Real-Time Communication

A Decision Theoretic Framework

Aditya Mahajan
Research Adviser: Prof. Teneketzis

Department of EECS,
University of Michigan,
Ann Arbor, MI

13 May, 2005

Why Real-Time Communication?

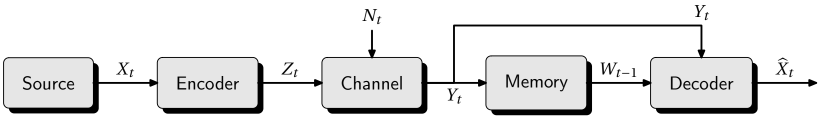
- Source output must be encoded, transmitted and decoded in real-time or small finite delay.
- Motivated by informationally decentralized systems:
  - QoS requirements in communication networks,
  - Sensor networks,
  - Traffic flow control in transportation networks,
  - Decentralized resource allocation

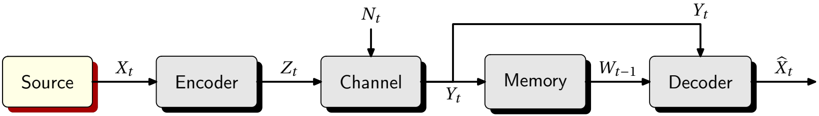
Why Real-Time Communication?

- Source output must be encoded, transmitted and decoded in real-time or small finite delay.
- Motivated by informationally decentralized systems:
  - QoS requirements in communication networks,
  - Sensor networks,
  - Traffic flow control in transportation networks,
  - Decentralized resource allocation

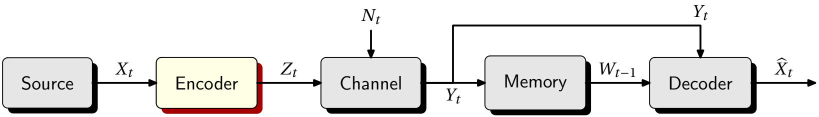
Difference from Information Theory

- Fundamental results of Information Theory are asymptotic.
- Encoding/Decoding based on typical sequences.
- Error exponents are of limited value for short block lengths.
- Encoding/Decoding schemes lead to long delays.


Source
X_t
Encoder
Z_t
N_t
Channel
Y_t
Y_t
Memory
W_{t-1}
Decoder
$\widehat{X}_t$

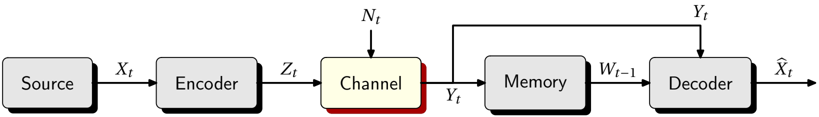


- Source is first order Markov with known statistics.



- Source is first order Markov with known statistics.
- Encoder is causal.

$$Z_t = f_t(X_1, X_2, \ldots, X_t)$$

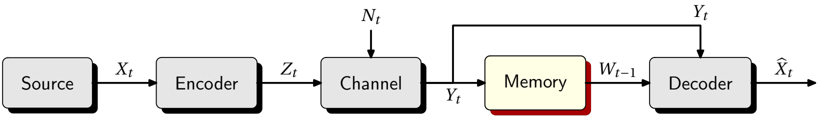


- Source is first order Markov with known statistics.
- Encoder is causal.

$$Z_t = f_t(X_1, X_2, \ldots, X_t)$$

- Discrete Memoryless Channel (not necessarily time-invariant), known statistics.

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



- Source is first order Markov with known statistics.
- Encoder is causal.

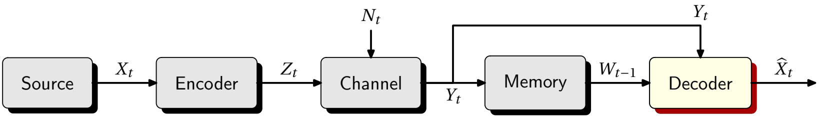
$$Z_t = f_t(X_1, X_2, \dots, X_t)$$

- Discrete Memoryless Channel (not necessarily time-invariant), known statistics.

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

- Finite memory receiver.

$$W_t = l_t(Y_t, W_{t-1})$$



- Source is first order Markov with known statistics.
- Encoder is causal.

$$Z_t = f_t(X_1, X_2, \ldots, X_t)$$

- Discrete Memoryless Channel (not necessarily time-invariant), known statistics.

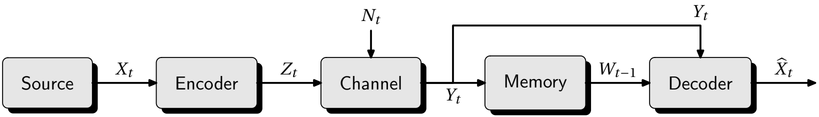
$$Y_t = h_t(Z_t, N_t)$$

- Finite memory receiver.

$$W_t = l_t(Y_t, W_{t-1})$$

- Real-time decoder.

$$\widehat{X}_t = g_t(Y_t, W_{t-1})$$



- Distortion measure

$$\rho_t : \mathscr{D} \times \mathscr{D} \to \mathbb{R}^+$$

- Average distortion at time t

$$\mathscr{J}_t = \mathbb{E}\{\rho_t(X_t, \widehat{X}_t)\}$$

- Design: Choice of $f \triangleq (f_1, f_2, \dots, f_T)$, $g \triangleq (g_1, g_2, \dots, g_T)$ and $l \triangleq (l_1, l_2, \dots, l_T)$.
- Performance measure

$$\mathscr{J}(f, g, l) = \sum_{t=1}^{T} \mathscr{J}_t = \mathbb{E}\left\{ \sum_{t=1}^{T} \rho_t(X_t, \widehat{X}_t) \right\}$$

Problem 1

Given

- statistics of the source
- statistics of the channel
- time horizon T

choose an optimal design (f^*, g^*, l^*) such that,

$$\mathscr{J}^* = \mathscr{J}(f^*, g^*, l^*) = \min_{(f,g,l)} \mathscr{J}(f, g, l)$$

PROBLEM 1

Given

- statistics of the source
- statistics of the channel
- time horizon T

choose an optimal design (f^*, g^*, l^*) such that,

$$\mathscr{J}^* = \mathscr{J}(f^*, g^*, l^*) = \min_{(f,g,l)} \mathscr{J}(f, g, l)$$

SALIENT FEATURES

- decentralized dynamic team problem
- non-classical information pattern — information of one agent depends on the decision rules of others.
- functional (rather than parameter) optimization problem.

Problem 1

Given

- statistics of the source
- statistics of the channel
- time horizon T

choose an optimal design (f^*, g^*, l^*) such that,

$$\mathscr{J}^* = \mathscr{J}(f^*, g^*, l^*) = \min_{(f,g,l)} \mathscr{J}(f, g, l)$$

Salient Features

- decentralized dynamic team problem
- non-classical information pattern — information of one agent depends on the decision rules of others.
- functional (rather than parameter) optimization problem.

What class does this problem belong to?

Standard Form for Sequential Control

H. S. Witsenhausen,
A standard form for sequential stochastic control.
Mathematical Systems Theory, vol. 7, no. 1, pp. 5–11, 1973.

Sequential Control Problem

- The ordering of control actions can be indexed in advance.
- The information available for action at time t depends at most up on the past control actions and realization of the random quantities involved.

Standard Form for Sequential Control

H. S. Witsenhausen,
A standard form for sequential stochastic control.
Mathematical Systems Theory, vol. 7, no. 1, pp. 5–11, 1973.

Sequential Control Problem

- The ordering of control actions can be indexed in advance.
- The information available for action at time t depends at most up on the past control actions and realization of the random quantities involved.

Our Problem as Sequential Control Problem

There are three agents — the encoder, decoder and memory update.

- At time t^+, the encoder chooses Z_t and transmits it.
- At time $(t + \frac{1}{2})$, the decoder chooses $\widehat{X}_t$.
- At time $(t + 1)^-$, the memory generates W_t.

Why not use the Standard Form?

- Standard form is applicable to all sequential control problems — so the solution is "too complicated".
- This problem has a rich structure. We can exploit this structure to come up with simpler solution.
- Standard form methodology can not be extended to infinite horizon problems.

WHY NOT USE THE STANDARD FORM?

- Standard form is applicable to all sequential control problems — so the solution is "too complicated".
- This problem has a rich structure. We can exploit this structure to come up with simpler solution.
- Standard form methodology can not be extended to infinite horizon problems.

What is this rich structure in our problem?

Information State — the Motivation

Suppose the memory update rule is fixed and decoder is fixed.

PRINCIPLE OF OPTIMALITY

$$\sum_{s=1}^{T} \mathbb{E}\{\rho_s(X_s, \widehat{X}_s)\} = \underbrace{\sum_{s=1}^{t-1} \mathbb{E}\{\rho_s(X_s, \widehat{X}_s)\}}_{\text{past cost already incurred}} + \underbrace{\sum_{s=t}^{T} \mathbb{E}\{\rho_s(X_s, \widehat{X}_s)\}}_{\text{cost to go}}$$

Cost to go depends only on $f_{t+1}, \dots, f_T$

MichiganEngineering

Information State — the Motivation

Suppose the memory update rule is fixed and decoder is fixed.

PRINCIPLE OF OPTIMALITY

$$\sum_{s=1}^{T} \mathbb{E}\{\rho_s(X_s, \widehat{X}_s)\} = \underbrace{\sum_{s=1}^{t-1} \mathbb{E}\{\rho_s(X_s, \widehat{X}_s)\}}_{\text{past cost already incurred}} + \underbrace{\sum_{s=t}^{T} \mathbb{E}\{\rho_s(X_s, \widehat{X}_s)\}}_{\text{cost to go}}$$

Cost to go depends only on $f_{t+1}, \dots, f_T$??
At time t, the encoder knows x^t, f^{t-1}.

$$\sum_{s=t}^{T} \mathbb{E}\{\rho_s(X_s, \widehat{X}_s)\} = \mathbb{E}\left\{ \sum_{s=t}^{T} \mathbb{E}\left\{ \rho_s(X_s, \widehat{X}_s) \,\middle|\, x^t, f^{t-1} \right\} \right\}$$

MichiganEngineering

π_t is an information state if (single agent system)

- $\pi_t = F_t(x^t, f^{t-1})$.
- $\pi_{t+1} = G_t(\pi_t, x_{t+1}, f_t)$.
- For $s \geq t$,
  $\mathbb{E}\{\rho_s(X_s, \widehat{X}_s) \mid X^t, f^{t-1}\} = \mathbb{E}\{\rho_s(X_s, \widehat{X}_s) \mid \pi_t\} = c_s(\pi_t)$

π_t is an information state if (single agent system)

- $\pi_t = F_t(x^t, f^{t-1})$.
- $\pi_{t+1} = G_t(\pi_t, x_{t+1}, f_t)$.
- For $s \geq t$,
  $\mathbb{E}\{\rho_s(X_s, \widehat{X}_s) \,|\, X^t, f^{t-1}\} = \mathbb{E}\{\rho_s(X_s, \widehat{X}_s) \,|\, \pi_t\} = c_s(\pi_t)$

Note: This definition is more general than the standard definition c.f.

P. R. Kumar and P. Varaiya,
Stochastic Systems: Estimation Identification and Adaptive Control.
Prentice Hall, 1986.

Information State

π_t is an information state if (single agent system)

- $\pi_t = F_t(x^t, f^{t-1})$.
- $\pi_{t+1} = G_t(\pi_t, x_{t+1}, f_t)$.
- For $s \geq t$,
  $\mathbb{E}\left\{\rho_s(X_s, \widehat{X}_s) \,\middle|\, X^t, f^{t-1}\right\} = \mathbb{E}\left\{\rho_s(X_s, \widehat{X}_s) \,\middle|\, \pi_t\right\} = c_s(\pi_t)$

Note: This definition is more general than the standard definition c.f.

P. R. Kumar and P. Varaiya,
Stochastic Systems: Estimation Identification and Adaptive Control.
Prentice Hall, 1986.

- $\pi_t = F_t(x^t, z^{t-1})$.
- $\pi_{t+1} = G_t(\pi_t, x_{t+1}, f_t(\pi_t))$.
- For $s \geq t$,
  $\mathbb{E}^f\left\{\rho_s(X_s, \widehat{X}_s) \,\middle|\, x^t, z^{t-1}\right\} = \mathbb{E}^{\not f}\left\{\rho_s(X_s, \widehat{X}_s) \,\middle|\, \pi_t\right\} = c_s(\pi_t)$

Implications of an Information State

- Absorbs/summarizes the effect of all the past actions and all the past decision rules on future decisions.
- The cost to go is a function of the information state.
- The decision rule is a function of information state.
- Can obtain a sequential decomposition of the problem.

Dynamic Program

$$V_{T+1}(\pi) = 0$$

$$V_t(\pi) = \min_{f_t} \Big\{ c_t(\pi) + \mathbb{E}\big\{ V_{t+1}\big(G_t(\pi, f_t)\big) \,\big|\, \pi \big\} \Big\}$$

Dynamic Program

$$V_{T+1}(\pi) = 0$$
$$V_t(\pi) = \min_{f_t} \left\{ c_t(\pi) + \mathbb{E}\left\{ V_{t+1}\left(G_t(\pi, f_t)\right) \,\middle|\, \pi \right\} \right\}$$

Backwards Recursion

- Set $t = T$.

Dynamic Program

$$V_{T+1}(\pi) = 0$$
$$V_t(\pi) = \min_{f_t} \Big\{ c_t(\pi) + \mathbb{E}\big\{ V_{t+1}\big(G_t(\pi, f_t)\big) \,\big|\, \pi \big\} \Big\}$$

Backwards Recursion

- Set $t = T$.
- For all π, evaluate $V_t(\pi)$ and f_t (or $f_t(\pi)$).

Dynamic Program

$$V_{T+1}(\pi) = 0$$

$$V_t(\pi) = \min_{f_t} \left\{ c_t(\pi) + \mathbb{E}\left\{ V_{t+1}\big(G_t(\pi, f_t)\big) \,\middle|\, \pi \right\} \right\}$$

Backwards Recursion

- Set $t = T$.
- For all π, evaluate $V_t(\pi)$ and f_t (or $f_t(\pi)$).
- Decrement t and repeat last step until $t = 1$.

Obtain $f_T, f_{T-1}, \ldots, f_1$ recursively.

D. Teneketzis.
On the structure of optimal real-time encoders and decoders in noisy communication.
submitted for publication in IEEE Trans. Inform. Theory.

Structure of Optimal Encoder

Fix (arbitrarily) g and l.

D. Teneketzis.
On the structure of optimal real-time encoders and decoders in noisy communication.
submitted for publication in IEEE Trans. Inform. Theory.

Structure of Optimal Encoder

Fix (arbitrarily) g and l. Then,

- Let $P_{W_t}(w) = \Pr\left(W_t = w \,\middle|\, X^t, Z^t, f^t, l^{t-1}\right)$ be the encoder's belief about the memory of the decoder. Then $P_{W_t} = \mu_t(P_{W_{t-1}}, Z_t; l_t)$

D. Teneketzis.
On the structure of optimal real-time encoders and decoders in noisy communication.
submitted for publication in IEEE Trans. Inform. Theory.

Structure of Optimal Encoder

Fix (arbitrarily) g and l. Then,

- Let $P_{W_t}(w) = \Pr\left(W_t = w \,\middle|\, X^t, Z^t, f^t, l^{t-1}\right)$ be the encoder's belief about the memory of the decoder. Then $P_{W_t} = \mu_t(P_{W_{t-1}}, Z_t; l_t)$

- The process $\{R_t \triangleq (X_t, P_{W_{t-1}}),\ t = 1, 2, \dots, T\}$ is an information state for the encoder.
- Allows sequential decomposition of the problem.

D. Teneketzis.
On the structure of optimal real-time encoders and decoders in noisy communication.
submitted for publication in IEEE Trans. Inform. Theory.

Structure of Optimal Encoder

Fix (arbitrarily) g and l. Then,

- Let $P_{W_t}(w) = \Pr\left(W_t = w \,\middle|\, X^t, Z^t, f^t, l^{t-1}\right)$ be the encoder's belief about the memory of the decoder. Then $P_{W_t} = \mu_t(P_{W_{t-1}}, Z_t; l_t)$

- The process $\{R_t \triangleq (X_t, P_{W_{t-1}}), t = 1, 2, \ldots, T\}$ is an information state for the encoder.
- Allows sequential decomposition of the problem.

- There is no loss of optimality in restricting attention to encoding rules of the form

$$Z_t = f_t(X_t, P_{W_{t-1}}), \quad t = 2, 3, \ldots, T$$

D. Teneketzis.
On the structure of optimal real-time encoders and decoders in noisy communication.
submitted for publication in IEEE Trans. Inform. Theory.

Structure of Optimal Decoder

Fix (arbitrarily) f and l.

D. Teneketzis.
On the structure of optimal real-time encoders and decoders in noisy communication.
submitted for publication in IEEE Trans. Inform. Theory.

Structure of Optimal Decoder

Fix (arbitrarily) f and l. Then

- Obtaining the optimal decoder is a filtering problem — At each t obtain g_t to minimize $\mathscr{J}_t$.
- An optimal decoding rule $g^* \triangleq (g_1^*, g_2^*, \dots, g_T^*)$ is given by

$$g_t^*(y_t, w_{t-1}) = \tau_t\big(\xi_t(y_t, w_{t-1})\big)$$

where

$$\xi_t^{f,l}(y, w)(x) = \Pr\left(X_t = x \,|\, Y_t = y, W_{t-1} = w\right)$$

and

$$\tau_t\big(\xi_t(y, w)\big) = \arg\min_a \sum_x \rho_t(x, a)\xi_t(y, w)(x)$$

Implication of Structural Results

1. Without loss of optimality we can restrict attention to encoders of the form $Z_t = f_t(X_t, P_{W_{t-1}})$.
2. Optimal decoder can be expressed in terms of encoder and memory update rule, i.e., $g^* = g^*(f, l) \implies$

$$\min_{(f,g,l)} \mathscr{J}(f, g, l) = \min_{f,l} \mathscr{J}\big(f, g^*(f, l), l\big)$$

Implication of Structural Results

1. Without loss of optimality we can restrict attention to encoders of the form $Z_t = f_t(X_t, P_{W_{t-1}})$.
2. Optimal decoder can be expressed in terms of encoder and memory update rule, i.e., $g^* = g^*(f, l) \implies$

$$\min_{(f,g,l)} \mathscr{J}(f, g, l) = \min_{f,l} \mathscr{J}\big(f, g^*(f, l), l\big)$$

Sequential Decomposition

No straight forward sequential decomposition of the problem.

- g_t depends on $f_1, f_2, \ldots, f_t$ and $l_1, l_2, \ldots, l_{t-1}$.
- Memory update does not have perfect recall — selective shedding of information.
- What is the information state of each agent?

Consider the random vectors

$$P_{W_t}(w) = \Pr\left(W_t = w \,\middle|\, X^t, Z^t, f^t, l^{t-1}\right)$$
$$P_{Y_t,W_{t-1}}(y,w) = \Pr\left(Y_t = y, W_{t-1} = w \,\middle|\, X^t, Z^t, f^t, l^{t-1}\right)$$

INFORMATION STATES

$$\pi_t = \Pr\left(X_t, P_{W_{t-1}}\right), \qquad \text{(Info. state for Encoder)}$$
$$\varphi_t = \Pr\left(X_t, P_{Y_t,W_{t-1}}\right), \qquad \text{(Info. state for Receiver)}$$

RECURSIVE UPDATES

There exist linear transformations $Q_t(f_t)$ and $\widehat{Q}_t(l_t)$ such that

- $\varphi_t = Q_t(f_t)\pi_t$.
- $\pi_{t+1} = \widehat{Q}_t(l_t)\varphi_t$.

Sequential Decomposition

For a given choice of f and l,

- π_t and φ_t are deterministic sequences.
- from the decoder's perspective,

$$\xi_t^{f,l}(y,w) = M(y,w,\varphi_t)$$

- from the encoder's perspective,

$$P_{\widehat{X}_t}(\hat{x}) = \Pr\left(\widehat{X}_t = \hat{x} \,\middle|\, X^t, Z^t, f^t, l^{t-1}\right)$$

  satisfies

$$P_{\widehat{X}_t} = \widehat{M}_t(\varphi_t)$$

- $\mathscr{J}_t = \mathbb{E}\left\{\rho_t(X_t, \widehat{X}_t)\right\} = c_t(\varphi_t)$

Theorem

An optimal design (f^*, l^*) for give problem is determined by the following nested optimality equations

$$\begin{aligned} V_{T+1}(\pi) &= 0 \\ V_t(\pi) &= \inf_{f_t \in \mathscr{F}_S} \left[c_t(Q_t(f_t)\pi) + \widehat{V}_t\big(Q_t(f_t)\pi\big) \right], \qquad t = 1, 2, \ldots, T \\ \widehat{V}_t(\varphi) &= \inf_{l_t \in \mathscr{L}} \left[V_{t+1}\big(\widehat{Q}_t(l_t)\varphi\big) \right], \qquad t = 1, 2, \ldots, T. \end{aligned}$$

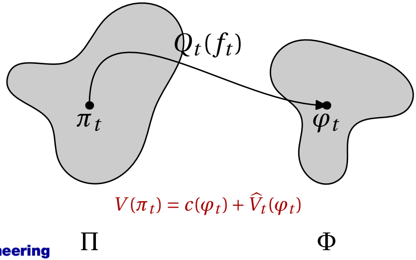
Theorem

An optimal design (f^*, l^*) for give problem is determined by the following nested optimality equations

$$V_{T+1}(\pi) = 0$$

$$V_t(\pi) = \inf_{f_t \in \mathscr{F}_S} \left[c_t(Q_t(f_t)\pi) + \widehat{V}_t\big(Q_t(f_t)\pi\big) \right], \qquad t = 1, 2, \ldots, T$$

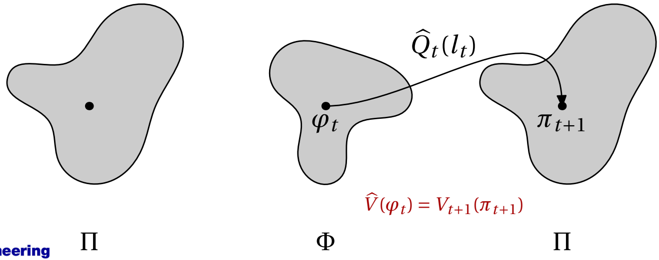
$$\widehat{V}_t(\varphi) = \inf_{l_t \in \mathscr{L}} \left[V_{t+1}\big(\widehat{Q}_t(l_t)\varphi\big) \right], \qquad t = 1, 2, \ldots, T.$$



Theorem

An optimal design (f^*, l^*) for give problem is determined by the following nested optimality equations

$$
\begin{aligned}
V_{T+1}(\pi) &= 0 \\
V_t(\pi) &= \inf_{f_t \in \mathscr{F}_S} \left[c_t(Q_t(f_t)\pi) + \widehat{V}_t\big(Q_t(f_t)\pi\big) \right], \qquad t = 1, 2, \ldots, T \\
\widehat{V}_t(\varphi) &= \inf_{l_t \in \mathscr{L}} \left[V_{t+1}\big(\widehat{Q}_t(l_t)\varphi\big) \right], \qquad t = 1, 2, \ldots, T.
\end{aligned}
$$



1. Time homogenous case.
2. Continuous state source and channel noise.
3. k-th order Markov Source.
4. Finite delay case.
5. Observation noise.
6. Infinite horizon discounted cost problem

Time Homogenous Case

- source transition matrix
- channel
- noise statistics
- distortion measure

are time invariant.

Time Homogenous Case

- source transition matrix
- channel
- noise statistics
- distortion measure

are time invariant. Then,

- The Dynamic program is time-homogenous.

Time Homogenous Case

- source transition matrix
- channel
- noise statistics
- distortion measure

are time invariant. Then,

- The Dynamic program is time-homogenous.
- Time-homogeneity is different from time-invariant.
- Simplifies computations.

Time Homogenous Case

- source transition matrix
- noise statistics
- channel
- distortion measure

are time invariant. Then,

- The Dynamic program is time-homogenous.
- Time-homogeneity is different from time-invariant.
- Simplifies computations.

Continuous State Source and Channel

- Source has continuous state.
- Channel has continuous state.

Results still hold in this case.

Sliding window Repackaging of Source

k-TH ORDER MARKOV SOURCE

- Can be converted into first-order Markov source.
- We can restrict attention to encoding rules of the form

$$Z_t = f_t(X_{t-k+1}, \ldots, X_t, P_{W_{t-1}}), \quad t = k, \ldots, T$$

- Can use same technical approach for joint optimization.

Sliding window Repackaging of Source

k-th order Markov Source

- Can be converted into first-order Markov source.
- We can restrict attention to encoding rules of the form

$$Z_t = f_t(X_{t-k+1}, \dots, X_t, P_{W_{t-1}}), \quad t = k, \dots, T$$

- Can use same technical approach for joint optimization.

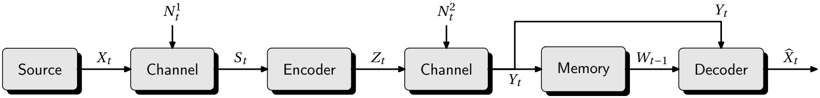
Finite Delay Case

Suppose the distortion measure can accept a delay of δ, i.e. for $t > \delta$ the distortion is given by $\rho_t(X_{t-\delta}, \widehat{X}_t)$.

- Can be converted into real-time distortion.
- We can restrict attention to encoding rules of the form

$$Z_t = f_t(X_{t-\delta+1}, \dots, X_t, P_{W_{t-1}}), \quad t = \delta, \dots, T$$

- Can use same technical approach for joint optimization.



- The encoder can observe noisy output of the source.
- Can be reduced to the problem with complete observation.

Time homogenous case where the design objective is to minimize

$$\mathscr{J}(f, g, l) = \mathbb{E}\left\{ \sum_{t=1}^{\infty} \beta^t \rho(X_t, \widehat{X}_t) \right\}$$

We show that without loss of optimality (in the sense of ε-optimality), we can restrict attention to stationary policies.

- We provide a decision theoretic framework to study real-time and finite-delay communication.
- Show how to use the structural results to obtain joint optimization.
- Solution methodology considerably simpler than "Standard Form".

Summary

- We provide a decision theoretic framework to study real-time and finite-delay communication.
- Show how to use the structural results to obtain joint optimization.
- Solution methodology considerably simpler than "Standard Form".

FUTURE WORK

- Extend the methodology to multi-terminal scenarios.
- "Simple" encoders and more complicated decoders.
- Performance bounds.
- Computational issues.

H. S. Witsenhausen,
On performance bounds for uncertain systems,
SIAM Journal of Control, vol. 8, pp. 55–89, 1970.