Sequential decomposition of sequential teams

applications to real-time communication and optimal control in decentralized multi-agent systems

Aditya Mahajan

Thesis Committee

Prof. Demosthenis Teneketzis,
Prof. Jeffrey C. Lagarias,
Prof. Achilleas Anastasopoulos,
Prof. S. Sandeep Pradhan.

Thesis defense
September 4th, 2008

Outline of the thesis

1. Overview of multi-agent systems
   - ▷ Classification : games vs. teams and single-state vs. multi-state
   - ▷ Sequential decomposition : what and why
2. Optimal design of two-agent teams
   - ▷ The notion of information state
   - ▷ Sequential decomposition for general finite horizon problem
   - ▷ . . . and for three variations for infinite horizon problems
3. Real-time communication
4. Networked control systems
5. Conclusion

Multi-agent decentralized systems

- Applications

  - telecommunication networks
  - sensor networks
  - surveillance networks
  - transportation networks
  - control systems
  - monitoring and diagnostic systems
  - multi-robot systems
  - multi-core CPUs
  - . . .

- Salient features

  - System has different components
  - These components know different information
  - The components need to cooperate and coordinate

Question

How do we approach the design of a decentralized multi-agent system?

Classification of multi-agent systems

- Teams vs. Games
- Single-stage vs. Multi-stage
- Sequential vs. non-sequential systems
- Classical vs. non-classical information structures

Sequential multi-stage teams with
non-classical information structures

Sequential Decomposition

- Divide and conquer :

  Exploit sequential and multi-stage nature of the problem

Convert a one-shot optimization problem into a sequence of nested optimization problems

- Classical information structure (MDP, POMDP)

  Dynamic programming

- Non-classical information structure

  Witsenhausen's standard form (not applicable to infinite horizon problems)

Why consider sequential decomposition

- Finite horizon
  - Brute force search always possible but has high complexity
  - ***Provides a systematic way to search for an optimal solution efficiently***
- Infinite horizon
  - Brute force search not possible
  - An arbitrary solution cannot be implemented
  - ***Identify qualitative properties to search and implement optimal designs compactly***
- May help in identifying (and proving) other qualitative properties

Can we obtain a sequential decomposition of decentralized finite and infinite horizon sequential teams?

YES

Result of this thesis

- General two-agent team : finite horizon
  - Identify an *information state* sufficient for performance analysis
  - Optimally control the evolution of information state

$$\Updownarrow$$

Markov decision process where

$$\begin{aligned}\text{State space} &: \text{space of probability measures}\\ \text{Action space} &: \text{space of functions}\end{aligned}$$

Result of this thesis

- Two-agent team : infinite horizon
  - ***Cannot restrict attention to time invariant designs***
  - ***Can restrict attention to time invariant meta-designs***
  - Consider three specific variations

Variation V1 $\equiv$ POMDP with finite (unobserved) state and action spaces

Variations V2 and V3 $\equiv$ POMDP with uncountable (unobserved) state and action spaces

Contributions –
conceptual and technical

Conceptual contributions

- How to approach the sequential decomposition of a decentralized sequential multi-stage team.

- Properties of information states for systems with non-classical information structures

Technical contributions

1. In the context of general two-agent teams

2. In the context of real-time communication

3. In the context of networked control systems

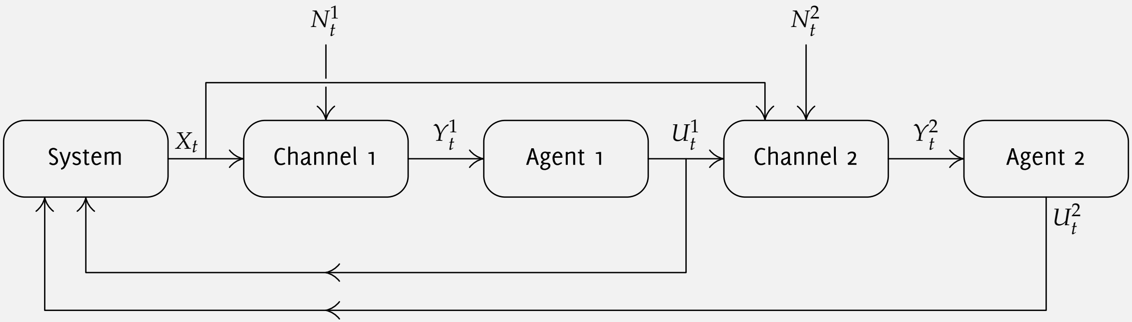
- General contributions for all three problems

  - Sequential decomposition of a general finite horizon problem
  - Sequential decomposition for three general instances for infinite horizon problems.
  - Obtained qualitative properties of optimal decision rules under general conditions

This talk . . .

A general two-agent team

Model



System/Plant $X_{t+1} = f_t(X_t, U_t^1, U_t^2, W_t)$

Observations
$$Y_t^1 = h_t^1(X_t, N_t^1)$$
$$Y_t^2 = h_t^2(X_t, U_t^1, N_t^2)$$

Control
$$U_t^1 = g_t^1(Y_t^1, S_{t-1}^1)$$
$$U_t^2 = g_t^2(Y_t^2, S_{t-1}^2)$$

State Update
$$S_t^1 = l_t^1(Y_t^1, U_t^1, S_{t-1}^1)$$
$$S_t^2 = l_t^2(Y_t^2, U_t^2, S_{t-1}^2)$$

Performance

- Control strategy

$$G^k := (g_1^k, \dots, g_T^k), \qquad k = 1, 2$$

- State update strategy

$$L^k := (l_1^k, \dots, l_T^k), \qquad k = 1, 2$$

- System design

$$(G^1, L^1, G^2, L^2)$$

- Performance

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

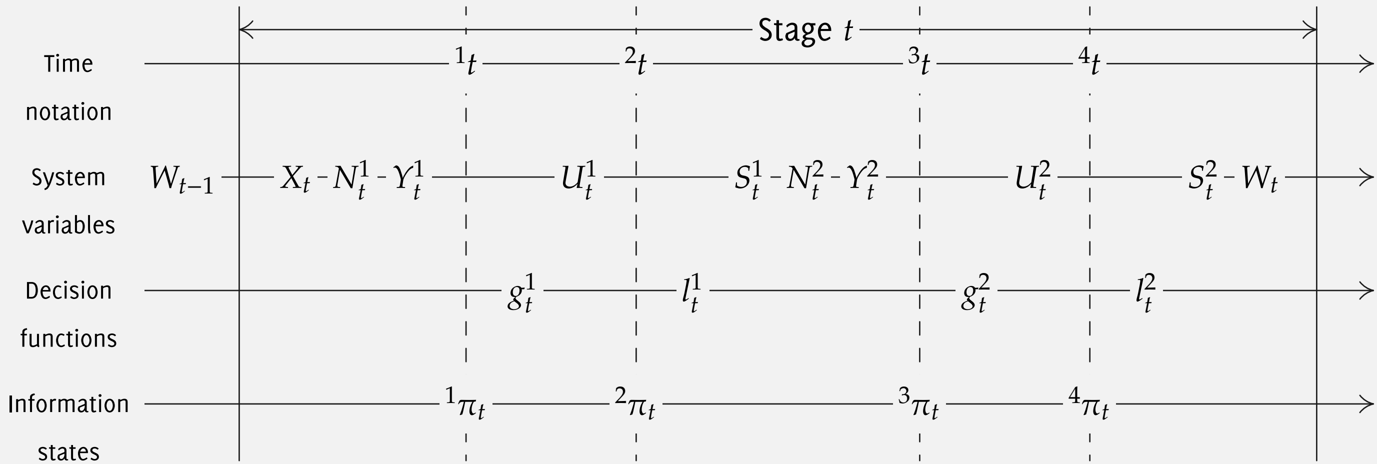
Problem formulation (Finite horizon)

- Given
  - Alphabets $(\mathcal{X}_t, \mathcal{Y}^1_t, \mathcal{Y}^2_t, \mathcal{U}^1_t, \mathcal{U}^2_t, \mathcal{S}^1_t, \mathcal{S}^2_t)$
  - Plant functions f_t
  - Observation functions h^1_t and h^2_t
  - Cost functions ρ_t
  - PMF of N^1_t, N^2_t, and W_t
- Determine
  - An optimal design $(G^{1,*}, L^{1,*}, G^{2,*}, L^{2,*})$

$$\mathcal{J}_T(G^{1,*}, L^{1,*}, G^{2,*}, L^{2,*}) = \mathcal{J}^*_T \coloneqq \min_{(G^1,L^1,G^2,L^2)} \mathcal{J}_T(G^1, L^1, G^2, L^2)$$

Salient Features

- Sequential nature of the problem



- Data, information and non-classical information structure

$$\sigma(Y_t^1, S_{t-1}^1; {}^1\varphi^{t-1}) \underset{\not\supseteq}{\not\subseteq} \sigma(S_{t-1}^2; {}^1\varphi^{t-1})$$

Solution Approach:

Sequential Decomposition

Key Idea:

Information State

Properties of information state?

- Well understood for systems with classical information structures

- Not known for systems with non-classical information structures

- Solution Approach

  - All the agents decide what to do before the system starts operating

  - At any given time, they can adapt their strategies only based on information that is common to everyone—the past decision rules

  - Store and process the past decision rules more efficiently

Properties on information state

P1 Sufficient summary of past information

P2 Common knowledge and sequential update

P3 Time invariant domain

P4 Minimality

P1. Sufficient summary of past information

- Define an equivalence relation on past decision rules:

$^1\varphi^{t-1,(1)} \sim {}^1\varphi^{t-1,(2)} \implies \forall\, {}^1\varphi_t^T = (g_t^{1,T}, l_t^{1,T},\ g_t^{2,T}, l_t^{2,T})$, we have

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

Two choices of past decision rules are equivalent if they give the same performance for any choice of future decision rules

- $^1\varphi^{t-1,(1)} \sim {}^1\varphi^{t-1,(2)} \implies$ both have same optimal future decision rules

P1. Sufficient summary of past information

- ${}^i\pi_t : {}^i\Phi^{t-1} \to {}^i\Pi_t$ is an information state if $\forall\, {}^i\varphi^{t-1,(1)}, {}^i\varphi^{t-1,(2)} \in {}^i\Phi^{t-1}$,

$$ {}^i\pi_t({}^i\varphi^{t-1,(1)}) = {}^i\pi_t({}^i\varphi^{t-1,(2)}) \implies {}^i\varphi^{t-1,(1)} \sim {}^i\varphi^{t-1,(2)} $$

${}^i\Phi^{t-1}$	space of realization of all past decision rules
${}^i\Pi_t$	any arbitrary space

Information state is a many-to-one function which assigns same values to two past choices of information states only if they are equivalent for performance evaluation

P2. Common knowledge and sequential update

- All agents should be able to perform the sequential decomposition
- Information state should be common knowledge (in the sense of Aumann)

Information state cannot depend on local data

- For any realization of information state and any choice of current decision rule, all agents should be able to determine to next information state

Information state update cannot depend on local data

$$\dots {}^1\pi_t \xrightarrow{g_t^1} {}^2\pi_t \xrightarrow{l_t^1} {}^3\pi_t \xrightarrow{g_t^2} {}^4\pi_t \xrightarrow{l_t^2} \dots$$

Properties of information state

- Sequence of functions $\{{}^{i}\pi_t,\ i = 1, 2, 3, 4,\ t = 1, \ldots, T\}$ with properties (P1) and (P2) $\implies$ information state for finite horizon problem.

- We want to find conditions under which the methodology can be extended to infinite horizon problems

P3. Time invariant domain

$${}^{i}\pi_t : {}^{i}\Phi^{t-1} \to {}^{i}\Pi \qquad \text{(Further } {}^{i}\Pi \text{ should not depend on time horizon T).}$$

P4. Minimality

- Information states should be *efficient*, both from the point of view of storage and computation.

Identifying appropriate

information states

is highly non-trivial

Information states

$$
\begin{aligned}
{}^1\pi_t &:= \Pr\left(X_t, Y_t^1, S_{t-1}^1, S_{t-1}^2 \,\middle|\, {}^1\varphi^{t-1}\right), \\
{}^2\pi_t &:= \Pr\left(X_t, Y_t^1, U_t^1, S_{t-1}^1, S_{t-1}^2 \,\middle|\, {}^2\varphi^{t-1}\right), \\
{}^3\pi_t &:= \Pr\left(X_t, Y_t^2, U_t^1, S_t^1, S_{t-1}^2 \,\middle|\, {}^3\varphi^{t-1}\right), \\
{}^4\pi_t &:= \Pr\left(X_t, Y_t^2, U_t^1, U_t^2, S_t^1, S_{t-1}^2 \,\middle|\, {}^4\varphi^{t-1}\right).
\end{aligned}
$$

Probability measure on the state for input output mapping

Properties of information state

- State update

$$\dots {}^1\pi_t \xrightarrow{g_t^1} {}^2\pi_t \xrightarrow{l_t^1} {}^3\pi_t \xrightarrow{g_t^2} {}^4\pi_t \xrightarrow{l_t^2} \dots$$

- Expected instantaneous cost

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

Main Idea :

Optimally control
the evolution of the
information state

Equivalent optimization problem

- System

$$
\begin{aligned}
g_t^1 &= {}^1\Delta_t({}^1\pi_t), & {}^2\pi_t &= {}^1Q_t(g_t^1)\,{}^1\pi_t, \\
l_t^1 &= {}^2\Delta_t({}^2\pi_t), & {}^3\pi_t &= {}^2Q_t(l_t^1)\,{}^2\pi_t, \\
g_t^2 &= {}^3\Delta_t({}^3\pi_t), & {}^3\pi_t &= {}^3Q_t(g_t^2)\,{}^1\pi_t, \\
l_t^2 &= {}^4\Delta_t({}^4\pi_t), & {}^1\pi_{t+1} &= {}^4Q_t(l_t^2)\,{}^4\pi_t.
\end{aligned}
$$

${}^1\Delta_t$, ${}^2\Delta_t$, ${}^3\Delta_t$ and ${}^4\Delta_t$ are called *meta-function*

Equivalent optimization problem

- Equivalent cost

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

where $\Delta^T := ({}^1\Delta_1, {}^2\Delta_1, {}^3\Delta_1, {}^4\Delta_1, \dots, {}^1\Delta_T, {}^2\Delta_T, {}^3\Delta_T, {}^4\Delta_T)$ is *meta-strategy*

Optimally controlling
the evolution of the
information state
is a classical MDP

Sequential decomposition

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

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

$$ {}^{1}V_t({}^{1}\pi_t) = \inf_{g_t^1 \in \mathcal{G}_t^1} {}^{2}V_t\big({}^{1}Q_t(g_t^1)\,{}^{1}\pi_t\big), $$

$$ {}^{2}V_t({}^{2}\pi_t) = \inf_{l_t^1 \in \mathcal{L}_t^1} {}^{3}V_t\big({}^{2}Q_t(l_t^1)\,{}^{2}\pi_t\big), $$

$$ {}^{3}V_t({}^{3}\pi_t) = \inf_{g_t^2 \in \mathcal{G}_t^2} {}^{4}V_t\big({}^{3}Q_t(g_t^2)\,{}^{3}\pi_t\big), $$

$$ {}^{4}V_t({}^{4}\pi_t) = \hat{\rho}_t({}^{4}\pi_t) + \inf_{l_t^2 \in \mathcal{L}_t^2} {}^{1}V_{t+1}\big({}^{4}Q_t(l_t^2)\,{}^{4}\pi_t\big). $$

Each step is a functional optimization problem

Computing the optimal design

1. The backward step — computations
   - ▷ For each time instant t and each ${}^{i}\pi_t$
     - Evaluate the value function ${}^{i}V_t({}^{i}\pi_t)$
     - and store the corresponding arg inf ${}^{i}\Gamma_t({}^{i}\pi_t)$
   - ▷ $\mathcal{J}^*_T = {}^{1}V_1({}^{1}\pi^*_1)$

2. The forward step — implementation
   - ▷ Start at $t = 1$. We know ${}^{1}\pi^*_1$. Look-up $g^{1,*}_1 = {}^{1}\Gamma_1({}^{1}\pi^*_1)$
   - ▷ ${}^{1}\pi^*_1$ and $g^{1,*}$ determine ${}^{2}\pi^*_1$. Look-up $l^{1,*}_1 = {}^{2}\Gamma_1({}^{2}\pi^*_1)$.
   - ▷ And so on ...

Determine optimal $(g^{1,*}_1, l^{1,*}_1, g^{2,*}_1, l^{2,*}_1, \dots, g^{1,*}_T, l^{1,*}_T, g^{2,*}_T, l^{2,*}_T)$

Numerical computations

- Reachability Analysis
  - Find all reachable ${}^{i}\pi_t$ and solve the nested optimality equations for them
- Smallwood & Sondik-like algorithms
  - ${}^{i}V_t(\cdot)$ is piecewise linear and convex
  - Can be represented as a pointwise minimum of a finite family of affine functions
  - These affine functions can be computed recursively by linear programming

Numerical computations

- Approximation algorithms
  - Grid based solutions
  - Point based solutions
  - Randomized algorithms
- Specialized algorithms ??
  - Exploit *open loop* nature of optimality equations

Extensions to infinite horizon

Infinite horizon systems

- Assumption
  - Time-homogeneous system: The following do not depend on t:
    - Alphabets $(\mathcal{X}_t, \mathcal{Y}^1_t, \mathcal{Y}^2_t, \mathcal{U}^1_t, \mathcal{U}^2_t)$
    - Plant functions f_t
    - Observation functions h^1_t and h^2_t
    - Cost functions ρ_t
    - PMF of N^1_t, N^2_t, and W_t
- Four variations

V1 $(\mathcal{S}^1_t, \mathcal{S}^2_t)$ do not depend on t

V2 $\mathcal{S}^1_t = (\mathcal{Y}^{1,t} \times \mathcal{U}^{1,t})$ and $\mathcal{S}^2_t$ does not depend on t.

V3 $\mathcal{S}^1_t$ does not depend on t, $\mathcal{S}^2_t = (\mathcal{Y}^{2,t} \times \mathcal{U}^{2,t})$

V4 $\mathcal{S}^1_t = (\mathcal{Y}^{1,t} \times \mathcal{U}^{1,t})$ and $\mathcal{S}^2_t = (\mathcal{Y}^{2,t} \times \mathcal{U}^{2,t})$

Performance criteria

1. The expected discounted cost criterion

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

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

2. The average cost per unit time criterion

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

V1. Both agents have time-invariant state space

- Information states belong to time invariant spaces

- Can be extended to infinite horizon horizon using standard techniques

$$
\begin{aligned}
{}^1\pi_t &:= \Pr\left(X_t, Y_t^1, S_{t-1}^1, S_{t-1}^2 \,\middle|\, {}^1\varphi^{t-1}\right), \\
{}^2\pi_t &:= \Pr\left(X_t, Y_t^1, U_t^1, S_{t-1}^1, S_{t-1}^2 \,\middle|\, {}^2\varphi^{t-1}\right), \\
{}^3\pi_t &:= \Pr\left(X_t, Y_t^2, U_t^1, S_t^1, S_{t-1}^2 \,\middle|\, {}^3\varphi^{t-1}\right), \\
{}^4\pi_t &:= \Pr\left(X_t, Y_t^2, U_t^1, U_t^2, S_t^1, S_{t-1}^2 \,\middle|\, {}^4\varphi^{t-1}\right).
\end{aligned}
$$

- Restricting attention to stationary strategies is not necessarily optimal.

- There is no loss of optimality in restricting attention to *stationary meta-strategies*

- Optimal stationary meta-function can be determined by the fixed point of a functional equation

V2 and V3. One agent has time-invariant state space and the other has perfect recall

- The results of the finite horizon problem *cannot* be directly extended to infinite horizon
- Structural property of optimal design

The agent with perfect recall can choose a control action based on its belief on the state of the system and the state of the other agent

Importance of structural properties

- The structural property provides a compact representation of optimal control laws

$$U^1_t = g^1_t(Y^{1,t}, U^{1,t-1}) \qquad \text{vs} \qquad U^1_t = g^1_t(B^1_t)$$

$$\text{where } B^1_t \coloneqq \Pr\left(X_t, S^2_{t-1} \,\middle|\, Y^{1,t}, U^{1,t-1}\right)$$

- With this compact representation we can obtain a sequential decomposition for infinite horizon problems

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

Summary of results

- Finite horizon
  - Identify an appropriate information state
  - Consider optimally controlling the evolution of the information state
    - Equivalent to original problem
    - Equivalent to centralized MDP
  - Results in a sequential decomposition

- Infinite horizon
  - Two different performance criteria
  - Three variations based on the state space of the agents
  - Near-optimal designs given by solution of fixed point of functional equations

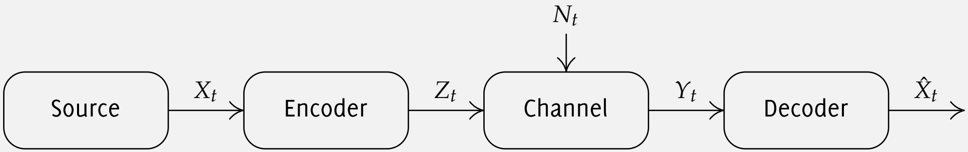
Applications

Real-time communication

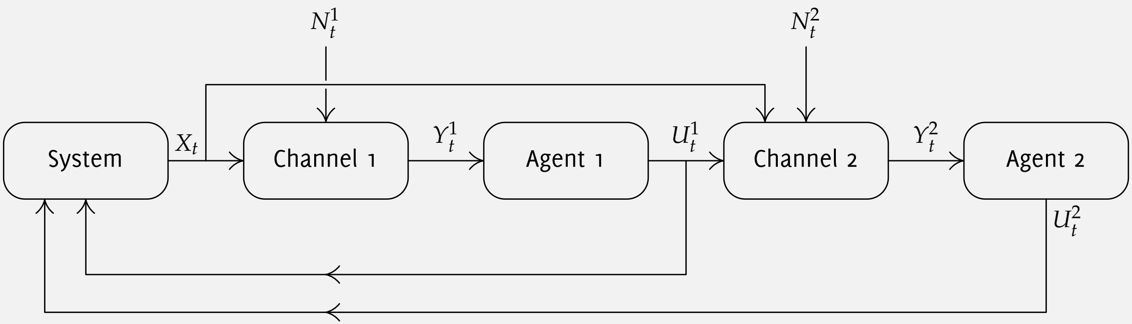
Communication systems in which information should be transmitted and decoded within a fixed delay constraint

- Asymptotic results of information theory not applicable
- Can be formulated as a two-agent sequential team
- Identify structural properties and obtain sequential decomposition
- Results can be extended to sources and channels with memory

Real-time communication as a two-agent team

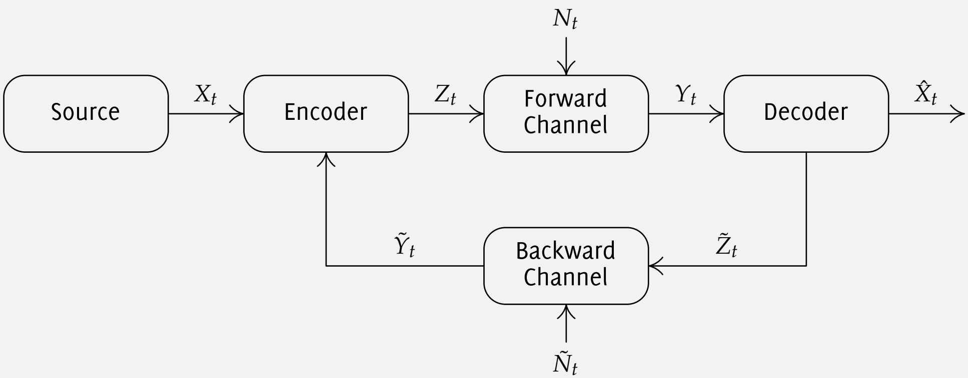


Real-time communication with no feedback

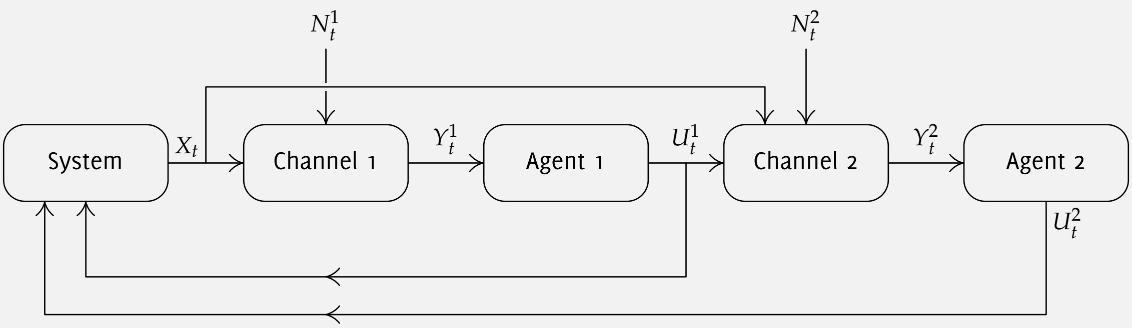


Two agent system

Real-time communication as a two-agent team



Real-time communication with noisy communication



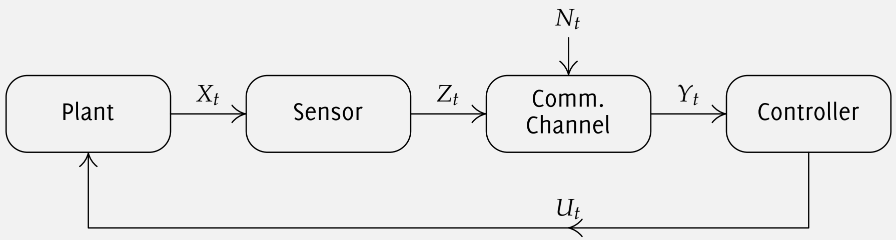
Two agent system

Networked control systems

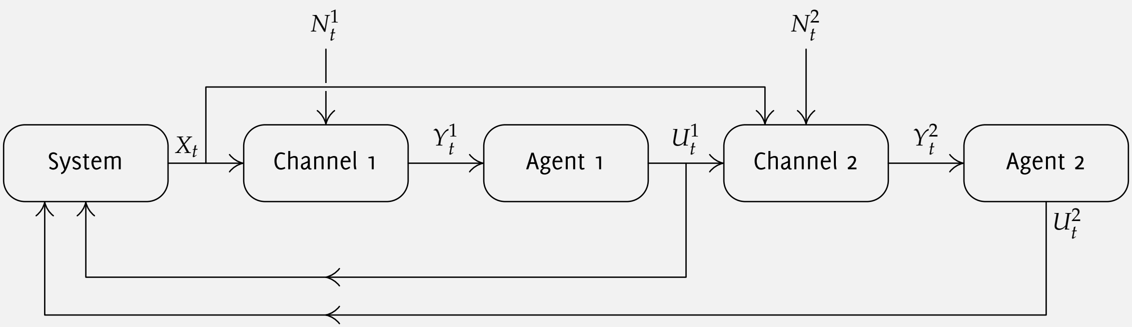
Controller and plant located at different locations and connected over a network

- NCS: Stability vs. performance
- Asymptotic results of stability and information theory not applicable
- Can be formulated as a two-agent sequential team
- Identify structural properties and obtain sequential decomposition

Networked control system as a two-agent team



A simple NCS



Two agent system

Specific applications

can be solved “easily”

Reflections

- Going from one agent systems to two agent systems requires a paradigm shift

- The success of POMDPs has led to the failure to understand multi-agent systems

Do not think in terms of data

- Do not think in terms of ***one*** agent,

  think in terms of ***every*** agent

Future directions

- Computational algorithms

  The theory will not be practical until efficient algorithms to solve nested optimality equations are identified

- Extension to multi-agent systems

  - Scaling with number of agents
  - Agents with perfect recall

- Extensions to non-stochastic systems

  - Appear feasible
  - Use σ-algebra rather than the measure induced by the σ-algebra

Thank you