Markov Chains, Feedback, and Directed Information

Todd P. Coleman
UIUC
Department of ECE
ITMANET Focus Talk
May 24, 2010
Outline

• Formulation and solution of a decentralized control problem
  – communication over networks with feedback
    • Memoryless/broadcast/finite-state channels
  – Causal joint source-channel coding of a Markov process with feedback

• Generalizes “to code or not to code” results to random process when feedback is present

• Relates Markov chain time reversibility as a sufficient property to attain fundamental limits
Motivation

- Links control/thermo/info-theory
- Gives new insights into solved problems and provides solutions to new ones
  - Potential for insight into coding with noisy feedback
  - Networks and feedback
- Importance of formulating problems carefully, as compared to trying to solve other ones
Problem setup

- $W$ is a Markov process
Problem setup

- $W$ is a Markov process
- Causal encoder $X_i = \mu_i(W^i, Y^{i-1})$
- Causal decoder $Z_i = d_i(Y^i)$
Problem setup

- $W$ is a Markov process
- Causal encoder $X_i = \mu_i(W_i, Y_{i-1})$
- Causal decoder $Z_i = d_i(Y^i)$
- Cost: $J_n(\mu, d) = \mathbb{E}_{\mu,d} \left[ \frac{1}{n} \sum_{i=1}^{n} J(W_i, Z_{i-1}, Z_i, X_i) \right]$
- Objective: find optimal causal encoder/decoders

\[ J_n(\mu^*, d^*) \leq J_n(\mu, d) \quad \text{for all } \mu, d \]
Main Results (Gorantla, Coleman ‘10)

Main Theorem: there always exist optimal \((\mu^*, d^*)\) policies of form

\[
\mu_i^*(W_i, Y_{i-1}) \equiv \tilde{\mu}_i^*(W_i, B_{i|i-1}, Z_{i-1})
\]

\[
Z_i = d_{i}^{\tilde{\mu}^*}(y^i) \equiv \tilde{d}_i^*(Z_{i-1}, B_{i|i})
\]
Main Theorem: there always exist optimal $(\mu^*, d^*)$ policies of form

\[ \mu^*_i(W_i, Y_{i-1}) \equiv \tilde{\mu}^*_i(W_i, B_{i|i-1}, Z_{i-1}) \]
\[ Z_i = d^*_i(y^i) \equiv \tilde{d}^*_i(Z_{i-1}, B_{i|i}) \]

- A form of separation theorem (sufficient statistics)
- $W$ need not be stationary nor ergodic (mobility)
- Proof uses 2D dynamic programming argument
Relationship with Other Results

Relationship with Other Results


**Differences**

- Cost function
  \[ J_n(\mu, d) = \mathbb{E}_{\mu,d} \left[ \frac{1}{n} \sum_{i=1}^{n} J(W_i, Z_{i-1}, Z_i, X_i) \right] \]
- \( W \) need not be discrete; \( Z \) need not be \( W \)
Example of the importance of this key difference

\[
I(W_0; Y^n) = \sum_{i=1}^{n} \mathbb{E} \left[ \log \frac{f_{W_0|Y^i}(W_0|Y^i)}{f_{W_0|Y^{i-1}}(W_0|Y^{i-1})} \right] \\
= \sum_{i=1}^{n} \mathbb{E} \left[ \log \frac{dB_{0|i}(W_0)}{dB_{0|i-1}(W_0)} \right]
\]
Example of the importance of this key difference

$$I(W_0; Y^n) = \sum_{i=1}^{n} \mathbb{E} \left[ \log \frac{f_{W_0|Y^i}(W_0|Y^i)}{f_{W_0|Y^{i-1}}(W_0|Y^{i-1})} \right]$$

$$= \sum_{i=1}^{n} \mathbb{E} \left[ \log \frac{d B_{0|i}}{d B_{0|i-1}}(W_0) \right]$$
Example of the importance of this key difference

\[ I(W_0; Y^n) = \sum_{i=1}^{n} \mathbb{E} \left[ \log \frac{f_{W_0|Y^i}(W_0|Y^i)}{f_{W_0|Y^{i-1}}(W_0|Y^{i-1})} \right] \]
\[ = \sum_{i=1}^{n} \mathbb{E} \left[ \log \frac{dB_0|i}{dB_0|i-1}(W_0) \right] \]

- **Alphabets** \( \mathcal{W} = [0, 1], \mathcal{Z} = \mathcal{M}([0, 1]) \)
- **W is Markov** \( W_i = W_{i-1} : i \geq 1 \), \( W_0 \sim \text{unif}[0, 1] \)
- **Example Cost:** \( J(w_i, z_{i-1}, z_i, x_i) = -\log \frac{dz_i}{dz_{i-1}}(w_i) + \lambda \eta(x) \)
Posterior Matching Scheme (Shayevitz-Feder ‘09)

\[ F_n(\cdot) \triangleq F_{W|Y^n}(\cdot|Y^n) \]

\[ X_{n+1} = F_X^{-1}\left(\overbrace{F_n(W)}^{\text{what is missing}}\right) \]
Posterior Matching Scheme (Shayevitz-Feder ‘09)

- Non-standard notion of communication
  - Message point $W$ on $[0,1]$ line
  - No block length, no forward error correction
  - Achieves capacity on arbitrary memoryless channels
  - Attains Shannon’s converse $\frac{1}{n}I(W;Y^n) = C$ for all $n$!
Observation: \[ I(W_0 \rightarrow Y^n) = \sum_{i=1}^{n} \mathbb{E} \left[ \log \frac{dB_{0|i}}{dB_{0|i-1}}(W_i) \right] \]
**Explanation from our Viewpoint**

\[ J(w_i; z_{i-1}, z_i; x_i) = -\log \frac{dz_i}{dz_{i-1}}(w_i) + \lambda \eta(x) \]

- **Observation:**
  \[ I(W_0 \to Y^n) = \sum_{i=1}^{n} \mathbb{E} \left[ \log \frac{dB_{0|i}}{dB_{0|i-1}}(W_i) \right] \]

- **Cost setup:**

![Diagram](image)
Explanation from our Viewpoint

\[ J(w_i; z_{i-1}; z_i; x_i) = -\log \frac{dz_i}{dz_{i-1}}(w_i) + \lambda \eta(x) \]

- Observation:
  \[ I(W_0 \rightarrow Y^n) = \sum_{i=1}^{n} \mathbb{E} \left[ \log \frac{dB_{0|i}}{dB_{0|i-1}}(W_i) \right] \]

- Cost setup:
  \[ J(w_i, z_{i-1}, z_i, x_i) = -\log \frac{dz_i}{dz_{i-1}}(w_i) + \lambda \eta(x) \]

- Lemma: for \((W_i = W_{i-1} : i \geq 1)\)

- Posterior Matching explained from suff. stat thm:
  \[ X_i = F_X^{-1} \left( F_{W_0|Y^{i-1}}(W_0|Y^{i-1}) \right) = \mu_i^* \left( W_0, B_{0|i-1} \right) \]
sequential joint source channel coding w/ feedback

- W non-stationary: $W_i = W_{i-1} + \tilde{W}_i$
- LTI encoder/decoder pair:

  $X_i = \tilde{X}_i = X_{i-1} + c_2 Z_{i-1} + c_3 Y_{i-1}$

  $Z_i = Z_{i-1} + \tilde{c}_3 Y_i$

Objective

$$J(w_i, z_{i-1}, z_i) = \rho(w_i, z_{i-1}, z_i) + \lambda \eta(x_i)$$
sequential joint source channel coding w/ feedback

• W non-stationary: \( W_i = W_{i-1} + \tilde{W}_i \)
• LTI encoder/decoder pair:
  \[
  X_i = X_{i-1} + c_2 Z_{i-1} + c_3 Y_{i-1}
  \]
  \[
  Z_i = Z_{i-1} + \tilde{c}_3 Y_i
  \]

Objective

\[
J(w_i, z_{i-1}, z_i) = \rho(w_i, z_{i-1}, z_i) + \lambda \eta(x_i)
\]

Sequential rate distortion function (Tatikonda PhD, 06)

\[
\mathbf{R}_n(D) \triangleq \inf_{P_{Z_i|Z_{i-1},W_{i-1}}:E[\rho] \leq D} \frac{1}{n} I(W^n \rightarrow Z^n)
\]

Fundamental bounds

\[
\mathbf{R}_n(D) \leq \frac{1}{n} I(W^n \rightarrow Z^n) \leq C(\eta, b)
\]

If tight, then system is optimal
**Sequential joint source channel coding w/ feedback**

![Diagram](image)

- W non-stationary: \( W_i = W_{i-1} + \tilde{W}_i \)
- LTI encoder/decoder pair:
  \[
  X_i = X_{i-1} + c_2 Z_{i-1} + c_3 Y_{i-1} \\
  \hat{X}_i \\
  Z_i = Z_{i-1} + \tilde{c}_3 Y_i
  \]

**Objective**

\[
J(w_i, z_{i-1}, z_i) = \rho(w_i, z_{i-1}, z_i) + \lambda \eta(x_i)
\]

**Sequential rate distortion function (Tatikonda PhD, 06)**

\[
\overrightarrow{R}_n(D) \triangleq \inf_{P_{Z_i|Z_{i-1}, W_{i-1}}: E[\rho] \leq D} \frac{1}{n} I(W^n \to Z^n)
\]

**Fundamental bounds**

\[
\overrightarrow{R}_n(D) \leq \frac{1}{n} I(W^n \to Z^n) \leq C(\eta, b)
\]

If tight, then system is optimal.

**Theorem 2 (Gorantla, Coleman ‘10)**

If the LTI enc-dec pair \((\mu, d)\) results in \(X_i\) indep \(Y_{i-1}\) then it is optimal for

(a.) \( \eta(x) = \alpha_1 D(P_{Y|X=x} || P_Y) + \eta_0 \)

(b.) \( \rho(w, z, z') = -\alpha_2 \log P_{Y|X} \left( \left( \frac{z-z'}{c_3} \right) || (w - z') \right) + \rho'(w) \)
Objective \( J(w, z_{i-1}, z_i) = \rho(w, z_{i-1}, z_i) + \lambda \eta(x_i) \)

Sequential rate distortion function (Tatikonda PhD, 06)
\[
\overline{R}_n(D) \triangleq \inf_{P_{Z_i|Z_{i-1},W_{i-1}:E[\rho]\leq D}} \frac{1}{n} I(W^n \rightarrow Z^n)
\]

Theorem 2 (Gorantla, Coleman ‘10)
If the LTI enc-dec pair \((\mu, d)\) results in \(X_i\) indep \(Y^{i-1}\) then it is optimal for
(a.) \(\eta(x) = \alpha_1 D(P_{Y|X=x}||P_Y) + \eta_0\)
(b.) \(\rho(w, z, z') = -\alpha_2 \log P_{Y|X} \left( \left( \frac{z-z'}{c_3} \right) | (w - z') \right) + \rho'(w)\)

Corollary: If \((X_i, \tilde{X}_i)\) is reversible then \(X_i\) indep \(Y^{i-1}\), Theorem 2 holds, and
\[
\frac{1}{n} I(W^n \rightarrow Z^n) = C(\eta, b)
\]
Theoretical Unification

• Simple and generalizable proof of
Theoretical Unification

- Simple and generalizable proof of

- Generalizes *Gastpar’s “to code or not to code”* result about source-channel measure matching to consider **Markov processes and feedback**
Theoretical Unification

• Simple and generalizable proof of

• Generalizes *Gastpar’s “to code or not to code”* result about source-channel measure matching to consider *Markov processes and feedback*

• Interesting new consequences:
  • Uncoded communication of a Poisson process over an exponential server timing channel is optimal (Gorantla, Coleman, *ISIT* 2010).
  • Analogous statement for Blackwell’s chemical “trapdoor” channel
Future Paradigms this Can Enable

- Extension to networks

S. K. Gorantla and T. P. Coleman, "A Stochastic Control Approach to Coding with Feedback over Degraded Broadcast Channels", IEEE Conference on Decision and Control (CDC), submitted April 2010
Future Paradigms this Can Enable

• Extension to networks

S. K. Gorantla and T. P. Coleman, "A Stochastic Control Approach to Coding with Feedback over Degraded Broadcast Channels", *IEEE Conference on Decision and Control (CDC)*, submitted April 2010

• Dynamical systems viewpoint on general feedback coding
  **Arbitrary (not nec. memoryless) channels; replaces info. density w/ Lyapunov exponent density**
  **Cts. alphabets; exploits Lyapunov exponent of PM scheme and continuity of distortion measure**
Future Paradigms this Can Enable

- The $64,000 question: **noisy** feedback
  - Insight: recent paper identifies suff. statistics
Future Paradigms this Can Enable

• The $64,000 question: **noisy** feedback
  – Insight: recent paper identifies suff. statistics

• Augment their approach to be PM-like:
  – Message point on [0,1]
  – Cost function that encodes *information*

\[
J_n(\mu, d) = \mathbb{E}_{\mu, d} \left[ \frac{1}{n} \sum_{i=1}^{n} J(W_i, Z_{i-1}, Z_i, X_i) \right]
\]
New Theoretical Avenues

• Connection btwn thermodynamics, directed information, game theory, and decentralized control
  – related: see poster tomorrow
    • T. P. Coleman, M. Raginsky, “Mutual information saddle points for channels of exponential family type”, to appear, ISIT 2010
New Theoretical Avenues

• Connection btwn thermodynamics, directed information, game theory, and decentralized control
  – related: see poster tomorrow
    • T. P. Coleman, M. Raginsky, “Mutual information saddle points for channels of exponential family type”, to appear, ISIT 2010

• Studies of communication through the state of interacting stochastic dynamical systems
  – Resonates w/ Tom Cover’s viewpoint:
  – networks with interacting agents, whose observations and decisions are correlated through their sequential actions
  – what is set of all possible joint distributions?