

*Fixed Delay Joint Source
Channel Coding for
Finite Memory Systems*

Aditya Mahajan and Demosthenis Teneketzis

Dept. of EECS, University of Michigan,
Ann Arbor, MI-48109

ISIT 2006–July 13, 2006

Fixed Delay & Fixed Complexity

Motivation

- Classical Information Theory does not take **delay and complexity** into account.
- Why consider delay and complexity?
- **Delay:**
 - QoS (end-to-end delay) in communication networks
 - Control over communication channels.
 - Decentralized detection in sensor networks.
- **Complexity:** (size of **lookup table**)
 - cost
 - power consumption

Finite Delay Communication

- Separation Theorem: distortion d is **feasible** is

Rate Distortion of Source $<$ Channel Capacity

$$R(d) < C$$

- For finite delay system **Separation Theorem** does not hold.
- What is equivalent of **rate distortion** and **channel capacity**?
- Find a metric to check whether **distortion level d is feasible** or not.
- Metric will depend on the source **and** the channel.

Problem Formulation

Objective

Evaluate optimal performance $R^{-1}(C)$ for the simplest non-trivial system

- Markov Source
- memoryless noisy channel
- additive distortion

Constraints

- Use stationary encoding and decoding schemes.
- Fixed memory available at the encoder and the decoder.

Model

- **Markov Source:**

- Source Output $\{X_1, X_2, \dots\}$, $X_n \in \mathcal{X}$.
- Transition probability matrix P

- **Finite State Encoder:**

- Input X_n , State S_n , Output Z_n

$$Z_n = f(X_n, S_{n-1}), \quad Z_n \in \mathcal{Z}$$

$$S_n = h(X_n, S_{n-1}), \quad S_n \in \mathcal{S}$$

- **Memoryless Channel:**

$$\Pr(Y_n | Z^n, Y^{n-1}) = \Pr(Y_n | Z_n) = Q(Y_n, Z_n)$$

- **Finite State Decoder:**

- Input Y_n , State M_n , Output \hat{X}_n

$$\hat{X}_n = g(Y_n, M_{n-1}), \quad \hat{X}_n \in \mathcal{X}$$

$$M_n = h(Y_n, M_{n-1}), \quad M_n \in \mathcal{M}$$

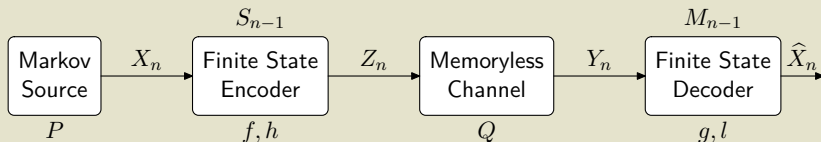
- **Distortion Metric:**

$$\rho : \mathcal{X} \times \mathcal{X} \rightarrow [0, K], \quad K < \infty$$

- **D step delay**

$$\rho(X_{n-D}, \hat{X}_n)$$

Problem Formulation



Problem (P1)

Given source (\mathcal{X}, P) , channel $(\mathcal{Z}, \mathcal{Y}, Q)$, memory $(\mathcal{S}, \mathcal{M})$ and distortion (ρ, D) , determine encoder (f, h) and decoder (g, l) so as to minimize

$$\mathcal{J}(f, h, g, l) \triangleq \limsup_{N \rightarrow \infty} \frac{1}{\tilde{N}} \mathbb{E} \left\{ \sum_{n=D+1}^N \rho(X_{n-D}, \hat{X}_n) \mid f, h, g, l \right\}$$

where $\tilde{N} = N - D + 1$

Literature Overview

- Transmitting Markovian source through finite-state machines as encoders and decoders.
- Problem considered by Gaarder and Slepian in mid 70's.
- N. T. Gaarder and D. Slepian
On optimal finite-state digital communication systems,
ISIT, Grignano, Italy, 1979
TIT, vol. 28, no. 2, pp. 167–186, 1982.

Our Approach

- Start with a simpler (to analyze) problem
 - Finite horizon
 - zero delay
 - time-varying design
- dynamic team problem—solved using Stochastic Optimization Techniques
- finite delay problem
- infinite horizon problem
- Find conditions under which time invariant (stationary) designs are optimal.
- Low complexity algorithms to obtain optimal performance and optimal design.

Finite Horizon Problem

Finite Horizon Case — Model

- Encoder and Tx Memory Update

$$Z_n = f_n(X_n, S_{n-1}) \quad f \triangleq (f_1, \dots, f_N)$$

$$S_n = h_n(X_n, S_{n-1}) \quad h \triangleq (h_1, \dots, h_N)$$

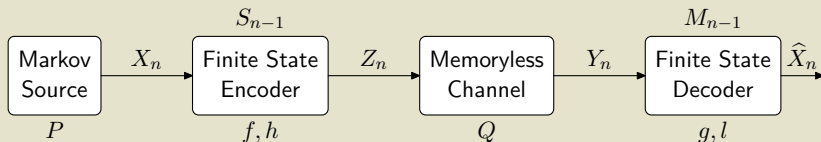
- Decoder and Rx Memory Update

$$\hat{X}_n = g_n(Y_n, M_{n-1}) \quad g \triangleq (g_1, \dots, g_N)$$

$$M_n = l_n(Y_n, M_{n-1}) \quad l \triangleq (l_1, \dots, l_N)$$

- Delay $D = 0$

Finite Horizon Problem Formulation



Problem (P2)

Given source (\mathcal{X}, P) , channel $(\mathcal{Z}, \mathcal{Y}, Q)$, memory $(\mathcal{S}, \mathcal{M})$, distortion $(\rho, D = 0)$ and horizon N , determine encoder (f, h) and decoder (g, l) so as to minimize

$$\mathcal{J}_N(f, h, g, l) \triangleq E \left\{ \sum_{n=1}^N \rho(X_n, \hat{X}_n) \mid f, h, g, l \right\}$$

where $f \triangleq (f_1, \dots, f_N)$, and so on for h, g, l .

Solution Concept in Seq. Stoch. Opt

- One Step Optimization

$$\min_{\substack{f_1, f_2, \dots, f_N \\ h_1, h_2, \dots, h_N \\ g_1, g_2, \dots, g_N \\ l_1, l_2, \dots, l_N}} \mathbb{E} \left\{ \sum_{n=1}^N \rho(X_n, \hat{X}_n) \mid f^N, h^N, g^N, l^N \right\}$$

- 4N Step Optimization—Sequential Decomposition

$$\min_{f_1} \left\{ \min_{g_1} \left\{ \min_{l_1} \left\{ \min_{h_1} \left\{ \dots \right. \right. \right. \right. \right. \\ \left. \left. \left. \left. \dots \min_{f_N} \left\{ \min_{g_N} \left\{ \min_{l_N} \left\{ \min_{h_N} \left\{ \square \right\} \right\} \dots \right\} \right\} \right\} \right\} \right\} \right\}$$

Dynamic Team Problems

- **Team Decision Theory:** distributed agents with common objective
 - Marshak and Radner
 - Witsenhausen
- Decentralized of information—encoder and decoder have different **view of the world**.
- Non-classical information pattern
- Non-convex functional optimization problem
- Most important step is **identifying information state**

Information State

If φ_{n-1} is the information state at n^- (and $\gamma = (f, h, g, l)$)

- **State** in the sense of

$$\longrightarrow \varphi_{n-1} \xrightarrow{T_{n-1}(\gamma_n)} \varphi_n \xrightarrow{T_n(\gamma_{n+1})} \varphi_{n+1} \longrightarrow$$

- **Absorbs** the effect of past decision rules on future performance.

$$\begin{aligned} E \left\{ \sum_{i=n}^N \rho(X_i, \hat{X}_i) \mid \gamma_1^N \right\} \\ = E \left\{ \sum_{i=n}^N \rho(X_i, \hat{X}_i) \mid \pi_{n-1}^0, \gamma_n^N \right\} \end{aligned}$$

*Find an information state
for Problem (P2)*

*Find an information state
for Problem (P2)*

Guess & Verify

Information State for (P2)

- Definition

$$\pi_n^1 \triangleq \Pr(X_n, Y_n, S_{n-1}, M_{n-1})$$

$$\pi_n^2 \triangleq \Pr(X_n, S_{n-1}, M_n)$$

$$\pi_n^0 \triangleq \Pr(X_n, S_n, M_n)$$

Information State for (P2)

- **Definition**

$$\pi_n^1 \triangleq \Pr(X_n, Y_n, S_{n-1}, M_{n-1})$$

$$\pi_n^2 \triangleq \Pr(X_n, S_{n-1}, M_n)$$

$$\pi_n^0 \triangleq \Pr(X_n, S_n, M_n)$$

Lemma

For all $n = 1, \dots, N$,

- there exist linear transforms T^0, T^1, T^2 such that

$$\longrightarrow \pi_{n-1}^0 \xrightarrow{T_{n-1}^0(f_n)} \pi_n^1 \xrightarrow{T_n^1(l_n)} \pi_n^2 \xrightarrow{T_n^2(h_n)} \pi_n^0 \longrightarrow$$

Information State for (P2)

- **Definition**

$$\pi_n^1 \triangleq \Pr(X_n, Y_n, S_{n-1}, M_{n-1})$$

$$\pi_n^2 \triangleq \Pr(X_n, S_{n-1}, M_n)$$

$$\pi_n^0 \triangleq \Pr(X_n, S_n, M_n)$$

Lemma (cont. . .)

For all $n = 1, \dots, N$,

- the expected instantaneous cost can be written as

$$\mathbb{E} \left\{ \rho(X_n, \hat{X}_n) \mid f^n, h^n, g^n, l^n \right\} = \tilde{\rho}(\pi_n^1, g_n)$$

Solution of (P2)

Dynamic Program

- For $n = 1, \dots, N$

$$V_{n-1}^0(\pi_{n-1}^0) = \min_{f_n} \{V_n^1(T^0(f_n)\pi_{n-1}^0)\},$$

$$V_n^1(\pi_n^1) = \bar{V}_n(\pi_n^1) + \min_{l_n} \{V_n^2(T^1(l_n)\pi_n^1)\},$$

$$\bar{V}_n(\pi_n^1) = \min_{g_n} \{\tilde{\rho}(\pi_n^1, g_n)\},$$

$$V_n^2(\pi_n^2) = \min_{h_n} \{V_n^0(T^2(h_n)\pi_n^2)\},$$

and

$$V_N^0(\pi_N^0) \triangleq 0.$$

Solution of (P2)

- The arg min at each step determines the corresponding optimal design rule.
- The optimal performance is given by

$$J_N^* = V_0^0(\pi_0^0)$$

- **Computations:** Numerical methods from Markov decision theory can be used.

Next steps . . .

Finite Delay Problem

- Delay $D \neq 0$
- Sliding window transformation of the source

$$\bar{X}_n = (X_{n-D}, \dots, X_n)$$

$$\bar{\rho}(\bar{X}_n, \hat{X}_n) = \rho(X_{n-D}, \hat{X}_n)$$

- Reduces to problem (P2).

Infinite Horizon Problem

- First consider delay $D = 0$
- Two related ways to making the horizon $N \rightarrow \infty$.
- **Expected Discounted Cost Problem**

$$J^\beta(f, h, g, l) \triangleq E \left\{ \sum_{n=1}^{\infty} \beta^{n-1} \rho(X_n, \hat{X}_n) \mid f, h, g, l \right\}$$

- **Average Cost Per Unit Time Problem**

$$\bar{J}(f, h, g, l) = \limsup_{N \rightarrow \infty} \frac{1}{N} E \left\{ \sum_{n=1}^N \rho(X_n, \hat{X}_n) \mid f, h, g, l \right\}$$

Expected Discounted Cost Problem

- $J^\beta(f, h, g, l) \triangleq E \left\{ \sum_{n=1}^{\infty} \beta^{n-1} \rho(X_n, \hat{X}_n) \mid f, h, g, l \right\}$

- Find a fixed point (V^0, V^1, \bar{V}, V^2) of

$$V^0(\pi^0) = \min_f \{V^1(T^0(f)\pi_{n-1}^0)\},$$

$$V^1(\pi^1) = \beta \bar{V}(\pi^1) + \min_l \{V^2(T^1(l)\pi^1)\},$$

$$\bar{V}(\pi^1) = \min_g \{\tilde{\rho}(\pi^1, g)\},$$

$$V^2(\pi^2) = \min_h \{V^0(T^2(h)\pi^2)\},$$

- Fixed point exists and is **unique** provided the distortion ρ is uniformly bounded.

Average Cost per Unit Time

- Average Cost

$$\bar{J}(f, h, g, l) = \limsup_{N \rightarrow \infty} \frac{1}{N} \mathbb{E} \left\{ \sum_{n=1}^N \rho(X_n, \hat{X}_n) \mid f, h, g, l \right\}$$

- Define

$$\gamma_n \triangleq (f, h, g, l)$$

$$\hat{T}(\gamma) \triangleq T^0(f) \circ T^1(l) \circ T^2(h)$$

$$\hat{\rho}(\pi_{n-1}^0, \gamma_n) \triangleq \tilde{\rho}(T^0(f_n)\pi_{n-1}^0, g_n)$$

Average Cost per Unit Time

- **Assumption (A1):** for some $\epsilon > 0$ there exist bounded measurable functions $v(\cdot)$ and $r(\cdot)$ and design γ_0 such that for all π^0

$$v(\pi^0) = \min_{\gamma} \left\{ v(\widehat{T}(\gamma)\pi^0) \right\} = v(\widehat{T}(\gamma_0)\pi^0)$$

$$\begin{aligned} \min_{\gamma} \left\{ \widehat{\rho}(\pi^0, \gamma) + r(\widehat{T}(\gamma)\pi^0) \right\} &\leq v(\pi^0) + r(\pi^0) \\ &\leq \widehat{\rho}(\pi^0, \gamma_0) + r(\widehat{T}(\gamma_0)\pi^0) + \epsilon \end{aligned}$$

Average Cost per Unit Time

- If (A1) holds then, $\gamma_0^\infty \triangleq (\gamma_0, \gamma_0, \dots)$ is ϵ -optimal, that is, for any other design γ'

$$\bar{J}(\gamma_0^\infty) = v(\pi_0^0) \leq \underline{J}(\gamma') + \epsilon$$

where

$$\underline{J}(\gamma') \triangleq \liminf_{N \rightarrow \infty} \frac{1}{N} \sum_{n=1}^N \hat{p}(\pi_{n-1}^0, \gamma'_n).$$

- Conditions sufficient to ensure (A1) are known.
- Not easy to translate them into conditions on the problem

Some Comments

- For the expected discounted cost problem, time-invariant designs are optimal.
- For the average cost per unit time problem, time-invariant designs are optimal under certain conditions.
- The two problems are related via a Tauberian theorem

$$\begin{aligned} \liminf_{n \rightarrow \infty} \frac{\sum_{i=1}^n a_i}{n} &\leq \liminf_{\beta \rightarrow 1^-} (1 - \beta) \sum_{i=1}^{\infty} \beta^{i-1} a_i \\ &\leq \limsup_{\beta \rightarrow 1^-} (1 - \beta) \sum_{i=1}^{\infty} \beta^{i-1} a_i \leq \limsup_{n \rightarrow \infty} \frac{\sum_{i=1}^n a_i}{n} \end{aligned}$$

Solution Framework . . .

General Methodology

- Given source (\mathcal{X}, P) , channel $(\mathcal{Z}, \mathcal{Y}, Q)$, memory $(\mathcal{S}, \mathcal{M})$ and distortion (ρ, D) .
- Convert to zero delay problem
- Find ϵ -optimal design and performance for the discounted cost problem for β close to 1.
- This can be done using a polynomial complexity algorithms.
- The resultant design is ϵ -optimal for the average cost per unit time problem (if an ϵ -optimal design for the average cost per unit time problem exists)

Some Interesting Cases

- **Fixed Delay Source Coding Problem**
- The technique presented here can be extended to non-stochastic min-max problems.
- Used to study fixed delay encoding/decoding of individual sequences.
- Interesting to compare the results with “standard” fixed delay source coding techniques.

Some Interesting Cases

- **Fixed Delay Channel Coding Problem**
- Fixed delay decoding of convolutional codes.
- Most researchers focus on computationally efficient algorithms to determine the MAP bit decoding rule.
- The problem of efficiently **storing** the observations has not been considered.
- If receiver memory $|\mathcal{M}| = k|\mathcal{Y}|$, should one store the previous k channel observations?
- Can all the past observations be **compressed** in $k|\mathcal{Y}|$ to get better performance.
- How can such “compression” functions be found.
- This problem fits naturally in the framework presented here.

Conclusion.

Summary

- Consider fixed delay, fixed complexity communication system
- Markov source and noisy memoryless channel
- Objective: Minimize total (or discounted or average) distortion
- Provide a systematic methodology for determining optimal encoding–decoding strategies and optimal performance
- There exist low complexity algorithms to find such solutions
- Interesting special cases of the framework

Thank You

Gaarder and Slepian's Approach

- Fix a design (f, h, g, l) .
- $\{X_{n-D}^n, S_n, Z_n, Y_n, M_n, \hat{X}_n\}$ forms a Markov chain.
- Find its **steady-state distribution**.
- Find the **steady-state distortion**

$$\lim_{n \rightarrow \infty} E \left\{ \rho(X_{n-D}, \hat{X}_n) \right\}.$$

- **Cezáro Mean:** For any sequence of real numbers (a_n) ,

$$\text{If } \lim_{n \rightarrow \infty} a_n = a \quad \text{then} \quad \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n a_i = a$$

- Repeat for all designs (f, h, g, l) .

Gaarder and Slepian's Approach

- **Difficulty:** Evaluating asymptotic (steady-state) performance is difficult.

$$\min_{f,h,g,l} \lim_{n \rightarrow \infty} E \left\{ \rho(X_{n-D}, \hat{X}) \right\}$$

- "A sore point here is the very complicated way in which the stationary distribution of a Markov chain depends on the elements of its transition matrix "
- The matrix elements change discontinuously with a change in design (f, h, g, l).
- The resultant Markov chain can have several recurrence classes, be periodic, have several transient states etc., depending on the nature of the design (f, h, g, l).