producing a channel output Y_t which belongs to a finite alphabet $\mathcal{Y}$. The channel can be described by

$$Y_t = h_t(Z_t, N_t), \tag{3.18}$$

where $h_t(\cdot)$ denotes the channel function at time t, and N_t, which belongs to $\mathcal{N}$, denotes the channel noise at time t. We assume that $\{N_t, t = 1, \ldots, T\}$ is a sequence of independent random variables and denote the PMF (probability mass function) of N_t by P_{N_t}. We also assume that $\{N_t, t = 1, \ldots, T\}$ is independent of the source output $\{X_t, t = 1, \ldots, T\}$.

The receiver generates an estimate $\hat{X}_t$ of the source according to

$$\hat{X}_t = g_t(Y_t, M_{t-1}), \tag{3.19}$$

where $\hat{X}_t \in \hat{\mathcal{X}}$, g_t is the ***decoding function*** and $M_{t-1} \in \mathcal{M}_{t-1}$ is the state or memory of the receiver. The size of the receiver's memory can increase with time; so, this model includes the case when the receiver has perfect recall. The receiver generates a feedback symbol according to

$$\tilde{Z}_t = \tilde{c}_t(Y_t, M_{t-1}), \tag{3.20}$$

where $\tilde{c}_t$ is the ***feedback function***. The receiver then updates its memory according to

$$M_t = l_t(Y_t, M_{t-1}), \tag{3.21}$$

where l_t is the ***receiver's memory-update rule***.

The feedback symbol is transmitted over a $|\tilde{\mathcal{Z}}|$-input $|\tilde{\mathcal{Y}}|$ output DMC producing a channel output $\tilde{Y}_t$ which belongs to a finite alphabet $\tilde{\mathcal{Y}}$. The backward channel is described by

$$\tilde{Y}_t = \tilde{h}_t(\tilde{Z}_t, \tilde{N}_t), \tag{3.22}$$

where $\tilde{h}_t(\cdot)$ denotes the backward channel at time t, and $\tilde{N}_t$, which belongs to $\tilde{\mathcal{N}}$, denotes the channel noise at time t. We assume that $\{\tilde{N}_t, t = 1, \ldots, T\}$ is a sequence of independent random variables with PMF $P_{\tilde{N}_t}$. We also assume that $\{\tilde{N}_t, t = 1, \ldots, T\}$ is independent of the noise in the forward channel $\{N_t, t = 1, \ldots, T\}$ and the source output $\{X_t, t = 1, \ldots, T\}$.

The performance of the system is determined by a sequence of distortion functions, $\rho_t : \mathcal{X} \times \hat{\mathcal{X}} \to [0, \rho_{\max}]$, where $\rho_{\max} < \infty$. The function $\rho_t(X_t, \hat{X}_t)$ measures the distortion at stage t.

The collection $C := (c_1, \ldots, c_T)$ of the encoding rules for the entire horizon is called an ***encoding strategy***; the collection $D := (d_1, \ldots, d_T)$ of the encoder's memory-update rules is called an ***encoder's memory-update strategy***. Similarly, $G := (g_1, \ldots, g_T)$ is called a ***decoding strategy***, $\tilde{C} := (\tilde{c}_1, \ldots, \tilde{c}_T)$ is called a ***feedback strategy***, and $L := (l_1, \ldots, l_T)$ is called a ***receiver's memory update strategy***. Further, the choice $(C, D, G, \tilde{C}, L)$ of communication rules for the entire horizon is called a ***communication strategy*** or a ***design***. The performance of a communication strategy is quantified by the expected total distortion under that strategy and is given by

$$\mathcal{J}_T(C, D, G, \tilde{C}, L) \coloneqq \mathbb{E}\left\{ \sum_{t=1}^{T} \rho_t(X_t, \hat{X}_t) \,\middle|\, C, D, G, \tilde{C}, L \right\}. \tag{3.23}$$

We are interested in the following optimization problem

Problem 3.2. *Assume that the encoder and the receiver know the time horizon T, the statistics of the source (i.e., the PMF of X_1 and the transition probabilities $P_{X_{t+1}|X_t}$), the channel functions h_t and $\tilde{h}_t$, the statistics P_{N_t} and $P_{\tilde{N}_t}$ of the noise, the distortion function $\rho_t(\cdot,\cdot)$, $t = 1, \dots, T$. Determine a communication strategy $(C^*, D^*, G^*, \tilde{C}^*, L^*)$ that is optimal with respect to the performance criterion of* (3.23)*, i.e.,*

$$\mathcal{J}_T(C^*, D^*, G^*, \tilde{C}^*, L^*) = \mathcal{J}_T^* \coloneqq \min_{\substack{C \in \mathscr{C}^T \\ D \in \mathscr{D}^T \\ \tilde{C} \in \tilde{\mathscr{C}}^T \\ G \in \mathscr{G}^T \\ L \in \mathscr{L}^T}} \mathcal{J}_T(C, D, G, \tilde{C}, L), \tag{3.24}$$

where $\mathscr{C}^T \coloneqq \mathscr{C}_1 \times \cdots \times \mathscr{C}_T$; $\mathscr{C}_t$ is the family of functions from $\mathcal{X}_t \times \mathcal{S}_{t-1}$ to $\mathcal{Z}$; $\mathscr{D}^T \coloneqq \mathscr{D}_1 \times \cdots \times \mathscr{D}_T$; $\mathscr{D}_t$ is the family of functions from $\mathcal{X}_t \times \mathcal{S}_{t-1}$ to $\mathcal{S}_t$; $\mathscr{G}^T \coloneqq \mathscr{G}_1 \times \cdots \times \mathscr{G}_T$; $\mathscr{G}_t$ is the family of functions from $\mathcal{Y}_t \times \mathcal{M}_{t-1}$ to $\hat{\mathcal{X}}$; $\tilde{\mathscr{C}}^T \coloneqq \tilde{\mathscr{C}}_1 \times \cdots \times \tilde{\mathscr{C}}_T$; $\tilde{\mathscr{C}}_t$ is the family of functions from $\mathcal{Y}_t \times \mathcal{M}_{t-1}$ to $\tilde{\mathcal{Z}}$; $\mathscr{L}^T \coloneqq \mathscr{L}_1 \times \cdots \times \mathscr{L}_T$; and $\mathscr{L}_t$ is the family of functions from $\mathcal{Y}_t \times \mathcal{M}_{t-1}$ to $\mathcal{M}_t$.

Reduction to the model of Chapter 2

Consider an instance of a two-agent team of Chapter 2 with the following restrictions:

1. The state of the plant consists of two components X_t and $\tilde{X}_t$; the control action of agent 2 consists of two components U_t^2 and $\tilde{U}_t^2$; and the plant disturbance consists of two components W_t and $\tilde{N}_t$. The plant update function does not depend on the control action of agent 1. Further, the two components of the state of the plant evolve as follows:

$$\begin{aligned}(X_{t+1}, \tilde{X}_{t+1}) &= f_t\Big((X_t, \tilde{X}_t), U_t^1, (U_t^2, \tilde{U}_t^2), (W_t, \tilde{N}_t)\Big) \\ &\coloneqq \Big(\tilde{f}_t(X_t, W_t), \tilde{h}_t(U_t^2, \tilde{N}_t)\Big)\end{aligned} \tag{3.25a}$$

2. The observation channel of agent 1 is noiseless; the observation channel of agent 2 does not depend on the state of the plant, i.e.,

$$Y_t^1 = (X_t, \tilde{X}_t), \qquad Y_t^2 = h_t^2(U_t^1, N_t^2). \tag{3.25b}$$

and therefore the control laws can be written as

$$U_t^1 = g_t^1(X_t, \tilde{X}_t, S_{t-1}^1), \qquad (U_t^2, \tilde{U}_t^2) = g_t^2(Y_t^2, S_{t-1}^2). \tag{3.25c}$$

3. The state-update functions of both agents do not depend on the control actions of the agents, i.e.,

$$S_t^1 = l_t^1(X_t, S_{t-1}^1), \qquad S_t^2 = l_t^2(Y_t^2, S_{t-1}^2). \tag{3.25d}$$

4. The instantaneous cost does not depend on the control action of agent 1, and is given by $\rho_t(X_t, U_t^2)$.

This instance of the two-agent problem of Section 2.1 is equivalent to model R4 of real-time communication over noisy channels with noisy feedback: the Markov source and the feedback from the receiver is the plant, the encoder is agent 1, and the receiver is agent 2, respectively. The relation between the variables of the models is shown in Table 3.2.

Component	**Variable**	**Two-agent team**	**Model R2**
Plant	State	X_t	$(X_t, \tilde{Y}_{t-1})$
Agent 1	Observation	Y_t^1	$(X_t, \tilde{Y}_{t-1})$
	Control action	U_t^1	Z_t
	State	S_t^1	S_t
Agent 2	Observation	Y_t^2	Y_t
	Control action	U_t^2	$(\hat{X}_t, \tilde{Z}_t)$
	State	S_t^2	M_t

Table 3.2: Model R4 as an instance of two-agent team. In model R4, the Markov source and the backward channel are the plant, the encoder is agent 1, and the receiver is agent 2.

Information states and global optimization

Substituting the above described reduction in Definition 2.3, we get that the information states for Model R4 are given by

$$^1\pi_t = \Pr\left(X_t, \tilde{Y}_{t-1}, S_{t-1}, M_{t-1} \,\middle|\, ^1\varphi^{t-1}\right) \tag{3.26a}$$

$$^2\pi_t = \Pr\left(X_t, \tilde{Y}_{t-1}, Z_t, S_{t-1}, M_{t-1} \,\middle|\, ^2\varphi^{t-1}\right) \tag{3.26b}$$

$$^3\pi_t = \Pr\left(X_t, \tilde{Y}_{t-1}, Y_t, S_t, M_{t-1} \,\middle|\, ^3\varphi^{t-1}\right) \tag{3.26c}$$

$$^4\pi_t = \Pr\left(X_t, \tilde{Y}_{t-1}, Y_t, \hat{X}_t, \tilde{Y}_t, S_t, M_{t-1} \,\middle|\, ^4\varphi^{t-1}\right) \tag{3.26d}$$

In the system equations (3.25), the component $\tilde{Y}_{t-1}$ of the state $\tilde{X}_t$ only affects the observation of agent 1. So, it can be discarded after time ^2t, when agent 1 has updated its state. Thus, the information states can be further simplified to

$$^1\pi_t = \Pr\left(X_t, \tilde{Y}_{t-1}, S_{t-1}, M_{t-1} \,\middle|\, ^1\varphi^{t-1}\right) \tag{3.27a}$$

$$^2\pi_t = \Pr\left(X_t, \tilde{Y}_{t-1}, Z_t, S_{t-1}, M_{t-1} \,\middle|\, ^2\varphi^{t-1}\right) \tag{3.27b}$$

$$^3\pi_t = \Pr\left(X_t, Y_t, S_t, M_{t-1} \,\middle|\, ^3\varphi^{t-1}\right) \tag{3.27c}$$

$$^4\pi_t = \Pr\left(X_t, Y_t, \hat{X}_t, \tilde{Y}_t, S_t, M_{t-1} \,\middle|\, ^4\varphi^{t-1}\right) \tag{3.27d}$$

The time evolution of these information states is given by Lemma 2.1; the transformations ^1Q_t, ^2Q_t, ^3Q_t and ^4Q_t and the cost $\hat{\rho}_t$ in Lemma 2.1 can be simplified by incorporating the above described reduction in their definitions. Consequently, the sequential decomposition of Problem 3.1 is given by the nested optimality equations of Theorem 2.1.

Infinite horizon problems

We can consider the following three variations of the time-homogeneous infinite-horizon version of Model R4

- V1. Both the encoder and receiver have finite memory;
- V2. The encoder has perfect recall while the receiver has finite memory;
- V3. The encoder has finite memory while the receiver has perfect recall.

These are equivalent to the variations of the two-agent teams considered in Chapter 2. Thus, the fixed-point equations for the two infinite horizon criteria derived

for variations v1, v2, and v3 in Chapter 2 are also applicable to the corresponding variations of the real-time communication problem of model R4.

Variation v2

For variation v2, the results of the finite and infinite horizon can be simplified further. The encoder (agent 1) perfectly observes $(X_t, \tilde{Y}_{t-1})$ which is equivalent to the state of the plant. Therefore, the beliefs of agent 1 given by Definition 2.6 simplify to

$$ {}^{i}B^1_t = \Pr\Big(X_t, \tilde{Y}_{t-1}, M_{t-1} \Big| {}^{i}\mathfrak{I}^1_t\Big) \equiv \Big(X_t, \tilde{Y}_{t-1}, \Pr\Big(M_{t-1} \Big| {}^{i}\mathfrak{I}^1_t\Big)\Big). $$

Consequently, the structural results of variation v2 (Theorem 2.5) imply that there is no loss of optimality to restrict attention to encoders of the form

$$ Z_t = c_t\Big(X_t, \tilde{Y}_{t-1}, \Pr\Big(M_{t-1} \Big| {}^{1}\mathfrak{I}^1_t\Big)\Big) \tag{3.28} $$

Hence, for variation v2, the information states simplify to

$$ {}^{1}\pi_t = \Pr\Big(X_t, \tilde{Y}_{t-1}, {}^{1}\tilde{B}^1_t, M_{t-1} \Big| {}^{1}\varphi^{t-1}\Big) \tag{3.29a} $$

$$ {}^{2}\pi_t = \Pr\Big(X_t, \tilde{Y}_{t-1}, Z_t, {}^{2}\tilde{B}^1_t, M_{t-1} \Big| {}^{2}\varphi^{t-1}\Big) \tag{3.29b} $$

$$ {}^{3}\pi_t = \Pr\Big(X_t, Y_t, {}^{3}\tilde{B}^1_t, M_{t-1} \Big| {}^{3}\varphi^{t-1}\Big) \tag{3.29c} $$

$$ {}^{4}\pi_t = \Pr\Big(X_t, Y_t, \hat{X}_t, \tilde{Z}_t, {}^{4}\tilde{B}^1_t, M_{t-1} \Big| {}^{4}\varphi^{t-1}\Big) \tag{3.29d} $$

where ${}^{i}\tilde{B}^1_t \coloneqq \Pr\Big(M_{t-1} \Big| {}^{i}\mathfrak{I}^1_t\Big)$. This leads to the corresponding simplification of nested optimality equations of Theorem 2.1.

3.5 Model R3: real-time communication over noisy channels with noiseless feedback

Finite horizon problem

Model R3, which is shown in Figure 3.2, is identical to model R4 except that $\tilde{\mathcal{Y}} = \tilde{\mathcal{Z}} = \mathcal{Y}$, the output of the forward channel is fed back into the backward channel, and the backward channel $\tilde{h}_t$ is noiseless, i.e., in model R4 the backward channel is given by

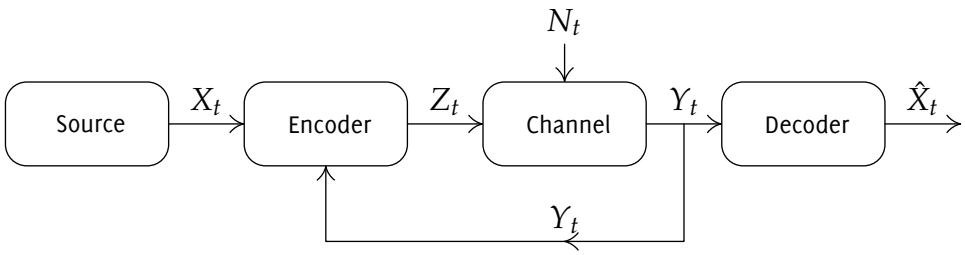


Figure 3.5: Real-time communication over noisy forward and noiseless backward channels

$$\tilde{Z}_t = \tilde{c}_t(Y_t, M_{t-1}) \coloneqq Y_t \tag{3.30}$$

instead of (3.20) and

$$\tilde{Y}_t = \tilde{h}_t(\tilde{Z}_t, \tilde{N}_t) \coloneqq \tilde{Z}_t \tag{3.31}$$

instead of (3.22). Thus, $\tilde{Y}_t = Y_t$.

For this model the information states of model R4, given by (3.27), simplify to

$$^1\pi_t = \Pr\left(X_t, Y_{t-1}, S_{t-1}, M_{t-1} \,\middle|\, ^1\varphi^{t-1}\right) \tag{3.32a}$$

$$^2\pi_t = \Pr\left(X_t, Y_{t-1}, Z_t, S_{t-1}, M_{t-1} \,\middle|\, ^2\varphi^{t-1}\right) \tag{3.32b}$$

$$^3\pi_t = \Pr\left(X_t, Y_t, S_t, M_{t-1} \,\middle|\, ^3\varphi^{t-1}\right) \tag{3.32c}$$

$$^4\pi_t = \Pr\left(X_t, Y_t, \hat{X}_t, S_t, M_{t-1} \,\middle|\, ^4\varphi^{t-1}\right) \tag{3.32d}$$

The time evolution of the information states (Lemma 2.1) simplify further due to the noiseless nature of $\tilde{h}_t$. Optimal encoding and decoding strategies can be determined by the nested optimality equations of Theorem 2.1.

Infinite horizon problem

For the infinite horizon problem we can consider variations V1, V2, and V3 as in model R4; optimal encoding and decoding strategies are given by the fixed point equations for variations V1, V2, and V3 derived in Chapter 2. For variation V2, a further simplification can be made. Agent 1 (the encoder) perfectly observes the state of the plant (the Markov source and the output of the channel) and consequently observes the observations of agent 2 (the receiver) after one unit of delay. As a result, agent 1 knows the state of the agent 2. Consequently, the belief of agent 1 (see Definition 2.6) simplify to

$$ {}^{i}B_t^1 = \Pr\left(X_t, \tilde{Z}_{t-1}, M_{t-1} \,\middle|\, {}^{i}\mathfrak{I}_t^1\right) \equiv (X_t, Y_{t-1}, M_{t-1}). $$

Therefore, the structural results of variation v2 (Theorem 2.5) imply that there is no loss of optimality to restrict attention

$$ Z_t = c_t(X_t, Y_{t-1}, M_{t-1}). \tag{3.33} $$

Hence, for variation v2, the information states simplify to

$$ {}^{1}\pi_t = \Pr\left(X_t, Y_{t-1}, M_{t-1} \,\middle|\, {}^{1}\varphi^{t-1}\right) \tag{3.34a} $$

$$ {}^{2}\pi_t = \Pr\left(X_t, Y_{t-1}, Z_t, M_{t-1} \,\middle|\, {}^{2}\varphi^{t-1}\right) \tag{3.34b} $$

$$ {}^{3}\pi_t = \Pr\left(X_t, Y_t, M_{t-1} \,\middle|\, {}^{3}\varphi^{t-1}\right) = {}^{2}\pi_t \tag{3.34c} $$

$$ {}^{4}\pi_t = \Pr\left(X_t, Y_t, \hat{X}_t, M_{t-1} \,\middle|\, {}^{4}\varphi^{t-1}\right) \tag{3.34d} $$

This leads to the corresponding simplification of the nested optimality equations of Theorem 2.1.

3.6 Comparison with the philosophy of information theory and coding theory

This chapter takes a drastically different approach to the design of a communication system than the traditional approach of information theory and coding theory. In this section we explain the reason for taking this different approach; we also explain the step that needs to be added to our approach in order to provide a complete solution methodology to determining good communication strategies for real-time communication systems.

The objective of the design of a communication system is to find communication strategies that perform nearly optimally and are easy to implement. For communication systems with no restriction on communication delay, information theory and coding theory break down the design of a communication system into two steps:

1. First, information theory is used to determine the fundamental limits of performance of a communication system.

2. Then, coding theory investigates codes that are easy to implement and perform close to the fundamental performance limits determined by information theory.

This approach works even for communication systems with finite but sufficiently large delay constraints. However, this approach fails for communication systems with small delay constraints because for information theoretic bounds are not tight for small values of delay and consequently, fundamental limits of performance are not known. As a result, there is no benchmark for performance evaluation of communication strategies, and we cannot determine whether or not a particular family of codes performs close to optimal.

Given the current state of knowledge, one can take two approaches to the design of real-time communication systems: either determine tight bounds on optimal performance (and then find codes that come close to those bounds), or use some other technique to find good codes. In this chapter we follow the second approach. We formulate the real-time communication problem as a decentralized stochastic optimization problem and develop a methodology to systematically search for an optimal communication strategy. This methodology drastically simplifies the search for an optimal solution. In spite of this simplification, numerically solving the resultant optimality equations is a formidable task. As explained in Chapter 2, for variation v1 we can efficiently approximate the optimal solution; for variation v2 and v3, we are not aware of any good approximation techniques. If such approximation techniques are discovered, only then would the results of this chapter along with those techniques provide a complete methodology to determining communication strategies that perform well for small delays.

3.7 Conclusion

We considered two models of point-to-point real-time communication systems: real-time communication over noisy channels, and real-time communication over noisy channels with noisy feedback. (The other two models of real-time communication, viz., real-time communication over noiseless channels, and real-time communication over noisy channels with noiseless feedback have already been considered in the literature and are special cases of the models considered in this chapter.)

We showed that both these models are special instances of two-agent teams considered in Chapter 2; hence, the results on sequential decomposition of finite and infinite horizon two-agent team problems derived in Chapter 2 are also applicable to real-time communication systems.

The qualitative/structural properties of optimal encoders for variation v2 of models R1, R2, and R3 have been considered in Witsenhausen (1979), Walrand and Varaiya (1983a) and Teneketzis (2006), respectively. In this chapter we showed that the structural properties of general two agent teams can be used to obtain these qualitative properties directly. This shows the usefulness of the structural results for variations v2 and v3 derived in Chapter 2.

Parts of the results of this chapter have appeared in different publications. For model R2, variation v1 was considered in Mahajan and Teneketzis (2006b) and variation v2 along with several extensions was considered in Mahajan and Teneketzis (2006a); for model R4 variation v2 was considered in Mahajan and Teneketzis (2007, 2008).

In models R1 and R3, the decisions of the receiver (agent 2) do not affect the observations of the encoder (agent 1) or the time-evolution of the source. In models R2 and R4, the decisions of the agent affect the observations of the encoder (agent 1), but they do not affect the time-evolution of the source. Thus, in all four cases, the decisions of the receiver do not affect the time evolution of the source. In the next chapter we consider networked control systems where the decisions of agent 2 affect the time-evolution of the plant.

Chapter 4

Optimal feedback control over noisy communication

4.1 Introduction

Motivation

In today's world, networking capabilities—both wired and wireless—are ubiquitous. This has motivated the study of control systems where the plant and the controller are located geographically apart and connected over a network. Such systems, called networked control system (NCS), have given rise to new and interesting problems in control and communications and have spurred considerable research interest including special issues in IEEE control systems magazine (Bushnell, 2001), IEEE transactions on automatic control (Antsaklis and Baillieul, 2004), proceedings of the IEEE (Antsaklis and Baillieul, 2007) and IEEE journal on special areas in communication (Franceshchetti et al., 2008).

In the simplest setup, a NCS consists of a sensor located at a plant and a remotely located controller. The sensor can communicate with the controller over a (possibly noisy) communication channel and the controller can send its actions to the plant over a (possibly noisy) communication channel. Depending on the application, the design objective can either be stability or optimal performance. Stability is an asymptotic concept where we want to ensure that eventually the state of the plant will belong to a safe region. Thus, for stability analysis only the steady state behavior of the system is important. For optimal performance, both the transient and the steady state behavior is important; usually optimal performance problems

assume that an instantaneous cost in incurred at each time step and the objective is to minimize an expected total cost.

Literature overview

Stability analysis of NCS has received considerable attention in the literature. The problem of stabilization of a plant with finite data rate feedback was investigated in Delchamps (1990), Wong and Brockett (1999), Baillieul (2001, 2002), Elia and Mitter (2001), Brockett and Liberzon (2000), Peterson and Savkin (2001), Ishii and Francis (2003), Liberzon (2003), Nair and Evans (2000, 2003, 2004), Nair et al. (2004) and Martins et al. (2004). A unified overview of stabilization with finite data rate feedback is presented in Nair et al. (2007). LQG stability of various systems (deterministic LQ, stochastic, stable, unstable) under various kinds of communication constraints (noisy and noiseless channel) was considered in Tatikonda (2000), Tatikonda and Mitter (2004a, 2004b) and Tatikonda et al. (2004). Stability of an unstable plant over AWGN channel subject to input power constraints was considered in Braslavsky et al. (2005). Fundamental asymptotic limitations of feedback for a linear time invariant plant and arbitrary time-invariant causal feedback were investigated in Martins and Dahleh (2004) and Martins et al. (2005) using an information theoretic formulation.

The problem of optimal performance has received less attention than stabilization in the literature. The problems considered in the literature can be classified on the basis of their plant dynamics (linear or non-linear), the nature of the communication channel (rate-limited noiseless channel or noisy channel), and the information structure (classical or non-classical information structures). Optimal performance of a linear plant with rate-limited noiseless communication channel was considered in Matveev and Savkin (2004): in Matveev and Savkin (2004) the plant disturbance is Gaussian and the controller is memoryless; in Savkin (2006) the plant is undisturbed and the controller has perfect recall. Optimal performance of a linear plant with Gaussian disturbance, either a rate-limited noiseless channel or a Gaussian memoryless channel, and various information structures at the encoder was considered in Tatikonda et al. (2004). Optimal performance of a non-linear plant with a noisy forward channel and noiseless feedback from the output of the channel to the encoder was considered in Walrand and Varaiya (1983b).

Conceptual difficulties

While analyzing stability, most results show a connection between the rate or capacity of the channel and the sum of unstable eigenvalues of the plant. In retrospect, the connection between stability and information theory is not surprising since stability as well as the information theoretic notions of source entropy and channel capacity are asymptotic concepts. This however, is not the case with optimal performance where the asymptotic notions of source entropy and channel capacity, and the asymptotic results on stability are not appropriate.

The most important feature in problems of optimal performance of NCS is whether the encoder/sensor knows the information available at the decoder/controller or not. We can classify problems into two cases on the basis of the presence or absence of this feature: case 1, when the encoder has access to all the information available at the decoder/controller, and case 2, when it does not. In case 1 the problem of determining optimal performance can be reduced to a centralized stochastic control problem from the encoder's point of view. Such a reduction is not possible in case 2. In case 1 the encoder knows how the decoder/controller will interpret its messages; in case 2, it does not. Therefore, efficient communication between the encoder and decoder/controller is easier in case 1 than in case 2. Hence, determining optimal strategies for the encoder and the controller in case 2 is a considerably more difficult problem than in case 1.

The models of Matveev and Savkin (2004) and Walrand and Varaiya (2006, 1983b) and the instances in Tatikonda et al. (2004) where there are noiseless channels as well as the instance of information pattern A (see Tatikonda et al. (2004, pg. 1550) for definition of information pattern A) belong to case 1. In all these situations optimal encoding and control strategies have been determined. The model in Tatikonda et al. (2004) with information pattern B (see Tatikonda et al. (2004, pg. 1550) for definition of information pattern B) belong to case 2. In this situation only sub-optimal encoding and control strategies have been proposed. Thus, the optimal strategies for case 2 are not known.

Outline of the approach

In this chapter we consider a non-linear plant with a noisy communication channel. Our model belongs to case 2. We study the simplest NCS—a network with only two nodes with a noisy communication link between them. We show that optimal

performance of NCS is a special case of two-agent team considered in Chapter 2. As such, the results of Chapter 2 can be used to obtain optimal communication and control strategies for the NCS under consideration.

4.2 Optimal performance of NCS

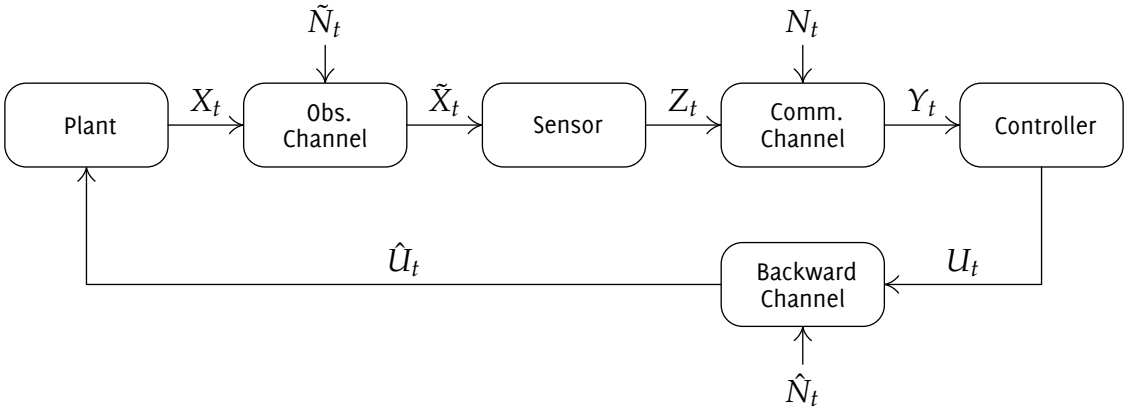


Figure 4.1: A simple two-node networked control system

Problem formulation

Consider a system, shown in Figure 4.1, which operates in discrete time for a finite horizon T. For the ease of exposition, we assume that all system variables are finite valued. The state of the plant at time t is X_t and takes values in $\mathcal{X}$. The initial state X_1 has a PMF P_{X_1}. The plant evolves as follows:

$$X_{t+1} = f_t(X_t, \tilde{U}_t, W_t) \tag{4.1}$$

where f_t denotes the ***plant function*** and $\tilde{U}_t \in \tilde{\mathcal{U}}$ denotes the control input to the plant (which may be different from the control action taken by the controller) and $W_t \in \mathcal{W}$ denotes the plant disturbance. The PMF of W_t is given by P_{W_t}. We assume that $\{W_t, t = 1, \dots, T\}$ is a sequence of independent random variables that are also independent of X_1.

A sensor is co-located with the plant and observes the state of the plant in a noisy manner. The observations $\tilde{X}_t$ of the sensor take values in $\tilde{\mathcal{X}}$ and are generated according to

$$\tilde{X}_t = \tilde{h}_t(X_t, \tilde{N}_t) \tag{4.2}$$

where $\tilde{h}_t$ denotes the observation channel and $\tilde{N}_t \in \tilde{\mathcal{N}}$ denotes the observation noise. The PMF of $\tilde{N}_t$ is given by $P_{\tilde{N}_t}$. We assume that $\{\tilde{N}_t, t = 1, \dots, T\}$ is a sequence of independent random variables that is also independent of $\{W_t, t = 1, \dots, T\}$ and X_1.

The sensor encodes its observations and transmits the encoded symbol over a noisy communication channel to the controller. The sensor has a memory $\mathcal{M}_t$ at time t. The size of the sensor's memory can increase with time; so this model includes the case when the sensor has perfect recall. The encoded symbol Z_t takes values in $\mathcal{Z}$ and is generated as follows

$$Z_t = c_t(\tilde{X}_t, M_{t-1}) \tag{4.3}$$

where c_t is the ***encoding function*** of the sensor. The sensor then updates its memory according to

$$M_t = d_t(\tilde{X}_t, M_{t-1}) \tag{4.4}$$

where d_t is the sensor's ***memory update function***.

The encoded symbol Z_t is transmitted over a $|\mathcal{Z}|$-input $|\mathcal{Y}|$-output discrete memoryless channel to produce a channel output Y_t according to

$$Y_t = h_t(Z_t, N_t) \tag{4.5}$$

where h_t denotes the communication channel between the sensor and the controller, and N_t denotes the channel noise. The PMF of N_t is P_{N_t}. We assume that $\{N_t, t = 1, \dots, T\}$ is independent of $\{W_t, \tilde{N}_t, t = 1, \dots T\}$ and X_1.

The controller observes Y_t and takes a control action $U_t \in \mathcal{U}$ as follows:

$$U_t = g_t(Y_t, S_{t-1}) \tag{4.6}$$

where g_t is the ***control law*** and S_{t-1} denotes the controller's memory contents at time $t - 1$. S_t takes values in $\mathcal{S}_t$. The size of the controller's memory can increase with time; so, this model includes the case when the controller has perfect recall. The controller's memory is updated according to

$$S_t = l_t(Y_t, S_{t-1}) \tag{4.7}$$

where l_t is the ***controller's memory update rule***.

The control action U_t is transmitted over a backward channel which is a $|\mathcal{U}|$-input $|\hat{\mathcal{U}}|$-output DMC and generates an output $\hat{U}_t \in \hat{\mathcal{U}}$ as follows

$$\hat{U}_t = \hat{h}_t(U_t, \hat{N}_t) \tag{4.8}$$

where $\hat{h}_t$ denotes the backward communication channel and $\hat{N}_t$ denotes the channel noise. The PMF of $\hat{N}_t$ is given by $P_{\hat{N}_t}$. We assume that $\{N_t,\ t = 1, \dots, T\}_t$ is a sequence of independent random variables that are also independent of $\{W_t, \tilde{N}_t, N_t,\ t = 1, \dots T\}$ and X_1.

The output of the backward channel acts as a control input to the plant and the state of the plant gets updated according to (4.1). At each instant of time an instantaneous cost $\rho_t(X_t, Z_t, U_t)$ is incurred.

The collection (C, D, G, L), where $C \coloneqq (c_1, \dots, c_T)$, $D \coloneqq (d_1, \dots, d_T)$, $G \coloneqq (g_1, \dots, g_T)$, and $L \coloneqq (l_1, \dots, l_T)$, is called a ***design*** of the system. The performance of a design is quantified by the expected total cost under that design and is given by

$$\mathcal{J}_T(C, D, G, L) \coloneqq \mathbb{E}\left\{ \sum_{t=1}^{T} \rho_t(X_t, Z_t, U_t) \,\middle|\, C, D, G, L \right\} \tag{4.9}$$

We are interested in the following optimization problem.

Problem 4.1. *Assume that the sensor and the controller know the plant function f_t, the observation, forward and backward channels $\tilde{h}_t$, h_t and $\hat{h}_t$, respectively, the statistics of plant disturbance and noise in the observation, forward, and backward channels, and the distortion function ρ_t. Determine a design (C^*, D^*, G^*, L^*) that is optimal with respect to the performance criterion of* (4.9)*, i.e.,*

$$\mathcal{J}_T(C^*, D^*, G^*, L^*) = \mathcal{J}_T^* \coloneqq \min_{\substack{C \in \mathscr{C}^T \\ D \in \mathscr{D}^T \\ G \in \mathscr{G}^T \\ L \in \mathscr{L}^T}} \mathcal{J}_T(C, D, G, L), \tag{4.10}$$

where $\mathscr{C}^T \coloneqq \mathscr{C}_1 \times \cdots \times \mathscr{C}_T$; $\mathscr{C}_t$ is the family of functions from $\mathcal{X}_t \times \mathcal{S}_{t-1}$ to $\mathcal{Z}$; $\mathscr{D}^T \coloneqq \mathscr{D}_1 \times \cdots \times \mathscr{D}_T$; $\mathscr{D}_t$ is the family of functions from $\mathcal{X}_t \times \mathcal{S}_{t-1}$ to $\mathcal{S}_t$; $\mathscr{G}^T \coloneqq \mathscr{G}_1 \times \cdots \times \mathscr{G}_T$; $\mathscr{G}_t$ is the family of functions from $\mathcal{Y}_t \times \mathcal{M}_{t-1}$ to $\hat{\mathcal{X}}$; $\mathscr{L}^T \coloneqq \mathscr{L}_1 \times \cdots \times \mathscr{L}_T$; and $\mathscr{L}_t$ is the family of functions from $\mathcal{Y}_t \times \mathcal{M}_{t-1}$ to $\mathcal{M}_t$.

Salient features of the model

The above described model captures the realistic features of NCS. In particular,

1. The plant can be non-linear and its evolution can be stochastic.
2. The sensor observes the state of the plant in a noisy manner which is usually the case in practice.
3. The sensor can have a small memory; this models the situation when the sensor is a low-complexity device.
4. The forward channel is noisy; this models the situation when the sensor has limited power and therefore must communicate with limited power.
5. Encoding and control are done in a causal manner.
6. The instantaneous cost depends on the encoded symbol. This can be used to model average power constraint on the communication channel.

Reduction to the model of Chapter 2

We can reduce the NCS modeled above into the two-agent team considered in Chapter 2, Section 2.1 by assuming that the backward channel is part of the plant. That is, we can combine (4.1) and (4.8) to give

$$\begin{aligned} X_{t+1} &= f_t(X_t, \hat{h}_t(U_t, \hat{N}_t), W_t) \\ &=: \hat{f}_t(X_t, U_t, (\hat{N}_t, W_t)) \end{aligned} \tag{4.11}$$

The system given by (4.11) and (4.2)—(4.7) is a special case of the model of the two agent team of Chapter 2 in which the sensor is the first agent, the controller is the second agent, and the plant update does not depend on the action of agent 1. The relation between the variables of the two models is shown in Table 4.1.

Information states and global optimization

Substituting the above described reduction in Definition 2.3, we get that the information states for this model of NCS are given by

$${}^1\pi_t \coloneqq \Pr\left(X_t, \tilde{X}_t, M_{t-1}, S_{t-1} \,\middle|\, {}^1\varphi^{t-1}\right), \tag{4.12a}$$

$${}^2\pi_t \coloneqq \Pr\left(X_t, \tilde{X}_t, Z_t, M_{t-1}, S_{t-1} \,\middle|\, {}^2\varphi^{t-1}\right), \tag{4.12b}$$

$${}^3\pi_t \coloneqq \Pr\left(X_t, Y_t, Z_t, M_t, S_{t-1} \,\middle|\, {}^3\varphi^{t-1}\right), \tag{4.12c}$$

$${}^4\pi_t \coloneqq \Pr\left(X_t, Y_t, Z_t, U_t, M_t, S_{t-1} \,\middle|\, {}^4\varphi^{t-1}\right). \tag{4.12d}$$

Component	Variable	Two-agent team	Model R2
Plant	State	X_t	X_t
Agent 1	Observation	Y_t^1	$\tilde{X}_t$
	Control action	U_t^1	Z_t
	State	S_t^1	M_t
Agent 2	Observation	Y_t^2	Y_t
	Control action	U_t^2	U_t
	State	S_t^2	S_t

Table 4.1: The simple NCS model as an instance of two-agent team. In the NCS model, the plant and the backward channel corresponds to the plant of two-agent team, the sensor corresponds to agent 1, and the controller corresponds to agent 2.

The time evolution of these information states is given by Lemma 2.1; the transformations ${}^1Q_t$, ${}^2Q_t$ and ${}^3Q_t$ and the cost $\hat{\rho}_t$ in Lemma 2.1 can be simplified by incorporating the above described reduction in their definitions. Consequently, the sequential decomposition of Problem 3.1 is given by the nested optimality equations of Theorem 2.1.

Infinite horizon problems

We can consider the following three variations of the time-homogeneous infinite-horizon version of the NCS model:

v1. Both the sensor and the controller have finite memory;
v2. The encoder has perfect recall while the controller has finite memory;
v3. The encoder has finite memory while the controller has perfect recall.

These are equivalent to the variations of the two-agent teams considered in Chapter 2. Thus, the fixed-point equations for the two infinite horizon criteria derived for variations v1, v2, and v3 in Chapter 2 are also applicable to the corresponding variations of the networked control system modeled above.

4.3 Conclusion

We considered a simple networked control system: a sensor co-located with a plant that communicates over a noisy channel to a remote controller. We showed that this system is a special instance of the two-agent team considered in Chapter 2; hence, the results on sequential decomposition of finite and infinite horizon two-agent team problems derived in Chapter 2 are also applicable to networked control system. Results of this chapter for variation v3 have appeared in Mahajan and Teneketzis (2006c, 2006d).

Chapter 5

Conclusion

In this thesis we studied two-agent teams with strictly non-classical information structures and showed that an appropriate choice of an information state results in a sequential decomposition of the problem. We described the properties that appropriate information states should satisfy and provided some guidelines on how to identify information states with these properties. A sequential decomposition converts a one shot optimization problem into a sequence of nested optimization problems. For some instances of the problem, this sequential decomposition provides computationally tractable approximate solutions; for others, it provides a potential approach for obtaining efficient computational methods. The sequential decomposition can also potentially help in identifying additional qualitative properties of optimal designs. We also considered real-time communication and networked control systems and showed that they can be considered as special instances of two-agent teams. The results of two-agent teams can be used to provide a solution methodology for these applications.

In this chapter we conclude with some reflections on the solution framework developed in this thesis and some possible future directions.

5.1 Reflections

Decentralized teams have been studied since the early '60's but so far there was no solution framework to obtain a sequential decomposition for both finite and infinite horizon problems. This thesis provides such a solution framework. We highlight the philosophy of thinking and the modeling assumptions that enabled us to obtain a solution framework, and some of the strengths and weaknesses of

our solution framework. In this section, we reflect on the conceptual and practical difficulties associated with sequential decomposition of decentralized teams. We begin by explaining how we resolve the main conceptual difficulty in the design of decentralized teams

The philosophy of designing decentralized teams — resolving the second guessing argument

In centralized system with partial observations (POMDPs), the control action is based on the control agent's belief about the rest of the system. In decentralized systems, the control action of an agent cannot be based on its belief since other agents observe different data. Each agent can form a belief on other agents' data, but then each agent will not know the other agents' belief on its own data. If each agent forms a belief on the other agents' belief on its own data, it will not know the other agents' beliefs on its belief on the other agents' data. This process of forming the belief on what the other agents are "thinking" will continue until the agents agree on what everyone is thinking. This process is called the ***second guessing argument***,[3] and it has been the main conceptual difficulty in designing decentralized dynamic systems.

Aumann (1976) showed that two agents with inconsistent beliefs can only agree on the ***common knowledge*** between them. Thus, the solution of the second guessing argument will be the common knowledge between the agents. We resolve the second guessing argument by starting from the common knowledge between two agents and coarsening it to come up with appropriate information states.

In conclusion, optimal design of decentralized teams requires a paradigm shift from the philosophy of optimal design of centralized dynamic systems. ***Instead of reasoning in terms of the information available to individual agent, we must reason in terms of the information that is common knowledge to all agents.***

In the next section we revisit the standard form of Witsenhausen (1973), which provided a sequential decomposition for finite horizon sequential teams.

Comparison with Witsenhausen's standard form

Witsenhausen (1973) proposed a standard form for sequential stochastic control. He showed that *any* sequential team can be converted into a standard form and

[3] The term "second guessing argument" is due to Hans Witsenhausen.

developed a solution methodology to obtain a sequential decomposition of any problem in standard form; thereby providing a solution methodology for obtaining a sequential decomposition of *any* sequential team. To the best of our knowledge, this was the first and so far the only result that shows how to obtain a sequential decomposition of *any* sequential team problem.

For finite horizon problems, the model of two-agent team considered in this thesis can be considered as a special case of the model of standard form. Compared Witsenhausen's model (1973), which does not make any assumptions other than assuming sequentiality and a common objective, our model makes the following "simplifying" assumptions:

A1. The system has two agents that act cyclically (that is first agent 1 acts, then agent 2, then agent 1, and so on). In Witsenhausen's model (1973), each action is assumed to be taken by a different agent.

A2. In our model, the state X_t of the system is a controlled Markov process controlled by past control actions U^1_{t-1} and U^2_{t-1} (this follows from the plant update equation (2.1)). Witsenhausen (1973) does not make a Markov assumption on the state evolution of the system.

A3. In our model the primitive random variables of the system are independent. Witsenhausen (1973) assumes that all randomness is due to one intrinsic random variable (or intrinsic event); no assumption is made regarding the probability space on which the intrinsic random variable is defined. In our model, this would be equivalent to assuming that the primitive random variables are correlated.

A4. In our model the cost is additive; at each time an instantaneous cost, which depends on the current state of the system and the current control actions of the agents, is incurred. The total cost is the sum of instantaneous costs for the entire horizon. Witsenhausen (1973) does not assume an instantaneous cost; rather assumes that a terminal cost, which depends on the intrinsic random variable and all the control actions taking in the entire horizon, is incurred at the end of the horizon.

In (A1) assuming that there are only two agents is a restriction (although our framework could be extended to multiple agents to remove this restriction). Assuming

that the agents act cyclically is not a restriction; any two-agent sequential systems can be made cyclic by introducing dummy actions for the agents.

Assumptions (A2) and (A3) are also not real restrictions. Any dynamic system can be made to satisfy these assumptions by a suitable expansion of the state spaces.

The real restriction in our model is (A4). However, assumption A4, which is a standard assumption in Markov decision theory, is a mild assumption; most applications satisfy this assumption. This assumption buys us the following two simplifications:

1. Our information states take values in a smaller space as compared to the information states of the standard form. Due to (A2) and (A4), in our model the information states should be sufficient to determine a measure on the current state of the system and the current control actions of the agents; in the standard form the information states should be sufficient to determine a measure on the intrinsic random variable and all the past control actions of all agents.
2. In our model we can formulate and solve infinite horizon problems. However, Witsenhausen's model (1973) cannot be extended to infinite horizon due to the assumption of a terminal cost.

Nevertheless, Witsenhausen (1973) was the first to provide a solution framework for optimally designing finite-horizon decentralized sequential teams. Surprisingly, the result was virtually ignored in the literature. Most publications that appeared in the literature in the subsequent 35 years assumed that a general sequential team cannot be solved. We believe that this was due to two reasons. Firstly, Witsenhausen (1973) was a difficult and tersely written paper; there was no explanation of *why* we are able to obtain a sequential decomposition of a system in standard form. Secondly, and perhaps more importantly, the computational solution of the nested optimality equations of Witsenhausen (1973) was not explored in any subsequent paper. As a result, it was difficult to appreciate the practical significance of the sequential decomposition proposed by Witsenhausen.

In this thesis, we have shown how existing computational methods for POMDPs could be used for solving the nested optimality equations that arise in our models. Next we present the salient features of the numerical results.

Numerical solution of the optimality equations

As mentioned earlier, two-agent teams can be thought of as POMDPs where the state sufficient for input-output mapping is the unobserved state of the POMDP and the control and state-update functions are the control actions of the POMDP. Thus, we can leverage on the existing computational techniques for POMDPs to numerically solve the optimality equations of two-agent teams. However, there is one fundamental difference between the optimality equations of POMDPs and those of two-agent teams. Each step of the nested optimality equations is a parametric (scalar) optimization problem in POMDPs while it is a functional optimization problem in two-agent teams. This difference increases the complexity of solving the optimality equations of two-agent teams using the computational techniques for POMDPs.

Suppose that all the system variables are finite valued. Then finite horizon problems and variation v1 of infinite horizon problem for two-agent teams are equivalent to POMDPs with finite state and action spaces. Such POMDPs can be solved efficiently for both finite and infinite horizon (see Cassandra et al. (1997) and Rust (1997)). The solution algorithms for POMDPs are polynomial in the size of the state and action spaces. When we look at two-agent teams as POMDPs, the action space is a space of functions; although it is finite, its size is exponential in the system variables. Therefore, even good computational algorithms for POMDPs may not be efficient for two-agent teams with large alphabets.

Variations v2 and v3 of the infinite horizon problems for two-agent teams are equivalent to POMDPs with uncountable state and action spaces. There are not many efficient algorithms to numerically solve such POMDPs.

Limitations of the existing computational techniques motivate the study of numerical algorithms for POMDPs (and MDPs) with specific features that are present in optimality equations of two-agent teams. These features include the deterministic evolution of the information states, lack of observations by the "controller" (or the designer), and the optimality equation at each step being a functional optimization problem. It is possible that some of the computational techniques for deterministic *open loop* optimization problems (e.g., traveling salesman or shortest path in networks) could be extended to the case of uncountable state space and be useful for computationally solving optimality equations of two- and multi-agent teams.

There are two sources of difficulty in the numerical solution of the optimality equations of two-agent teams. The first is the functional rather than parametric nature of the each step of the nested optimality equations. The second is the uncountable nature of the state for input-output mapping that arise in variations v2 and v3 of the infinite horizon problems. In the next two sections, we take a critical look at the reason behind these difficulties.

The communication aspect of control

We have shown that an appropriate choice of an information state converts a decentralized team into a MDP. However, we end up with optimality equations where each step is a functional optimization problem. This makes the numerical solution of the optimality equations more complicated. The reader may wonder if we could have chosen a different information state to come up with a sequential decomposition where each step was a parametric optimization problem. We believe that such an information state cannot be found due to the communication aspect of control.

The fundamental difficulty in communication between two agents in a decentralized system is the following. Consider a time instant when agent 1 communicates with agent 2 over a noisy communication channel. In order to determine the effect of his control action of the instantaneous cost, agent 2 needs to form a belief on agent 1's data; to form such a belief, he needs to know control actions that agent 1 would have chosen for all possible values of his (agent 1's) data. Thus, ***agent 2 needs to know agent 1's control law.*** Therefore, a sequential decomposition for a decentralized team with strictly non-classical information structure will always result in functional optimization problems. Depending on the specifics of the model it may be possible to obtain simpler optimality equations. For example, see Walrand and Varaiya (1983a) and Mahajan et al. (2008).

In the next section we explain the modelling assumptions that lead to uncountable state spaces in variations v2 and v3.

The assumption of perfect recall

Perfect recall at an agent is an impractical assumption. In single-agent stochastic control problems, perfect recall at the controller implies that the system has classical information structure, which makes the analysis of the problem simpler. However, in two-agent (or multi-agent) teams with noisy channels between the agents,

assuming perfect recall at each agent does not imply a classical information structure. Further, it makes the problem analytically and computationally more difficult. For example, compare variations v1 and v2. In variation v1, we assume that both agents have fixed finite memory; in variation v2, we assume that agent 1 has perfect recall. The analysis of variation v2 is more difficult than that of variation v1. In variation v2, we need to first establish qualitative properties of optimal control laws of agent 1 which transform the model of variation v2 into one where the size of the state space of agent 1 is not changing with time. In contrast, in variation v1 we start with an assumption that the state space of both agents is not changing with time. Further, the optimality equations of variation v1 are equivalent to those of a POMDP with finite (unobserved) state and action space while those of variation v2 are equivalent to a POMDP with uncountable (unobserved) state and action spaces. So, when we go from the realistic assumption of fixed finite memory at both agents (variation v1) to unrealistic assumption of perfect recall at agent 1 (variation v2) or at agent 2 (variation v3), the problem becomes harder to analyze and harder to solve computationally. Further, when we assume perfect recall at both agents (variation v4) we do not know how to solve a general infinite horizon problem. ***Therefore, we must carefully reconsider the modeling assumptions of single-agent centralized stochastic control problems before adopting them for multi-agent decentralized team problems.***

In the next section, we mention some possible research problems that can be considered in the future, based on the results of this thesis.

5.2 *Future Directions*

As mentioned earlier, there are two difficulties associated with multi-agent teams with strictly non-classical information structures: conceptual and computational. This thesis helps resolve some of the conceptual difficulties and suggests a few possibilities of resolving the computational difficulties. However, many computational and conceptual difficulties remain unresolved. We highlight some of them in this section.

Scalable sequential decomposition for multi-agent teams

In this thesis we have focussed on two-agent teams. The intuition behind the choice of information states presented in Section 2.8 make it evident that the idea can be easily extended to multi-agent team problems. We need to find a state sufficient for input output mapping, and consider the joint measure on this state as the information state. This will also work for infinite horizon problems when all the agents have finite memory. However, the complexity of solving the optimality equations of multi-agent systems would increase exponentially with the number of agents. So, the solution methodology presented in this chapter would not scale with the number of agents. We believe that, in general, it is not possible to obtain sequential decomposition that scale well with the number of agents. Nevertheless, it should be possible to obtain a scalable solution methodologies under certain modelling assumptions. Two such possibilities are systems with symmetric agents and systems with asymptotically large number of agents. For both these systems, it should be possible to come up with a more compact representation of the state sufficient for input-output mapping, and thereby a more compact representation of the information states. It would be worthwhile to investigate the above-mentioned or any other modelling assumptions that would lead to a sequential decomposition that is scalable with the number of agents.

Structural properties for multi-agent systems

To obtain a sequential decomposition for infinite horizon teams, we need to have a time-invariant state for each agent. In the four instances of infinite horizon problems that we considered for two-agent teams, variation v1 has a time-invariant state for both agents, variations v2 and v3 has a time-invariant state for one agent, and variation v4 does not have a time-invariant state for any agent. For variations v2 and v3 we obtained structural/qualitative properties of optimal control laws for the agent that does not have a time-invariant state; these structural properties provided a time-invariant state representation of optimal control strategies for these agents. However, in general, it is not easy to obtain a structural properties for a general multi-agent system. A cast in point is variation v4 of two-agent teams where we could not obtain structural properties for either agent. It will be useful to find examples of multi-agent systems where we can obtain structural properties of optimal designs and use these examples to determine general models of multi-agent

systems where structural properties of optimal control laws can be obtained. One such is example appears in Nayyar and Teneketzis (2008) who consider a three-agent team that arises in real-time communication and obtain structural properties of optimal control strategies for all agents; these structural properties are used to obtain a sequential decomposition of the optimization problem associated with the three agent team.

Logical systems with minimax cost criterion

In this thesis, we considered stochastic systems where the performance criterion is given by an expected cost. Dynamic systems can also be modelled as logical systems with a (worst-case) minimax cost criterion. For centralized systems, Markov decision theory also provides a sequential decomposition for logical systems. We believe that it is possible to extend the results of this thesis, at least for finite horizon problem and variation v1 of infinite horizon problems, to logical systems. This is because the key concepts of our solution framework—the notion of information state and common knowledge—are fundamental ideas in dynamic systems that are independent of the modelling framework. In logical systems the information states would be the reachable set of the state sufficient for input-output mapping.

Class of problems with parametric information states

It is important to identify special cases in which the information states can be restricted to a parametric family of distributions. In centralized stochastic control problems, LQG systems possess such a property—the information state can be restricted to Gaussian distributions. This is because in LQG systems with classical information pattern, without any loss of optimality we can restrict attention to affine control laws, which implies that the state of the plant is always Gaussian. Thus, the information state—which is the conditional probability of the state of the plant, conditioned on all the past observations and all the past control actions of the controller—is also Gaussian and can be characterized only by its mean and variance (which is data independent). This simplifies the search for an optimal design. Unfortunately, in decentralized systems non-linear control laws can outperform affine control laws even in linear systems where all primitive random variables are Gaussian, as illustrated by the Witsenhausen counterexample (Witsenhausen, 1968). So, the state of the plant may not be Gaussian and hence the information state need

not be Gaussian. However, there may be other special cases for which information states in a decentralized system belong to a parametric family of distributions. Finding such special cases remains a challenging open problem.

5.3 *Final thoughts*

For two-agent teams this thesis has resolved all conceptual difficulties except the case of infinite horizon problem when both agents have perfect recall. For multi-agent teams finding structural properties of optimal control laws remains a challenge.

From a practical view-point there are two possible directions. One direction is to find bounds for optimal performance, which can be evaluated without identifying an optimal design. For centralized systems some initial results were presented in Witsenhausen (1966, 1970). Such bounds will make it easier to bound the degree of suboptimality of a heuristic policy. Another possible direction is to find good suboptimality algorithms where the degree of suboptimality can be bound.

For decentralized systems with large number of agents modular and hierarchical control architectures may not be desirable. To the best of our knowledge, the optimal design of modular and/or hierarchical architectures for sequential dynamic teams is still an open problem.

In general, decentralized systems is an intellectually stimulating area which is of great practical importance. We believe it will remain an exciting and challenging research field for years to come.

References

Andersland, M. S.,
1991. "Decoupling non-sequential stochastic control problems", in *Systems & Control Letters*, Volume 16, Number 1, pages 65–69.

Andersland, M. S. and D. Teneketzis,
1992. "Information structures, causality, and nonsequential stochastic control I: Design-independent properties", in *SIAM Journal on Control and Optimization*, Volume 30, Number 6, pages 1447–1475.

1994. "Information structures, causality, and nonsequential stochastic control. II: Design-dependent properties", in *SIAM Journal on Control and Optimization*, Volume 32, Number 6, pages 1726–1751.

Antsaklis, P. J. and J. Baillieul, editors
2004. *IEEE Transactions on Automatic Control: Special Issue on Networked Control Systems*, Volume 49, Number 9.

2007. *Proceedings of the IEEE: Specical issue on Technology of Networked Control Systems*, Volume 95, Number 1.

Arapostathis, A., V. S. Borkar, E. Fernandez-Gaucherand, M. K. Ghosh and S. I. Marcus,
1993. "Discrete-time controlled Markov processes with average cost criterion - a survey", in *SIAM Journal of Control and Optimization*, Volume 31, Number 2, pages 282–344.

Aumann, R. and S. Hart, editors
1992. *Handbook of Game Theory with Economic Applications* Number 1. Elsevier.

1994. *Handbook of Game Theory with Economic Applications*, Volume 2. Elsevier.

2002. *Handbook of Game Theory with Economic Applications*, Volume 3. Elsevier.

Aumann, R. J.,
1976. "Agreeing to disagree", in *Annals of Statistics*, pages 1236-39.

Baillieul, J.
2001. Feedback desings in information based control. In Pasik-Duncan, B., editor, *Proc. Workshop in Stochastic Theory and Control*, pages 35-27. Springer Verlag.

2002. Feedback coding for information based control — operating near the data-rate limit. In *Proceedings of 41st IEEE Conference on Decision and Control*, pages 3229-3236.

Basar, T. and R. Bansal,
1989. "Simultaneous design of measurement channels and control stategies for stochastic systems with feedback", in *Automatica*, Volume 25, pages 679-694.

1994. "Optimal design of measurement channels and control policies for linear-quadratic stochastic systems", in *European Journal of Operational Research*, Volume 93, pages 226–236.

Bernstein, D. S., S. Zilberstein and N. Immerman
2000. The complexity of decentralized control of Markov decision processes. In *Proceedings of the 16th International Conference on Uncertainty in Artificial Intelligence (UAI)*, pages 32-27, Stanford, CA.

Blondel, V. D. and J. N. Tsitsiklis,
2000. "A survey of computational complexity results in systems and control", in *Automatica*, Volume 36, Number 9, pages 1249-1274.

Borkar, V. S., S. K. Mitter and S. Tatikonda,
2001. "Optimal sequential vector quantization of Markov sources", in *SIAM Journal of Optimal Control*, Volume 40, Number 1, pages 135–148.

Braslavsky, J. H., R. H. Middleton and J. S. Freudenberg,
2005. "Feedback stabilization over signal-to-noise ratio constrained channels", in *submitted to the IEEE Transactions on Automatic Control*.

Brockett, R. W. and D. Liberzon,
2000. "Quantized feedback stabilization of linear systems", in *IEEE Transactions on Automatic Control*, Volume 45, pages 1279-1289.

Bushnell, L. G., editor
2001. *IEEE Control Systems Magazine: Special Section on Networks and Control*, Volume 21, Number 1.

Cassandra, A., M. L. Littman and N. L. Zhang
1997. Incremental pruning: A simple, fast, exact method for partially observable Markov decision processes. In *Proceedings of the Thirteenth Conference on Uncertainty in Artificial Intelligence*.

Chu, K.-C.,
1972. "Team decision theory and information structures in optimal control problems–part II", in *Automatic Control, IEEE Transactions on*, Volume 17, Number 1, pages 22–28.

Delchamps, D. F.,
1990. "Stabilizing a linear system with quantized state feedback", in *IEEE Transactions on Automatic Control*, Volume 35, pages 916-924.

d'Epenoux, F.,
1960. "Sur un prolème de production et de stockage dans l'aléatorie", in *Française de Recherche Opérationnelle*, Volume 4, Number 14.

Devore, J.,
1974. "A note on the observation of a Markov source through a noisy channel", in *IEEE Transactions on Information Theory*, Volume 20, Number 6, pages 762–764.

Drake, A. W.
1962. Observation of a Markov process through a noisy channel. Sc.D. Thesis, Department of EE, MIT, Cambridge, MA.

Dynkin, E. B. and A. A. Yushkevich,
1975. *Controlled Markov Processes* A Series of Comprehensive Studies in Mathematics. Springer-Verlag.

Elia, N. and S. K. Mitter,
2001. "Stabilization of linear systems with limited information", in *IEEE Transactions on Automatic Control*, Volume 46, pages 1384-1400.

Franceshchetti, M., T. Javidi, P. R. Kumar, S. K. Mitter and D. Teneketzis, editors
2008. *IEEE Journal on Selected Areas in Communications: Control and Communication*, Volume 26, Number 4.

Gaarder, N. T. and D. Slepian
1979. On optimal finite-state digital transmission systems. In *International Symposium on Information Theory*.

1982. "On optimal finite-state digital transmission systems", in *IEEE Transactions on Information Theory*, Volume 28, Number 2, pages 167–186.

Gihman, I. I. and A. V Skorohod,
1979. *Controlled Stochastic Processes* Springer-Verlag.

Gorbunov, A. K. and P. S. Pinsker,
1973. "Non-anticipatory and prognostic epsilon entropies and message generation rates", in *Problems in Information Transmission*, Volume 9, pages 1840191.

1974. "Prognostic epsilon entropy of a Gaussian message and a Gaussian source", in *Problems in Information Transmission*, Volume 10, pages 184-191.

Gyorgy, A., T. Linder and G. Lugosi,
2004. "Efficient adaptive algorithms and minimax bounds for zero-delay lossy source coding", in *IEEE Transactions on Signal Processing*, Volume 52, Number 8, pages 2337–2347.

Ho, Y.-C.,
1980. "Team decision theory and information structures", in *Proceedings of the IEEE*, Volume 68, Number 6, pages 644–654.

Ho, Y.-C. and K.-C. Chu,
1972. "Team decision theory and information structures in optimal control problems–part 1", in *IEEE Transactions on Automatic Control*, Volume 17, Number 1, pages 15–22.

Ho, Y.-C., M. Kastner and E. Wong,
1978. "Teams, signaling, and information theory", in *IEEE Transactions on Automatic Control*, Volume 23, Number 2, pages 305–312.

Ishii, H. and B. A. Francis,
2003. "Quadratic stabilization of sampled-data systems with quantization", in *Automatica*, Volume 39, Number 10, pages 1793-1800.

Karmarkar, N.,
1984. "A new polynomial time algorithm for linear programming", in *Combinatorica*, Volume 4, Number 4, pages 373–395.

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

Liberzon, D.,
2003. "On stabilization of linear systems with limited information", in *IEEE Transactions on Automatic Control*, Volume 48, Number 2, pages 304-307.

Linder, T. and G. Lugosi,
2001. "A zero-delay sequential scheme for lossy coding of individual sequences", in *IEEE Transactions on Information Theory*, Volume 47, Number 6, pages 2533–2538.

Linder, T. and R. Zamir
2001. Causal source coding for stationary sources with high resolution. In *International Symposium on Information Theory*.

Lipster, R. S. and A. N. Shiryayev,
1977. *Statistics of Random Processes, Vol. II:Applications*. Springer-Verlag.

Littman, M. L.
1996. *Algorithms for sequential decision making*. PhD thesis, Brown University.

Lloyd, S. P.,
1977. "Rate versus fidelity for binary source", in *Bell System Technical Journal*, Volume 56, pages 427-437.

Mahajan, A., A. Nayyar and D. Teneketzis
2008. Identifying tractable decentralized control problems on the basis of information structures. In *to appear in proceedings of the 46th Allerton conference on communication, control and computation*.

Mahajan, A. and D. Teneketzis,
2006a. "On globally optimal encoding, decoding and memory update for noisy real-time communication systems", in *submitted to IEEE Transactions on Information Theory*. Available as Control Group Report CGR-06-03, Department of EECS, University of Michigan, Ann Arbor, MI 48109-2122.

2006b. Fixed delay optimal joint source-channel coding for finite-memory systems. In *proceedings of the IEEE International Symposium on Information Theory*, pages 2319-2323, Seattle, WA.

2006c. "Optimal performance of feedback control systems with limited communication over noisy channels", in *submitted to SIAM Journal of Control and Optimization*. Available as Control Group Report CGR-06-07, Department of EECS, University of Michigan, Ann Arbor, MI 48109-2122.

2006d. Optimal performance of feedback control systems with communcication over noisy channels. In *proceedings of the 45th IEEE Conference of Decision and Control*, pages 3228–3235, San Diego, CA.

2007. Real-time communication systems with noisy feedback. In *proceedings of IEEE Information Theory Workshop*, pages 283-287, Lake Tahoe, CA.

2008. "On the design of globally optimal communication strategies for real-time communcation systems with noisy feedback", in *IEEE Journal on Selected Areas in Communications*, Volume 26, Number 4, pages 580-595.

Manne, A. S.,
1960. "Linear programming and sequential decisions", in *Management Science*, Volume 6, Number 3, pages 259–267.

Marschak, J. and R. Radner,
1972. *Economic Theory of Teams*. Yale University Press, New Haven.

Martins, N. C. and M. A. Dahleh,
2004. "Feedback control in the presence of noisy channels: "Bode-like" fundamental limitations of performance", in *submitted to IEEE Transactions on Automatic Control*.

Martins, N. C., M. A. Dahleh and J. C. Doyle,
2005. "Fundamental limitations of feedback in the presence of side information", in *submitted to IEEE Transactions on Automatic Control*.

Martins, N. C., M. A. Dahleh and N. Elia,
2004. "Stabilization of uncertain systems in the presence of a stochastic digital link", in *submitted to IEEE Transactions on Automatic Control*.

Matloub, S. and T. Weissman,
2006. "Universal zero-delay joint source-channel coding", in *IEEE Transactions on Information Theory*, Volume 52, Number 12, pages 5240-5250.

Matveev, A. S. and A. V. Savkin,
2004. "Prolem of LQG optimal control via a limited capacity communication channel", in *System and Control Letters*, pages 51-64.

Merhav, N. and I. Kontoyiannis,
2003. "Source coding exponents for zero-delay coding with finite memory", in *IEEE Transactions on Information Theory*, Volume 49, Number 3, pages 609–625.

Munson, G.
1981. *Causal Information Transmission with Feedback*. Ph.D. Thesis, Department of Electrical Engineering, Cornell University, Ithaca, NY.

Nair, G. N. and R. J. Evans,
2000. "Stabilization with data-rate limited feedback: Tightest attainable bounds", in *Systems Control Letters*, Volume 41, Number 1, pages 304-307.

2003. "Exponential stabilisability of finite-dimensional linear systems with limited data rates", in *Automatica*, Volume 39, Number 4, pages 585–593.

2004. "Stabilization of stochastic linear systems with finite feedback data rates", in *SIAM Journal of Optimal Control*, Volume 43, Number 2, pages 413-436.

Nair, G. N., R. J. Evans, I. M. Y. Mareels and W. Moran,
2004. "Topological feedback entropy and nonlinear stabilization", in *IEEE Transactions on Automatic Control*, Volume 49, Number 9, pages 1585-1597.

Nair, G. N., F. Fagnani, S. Zampieri and R. J. Evans,
2007. "Feedback control under data rate constraints: An overview", in *Proceedings of the IEEE*, Volume 95, Number 1, pages 108–137.

Nayyar, A. and D. Teneketzis
2008. On jointly optimal real-time encoding and decoding strategies in multiterminal communication systems. In *submitted to 47th IEEE Conference of Decision and Ccontrol*.

Neuhoff, D. L. and R. K. Gilbert,
1982. "Causal source codes", in *IEEE Transactions on Information Theory*, Volume 28, Number 6, pages 701-713.

Peterson, I. R. and A. V. Savkin
2001. Multi-rate stabilization of multivariable discrete-time linear systems via a limited capacity communication channel. In *Proceedings of 40th IEEE Conference on Decision and Control*, pages 304-309.

Pinsker, P. S. and A. K. Gorbunov,
1987. "Epsilon entropy with delay with small mean-square reproduction error", in *Problems in Information Transmission*, Volume 23, Number 2, pages 3-8.

Piret, P.,
1979. "Causal sliding block encoders with feedback", in *IEEE Transactions on Information Theory*, Volume 25, Number 2, pages 237–240.

Porta, J. M., N. Vlassis, M. T. J. Spaan and P. Pouparts,
2006. "Point-based value iteration for continuous POMDPs", in *Journal of Machine Learning Research*, Volume 7, pages 2329-2367.

Radner, R.,
1962. "Team decision problems", in *Annals of Mathmatical Statistics*, Volume 33, pages 857-881.

Rust, J.,
1997. "Using randomization to break the curse of dimensionality", in *Econometrica*, Volume 65, Number 3, pages 487–516.

Sandell Jr., N. R.
1974. *Control of Finite-State, Finite-Memory Stochastic Systems*. Ph.D. Thesis, Massachussets Institute of Technology, Cambridge, MA.

Savkin, A. V.,
2006. "Analysis and synthesis of networked control systems: Topological entropy, observability, robustness and optimal control", in *Automatica*, Volume 42, pages 51-62.

Sennott, L. I.,
1997a. "The computation of average optimal policies in denumerable state Markov decision chains", in *Advances in Applied Probability*, Volume 29, pages 114-137.

1997b. "On computing average cost optimal policies with application to routing parallel queues", in *ZOR Mathematical Methods in Operations Research*, Volume 45, pages 45-62.

1999. *Stochastic dynamic programming and the control of queueing systems*. Wiley, New York, NY, USA.

Shannon, C. E.,
1948. "A mathematical theory of communication", in *Bell System Technical Journal*, Volume 22, pages 379-423.

Smallwood, R. D. and E. J. Sondik,
1973. "The optimal control of partially observable Markov processes over a finite horizon", in *Operations Research*, Volume 11, pages 1071-1088.

Tatikonda, S.
2000. *Control Under Communication Constraints*. Ph.D. Thesis, Department of EECS, Masachussets Institute of Technology, Cambridge, MA.

Tatikonda, S. and S. K. Mitter,
2004a. "Control under communication constraints", in *IEEE Transactions on Automatic Control*, Volume 49, Number 7, pages 1056-1068.

2004b. "Control over noisy channels", in *IEEE Transactions on Automatic Control*, Volume 49, Number 7, pages 1196-1201.

Tatikonda, S., A. Sahai and S. K. Mitter,
2004. "Stochastic linear control over a communication channel", in *IEEE Transactions on Automatic Control*, Volume 49, Number 9, pages 1549–1561.

Teneketzis, D.
1979. *Communication in Decentralized Control*. Ph.D. Thesis, Department of EECS, MIT, Cambridge, MA.

1996. "On information structures and nonsequential stochastic control", in *CWI Quarterly*, Volume 9, Number 3, pages 241–260.

2006. "On the structure of optimal real-time encoders and decoders in noisy communication", in *IEEE Transactions on Information Theory*, pages 4017-4035.

Teneketzis, D. and M. S. Andersland,
2000. "On partial order characterization of information structures", in *Mathematics of Control, Signals and Systems*, Volume 13, pages 277-292.

Thrun, S.
2000. Monte carlo POMDPs. In Solla, S., T. Leen and K.-R. Müller, editors, *Advances in Neural Information Processing Systems 12*, pages 1064–1070. MIT Press.

von Neumann, J. and O. Morgenstern,
1944. *Theory of Games and Economic Behavior*. Princeton University Press.

Wald, A.,
1947. *Sequential hypothesis testing*. Wiley.

Walrand, J. C. and P. Varaiya,
1983a. "Optimal causal coding—decoding problems", in *IEEE Transactions on Information Theory*, Volume 29, Number 6, pages 814-820.

1983b. "Causal coding and control of Markov chains", in *System and Control Letters*, Volume 3, pages 189-192.

Weissman, T. and N. Merhav,
2002. "On limited-delay lossy coding and filtering of individual sequences", in *IEEE Transactions on Information Theory*, Volume 48, Number 3, pages 721–733.

Whittle, P.,
1983. *Optimization Over Time*, Volume 2, *Wiley Series in Probability and Mathematical Statistics*. John Wiley and Sons.

Witsenhausen, H. S.

1966. *Minimax Control of Uncertain Systems*. PhD thesis, Massachusetts Institute of Technology.

1968. "A counterexample in stochastic optimum control", in *SIAM Journal of Optimal Control*, Volume 6, Number 1, pages 131-147.

1970. "On performance bounds for uncertain systems", in *SIAM Journal of Control*, Volume 8, pages 55–89.

1971a. "Separation of estimation and control for discrete time systems", in *Proceedings of the IEEE*, Volume 59, Number 11, pages 1557–1566.

1971b. "On information structures, feedback and causality", in *SIAM Journal of Control*, Volume 9, Number 2, pages 149–160.

1973. "A standard form for sequential stochastic control", in *Mathematical Systems Theory*, Volume 7, Number 1, pages 5-11.

1975. The intrinsic model for stochastic control: Some open problems. In *Lecture Notes in Economics and Mathematical Systems*, pages 322–335. Springer Verlag.

1976. Some remarks on the concept of state. In Ho, Y. C. and S. K. Mitter, editors, *Directions in Large-Scale Systems*, pages 69-75. Plenum.

1978. Informational aspects of stochastic control. In *Proceedings of the Oxford Conference on Stochastic Optimization*.

1979. "On the structure of real-time source coders", in *Bell System Technical Journal*, Volume 58, Number 6, pages 1437-1451.

1988. "Equivalent stochastic control problems", in *Mathematics of Control Signals and Systems*, Volume 1, pages 3–11.

Wong, W. S. and R. W. Brockett,

1999. "Systems with finite communciation bandwidth constraints II: Stabilization with limited information feedback", in *IEEE Transactions on Automatic Control*, Volume 44, pages 1049-1053.

Yoshikawa, T.,
1978. "Decomposition of dynamic team decision problems", , Volume 23, Number 4, pages 627–632.