Distortion-transmission trade-off in real-time transmission of Markov sources

Jhelum Chakravorty and Aditya Mahajan

McGill University

IEEE Information Theory Workshop (ITW)
28 April, 2015
The system model

Markov Source $X_t$ Transmitter $U_t$ Receiver $Y_t$ $\hat{X}_t$
The system model

\[ X_{t+1} = X_t + W_t \]
The system model

\[ Y_t = \begin{cases} X_t, & \text{if } U_t = 1 \\ \epsilon, & \text{if } U_t = 0 \end{cases} \]

\[ X_{t+1} = X_t + W_t \]

\[ U_t = f_t(X_{1:t}, U_{1:t-1}) \]
The system model

\[ Y_t = \begin{cases} X_t, & \text{if } U_t = 1 \\ \varepsilon, & \text{if } U_t = 0 \end{cases} \]

\[ X_{t+1} = X_t + W_t \]

\[ U_t = f_t(X_{1:t}, U_{1:t-1}) \]

\[ \hat{X}_t = g_t(Y_{1:t}) \]
The system model

\[ Y_t = \begin{cases} X_t, & \text{if } U_t = 1 \\ \varepsilon, & \text{if } U_t = 0 \end{cases} \]

\[ X_{t+1} = X_t + W_t \]

\[ U_t = f_t(X_{1:t}, U_{1:t-1}) \]

\[ \hat{X}_t = g_t(Y_{1:t}) \]

Communication Strategies

- Transmission strategy \( f = \{f_t\}^\infty_{t=0} \).
- Estimation strategy \( g = \{g_t\}^\infty_{t=0} \).
The system model

$Y_t = \begin{cases} X_t, & \text{if } U_t = 1 \\ \varepsilon, & \text{if } U_t = 0 \end{cases}$

1. Discounted setup, $\beta \in (0, 1)$

$$
D_\beta(f, g) = (1 - \beta) \mathbb{E}_0^{(f,g)} \left[ \sum_{t=0}^{\infty} \beta^t d(X_t - \hat{X}_t) \right]; \quad N_\beta(f, g) = (1 - \beta) \mathbb{E}_0^{(f,g)} \left[ \sum_{t=0}^{\infty} \beta^t U_t \right]
$$

$X_{t+1} = X_t + W_t$

$U_t = f_t(X_{1:t}, U_{1:t-1})$

$\hat{X}_t = g_t(Y_{1:t})$
The system model

\[ Y_t = \begin{cases} X_t, & \text{if } U_t = 1 \\ \varepsilon, & \text{if } U_t = 0 \end{cases} \]

Distortion
\[ d(X_t, \hat{X}_t) \]

\[ X_{t+1} = X_t + W_t \quad U_t = f_t(X_{1:t}, U_{1:t-1}) \quad \hat{X}_t = g_t(Y_{1:t}) \]

1. Discounted setup, \( \beta \in (0, 1) \)
\[
D_\beta(f, g) = (1 - \beta) \mathbb{E}_0^{(f, g)} \left[ \sum_{t=0}^{\infty} \beta^t d(X_t - \hat{X}_t) \right]; \quad N_\beta(f, g) = (1 - \beta) \mathbb{E}_0^{(f, g)} \left[ \sum_{t=0}^{\infty} \beta^t U_t \right]
\]

2. Average cost setup, \( \beta = 1 \)
\[
D_1(f, g) = \limsup_{T \to \infty} \frac{1}{T} \mathbb{E}_0^{(f, g)} \left[ \sum_{t=0}^{T-1} d(X_t - \hat{X}_t) \right]; \quad N_1(f, g) = \limsup_{T \to \infty} \frac{1}{T} \mathbb{E}_0^{(f, g)} \left[ \sum_{t=0}^{T-1} U_t \right]
\]
Optimization problems

Costly communication

For any $\lambda \in \mathbb{R}_{>0}$, $C^*_\beta(\lambda) = C_\beta(f^*, g^*; \lambda) := \inf_{(f, g)} \{D_\beta(f, g) + \lambda N_\beta(f, g)\}$

Constrained communication

For any $\alpha \in (0, 1)$, $D^*_\beta(\alpha) := \inf_{(f, g)} \{D_\beta(f, g) : N_\beta(f, g) \leq \alpha\}$
Optimization problems

Costly communication

For any $\lambda \in \mathbb{R}_{>0}$, $C^*_\beta(\lambda) = C_\beta(f^*, g^*; \lambda) \coloneqq \inf_{(f, g)} \{D_\beta(f, g) + \lambda N_\beta(f, g)\}$

Constrained communication

For any $\alpha \in (0, 1)$, $D^*_\beta(\alpha) \coloneqq \inf_{(f, g)} \{D_\beta(f, g) : N_\beta(f, g) \leq \alpha\}$
Optimization problems

Costly communication

For any $\lambda \in \mathbb{R}_{>0}$, $C^*_\beta(\lambda) = C_\beta(f^*, g^*; \lambda) := \inf_{(f, g)} \{D_\beta(f, g) + \lambda N_\beta(f, g)\}$

Constrained communication

For any $\alpha \in (0, 1)$, $D^*_\beta(\alpha) := \inf_{(f, g)} \{D_\beta(f, g) : N_\beta(f, g) \leq \alpha\}$

$C^*_\beta$ is cts, inc, and concave
Optimization problems

Costly communication
For any \( \lambda \in \mathbb{R}_{>0} \),
\[
C^*_\beta(\lambda) = \inf_{(f,g)} \{ D_\beta(f, g) + \lambda N_\beta(f, g) \}
\]

Constrained communication
For any \( \alpha \in (0, 1) \),
\[
D^*_\beta(\alpha) := \inf_{(f,g)} \{ D_\beta(f, g) : N_\beta(f, g) \leq \alpha \}
\]

\( C^*_\beta \) is cts, inc, and concave
**Optimization problems**

**Costly communication**

For any $\lambda \in \mathbb{R}_{>0}$, $C^*_\beta(\lambda) = C_\beta(f^*, g^*; \lambda) := \inf_{(f,g)} \{D_\beta(f, g) + \lambda N_\beta(f, g)\}$

**Constrained communication**

For any $\alpha \in (0, 1)$, $D^*_\beta(\alpha) := \inf_{(f,g)} \{D_\beta(f, g) : N_\beta(f, g) \leq \alpha\}$

- $C^*_\beta$ is cts, inc, and concave
- $D^*_\beta$ is cts, dec, and convex
### Optimization problems

#### Costly communication

For any \( \lambda \in \mathbb{R}_{>0} \),

\[
C_\beta^*(\lambda) = C_\beta(f^*, g^*; \lambda) := \inf_{(f, g)} \{D_\beta(f, g) + \lambda N_\beta(f, g)\}
\]

#### Constrained communication

For any \( \alpha \in (0, 1) \),

\[
D_\beta^*(\alpha) := \inf_{(f, g)} \{D_\beta(f, g) : N_\beta(f, g) \leq \alpha\}
\]

\( C_\beta^* \) is cts, inc, and concave

\( D_\beta^* \) is cts, dec, and convex

Distortion-transmission trade-off

(Chakravorty and Mahajan)
Optimization problems

Costly communication

For any $\lambda \in \mathbb{R}_{>0}$, 
\[ C^*_\beta(\lambda) = C_\beta(f^*, g^*; \lambda) := \inf_{(f, g)} \{ D_\beta(f, g) + \lambda N_\beta(f, g) \} \]

Constrained communication

For any $\alpha \in (0, 1)$, 
\[ D^*_\beta(\alpha) := \inf_{(f, g)} \{ D_\beta(f, g) : N_\beta(f, g) \leq \alpha \} \]

We provide explicit computable expressions for both curves

Distortion-transmission trade-off

$C^*_\beta$ is concave

$D^*_\beta$ is convex
Comparison to Information Theory

- Costly communication is analogous to communication under power constraint.
- Distortion-transmission is analogous to distortion-rate trade-off.
Comparison to Information Theory

- **Costly communication** is analogous to communication under power constraint.
- **Distortion-transmission** is analogous to distortion-rate trade-off.
- The source reconstruction must be done in **real-time** (or with zero delay).
Comparison to Information Theory

- **Costly communication** is analogous to **communication under power constraint**.
- **Distortion-transmission** is analogous to **distortion-rate trade-off**.
- The source reconstruction must be done in **real-time** (or with zero delay).

Comparison to real-time communication

- Special case of the real-time communication model
- Existing results in the literature establish **structure** of optimal coding strategies and a **dynamic program** to identify optimal strategies.
- The resultant dynamic programs correspond to decentralized control problem and are hard to solve.
Comparison to Information Theory

- Costly communication is analogous to communication under power constraint.
- Distortion-transmission is analogous to distortion-rate trade-off.
- The source reconstruction must be done in real-time (or with zero delay).

Comparison to real-time communication

- Special case of the real-time communication model

- Existing results in the literature establish structure of optimal coding strategies and a dynamic program to identify optimal strategies.
- The resultant dynamic programs correspond to decentralized control problem and are hard to solve.

Our approach

- Previous results have established the structure of optimal strategies.
- Exploit the structural results to explicitly identify optimal strategies.
## Modeling assumptions

<table>
<thead>
<tr>
<th>Markov chain setup</th>
<th>Guass-Markov setup</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>State spaces</strong></td>
<td><strong>State spaces</strong></td>
</tr>
<tr>
<td>$X_t, W_t \in \mathbb{Z}$</td>
<td>$X_t, W_t \in \mathbb{R}$</td>
</tr>
</tbody>
</table>
Modeling assumptions

**Markov chain setup**
- State spaces: $X_t, W_t \in \mathbb{Z}$
- Noise distribution: Unimodal and symmetric
  \[ p_e = p_{-e} \geq p_{e+1} \]

**Guass-Markov setup**
- State spaces: $X_t, W_t \in \mathbb{R}$
- Noise distribution: Zero-mean Gaussian
  \[ \varphi_\sigma(\cdot) \]

Unimodal and symmetric distribution
Modeling assumptions

**Markov chain setup**
- **State spaces**: $X_t, W_t \in \mathbb{Z}$
- **Noise distribution**: Unimodal and symmetric
  - Probability: $p_e = p_{-e} \geq p_{e+1}$
- **Distortion**: Even and increasing
  - $d(e) = d(-e) \geq d(e+1)$

**Guass-Markov setup**
- **State spaces**: $X_t, W_t \in \mathbb{R}$
- **Noise distribution**: Zero-mean Gaussian
  - $\phi_{\sigma}(\cdot)$
- **Distortion**: Mean-squared
  - $d(e) = |e|^2$
Step 1  Structure of optimal strategies

Step 2  Performance of arbitrary threshold strategies $f^{(k)}$

Step 3  Values of $\lambda$ for which $f^{(k)}$ is optimal

Step 4  Distortion-transmission trade-off
Step 1: Structure of optimal strategies

Search space of strategies $(f, g)$

Step 2: Performance of arbitrary threshold strategies $f^{(k)}$

Step 3: Values of $\lambda$ for which $f^{(k)}$ is optimal

Step 4: Distortion-transmission trade-off
Step 1: Structure of optimal strategies

Search space of strategies \((f, g)\)

Step 2: Performance of arbitrary threshold strategies \(f^{(k)}\)

Step 3: Values of \(\lambda\) for which \(f^{(k)}\) is optimal

Step 4: Distortion-transmission trade-off
Step 1: Structure of optimal strategies

Search space of strategies \((f, g)\)

---

Step 2: Performance of arbitrary threshold strategies \(f^{(k)}\)

---

Step 3: Values of \(\lambda\) for which \(f^{(k)}\) is optimal

---

Step 4: Distortion-transmission trade-off
Step 1: Structure of optimal strategies

Search space of strategies \((f, g)\)

Step 2: Performance of arbitrary threshold strategies \(f^{(k)}\)

Step 3: Values of \(\lambda\) for which \(f^{(k)}\) is optimal

Step 4: Distortion-transmission trade-off
**Step 1** Structure of optimal strategies

Search space of strategies \((f, g)\)

**Step 2** Performance of arbitrary threshold strategies \(f^{(k)}\)

**Step 3** Values of \(\lambda\) for which \(f^{(k)}\) is optimal

**Step 4** Distortion-transmission trade-off

\[ D^*_\beta(\alpha) \]

\[ (N^\beta_{(k+1)}, D^\beta_{(k+1)}) \]

\[ (N^\beta_{(k)}, D^\beta_{(k)}) \]
Step 1 Structure of optimal strategies

Model the communication system as **decentralized stochastic control**

- Two decision makers: transmitter and receiver. Non-nested information.
- **Common-information approach** [Nayyar-Mahajan-Teneketzis 2013]
  
  Equivalent **centralized** problem from the point of view of a **coordinator**.

Choose **code functions** at each step (rather than **actions**).
Step 1 Structure of optimal strategies

Model the communication system as decentralized stochastic control

- Two decision makers: transmitter and receiver. Non-nested information.

- **Common-information approach** [Nayyar-Mahajan-Teneketzis 2013]
  Equivalent centralized problem from the point of view of a coordinator.

  Choose code functions at each step (rather than actions).

Previous results


- **Markov-chain setup** [Nayyar-Başar-Teneketzis-Veeravalli 2013]
Step 1  Structure of optimal strategies

Model the communication system as decentralized stochastic control

- Two decision makers: transmitter and receiver. Non-nested information.

- Common-information approach [Nayyar-Mahajan-Teneketzis 2013]
  
  Equivalent centralized problem from the point of view of a coordinator.

  Choose code functions at each step (rather than actions).

Previous results


- Markov-chain setup [Nayyar-Başar-Teneketzis-Veeravalli 2013]

  Proof idea: Majorization-based partial order on belief states.

  Prove that $\pi \succeq_m \varphi \implies V(\pi) \geq V(\varphi)$. 
Step 1  Structure of optimal estimator (Nayyar et al, 2013)

Transmitted Process  Let $Z_t$ denote the most recently transmitted value of the Markov source.

$$Z_0 = 0 \quad \text{and} \quad Z_t = \begin{cases} X_t & \text{if } U_t = 1; \\ Z_{t-1} & \text{if } U_t = 0. \end{cases}$$

The estimator can keep track of $Z_t$ as follows:

$$Z_0 = 0 \quad \text{and} \quad Z_t = \begin{cases} Y_t & \text{if } Y_t \neq \varepsilon; \\ Z_{t-1} & \text{if } Y_t = \varepsilon. \end{cases}$$
Step 1: Structure of optimal estimator (Nayyar et al, 2013)

Transmitted Process

Let $Z_t$ denote the most recently transmitted value of the Markov source.

$Z_0 = 0$ and $Z_t = \begin{cases} 
X_t & \text{if } U_t = 1; \\
Z_{t-1} & \text{if } U_t = 0.
\end{cases}$

The estimator can keep track of $Z_t$ as follows:

$Z_0 = 0$ and $Z_t = \begin{cases} 
Y_t & \text{if } Y_t \neq \varepsilon; \\
Z_{t-1} & \text{if } Y_t = \varepsilon.
\end{cases}$

Theorem 1

The process $\{Z_t\}_{t=0}^{\infty}$ is a sufficient statistic at the estimator and an optimal estimation strategy is given by

$\hat{X}_t = g_t^*(Z_t) = Z_t$ \quad (\ast)

Remark

The optimal estimation strategy is time-homogeneous and can be specified in closed form.
Step 1 Structure of optimal transmitter (Nayyar et al)

Error process

Let $E_t = X_t - Z_{t-1}$ denote the error process. $\{E_t\}_{t=0}^{\infty}$ is a controlled Markov process where

$$E_0 = 0 \quad \text{and} \quad \mathbb{P}(E_{t+1} = n \mid E_t = e, U_t = u) = \begin{cases} p_{|e-n|}, & \text{if } u = 0; \\ p_n, & \text{if } u = 1. \end{cases}$$
Step 1 Structure of optimal transmitter (Nayyar et al)

Error process

Let \( E_t = X_t - Z_{t-1} \) denote the error process. \( \{E_t\}_{t=0}^{\infty} \) is a controlled Markov process where

\[
E_0 = 0 \quad \text{and} \quad \mathbb{P}(E_{t+1} = n \mid E_t = e, U_t = u) = \begin{cases} \frac{p_{|e-n|}}{|e-n|}, & \text{if } u = 0; \\ p_n, & \text{if } u = 1. \end{cases}
\]

Theorem 2

When the estimation strategy is of the form (⋆), then \( \{E_t\}_{t=0}^{\infty} \) is a sufficient statistic at the transmitter.

Furthermore, an optimal transmission strategy is characterized by a time-varying threshold \( \{k_t\}_{t=0}^{\infty} \), i.e.,

\[
U_t = f_t(E_t) = \begin{cases} 1 & \text{if } |E_t| \geq k_t; \\ 0 & \text{if } |E_t| < k_t. \end{cases}
\]
Step 1 Main idea

Restrict attention to time-homogeneous estimation strategies of the form
\[ \hat{X}_t = g_t^*(Z_t) = Z_t. \]

Consider the problem of finding the best-response transmission strategy.

Under appropriate technical conditions, the best-response strategy is time-homogeneous.
Step 1: Main idea

Restrict attention to time-homogeneous estimation strategies of the form

\[ \hat{X}_t = g_t^*(Z_t) = Z_t. \]

Consider the problem of finding the best-response transmission strategy.

Under appropriate technical conditions, the best-response strategy is time-homogeneous.

Find the best threshold-based strategy within the class \( \mathcal{F} = \{ f^{(k)} : k \in \mathbb{Z}_{\geq 0} \} \) where

\[ f^{(k)}(e) = \begin{cases} 
1 & \text{if } |e| \geq k \\
0 & \text{otherwise}
\end{cases} \]

Search space of strategies \((f, g)\)
Step 1: Structure of optimal strategies

Search space of strategies \((f, g)\)

Step 2: Performance of arbitrary threshold strategies \(f^{(k)}\)

\(\tau^{(k)}\)

\(t\)

\(E_t\)

Step 3: Values of \(\lambda\) for which \(f^{(k)}\) is optimal

\(k^*\)

\(\beta(\lambda)\)

\(\lambda^{(k-1)}\)

\(\lambda^{(k)}\)

Step 4: Distortion-transmission trade-off

\(\beta(\alpha)\)

\(N^{(k)}_{\beta}, D^{(k)}_{\beta}\)

\(N^{(k+1)}_{\beta}, D^{(k+1)}_{\beta}\)

\(D^{\beta}_{\alpha}(\alpha)\)

\(\alpha_c\)
Step 2  Performance of threshold strategies

Consider a threshold-based strategy

\[ f^{(k)}(e) = \begin{cases} 
1 & \text{if } |e| \geq k \\
0 & \text{otherwise} 
\end{cases} \]
Step 2 Performance of threshold strategies

Consider a threshold-based strategy

\[ f^{(k)}(e) = \begin{cases} 
1 & \text{if } |e| \geq k \\
0 & \text{otherwise} 
\end{cases} \]

Let \( \tau^{(k)} \) denote the stopping time of first transmission (starting at \( E_0 = 0 \)).
Step 2 Performance of threshold strategies

Consider a threshold-based strategy

\[ f^{(k)}(e) = \begin{cases} 
1 & \text{if } |e| \geq k \\
0 & \text{otherwise}
\end{cases} \]

Let \( \tau^{(k)} \) denote the stopping time of first transmission (starting at \( E_0 = 0 \)).

Define

- \( L^{(k)}_\beta = (1 - \beta) \mathbb{E} \left[ \sum_{t=0}^{\tau^{(k)}-1} \beta^t d(E_t) \bigg| E_0 = 0 \right] \).
- \( M^{(k)}_\beta = (1 - \beta) \mathbb{E} \left[ \sum_{t=0}^{\tau^{(k)}-1} \beta^t |E_0 = 0 \right] \).
Step 2  Performance of threshold strategies

Consider a threshold-based strategy

\[ f^{(k)}(e) = \begin{cases} 1 & \text{if } |e| \geq k \\ 0 & \text{otherwise} \end{cases} \]

Let \( \tau^{(k)} \) denote the stopping time of first transmission (starting at \( E_0 = 0 \)).

\[ D^{(k)} := D_\beta(f^{(k)}, g^*) = \frac{L^{(k)}}{M^{(k)}} \quad \text{and} \quad N^{(k)} := N_\beta(f^{(k)}, g^*) = \frac{1}{M^{(k)}} - (1 - \beta). \]

**Proposition** \( \{E_t\}_{t=0}^\infty \) is a regenerative process and by renewal theory, we have that

\[ \mathbb{E} \left[ \sum_{t=0}^{\tau^{(k)}-1} \beta^t d(E_t) \mid E_0 = 0 \right]. \]
Step 2 Computing $D_{\beta}^{(k)}$ and $N_{\beta}^{(k)}$

Notation

- $S^{(k)} = \{- (k - 1), \ldots, k - 1\}$.
- $[P^{(k)}]_{ij} = p_{|i-j|}$, for $i, j \in S^{(k)}$.
- $[d^{(k)}]_i = d(i)$, for $i \in S^{(k)}$.
- $[1^{(k)}]_i = 1$, for $i \in S^{(k)}$. 
Step 2

Computing $D^{(k)}_{\beta}$ and $N^{(k)}_{\beta}$

Notation

1. $\mathcal{S}^{(k)} = \{-k, \ldots, k-1\}$.
2. $[P^{(k)}]_{ij} = p_{|i-j|}$, for $i, j \in \mathcal{S}^{(k)}$.
3. $[d^{(k)}]_i = d(i)$, for $i \in \mathcal{S}^{(k)}$.
4. $[1^{(k)}]_i = 1$, for $i \in \mathcal{S}^{(k)}$.

Proposition

1. $L^{(k)}_{\beta} = \left[[I - \beta P^{(k)}]^{-1} d^{(k)}\right]_0$.
2. $M^{(k)}_{\beta} = \left[[I - \beta P^{(k)}]^{-1} 1^{(k)}\right]_0$. 
Step 2 Computing $D^{(k)}_\beta$ and $N^{(k)}_\beta$

Notation
- $S^{(k)} = \{- (k-1), \ldots, k - 1\}$.
- $[P^{(k)}]_{ij} = p_{|i-j|}$, for $i, j \in S^{(k)}$.
- $[d^{(k)}]_i = d(i)$, for $i \in S^{(k)}$.
- $[1^{(k)}]_i = 1$, for $i \in S^{(k)}$.

Proposition
- $L^{(k)}_\beta = [I - \beta P^{(k)}]^{-1} d^{(k)} 0$.
- $M^{(k)}_\beta = [I - \beta P^{(k)}]^{-1} 1^{(k)} 0$.

$D^{(k)}_\beta$ and $N^{(k)}_\beta$ can be computed using these expressions.
**Step 1** Structure of optimal strategies

Search space of strategies $(f, g)$

**Step 2** Performance of arbitrary threshold strategies $f^{(k)}$

**Step 3** Values of $\lambda$ for which $f^{(k)}$ is optimal

**Step 4** Distortion-transmission trade-off
Step 3 Properties of optimal thresholds

Monotonicity

\[ L_{\beta}^{(k+1)} > L_{\beta}^{(k)} \quad \text{and} \quad M_{\beta}^{(k+1)} > M_{\beta}^{(k)} \]

Depends on unimodularity of noise
Step 3 Properties of optimal thresholds

Monotonicity

\[ L_{\beta}^{(k+1)} > L_{\beta}^{(k)} \quad \text{and} \quad M_{\beta}^{(k+1)} > M_{\beta}^{(k)} \]

Implication:

\[ D_{\beta}^{(k+1)} \geq D_{\beta}^{(k)} \quad \text{and} \quad N_{\beta}^{(k+1)} < N_{\beta}^{(k)} \]

Use DP and monotonicity of Bellman operator
Step 3 Properties of optimal thresholds

Monotonicity

\[ L^{(k+1)}_{\beta} > L^{(k)}_{\beta} \text{ and } M^{(k+1)}_{\beta} > M^{(k)}_{\beta} \]

Implication:

\[ D^{(k+1)}_{\beta} \geq D^{(k)}_{\beta} \text{ and } N^{(k+1)}_{\beta} < N^{(k)}_{\beta} \]

Submodularity

\[ C^{(k)}_{\beta}(\lambda) := D^{(k)}_{\beta} + \lambda N^{(k)}_{\beta} \text{ is submodular in } (k, \lambda). \]
**Step 3** Properties of optimal thresholds

**Monotonicity**
\[ L_{\beta}^{(k+1)} > L_{\beta}^{(k)} \text{ and } M_{\beta}^{(k+1)} > M_{\beta}^{(k)} \]

Implication:
\[ D_{\beta}^{(k+1)} \geq D_{\beta}^{(k)} \text{ and } N_{\beta}^{(k+1)} < N_{\beta}^{(k)} \]

**Submodularity**
\[ C_{\beta}^{(k)}(\lambda) := D_{\beta}^{(k)} + \lambda N_{\beta}^{(k)} \text{ is submodular in } (k, \lambda). \]

Proof: \[ C_{\beta}^{(k+1)}(\lambda) - C_{\beta}^{(k)}(\lambda) = D_{\beta}^{(k+1)}(\lambda) - D_{\beta}^{(k)}(\lambda) - \lambda (N_{\beta}^{(k)}(\lambda) - N_{\beta}^{(k+1)}(\lambda)) \geq 0 \]
Step 3 Properties of optimal thresholds

Monotonicity

\[ L_{\beta}^{(k+1)} > L_{\beta}^{(k)} \quad \text{and} \quad M_{\beta}^{(k+1)} > M_{\beta}^{(k)} \]

Implication:

\[ D_{\beta}^{(k+1)} \geq D_{\beta}^{(k)} \quad \text{and} \quad N_{\beta}^{(k+1)} < N_{\beta}^{(k)} \]

Submodularity

\[ C_{\beta}^{(k)}(\lambda) := D_{\beta}^{(k)} + \lambda N_{\beta}^{(k)} \text{ is submodular in } (k, \lambda). \]

Proposition

\[ k_{\beta}^*(\lambda) := \arg \min_{k \in \mathbb{Z}_{\geq 0}} C_{\beta}^{(k)}(\lambda) \text{ is increasing in } \lambda. \]
Step 3 Properties of optimal thresholds

Monotonicity

\[ L_{\beta}^{(k+1)} > L_{\beta}^{(k)} \quad \text{and} \quad M_{\beta}^{(k+1)} > M_{\beta}^{(k)} \]

Implication:

\[ D_{\beta}^{(k+1)} \geq D_{\beta}^{(k)} \quad \text{and} \quad N_{\beta}^{(k+1)} < N_{\beta}^{(k)} \]

Submodularity

\[ C_{\beta}^{(k)}(\lambda) := D_{\beta}^{(k)} + \lambda N_{\beta}^{(k)} \text{ is submodular in } (k, \lambda). \]

Proposition

\[ k_{\beta}^{*}(\lambda) := \arg \min_{k \in \mathbb{Z}_{\geq 0}} C_{\beta}^{(k)}(\lambda) \text{ is increasing in } \lambda. \]

Define \( \Lambda_{\beta}^{(k)} := \{ \lambda \in \mathbb{R}_{\geq 0} : k_{\beta}^{*}(\lambda) = k \} = [\lambda_{\beta}^{(k-1)}, \lambda_{\beta}^{(k)}]. \)
Step 3 Optimal costly communication

\[ C_{\beta}^{(k)} (\lambda_{\beta}^{(k)}) = C_{\beta}^{(k+1)} (\lambda_{\beta}^{(k)}) \]
Step 3 Optimal costly communication

\[ C_{\beta}^{(k)}(\lambda_{\beta}^{(k)}) = C_{\beta}^{(k+1)}(\lambda_{\beta}^{(k)}) \]
Step 3: Optimal costly communication

\[ C^{(k)}_{\beta}(\lambda^{(k)}_{\beta}) = C^{(k+1)}_{\beta}(\lambda^{(k)}_{\beta}) \]

\[ \lambda^{(k)}_{\beta} = \frac{D^{(k+1)}_{\beta} - D^{(k)}_{\beta}}{N^{(k)}_{\beta} - N^{(k+1)}_{\beta}} \]
Step 3

Optimal costly communication

\[ C^{(k)}_{\beta}(\lambda^{(k)}_{\beta}) = C^{(k+1)}_{\beta}(\lambda^{(k)}_{\beta}) \]

\[ \lambda^{(k)}_{\beta} = \frac{D^{(k+1)}_{\beta} - D^{(k)}_{\beta}}{N^{(k)}_{\beta} - N^{(k+1)}_{\beta}} \]
Step 3 Optimal costly communication

\[ C_{\beta}^{(k)} (\lambda_{\beta}^{(k)}) = C_{\beta}^{(k+1)} (\lambda_{\beta}^{(k)}) \]

\[ \lambda_{\beta}^{(k)} = \frac{D_{\beta}^{(k+1)} - D_{\beta}^{(k)}}{N_{\beta}^{(k)} - N_{\beta}^{(k+1)}} \]
Step 3  Optimal costly communication

\[ \lambda^{(k+1)}_{\beta} = \frac{D^{(k+1)}_{\beta} - D^{(k)}_{\beta}}{N^{(k)}_{\beta} - N^{(k+1)}_{\beta}} \]

\[ C^{(k)}_{\beta}(\lambda^{(k)}_{\beta}) = C^{(k+1)}_{\beta}(\lambda^{(k)}_{\beta}) \]

Theorem

- For all \( \lambda \in (\lambda^{(k)}_{\beta}, \lambda^{(k+1)}_{\beta}] \) the threshold strategy \( f^{(k+1)} \) is optimal.

- \( C^*_{\beta}(\lambda) = \min_{k \in \mathbb{Z}_{\geq 0}} C^{(k)}_{\beta} \) is piecewise linear, continuous, concave, and increasing function of \( \lambda \).
Step 3 Optimal costly communication

Theorem

For all \( \lambda \in (\lambda_{\beta}^{(k)}, \lambda_{\beta}^{(k+1)}) \) the threshold strategy \( f^{(k+1)} \) is optimal.

- \( C_{\beta}^*(\lambda) = \min_{k \in \mathbb{Z}_{\geq 0}} C_{\beta}^{(k)} \) is piecewise linear, continuous, concave, and increasing function of \( \lambda \).
Step 1: Structure of optimal strategies

Search space of strategies \((f, g)\)

Step 2: Performance of arbitrary threshold strategies \(f^{(k)}\)

Step 3: Values of \(\lambda\) for which \(f^{(k)}\) is optimal

\[ k^*_\beta(\lambda) \]

\[ \lambda^{(k-1)} \quad \lambda^{(k)} \]

Step 4: Distortion-transmission trade-off

\[ D_\beta^*(\alpha) \]

\[ (N_\beta^{(k+1)}, D_\beta^{(k+1)}) \quad (N_\beta^{(k)}, D_\beta^{(k)}) \]

\[ 0 \quad \alpha \quad \alpha_c \quad 1 \]
Sufficient conditions for constrained optimality

A strategy \((f^\circ, g^\circ)\) is optimal for the constrained communication problem if

\[(C1) \quad N_\beta(f^\circ, g^\circ) = \alpha\]

\[(C2) \quad \text{There exists} \; \lambda^\circ \geq 0 \; \text{such that} \; (f^\circ, g^\circ) \; \text{is optimal for} \; C_\beta(f, g; \lambda^\circ).\]
Step 4 Distortion-transmission trade-off

Sufficient conditions for constrained optimality

A strategy \((f^o, g^o)\) is optimal for the constrained communication problem if

(C1) \(N_\beta(f^o, g^o) = \alpha\)

(C2) There exists \(\lambda^o \geq 0\) such that \((f^o, g^o)\) is optimal for \(C_\beta(f, g; \lambda^o)\).

Let \(k^*_\beta\) be such that \(N_\beta^{(k^*_\beta)} > \alpha > N_\beta^{(k^*_\beta + 1)}\)
Sufficient conditions for constrained optimality

A strategy \((f^\circ, g^\circ)\) is optimal for the constrained communication problem if

1. \(N_\beta(f^\circ, g^\circ) = \alpha\)
2. There exists \(\lambda^\circ \geq 0\) such that \((f^\circ, g^\circ)\) is optimal for \(C_\beta(f, g; \lambda^\circ)\).

Let \(k^*_\beta\) be such that
\[N_\beta^{(k^*_\beta)} > \alpha > N_\beta^{(k^*_\beta + 1)}\]
A strategy \((f^*, g^*)\) is optimal for the constrained communication problem if

(C1) \(N_\beta(f^*, g^*) = \alpha\)

(C2) There exists \(\lambda^o \geq 0\) such that \((f^*, g^*)\) is optimal for \(C_\beta(f, g; \lambda^o)\).

Let \(k^*_\beta\) be such that
\[N_\beta^{(k^*_\beta)} > \alpha > N_\beta^{(k^*_\beta+1)}\]
Step 4 Distortion-transmission trade-off

Sufficient conditions for constrained optimality

A strategy \((f^\circ, g^\circ)\) is optimal for the constrained communication problem if

\[(C1) \quad N_\beta(f^\circ, g^\circ) = \alpha\]

\[(C2) \quad \text{There exists } \lambda^\circ \geq 0 \text{ such that } (f^\circ, g^\circ) \text{ is optimal for } C_\beta(f,g;\lambda^\circ).\]

Bernoulli randomized strategy \((\theta^\ast, f^{(k)}, f^{(k+1)})\) is optimal where

\[\theta^\ast N^{(k)}_\beta + (1 - \theta^\ast) N^{(k+1)}_\beta = \alpha\]
Step 4  Distortion-transmission trade-off

Sufficient conditions for constrained optimality

A strategy \((f^\circ, g^\circ)\) is optimal for the constrained communication problem if

1. \(D_\beta^*(\alpha)\) is optimal
2. Bernstein randomized strategy \((\theta^*, f^{(k)}, f^{(k+1)})\) is optimal where
   \[
   \theta^* N_\beta^{(k)} + (1 - \theta^*) N_\beta^{(k+1)} = \alpha
   \]
Step 4 Distortion-transmission trade-off

Sufficient conditions for constrained optimality

A strategy \((f^\circ, g^\circ)\) is optimal for the constrained communication problem if

1. \[ (C1) \quad N_\beta(f^\circ, g^\circ) = \alpha \]
2. There exists \(\lambda^\circ \geq 0\) such that \((f^\circ, g^\circ)\) is optimal for \(C_\beta(f, g; \lambda^\circ)\).

Bernoulli randomized strategy \((\theta^*, f^{(k)}, f^{(k+1)})\) is optimal where

\[ \theta^* N^{(k)}_\beta + (1 - \theta^*) N^{(k+1)}_\beta = \alpha \]
Optimal strategy

\[ f^*(e) = \begin{cases} 
1 & \text{if } |e| > k^*_\beta \\
1 & \text{w.p. } \theta^* \text{ if } |e| = k^*_\beta \\
0 & \text{w.p. } 1 - \theta^* \text{ if } |e| = k^*_\beta \\
0 & \text{if } |e| < k^*_\beta 
\end{cases} \]
Step 4 Features of optimal strategy

Optimal strategy

$$f^*(e) = \begin{cases} 
1 & \text{if } |e| > k^*_\beta \\
1 \text{ w.p. } \theta^* & \text{if } |e| = k^*_\beta \\
0 \text{ w.p. } 1 - \theta^* & \text{if } |e| = k^*_\beta \\
0 & \text{if } |e| < k^*_\beta 
\end{cases}$$

Randomized action at a single state
Step 4 Features of optimal strategy

Optimal strategy

\[ f^*(e) = \begin{cases} 
1 & \text{if } |e| > k^*_\beta \\
1 & \text{w.p. } \theta^* \text{ if } |e| = k^*_\beta \\
0 & \text{w.p. } 1 - \theta^* \text{ if } |e| = k^*_\beta \\
0 & \text{if } |e| < k^*_\beta 
\end{cases} \]

Randomized action at a single state

Deterministic implementation

Time-sharing strategies

- Assume \( \theta^* = a/(a+b) \).

- Choose strategy \( f^{(k^*)} \) for \( a \) visits to state zero and strategy \( f^{(k^*+1)} \) for \( b \) visits to state zero and so on.
Step 4 Features of optimal strategy

Optimal strategy

\[
 f^*(e) = \begin{cases} 
 1 & \text{if } |e| > k^*_\beta \\
 1 \text{ w.p. } \theta^* & \text{if } |e| = k^*_\beta \\
 0 \text{ w.p. } 1 - \theta^* & \text{if } |e| = k^*_\beta \\
 0 & \text{if } |e| < k^*_\beta 
\end{cases}
\]

Randomized action at a single state

Deterministic implementation

Time-sharing strategies

- Assume \( \theta^* = a/(a+b) \).
- Choose strategy \( f^{(k^*)} \) for \( a \) visits to state zero and strategy \( f^{(k^*+1)} \) for \( b \) visits to state zero and so on.

Steering strategies:

- Let \( a_i^t \) denote the number of times action \( i \) is chosen in the past.
- At states \( \{-k^*, k^*\} \) choose an action that steers the empirical frequency closer to the desired randomization probability.
An example: Symmetric birth-death Markov Chain

\[ P_{ij} = \begin{cases} 
  p, & \text{if } |i - j| = 1; \\
  1 - 2p, & \text{if } i = j; \\
  0, & \text{otherwise,} 
\end{cases} \]

where \( p \in (0, \frac{1}{2}) \), \( d(e) = |e| \)
### Discounted cost

Let $K_\beta = -2 - (1 - \beta)/\beta p$ and $m_\beta = \cosh^{-1}(-K_\beta/2)$.

\[
D^{(k)}_\beta = \frac{\sinh(km_\beta) - k \sinh(m_\beta)}{2 \sinh^2(km_\beta/2) \sinh(m_\beta)}
\]

\[
N^{(k)}_\beta = \frac{2\beta p \sinh^2(m_\beta/2) \cosh(km_\beta)}{\sinh^2(km_\beta/2)} - (1 - \beta)
\]

### Average cost

\[
D^{(k)}_1 = \frac{k^2 - 1}{3k} \quad \text{and} \quad N^{(k)}_1 = \frac{2p}{k^2}
\]
Discounted cost

Let $K_\beta = -2 - (1 - \beta)/\beta p$ and $m_\beta = \cosh^{-1}(-K_\beta/2)$.

\[
D^{(k)}_\beta = \frac{\sinh(km_\beta) - k \sinh(m_\beta)}{2 \sinh^2(km_\beta/2) \sinh(m_\beta)}
\]

\[
N^{(k)}_\beta = \frac{2\beta p \sinh^2(m_\beta/2) \cosh(km_\beta)}{\sinh^2(km_\beta/2)} - (1 - \beta)
\]

$\lambda^{(k)}_\beta$ can be computed in terms of $D^{(k)}_\beta$ and $N^{(k)}_\beta$.

Average cost

$D^{(k)}_1 = \frac{k^2 - 1}{3k}$ and $N^{(k)}_1 = \frac{2p}{k^2}$

$\lambda^{(k)}_1 = \frac{k(k + 1)(k^2 + k + 1)}{6p(2k + 1)}$
An example: Symmetric birth-death Markov Chain

Discounted cost

\[ D(k) = \frac{1}{2} \sinh\left(2k\beta\right) - k \sinh\left(\beta\right) \]

Average cost

\[ N(k) = 2\beta p \sinh^2\left(\frac{1}{2}k\beta\right) \cosh\left(k\beta\right) - (1-\beta) \]

\[ \lambda \]

\[ C^*_\beta(\lambda) \]

\[ p = 0.3 \]

\[ \beta = 1 \]

\[ \beta = 0.9 \]

\[ \beta = 0.8 \]
Discounted cost
Let $K_\beta = -2 - (1 - \beta)/\beta p$ and $m_\beta = \cosh^{-1}(-K_\beta/2)$.

\[
D^{(k)}_\beta = \frac{\sinh(km_\beta) - k \sinh(m_\beta)}{2 \sinh^2(km_\beta/2) \sinh(m_\beta)}
\]
\[
N^{(k)}_\beta = \frac{2\beta p \sinh^2(m_\beta/2) \cosh(km_\beta)}{\sinh^2(km_\beta/2)} - (1 - \beta)
\]

\[
k^*_\beta = \sup \left\{ k \in \mathbb{Z}_{\geq 0} : \frac{\sinh^2(m_\beta/2) \cosh(km_\beta)}{\sinh^2(km_\beta/2)} \geq \frac{1 + \alpha - \beta}{2\beta p} \right\}
\]

Average cost
$D^{(k)}_1 = \frac{k^2 - 1}{3k}$ and $N^{(k)}_1 = \frac{2p}{k^2}$

\[
k^*_1 = \left\lfloor \sqrt{\frac{2p}{\alpha}} \right\rfloor
\]
An example: Symmetric birth-death Markov Chain

Distortion-transmission trade-off—(Chakravorty and Mahajan)
Discounted cost

Let $K_\beta = -(1 - \beta) / \beta p$ and $m_\beta = \cosh^{-1}(-K_\beta/2)$.

$$D(k)_\beta = \sinh (km_\beta) - k \sinh (m_\beta)$$

$$N(k)_\beta = 2 \beta p \sinh^2 (m_\beta/2) \cosh (km_\beta) \sinh (km_\beta/2)$$

Average cost

$$D^*_\beta (\alpha) = k^2 - \frac{1}{3} k$$

$$N^*_\beta (\alpha) = 2 p k^2$$

An example: Symmetric birth-death Markov Chain

<table>
<thead>
<tr>
<th>$p$</th>
<th>0.3</th>
<th>0.3</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta$</td>
<td>1.0</td>
<td>0.9</td>
</tr>
</tbody>
</table>

Graph showing $D^*_\beta (\alpha)$ for $\beta = 0.9$.
## Summary

### The system model

The system model consists of a Markov source, a transmitter, a receiver, and a distortion function $d(X_t, \hat{X}_t)$.

- **Markov Source**: $X_t$
- **Transmitter**: $U_t = f_t(X_{1:t}, U_{1:t-1})$
- **Receiver**: $\hat{X}_t = g_t(Y_{1:t})$
- **Distortion**: $d(X_t, \hat{X}_t)$

### 1. Discounted setup, $\beta \in (0, 1)$

$$D_\beta(f, g) = (1 - \beta) \mathbb{E}_0^{(f, g)} \left[ \sum_{t=0}^{\infty} \beta^t d(X_t - \hat{X}_t) \right]; \quad N_\beta(f, g) = (1 - \beta) \mathbb{E}_0^{(f, g)} \left[ \sum_{t=0}^{\infty} \beta^t U_t \right]$$

### 2. Average cost setup, $\beta = 1$

$$D_1(f, g) = \limsup_{T \to \infty} \frac{1}{T} \mathbb{E}_0^{(f, g)} \left[ \sum_{t=0}^{T-1} d(X_t - \hat{X}_t) \right]; \quad N_1(f, g) = \limsup_{T \to \infty} \frac{1}{T} \mathbb{E}_0^{(f, g)} \left[ \sum_{t=0}^{T-1} U_t \right]$$
## Summary

### The system model

<table>
<thead>
<tr>
<th>X_t</th>
<th>U_t</th>
<th>Y_t</th>
<th>ˆX_t</th>
</tr>
</thead>
</table>

1. **Discounted setup**, $\beta \in (0, 1)$
   \[ D_\beta(f, g) = (1 - \beta) \mathbb{E}[d(X_t - ˆX_t)] \]
   \[ N_\beta(f, g) = (1 - \beta) \mathbb{E}[U_t] \]

2. **Average cost setup**, $\beta = 1$
   \[ D_1(f, g) = \limsup_{T \to \infty} \frac{1}{T} \mathbb{E}[d(X_t - ˆX_t)] \]
   \[ N_1(f, g) = \limsup_{T \to \infty} \frac{1}{T} \mathbb{E}[U_t] \]

### Optimization problems

#### Costly communication

For any $\lambda \in \mathbb{R}_{>0}$,
\[ C_\beta^*(\lambda) = C_\beta(f^*, g^*; \lambda) = \inf_{(f, g)} \{ D_\beta(f, g) + \lambda N_\beta(f, g) \} \]

#### Constrained communication

For any $\alpha \in (0, 1)$,
\[ D_\beta^*(\alpha) = \inf_{(f, g)} \{ D_\beta(f, g) : N_\beta(f, g) \leq \alpha \} \]

$C_\beta^*$ is cts, inc, and concave

$D_\beta^*$ is cts, dec, and convex
Summary

The system model

1. Discounted setup, $\beta \in (0,1)$
   \[ C_{1} \beta = \beta \limsup_{T \to \infty} \frac{1}{T} \mathbb{E} \left[ f(X_{1:T}, U_{1:T-1}) \right] \]

2. Average cost setup, $\beta = 1$
   \[ C_{1} = \limsup_{T \to \infty} \frac{1}{T} \mathbb{E} \left[ f(X_{1:T}, U_{1:T-1}) \right] \]

Optimization problems

1. Costly communication
   For any $\lambda \in \mathbb{R}^+$
   \[ C_{\lambda} \beta = \inf (f, g) \{ D_{\beta}(f, g) + \lambda N_{\beta}(f, g) \} \]

2. Constrained communication
   For any $\alpha \in (0,1)$
   \[ D_{\alpha} \beta = \inf (f, g) \{ D_{\beta}(f, g) : N_{\beta}(f, g) \leq \alpha \} \]

Step 1: Structure of optimal strategies
- Search space of strategies $(f, g)$

Step 2: Performance of arbitrary threshold strategies $f^{(k)}$

Step 3: Values of $\lambda$ for which $f^{(k)}$ is optimal
- $k_{\beta}^\lambda(\lambda)$

Step 4: Distortion-transmission trade-off
- $D_{\beta}(\alpha)$
  \[ (N_{\beta}^{(k+1)}, D_{\beta}^{(k+1)}) \]
  \[ (N_{\beta}^{(k)}, D_{\beta}^{(k)}) \]
Analyze a fundamental trade-off in real-time communication
Conclusion

Analyze a fundamental trade-off in real-time communication

Possible generalizations (where the proposed approach may work)
- Symmetric finite sources
- Erasure channels
Conclusion

Analyze a fundamental trade-off in real-time communication

Possible generalizations (where the proposed approach may work)
- Symmetric finite sources
- Erasure channels

More realistic models (where the proposed approach will fail)
- Non-symmetric sources
- Rate constraints (affect of quantization)
- Network delays
Conclusion

Analyze a fundamental trade-off in real-time communication

Possible generalizations (where the proposed approach may work)
- Symmetric finite sources
- Erasure channels

More realistic models (where the proposed approach will fail)
- Non-symmetric sources
- Rate constraints (affect of quantization)
- Network delays

Full version to be posted on arxiv soon.