On real-time communication systems with noisy feedback

Aditya Mahajan and Demosthenis Teneketzis

Dept. of EECS,
University of Michigan,
Ann Arbor, MI. USA.

## Problem Formulation

**Model**
- **Discrete time, discrete valued**
- **Source**: \( \{ X_t, t = 1, \ldots, T \} \)
- **Encoder**: \( Z_t = c_t(X^t, \tilde{Y}^{t-1}, Z^{t-1}) \)
- **Forward ch.**: \( Y_t = h(Z_t, N_t) \)
- **Backward ch.**: \( \tilde{Y}_t = \tilde{h}(Y_t, \tilde{N}_t) \)
- **Decoder**: \( \hat{X}_t = g_t(Y_t, M_{t-1}) \)
- **Memory Up.**: \( M_t = l_t(Y_t, M_{t-1}) \)

**Objective**
- Choose \((c_1, \ldots, c_T), (g_1, \ldots, g_T), (l_1, \ldots, l_T)\) to minimize

\[
\mathcal{J} := \mathbb{E} \left\{ \sum_{t=1}^{T} \rho(X_t, \hat{X}_t) \right\}
\]
• Literature Overview

• Salient Features

• Solution Methodology
  (i) Structural Properties
  (ii) Global Optimization

• Conclusions
real-time communication

- *Types of problem*
  - (i) Source coding
  - (ii) Joint source-channel coding

- *Literature classification*
  - (i) Performance bounds
  - (ii) Coding of individual sequences
  - (iii) Coding of Markov sources

with

noisy feedback

- *Types of problem*
  - (i) Channel coding

- *Literature classification*
  - (i) Capacity of channels with memory and noisy feedback
  - (ii) Error exponents of memoryless channels with noisy feedback
**Literature Overview**

- **Types of problem**
  - (i) Source coding
  - (ii) Joint source-channel coding

- **Literature classification**
  - (i) Performance bounds
  - (ii) Coding of individual sequences
  - (iii) Coding of Markov sources

---

**real-time communication** with **noisy feedback**

- **Types of problem**
  - (i) Channel coding

- **Literature classification**
  - (i) Capacity of channels with memory and noisy feedback
  - (ii) Error exponents of memoryless channels with noisy feedback

---

Graph:

- Source $\rightarrow$ Encoder $\rightarrow$ Fwd. Ch. $\rightarrow$ Decoder
- $X_t \rightarrow c_t \rightarrow Z_t \rightarrow Y_t \rightarrow Y_t$ (Bwd. Ch.)
- $g_t, l_t \rightarrow \hat{X}_t$
Salient Features

- Real-time Constraint
- Finite Memory
- Noisy Feedback
Salient Features

- Real-time Constraint
- Finite Memory
- Noisy Feedback

⇒

Do not know how to apply Information Theory
Salient Features

- Real-time Constraint
- Finite Memory
- Noisy Feedback

⇒

Do not know how to apply Information Theory

- Use Stochastic Optimization
Salient Features

- Real-time Constraint
- Finite Memory
- Noisy Feedback

Do not know how to apply Information Theory

- Use Stochastic Optimization

Which solution methodology?

- Markov Decision Theory
- Orthogonal Search
- Standard Form
- ??
Markov Decision Theory

- works only for one decision maker with perfect recall
Stochastic Optimization

- Markov Decision Theory (not applicable)
  - works only for one decision maker with perfect recall
- Orthogonal Search (not appropriate)
  - May not converge
  - only guarantees local optima
STOCHASTIC OPTIMIZATION

- Markov Decision Theory (not applicable)
  - works only for one decision maker with perfect recall
- Orthogonal Search (not appropriate)
  - May not converge
  - only guarantees local optima
- Standard Form — Witsenhausen, 1973 (not appropriate)
  - Does not extend to infinite horizon
Stochastic Optimization

- Markov Decision Theory (not applicable)
  - works only for one decision maker with perfect recall
- Orthogonal Search (not appropriate)
  - May not converge
  - only guarantees local optima
- Standard Form — Witsenhausen, 1973 (not appropriate)
  - Does not extend to infinite horizon

Our Methodology

(i) Identify structural properties
(ii) Use structural properties to solve the global optimization problem
Our Methodology
(i) Structural Properties


**Structural Properties**

**NEED**

- Encoder: \[ Z_t = c_t(X^t, Z^{t-1}, \tilde{Y}^{t-1}), \quad c_t \in C_t : X^t \times Z^{t-1} \times \tilde{Y}^{t-1} \rightarrow Z \]

- Domain changing with time.

- *Makes the infinite horizon design hard*

- *Can we simplify implementation?*

**PRELIMINARIES**

- Notion of Information
- Notion of Beliefs
Let \((X_1, \ldots, X_T, N_1, \ldots, N_T, \tilde{N}_1, \ldots, \tilde{N}_T)\) be defined on \((\Omega, \mathcal{F}, P)\).
Preliminaries

- Let \((X_1, \ldots, X_T, N_1, \ldots, N_T, \tilde{N}_1, \ldots, \tilde{N}_T)\) be defined on \((\Omega, \mathcal{F}, P)\).

Information

| Information                  | Everything that can be inferred from the data |
Preliminaries

- Let \((X_1, \ldots, X_T, N_1, \ldots, N_T, \tilde{N}_1, \ldots, \tilde{N}_T)\) be defined on \((\Omega, \mathcal{F}, P)\).

<table>
<thead>
<tr>
<th>Information at Encoder</th>
<th>Information at Decoder</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1_{E_t} := (X^t, Z^{t-1}, \tilde{Y}^{t-1})), (1_{E_t} := \sigma(1_{E_t}; 1_{\phi^{t-1}})),</td>
<td>(1_{R_t} := (M_{t-1})), (1_{R_t} := \sigma(1_{R_t}; 1_{\phi^{t-1}})),</td>
</tr>
<tr>
<td>(2_{E_t} := (X^t, Z^t, \tilde{Y}^{t-1})), (2_{E_t} := \sigma(2_{E_t}; 2_{\phi^{t-1}})),</td>
<td>(2_{R_t} := (Y_t, M_{t-1})), (2_{R_t} := \sigma(2_{R_t}; 2_{\phi^{t-1}})),</td>
</tr>
<tr>
<td>(3_{E_t} := (X^t, Z^t, \tilde{Y}^t)), (3_{E_t} := \sigma(3_{E_t}; 3_{\phi^{t-1}})).</td>
<td>(3_{R_t} := (Y_t, M_{t-1})), (3_{R_t} := \sigma(3_{R_t}; 3_{\phi^{t-1}})).</td>
</tr>
</tbody>
</table>
• Let \((X_1, \ldots, X_T, N_1, \ldots, N_T, \tilde{N}_1, \ldots, \tilde{N}_T)\) be defined on \((\Omega, \mathcal{F}, P)\).

### Information

**Information at Encoder**

\[
\begin{align*}
1E_t &:= (X^t, Z^{t-1}, \tilde{Y}^{t-1}), \\
2E_t &:= (X^t, Z^t, \tilde{Y}^{t-1}), \\
3E_t &:= (X^t, Z^t, \tilde{Y}^t),
\end{align*}
\]

\[
\begin{align*}
1\mathcal{E}_t &:= \sigma(1E_t; 1\phi^{t-1}), \\
2\mathcal{E}_t &:= \sigma(2E_t; 2\phi^{t-1}), \\
3\mathcal{E}_t &:= \sigma(3E_t; 3\phi^{t-1}).
\end{align*}
\]

### Information at Decoder

\[
\begin{align*}
1R_t &:= (M_{t-1}), \\
2R_t &:= (Y_t, M_{t-1}), \\
3R_t &:= (Y_t, M_{t-1}),
\end{align*}
\]

\[
\begin{align*}
1\mathcal{R}_t &:= \sigma(1R_t; 1\phi^{t-1}), \\
2\mathcal{R}_t &:= \sigma(2R_t; 2\phi^{t-1}), \\
3\mathcal{R}_t &:= \sigma(3R_t; 3\phi^{t-1}).
\end{align*}
\]

### Beliefs

**Beliefs**

What one decision maker thinks about the data at other nodes
Preliminaries

- Let \((X_1, \ldots, X_T, N_1, \ldots, N_T, \tilde{N}_1, \ldots, \tilde{N}_T)\) be defined on \((\Omega, \mathcal{F}, P)\).

### Information

- **Information**
  - Everything that can be inferred from the data

<table>
<thead>
<tr>
<th>Information at Encoder</th>
<th>Information at Decoder</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1E_t := (X^t, Z^{t-1}, \tilde{Y}^{t-1})), (1\mathcal{E}_t := \sigma(1E_t, 1\phi^{t-1})),</td>
<td>(1R_t := (M_{t-1})), (1\mathcal{R}_t := \sigma(1R_t, 1\phi^{t-1})),</td>
</tr>
<tr>
<td>(2E_t := (X^t, Z^t, \tilde{Y}^{t-1})), (2\mathcal{E}_t := \sigma(2E_t, 2\phi^{t-1})),</td>
<td>(2R_t := (Y_t, M_{t-1})), (2\mathcal{R}_t := \sigma(2R_t, 2\phi^{t-1})),</td>
</tr>
<tr>
<td>(3E_t := (X^t, Z^t, \tilde{Y}^t)), (3\mathcal{E}_t := \sigma(3E_t, 3\phi^{t-1})).</td>
<td>(3R_t := (Y_t, M_{t-1})), (3\mathcal{R}_t := \sigma(3R_t, 3\phi^{t-1})).</td>
</tr>
</tbody>
</table>

### Beliefs

- **Beliefs**
  - What one decision maker thinks about the data at other nodes

<table>
<thead>
<tr>
<th>Belief of the encoder</th>
<th>Belief of the decoder</th>
</tr>
</thead>
<tbody>
<tr>
<td>(iB_t(i\mathcal{r}) := \Pr(iR_t = i\mathcal{r}</td>
<td>i\mathcal{E}_t)).</td>
</tr>
</tbody>
</table>
Structure of Opt Encoders

- The main idea
  - Fix the decoding and memory update function
  - Look at the problem from the encoder’s point of view
  - Derive qualitative properties of optimal encoders
Structure of Optimal Encoders

- The main idea
  - Fix the decoding and memory update function
  - Look at the problem from the encoder’s point of view
  - Derive qualitative properties of optimal encoders

\[
Z_t = c_t(X_t, ^1B_t), \quad t = 2, \ldots, T
\]
Structure of Optimal Encoders

- **The main idea**
  - Fix the decoding and memory update function
  - Look at the problem from the encoder’s point of view
  - Derive qualitative properties of optimal encoders

\[ Z_t = c_t(X_t, B_t), \quad t = 2, \ldots, T \]

- We recover the structural results of previous models considered in literature
- **Systems where the encoder knows the decoder’s information**
  - Real-time source coding (Witsenhausen, 1979)
  - Real-time joint source-channel coding with noiseless feedback (Walrand and Varaiya, 1982)
- **Systems where the encoder does not know the decoder’s information**
  - Real-time joint source-channel coding with no feedback (Teneketzis, 2006)
Our Methodology

(ii) Global Optimization
**Global Optimization**

<table>
<thead>
<tr>
<th>Information at Encoder</th>
<th>Information at Decoder</th>
</tr>
</thead>
<tbody>
<tr>
<td>$i\mathcal{E}_t = \sigma(iE_t; i\phi^{t-1})$</td>
<td>$i\mathcal{R}_t = \sigma(iR_t; i\phi^{t-1})$</td>
</tr>
</tbody>
</table>
### Information at Encoder

\[ i\mathcal{E}_t = \sigma(i\mathcal{E}_t; i\phi^{t-1}) \]

### Information at Decoder

\[ i\mathcal{R}_t = \sigma(i\mathcal{R}_t; i\phi^{t-1}) \]

- Information is non nested. 

\[ i\mathcal{E}_t \not\subseteq i\mathcal{R}_t \quad i\mathcal{E}_t \not\supseteq i\mathcal{R}_t \]
Global Optimization

<table>
<thead>
<tr>
<th>Information at Encoder</th>
<th>Information at Decoder</th>
</tr>
</thead>
<tbody>
<tr>
<td>$i\mathcal{E}_t = \sigma(i\mathcal{E}_t; i\phi^{t-1})$</td>
<td>$i\mathcal{R}_t = \sigma(i\mathcal{R}_t; i\phi^{t-1})$</td>
</tr>
</tbody>
</table>

- Information is non nested. $i\mathcal{E}_t \not\subseteq i\mathcal{R}_t$ $i\mathcal{E}_t \not\supseteq i\mathcal{R}_t$

Aumann's notion on Common Knowledge

$ i\mathcal{K}_t := i\mathcal{E}_t \cap i\mathcal{R}_t $
Global Optimization

Information at Encoder

\[ iE_t = \sigma(iE_t; i\phi^{t-1}) \]

Information at Decoder

\[ iR_t = \sigma(iR_t; i\phi^{t-1}) \]

- Information is non nested.

\[ iE_t \not\subseteq iR_t \quad iE_t \not\supseteq iR_t \]

Aumann's notion on Common Knowledge

\[ iK_t := iE_t \cap iR_t \]

- Choose future decision rules based on common knowledge.
- Or, since we are only interested in performance, choose future decision rules based on common belief: \( P_{iK_t} \)
# Global Optimization

<table>
<thead>
<tr>
<th>Information at Encoder</th>
<th>Information at Decoder</th>
</tr>
</thead>
<tbody>
<tr>
<td>$iE_t = \sigma(iE_t; i\phi_{t-1})$</td>
<td>$iR_t = \sigma(iR_t; i\phi_{t-1})$</td>
</tr>
</tbody>
</table>

- Information is non nested. $iE_t \not\subseteq iR_t$ $iE_t \not\supseteq iR_t$

## Aumann’s notion on Common Knowledge

$\mathcal{K}_t := iE_t \cap iR_t$

- Choose future decision rules based on common knowledge.
- Or, since we are only interested in performance, choose future decision rules based on **common belief**: $P_{\mathcal{K}_t}$

- Need to find the set of all feasible realizations of common information (or common belief)
- $iE_t$ and $iR_t$ depend on past decision rules — so does $i\mathcal{K}_t$. 
Aumann's notion on Common Knowledge

\[ i\mathcal{K}_t := E_t \cap \mathcal{K}_t \]

- Can work with a super-set of the set of all feasible realizations of common knowledge
### Aumann’s notion on Common Knowledge


\[ i \mathcal{K}_t := i \mathcal{E}_t \cap i \mathcal{R}_t \]

- Can work with a super-set of the set of all feasible realizations of common knowledge
- Choose **Total Information:**  
  \[ i \mathcal{T}_t := \sigma(\mathcal{X}_t, i \mathcal{B}_t, i \mathcal{R}_t; i \phi_t) \supseteq i \mathcal{K}_t \]
- Set of all realizations depends on the past decision rules
Aumann’s notion on Common Knowledge

\[ iK_t := iE_t \cap iR_t \]

- Can work with a super-set of the set of all feasible realizations of common knowledge
- Choose **Total Information:** 
  \[ i\mathcal{T}_t := \sigma(X_t, iB_t, iR_t, i\phi_t) \supseteq iK_t \]
- Set of all realizations depends on the past decision rules

**The Image Space**

- \((X_t, iB_t, iR_t) : (\Omega, \mathcal{F}, P) \rightarrow (X \times iB \times iR, \mathcal{B}(X \times iB \times iR), P')\)
Global Optimization

Aumann’s notion on Common Knowledge

\[ i\mathcal{K}_t := i\mathcal{E}_t \cap i\mathcal{R}_t \]

- Can work with a super-set of the set of all feasible realizations of common knowledge
- Choose Total Information: \[ i\mathcal{T}_t := \sigma(X_t, iB_t, iR_t, i\phi_t) \supseteq i\mathcal{K}_t \]
- Set of all realizations depends on the past decision rules

The Image Space

- \((X_t, iB_t, iR_t) : (\Omega, \mathfrak{F}, P) \rightarrow (X \times i\mathcal{B} \times i\mathcal{R}, \mathfrak{B} (X \times i\mathcal{B} \times i\mathcal{R}), P')\)
  - Only the measure \(P'\) depends on \(i\phi^{t-1}\)
  - Hard to determine set of feasible realizations.
Aumann’s notion on Common Knowledge: 

\[ i\mathcal{K}_t := i\mathcal{E}_t \cap i\mathcal{R}_t \]

- Can work with a super-set of the set of all feasible realizations of common knowledge.
- Choose Total Information: 
  \[ i\mathcal{I}_t := \sigma(\mathcal{X}_t, i\mathcal{B}_t, i\mathcal{R}_t, i\phi_t) \supseteq i\mathcal{K}_t \]
- Set of all realizations depends on the past decision rules.

**The Image Space**

- \((\mathcal{X}_t, i\mathcal{B}_t, i\mathcal{R}_t) : (\Omega, \mathcal{F}, P) \rightarrow (\mathcal{X} \times i\mathcal{B} \times i\mathcal{R}, \mathcal{B}(\mathcal{X} \times i\mathcal{B} \times i\mathcal{R}), P')\)
  - Only the measure \(P'\) depends on \(i\phi^{t-1}\)
  - Hard to determine set of feasible realizations.
- Work with the set of all probability measures on 
  \((\mathcal{X} \times i\mathcal{B} \times i\mathcal{R}, \mathcal{B}(\mathcal{X} \times i\mathcal{B} \times i\mathcal{R}))\)
### Information States

\[
1\pi_t = \Pr \left( X_t, M_{t-1}, 1B_t \right),
\]

\[
2\pi_t = \Pr \left( X_t, Y_t, M_{t-1}, 2B_t \right),
\]

\[
3\pi_t = \Pr \left( X_t, Y_t, M_{t-1}, 3B_t \right).
\]
**Global Optimization**

**Information States**

\[ 1\pi_t = \Pr(X_t, M_{t-1}, 1B_t), \]
\[ 2\pi_t = \Pr(X_t, Y_t, M_{t-1}, 2B_t), \]
\[ 3\pi_t = \Pr(X_t, Y_t, M_{t-1}, 3B_t). \]

\[ \implies \]

**A Dynamic System**

\[ 2\pi_t = 1Q(c_t)1\pi_t, \]
\[ 3\pi_t = 2Q2\pi_t, \]
\[ 1\pi_{t+1} = 3Q(l_t)3\pi_t. \]

\[ E \left\{ \rho(X_t, \hat{X}_t) \mid c^t, g^t, l^{t-1} \right\} = 2\rho(2\pi_t, g_t) \]
Global Optimization

**Information States**

\[ 1\pi_t = \Pr(X_t, M_{t-1}, B_t), \]
\[ 2\pi_t = \Pr(X_t, Y_t, M_{t-1}, B_t), \]
\[ 3\pi_t = \Pr(X_t, Y_t, M_{t-1}, B_t). \]

\[ \implies \]

A Dynamic System

\[ 2\pi_t = 1Q(c_t)1\pi_t, \]
\[ 3\pi_t = 2Q2\pi_t, \]
\[ 1\pi_{t+1} = 3Q(l_t)3\pi_t. \]

\[ E\left\{\rho(X_t, \hat{X}_t) \mid c^t, g^t, t^{t-1}\right\} = 2\rho(2\pi_t, g_t) \]

**An equivalent problem**

Determine *meta-functions* \(1\Delta_t, 2\Delta_t, \text{and } 3\Delta_t\) such that:

\[ c_t = 1\Delta_t(1\pi_t), \quad 2\pi_t = 1Q(c_t)1\pi_t, \]
\[ g_t = 2\Delta_t(2\pi_t), \quad 3\pi_t = 2Q2\pi_t, \]
\[ l_t = 3\Delta_t(3\pi_t), \quad 1\pi_{t+1} = 3Q(l_t)3\pi_t. \]

To minimize a total cost

\[ J_T(\Delta^T, 1\pi_1) = \sum_{t=1}^{T} 2\rho(2\pi_t, g_t). \]
Sequential Decomposition

\[ 1V_t^{(1)} = \inf_{c \in \mathcal{C}} 2V_t^{(1)}Q(c)^{1\pi}, \]
\[ 2V_t^{(2)} = \min_{g \in \mathcal{G}} 2\rho^{(2\pi, g)} + 3V_t^{(2)}Q^{2\pi}, \]
\[ 3V_t^{(3)} = \min_{l \in \mathcal{L}} 1V_{t+1}^{(3)}Q(l)^{3\pi}. \]

The arg min at each step determines an optimal meta-function.
Sequential Decomposition

\[ \begin{align*}
1V_t(1\pi) &= \inf_{c \in C} 2V_t(1Q(c)1\pi), \\
2V_t(2\pi) &= \min_{g \in G} 2\rho(2\pi, g) + 3V_t(2Q2\pi), \\
3V_t(3\pi) &= \min_{l \in L} 1V_{t+1}(3Q(l)3\pi).
\end{align*} \]

The arg min at each step determines an optimal meta-function

- **Complexity**
  - Search complexity is linear in time (versus exponential in time for a brute force search)
  - Search complexity is exponential in the size of the alphabets
Extensions

- Infinite horizon
  - Expected discounted distortion
  - Average distortion per unit time
- Active feedback
- Channels with memory
- k-th order Markov source
- Distortion with d-step delay
CONCLUSION

◦ An alternative approach to real-time communication

◦ Use stochastic optimization
  (i) Derive structural results
  (ii) Use structural results for global optimization

◦ A systematic search algorithm to determine optimal design

FUTURE DIRECTIONS

◦ Performance bounds
◦ Numerical algorithms
◦ Multi-terminal systems
Thank You