Fundamental limits of remote-estimation under communication constraints

Aditya Mahajan
McGill University

Joint work with Jhelum Chakravorty

Mathematical Cybernetics: Hybrid, Stochastic and Decentralized Systems
Carleton University, Ottawa, 28–29 May 2015
Motivation

Many applications require:
- Sequential transmission of data
- Zero- (or finite-) delay reconstruction
Motivation

Many applications require:
- Sequential transmission of data
- Zero- (or finite-) delay reconstruction
Motivation

Many applications require:
- Sequential transmission of data
- Zero- (or finite-) delay reconstruction
Motivation

Many applications require:
- Sequential transmission of data
- Zero- (or finite-) delay reconstruction

Internet of Things
**Motivation**

Many applications require:
- Sequential transmission of data
- Zero- (or finite-) delay reconstruction

**Salient features**
- Sensing is cheap
- Transmission is expensive
- Size of data-packet is not critical

*Internet of Things*
Motivation

Many applications require:
- Sequential transmission of data
- Zero- (or finite-) delay reconstruction

Salient features
- Sensing is cheap
- Transmission is expensive
- Size of data-packet is not critical

Analyze a stylized model and evaluate fundamental trade-offs
A completely solved example of a “simple” decentralized system with non-classical information structure
The system model

Markov Process \( X_t \) → Transmitter \( U_t \) → Receiver \( Y_t \) → \( \hat{X}_t \)
The system model

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

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

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

\[ X_1 \rightarrow X_2 \rightarrow X_3 \]

\[ U_1 \rightarrow U_2 \rightarrow U_3 \rightarrow \ldots \]
The system model

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

Distortion \( d(X_t - \hat{X}_t) \)

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

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

The system model consists of a Markov Process, a Transmitter, a Receiver, and a distortion measure. The state processes are as follows:

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

Communication Strategies

- **Transmission strategy** \( f = \{ f_t \}_{t=0}^{\infty} \).
- **Estimation strategy** \( g = \{ g_t \}_{t=0}^{\infty} \).

Distortion

\[ d(X_t - \hat{X}_t) \]
The system model

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

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

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

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

Costly communication

For $\lambda \in \mathbb{R}_{>0}$, 

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

Constrained communication

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

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

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

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

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

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

$C_{\beta}^*$ is cts, inc, and concave
Optimization problems

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

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

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

Estimation under communication constraints–(Mahajan and Chakravorty)
Optimization problems

Costly communication

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

Constrained communication

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

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

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

Costly communication

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

Constrained communication

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

Distortion-transmission function

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

$D^*_\beta$ is cts, dec, and convex
Optimization problems

Costly communication
For $\lambda \in \mathbb{R}_{>0}$, $C^*_\beta(\lambda) = C(\beta, f^*, g^*; \lambda) := \inf_{(f, g)} \{D(\beta, f, g) + \lambda N(\beta, f, g)\}$

Constrained communication
For $\alpha \in (0, 1)$, $D^*_\beta(\alpha) := \inf_{(f, g)} \{D(\beta, f, g) : N(\beta, f, g) \leq \alpha\}$

Distortion-transmission function

We provide explicit computable expressions for both curves
$X_{t+1} = X_t + W_t, \ W_t \sim \mathcal{N}(0, 1)$
\[ X_{t+1} = X_t + W_t, \quad W_t \sim \mathcal{N}(0, 1) \]
Periodic transmission strategy
Periodic transmission strategy

Error process

Estimation under communication constraints–(Mahajan and Chakravorty)
Periodic transmission strategy

Error process

\[ D = 0.69 \quad N \approx 1/3 \]
An alternative strategy
An alternative strategy

Error process

Estimation under communication constraints—(Mahajan and Chakravorty)
An alternative strategy

Error process

$D = 0.24 \quad N \approx 1/3$
Distortion-transmission function

- Periodic transmission strategy
- Threshold based strategy

Estimation under communication constraints—(Mahajan and Chakravorty)
Identify strategies that achieve the optimal trade-off
Simple and intuitive threshold based strategies are optimal.

Provide computable expressions for distortion-transmission function
Based on simple matrix calculations for discrete Markov processes
Based on solving Fredholm integral equations for Gaussian processes
Identify strategies that achieve the optimal trade-off
   Simple and intuitive threshold based strategies are optimal.

Provide computable expressions for distortion-transmission function
   Based on simple matrix calculations for discrete Markov processes
   Based on solving Fredholm integral equations for Gaussian processes

Beautiful example of stochastics and optimization
   Decentralized stochastic control and POMDPs
   Stochastic orders and majorization
   Markov chain analysis, stopping times, and renewal theory
   Constrained MDPs and Lagrangian relaxations
So how do we start?
Decentralized stochastic control
Dealing with non-classical information structure
Dealing with non-classical information structure

Classical info. struct.

\[ f_t \quad X_t, Y_{1:t-1} \quad U_t \]

\[ g_t \quad Y_{1:t-1}, Y_t \quad \hat{X}_t \]
Dealing with non-classical information structure

\[ f_t \quad X_t, Y_{1:t-1} \quad U_t \]

\[ g_t \quad Y_{1:t-1}, Y_t \quad \hat{X}_t \]
Dealing with non-classical information structure

Non-Classical info. struct.

\[ f_t \quad X_t, Y_{1:t-1} \quad u_t \]

\[ g_t \quad Y_{1:t-1}, Y_t \quad \hat{X}_t \]

Estimation under communication constraints–(Mahajan and Chakravorty)
Dealing with non-classical information structure

Belongs to the class of **tractable non-classical information structures** (called **partial-history sharing**) identified in [Mahajan–Nayyar–Teneketzis 2013]

\[
C_t := \bigcap_{s \geq t} \bigcap_{i=1}^{n} I_{s}^{i}
\]

\[
L_t^i := I_t^i \setminus C_t
\]

\[
g(C, L) = \psi(C)(L)
\]

\[
f_t : X_t, Y_{1:t-1}, U_t
\]

\[
g_t : Y_{1:t-1}, Y_t, \hat{X}_t
\]

Dealing with non-classical information structure

Belongs to the class of **tractable non-classical information structures**
(called **partial-history sharing**) identified in [Mahajan-Nayyar-Teneketzis 2013]

\[
\begin{align*}
C_t &= \bigcap_{s \geq t} \bigcap_{i=1}^{n} I^i_s \\
L^i_t &= I^i_t \setminus C_t \\
g(C, L) &= \psi(C)(L)
\end{align*}
\]
Dealing with non-classical information structure

The coordinated system is a centralized (i.e., single-agent) partially observed system

Belongs to the class of tractable non-classical information structures (called partial-history sharing) identified in [Mahajan-Nayyar-Teneketzis 2013]

\[
f_t \quad X_t, Y_{1:t-1} \quad U_t \\
\Leftrightarrow \\
g_t \quad Y_{1:t-1}, Y_t \quad \hat{X}_t
\]

\[
\varphi_t \quad X_t \quad U_t \\
\quad \quad Y_{1:t-1} \quad (\varphi_t, \gamma_t) \\
\quad \quad \gamma_t \quad Y_t \quad \hat{X}_t
\]

Equiv.


Estimation under communication constraints–(Mahajan and Chakravorty)
Information states and dynamic program

Information states

Pre-transmission belief: \( \Pi_t(x) = \mathbb{P}(X_t = x \mid Y_{1:t-1}) \).

Post-transmission belief: \( \Xi_t(x) = \mathbb{P}(X_t = x \mid Y_{1:t}) \).
Information states and dynamic program

Information states

Pre-transmission belief: \( \Pi_t(x) = \mathbb{P}(X_t = x \mid Y_{1:t-1}) \).

Post-transmission belief: \( \Xi_t(x) = \mathbb{P}(X_t = x \mid Y_{1:t}) \).

Structural results

There is no loss of optimality in using

\[ U_t = f_t(X_t, \Pi_t) \quad \text{and} \quad \hat{X}_t = g_t(\Xi_t). \]
Information states and dynamic program

Information states

Pre-transmission belief: \( \Pi_t(x) = \mathbb{P}(X_t = x \mid Y_{1:t-1}) \).
Post-transmission belief: \( \Xi_t(x) = \mathbb{P}(X_t = x \mid Y_{1:t}) \).

Structural results

There is no loss of optimality in using
\[
U_t = f_t(X_t, \Pi_t) \quad \text{and} \quad \hat{X}_t = g_t(\Xi_t).
\]

Dynamic Program

\[
W_{T+1}(\pi) = 0
\]

and for \( t = T, \ldots, 0 \)

\[
V_t(\xi) = \min_{\hat{x} \in \mathcal{X}} \mathbb{E}[d(X_t - \hat{x}) + W_{t+1}(\Pi_{t+1}) \mid \Xi_t = \xi],
\]

\[
W_t(\pi) = \min_{\varphi : \mathcal{X} \to \{0, 1\}} \mathbb{E}[\lambda \varphi(X_t) + V_t(\Xi_t) \mid \Pi_t = \pi, \varphi_t = \varphi].
\]
Information states and dynamic program

**Information states**

- **Pre-transmission belief**: \( \Pi_t(x) = \mathbb{P}(X_t = x | Y_1:t-1) \)
- **Post-transmission belief**: \( \Xi_t(x) = \mathbb{P}(X_t = x | Y_1:t) \)

**Structural results**

There is no loss of optimality in using

\[
U_t = f_t(X_t, \Pi_t) \quad \text{and} \quad \hat{X}_t = g_t(\Xi_t).
\]

**Dynamic Program**

\[
W_{T+1}(\pi) = 0
\]

and for \( t = T, \ldots, 0 \)

\[
V_t(\xi) = \min_{\hat{x} \in \mathcal{X}} \mathbb{E}[d(X_t - \hat{x}) + W_{t+1}(\Pi_{t+1}) | \Xi_t = \xi],
\]

\[
W_t(\pi) = \min_{\varphi : \mathcal{X} \rightarrow \{0,1\}} \mathbb{E}[\lambda \varphi(X_t) + V_t(\Xi_t) | \Pi_t = \pi, \varphi_t = \varphi].
\]

“Standard” POMDP. Optimal strategies can be computed numerically.

Estimation under communication constraints–(Mahajan and Chakravorty)
Can we use the DP to say something more about the optimal strategy?
Simplifying modeling assumptions

Markov process

\[ X_{t+1} = X_t + W_t \]
## Simplifying modeling assumptions

<table>
<thead>
<tr>
<th>Markov process</th>
<th>( X_{t+1} = X_t + W_t )</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>State spaces</strong></td>
<td>( X_t, W_t \in \mathbb{Z} )</td>
</tr>
</tbody>
</table>
Simplifying modeling assumptions

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

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
  \( \phi_\sigma(\cdot) \)

Estimation under communication constraints–(Mahajan and Chakravorty)
Simplifying modeling assumptions

Markov process

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

Markov chain setup

State spaces

\[ X_t, W_t \in \mathbb{Z} \]

Noise distribution

Unimodal and symmetric

\[ p_e = p_{-e} \geq p_{e+1} \]

Distortion

Even and increasing

\[ d(e) = d(-e) \leq d(e + 1) \]

Guass-Markov setup

State spaces

\[ X_t, W_t \in \mathbb{R} \]

Noise distribution

Zero-mean Gaussian

\[ \varphi_{\sigma}( \cdot ) \]

Distortion

Mean-squared

\[ d(e) = |e|^2 \]
Step 1: Structure of optimal strategies

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

Step 3: Optimal costly comm.

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

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

Step 3 Optimal costly comm.

Step 4 Distortion-transmission trade-off

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 \(f^{(k)}\)

Step 3: Optimal costly comm.

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** Optimal costly comm.

**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  Optimal costly comm.

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  Optimal costly comm.

Step 4  Distortion-transmission trade-off
Almost uniform and unimodal (ASU) distribution about $\alpha$

$$\pi_\alpha \geq \pi_{\alpha+1} \geq \pi_{\alpha-1} \geq \pi_{\alpha+2} \geq \cdots$$
Preliminaries

Almost uniform and unimodal (ASU) distribution about $\alpha$

$$\pi_{a} \geq \pi_{a+1} \geq \pi_{a-1} \geq \pi_{a+2} \geq \cdots$$

ASU Rearrangement

Estimation under communication constraints–(Mahajan and Chakravorty)
Almost uniform and unimodal (ASU) distribution about $\alpha$

\[ \pi_{\alpha} \geq \pi_{\alpha+1} \geq \pi_{\alpha-1} \geq \pi_{\alpha+2} \geq \cdots \]

ASU Rearrangement

\[ \pi \gg \pi^+ \]

Majorization

\[ \pi \geq \xi \text{ iff } \sum_{i=-n}^{n} \pi_i^+ \geq \sum_{i=-n}^{n} \xi_i^+ \text{ and } \sum_{i=-n}^{n+1} \pi_i^+ \geq \sum_{i=-n}^{n+1} \xi_i^+ \]

Invariant to permutations.

Estimation under communication constraints–(Mahajan and Chakravorty)
Properties of the value functions

Definition: \( \xi \triangleright \bar{\xi} \) if \( \xi \geq \bar{\xi} \) and \( \bar{\xi} \) is ASU about some point \( a \)

[LM11, NBTV13]
Step 1 Properties of the value functions

Definition

\( \xi \succ \tilde{\xi} \) if \( \xi \succeq \tilde{\xi} \) and \( \tilde{\xi} \) is ASU about some point \( a \)

Lemma

Similar to Schur-concavity

- If \( \xi \succ \tilde{\xi} \) then \( W_t(\xi) \geq W_t(\tilde{\xi}) \).
- If \( \pi \succ \tilde{\pi} \) then \( V_t(\pi) \geq V_t(\tilde{\pi}) \).
**Step 1  Properties of the value functions**  

[LM11, NBTV13]

**Definition**

$\xi \succ \tilde{\xi}$, if $\xi \geq \tilde{\xi}$, and $\tilde{\xi}$, is ASU about some point $a$.

**Lemma**

- If $\xi \succ \tilde{\xi}$ then $W_t(\xi) \geq W_t(\tilde{\xi})$.
- If $\pi \succ \tilde{\pi}$ then $V_t(\pi) \geq V_t(\tilde{\pi})$.

**Lemma (Arg min of $W$)**

If $\xi$ is ASU about $a$ then $a$ is the arg min of

$$V_t(\xi) = \min_{\hat{x} \in \mathcal{X}} \mathbb{E}[d(X_t - \hat{x}) + W_{t+1}(\Pi_{t+1}) | \Xi_t = \xi],$$
Step 1: Properties of the value functions

Definition
\( \xi \succ \tilde{\xi} \) if \( \xi \succeq \tilde{\xi} \) and \( \tilde{\xi} \) is ASU about some point \( a \).

Lemma
- If \( \xi \succ \tilde{\xi} \) then \( W_t(\xi) \geq W_t(\tilde{\xi}) \).
- If \( \pi \succ \tilde{\pi} \) then \( V_t(\pi) \geq V_t(\tilde{\pi}) \).

Lemma (Arg min of \( W \))
If \( \xi \) is ASU about \( a \) then \( a \) is the arg min of
\[
V_t(\xi) = \min_{\hat{x} \in \mathcal{X}} \mathbb{E}[d(X_t - \hat{x}) + W_{t+1}(\Pi_{t+1}) | \Xi_t = \xi],
\]

Lemma (Arg min of \( V \))
If \( \pi \) is ASU about \( a \) then the arg min of
\[
W_t(\pi) = \min_{\varphi: \mathcal{X} \rightarrow \{0, 1\}} \mathbb{E}[\lambda \varphi(X_t) + V_t(\Xi_t) | \Pi_t = \pi, \varphi_t = \varphi]
\]
is of the form
\[
\varphi(x) = \begin{cases} 
1, & \text{if } |x - a| > k(\pi) \\
0, & \text{if } |x - a| < k(\pi) \\
q_+, & \text{if } x - a = k(\pi) \\
q_-, & \text{if } x - a = -k(\pi)
\end{cases}
\]
Step 1  Structure of optimal strategies

$\pi_1$ is ASU about $z_0$
Step 1  Structure of optimal strategies

\[ \pi_1 \text{ is ASU about } z_0 \]

Is \( |x_1 - z_0| > k_1 ? \)
Step 1  Structure of optimal strategies

π₁ is ASU about \( z₀ \)

Is \( |x₁ - z₀| > k₁ \)?

\[ \text{NO. } u₁ = ε, z₁ = z₀ \]

ξ₁ is ASU about \( z₁ \)
Step 1 Structure of optimal strategies

$\pi_1$ is ASU about $z_0$

Is $|x_1 - z_0| > k_1$?

YES. $u_1 = 1$, $z_1 = x_1$

NO. $u_1 = \varepsilon$, $z_1 = z_0$

$\xi_1$ is ASU about $z_1$

$\xi_1$ is ASU about $z_1$
Step 1 Structure of optimal strategies

\[\pi_1 \text{ is ASU about } z_0\]

**Is** \(|x_1 - z_0| > k_1?\)**

**YES.** \(u_1 = 1, z_1 = x_1\)  
**NO.** \(u_1 = \varepsilon, z_1 = z_0\)

\(\xi_1 \text{ is ASU about } z_1\)

In both cases: \(\hat{x}_1 = z_1\)
Step 1: Structure of optimal strategies

\[ \pi_1 \text{ is ASU about } z_0 \]

\[ \text{Is } |x_1 - z_0| > k_1? \]

**YES.** \( u_1 = 1, z_1 = x_1 \)

**NO.** \( u_1 = \varepsilon, z_1 = z_0 \)

\[ \xi_1 \text{ is ASU about } z_1 \]

In both cases: \( \hat{x}_1 = z_1 \)
Step 1: Structure of optimal strategies

\[ \pi_1 \text{ is ASU about } z_0 \]

Is \(|x_1 - z_0| > k_1\)?

**YES.** \(u_1 = 1, z_1 = x_1\)

**NO.** \(u_1 = \varepsilon, z_1 = z_0\)

\[ \xi_1 \text{ is ASU about } z_1 \]

\[ \hat{x}_1 = z_1 \]

\[ X_2 = X_1 + W_1 \implies \pi_1 = \xi_1 \ast p \]

\(\pi_1 \text{ is ASU about } z_1\)

etc. . . .
Transmitted Process

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

[LM11, NBTV13]
Step 1 Structure of optimal estimator

Transmitted Process

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

Lemma

$\Xi_t$ is ASU about $Z_t$
Step 1 Structure of optimal estimator

Transmitted Process

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

Lemma

$\Xi_t$ is ASU about $Z_t$

Theorem

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

Remark

The optimal estimation strategy is time-homogeneous and can be specified in closed form.

[LM11, NBTV13]
Lemma \( \Pi_t \) is ASU about \( Z_{t-1} \)
Step 1  Structure of optimal transmitter  [LM11, NBTV13]

Lemma

$\Pi_t$ is ASU about $Z_{t-1}$

Theorem

$U_t = f_t(X_t, \Pi_t) = \begin{cases} 1, & \text{if } |X_t - E_t| \geq k_t \\ 0, & \text{if } |X_t - E_t| < k_t \end{cases}$
**Step 1  Structure of optimal transmitter**  

**Lemma**  
$\Pi_t$ is ASU about $Z_{t-1}$

**Theorem**  
$U_t = f_t(X_t, \Pi_t) = \begin{cases} 
1, & \text{if } |X_t - E_t| \geq k_t \\
0, & \text{if } |X_t - E_t| < k_t
\end{cases}$

**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

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

**Remark**  
The optimal transmission strategy is a function of the error process.
The results extend to infinite horizon setup under appropriate regularity conditions.

Time-homogeneous threshold-based strategies are optimal.
How do we find the optimal threshold-based strategy?
Step 1  Structure of optimal strategies

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

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

\[
\begin{align*}
E_t & \quad \tau^{(k)} \quad t \\
\text{Step 3} & \quad \text{Optimal costly comm.}
\end{align*}
\]

\[
\begin{align*}
& D_{\beta}^{(k+2)} \\
& D_{\beta}^{(k+1)} \\
& D_{\beta}^{(k)} \\
\lambda & \quad \lambda^{(k)} \quad \lambda^{(k+1)}
\end{align*}
\]

Step 4  Distortion-transmission trade-off

\[
\begin{align*}
D_{\beta}^*(\alpha) & \quad (N_{\beta}^{(k+1)}, D_{\beta}^{(k+1)}) \\
& \quad (N_{\beta}^{(k)}, D_{\beta}^{(k)}) \\
\alpha & \quad \alpha_c \quad 1
\end{align*}
\]
Consider a **threshold-based** strategy

\[
f^{(k)}(e) = \begin{cases} 
  1 & \text{if } |e| \geq k \\
  0 & \text{otherwise}
\end{cases}
\]
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 \( E_0 = 0 \)).
Consider a threshold-based strategy

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

Define

\[ L_{\beta}^{(k)}(e) = (1 - \beta) \mathbb{E} \left[ \sum_{t=0}^{\tau^{(k)}-1} \beta^t d(E_t) \bigg| E_0 = e \right]. \]

\[ M_{\beta}^{(k)}(e) = (1 - \beta) \mathbb{E} \left[ \sum_{t=0}^{\tau^{(k)}-1} \beta^t \bigg| E_0 = e \right]. \]

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

\[ L^{(k)}_{\beta}(e) = (1 - \beta) \mathbb{E} \left[ \sum_{t=0}^{\tau^{(k)} - 1} \beta^t d(E_t) \middle| E_0 = e \right] . \]

\[ M^{(k)}_{\beta}(e) = (1 - \beta) \mathbb{E} \left[ \sum_{t=0}^{\tau^{(k)} - 1} \beta^t \middle| E_0 = e \right] . \]

Proposition \( \{E_t\}_{t=0}^{\infty} \) is a regenerative process. By renewal theory,

\[ D^{(k)}_{\beta} := D_{\beta}(f^{(k)}, g^*) = \frac{L^{(k)}_{\beta}(0)}{M^{(k)}_{\beta}(0)} \quad \text{and} \quad N^{(k)}_{\beta} := N_{\beta}(f^{(k)}, g^*) = \frac{1}{M^{(k)}_{\beta}(0)} - (1 - \beta). \]
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 the first transmission (starting at $E_0 = 0$).

Computing $L^{(k)}_\beta$ and $M^{(k)}_\beta$ is sufficient to compute the performance of $f^{(k)}$ (i.e., to compute $D^{(k)}_\beta$ and $N^{(k)}_\beta$).

Define

\[
L^{(k)}_\beta(e) = (1 - \beta) \mathbb{E} \left[ \sum_{t=0}^{\tau^{(k)}-1} \beta^t d(E_t) \middle| E_0 = e \right].
\]

\[
M^{(k)}_\beta(e) = (1 - \beta) \mathbb{E} \left[ \sum_{t=0}^{\tau^{(k)}-1} \beta^t E_0 = e \right].
\]

Proposition

\[
\{E_t\}_{t=0}^{\infty} \text{ is a regenerative process. By renewal theory,}
\]

\[
D^{(k)}_\beta := D_\beta(f^{(k)}, g^*) = \frac{L^{(k)}_\beta(0)}{M^{(k)}_\beta(0)} \quad \text{and} \quad N^{(k)}_\beta := N_\beta(f^{(k)}, g^*) = \frac{1}{M^{(k)}_\beta(0)} - (1 - \beta).
\]
Step 2 Computing $L^{(k)}_{\beta}$ and $M^{(k)}_{\beta}$

Markov chain setup

$\mathbf{L}^{(k)}_{\beta}(e) = d(e) + \beta \sum_{n=-k}^{k} p_{n-e} \mathbf{L}^{(k)}_{\beta}(n)$

$\mathbf{M}^{(k)}_{\beta}(e) = 1 + \beta \sum_{n=-k}^{k} p_{n-e} \mathbf{M}^{(k)}_{\beta}(n)$
Step 2 Computing $L^{(k)}_{\beta}$ and $M^{(k)}_{\beta}$

Markov chain setup

$L^{(k)}_{\beta}(e) = d(e) + \beta \sum_{n=-k}^{k} p_{n-e} L^{(k)}_{\beta}(n)$

$M^{(k)}_{\beta}(e) = 1 + \beta \sum_{n=-k}^{k} p_{n-e} M^{(k)}_{\beta}(n)$

Proposition

$L^{(k)}_{\beta} = \left[I - \beta P^{(k)}\right]^{-1} d^{(k)}$. $P^{(k)}$ is substochastic.

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

Markov chain setup

\[ L^{(k)}_{\beta}(e) = d(e) + \beta \sum_{n=-k}^{k} p_{n-e} L^{(k)}_{\beta}(n) \]
\[ M^{(k)}_{\beta}(e) = 1 + \beta \sum_{n=-k}^{k} p_{n-e} M^{(k)}_{\beta}(n) \]

Proposition

\[ L^{(k)}_{\beta} = \left[ [I - \beta P^{(k)}]^{-1} d^{(k)} \right] \quad \text{P}^{(k)} \text{ is substochastic.} \]
\[ M^{(k)}_{\beta} = \left[ [I - \beta P^{(k)}]^{-1} 1^{(k)} \right] \]

Gauss-Markov setup

\[ L^{(k)}_{\beta}(e) = d(e) + \beta \int_{-k}^{k} \varphi(n-e) L^{(k)}_{\beta}(n) dn \]
\[ M^{(k)}_{\beta}(e) = 1 + \beta \int_{-k}^{k} \varphi(n-e) M^{(k)}_{\beta}(n) dn \]
Step 2 \hspace{1cm} \textbf{Computing} \( L^{(k)}_\beta \) \textbf{and} \( M^{(k)}_\beta \)

\textbf{Markov chain setup}

\begin{align*}
L^{(k)}_\beta (e) &= d(e) + \beta \sum_{n=-k}^{k} p_{n-e} L^{(k)}_\beta (n) \\
M^{(k)}_\beta (e) &= 1 + \beta \sum_{n=-k}^{k} p_{n-e} M^{(k)}_\beta (n)
\end{align*}

\textbf{Proposition}

\begin{align*}
L^{(k)}_\beta &= \left[ I - \beta P^{(k)} \right]^{-1} d^{(k)} . \quad P^{(k)} \text{ is substochastic.} \\
M^{(k)}_\beta &= \left[ I - \beta P^{(k)} \right]^{-1} 1^{(k)}
\end{align*}

\textbf{Gauss-Markov setup}

\begin{align*}
L^{(k)}_\beta (e) &= d(e) + \beta \int_{-k}^{k} \varphi(n-e) L^{(k)}_\beta (n) dn \\
M^{(k)}_\beta (e) &= 1 + \beta \int_{-k}^{k} \varphi(n-e) M^{(k)}_\beta (n) dn
\end{align*}

Fredholm Integral Equations of the 2nd kind.

Solutions exist and are unique.
Step 2 Computing $L^{(k)}_{\beta}$ and $M^{(k)}_{\beta}$

$D^{(k)}_{\beta}$ and $N^{(k)}_{\beta}$ can be computed using these expressions.

**Proposition**

\[
L^{(k)}_{\beta} = \left[ (I - \beta P^{(k)})^{-1} d^{(k)} \right].
\]

$P^{(k)}$ is substochastic.

\[
M^{(k)}_{\beta} = \left[ (I - \beta P^{(k)})^{-1} 1^{(k)} \right].
\]

**Gauss-Markov setup**

\[
L^{(k)}_{\beta}(e) = d(e) + \beta \int_{-k}^{k} \varphi(n - e) L^{(k)}_{\beta}(n) \, dn
\]

\[
M^{(k)}_{\beta}(e) = 1 + \beta \int_{-k}^{k} \varphi(n - e) M^{(k)}_{\beta}(n) \, dn
\]

Fredholm Integral Equations of the 2nd kind.

Solutions exist and are unique.
We found the performance of a generic threshold-based strategy.

How does this lead to identifying an optimal strategy?
Step 1  Structure of optimal strategies

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

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

Step 3  Optimal costly comm.

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

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

Implication:

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

Submodularity

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

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

**Monotonicity**

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

**Implication:**

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

**Submodularity**

\[ C^{(k)}_\beta (\lambda) := D^{(k)}_\beta + \lambda N^{(k)}_\beta \] is submodular in \((k, \lambda)\).

**Proposition**

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

Thus, optimal threshold increases with increase in \(\lambda\).
Characterizing the optimal threshold for a given communication cost is tricky.

Instead, we will characterize the optimal communication cost for a given threshold.
Define $\Lambda^{(k)}_\beta := \{ \lambda \in \mathbb{R}_{\geq 0} : k^*_\beta(\lambda) = k \} = [\lambda^{(k-1)}_\beta, \lambda^{(k)}_\beta]$.

$c^{(k)}_\beta(\lambda^{(k)}_\beta) = c^{(k+1)}_\beta(\lambda^{(k)}_\beta)$.
Define $\Lambda^{(k)}_\beta := \{ \lambda \in \mathbb{R}_{\geq 0} : k^*_\beta (\lambda) = k \} = \left[ \lambda^{(k-1)}_\beta , \lambda^{(k)}_\beta \right]$. 

$C^{(k)}_\beta (\lambda^{(k)}_\beta ) = C^{(k+1)}_\beta (\lambda^{(k)}_\beta )$
Define $\Lambda^{(k)}_\beta := \{ \lambda \in \mathbb{R}_{\geq 0} : k^*_\beta(\lambda) = k \} = [\lambda^{(k-1)}_\beta, \lambda^{(k)}_\beta]$.

$$C^{(k)}_\beta(\lambda^{(k)}_\beta) = C^{(k+1)}_\beta(\lambda^{(k)}_\beta)$$
Step 3: Optimal costly communication: Markov chain

Define $\Lambda^{(k)}_\beta := \{\lambda \in \mathbb{R}_\geq 0 : k^*_\beta (\lambda) = k\} = [\lambda^{(k-1)}_\beta, \lambda^{(k)}_\beta]$. Then:

$$C^{(k)}_\beta (\lambda^{(k)}_\beta) = C^{(k+1)}_\beta (\lambda^{(k)}_\beta)$$
Step 3  Optimal costly communication: Markov chain

Define $\Lambda^{(k)}_\beta := \{\lambda \in \mathbb{R}_{\geq 0} : k^*_\beta(\lambda) = k\}$

$$= [\lambda^{(k-1)}_\beta, \lambda^{(k)}_\beta].$$

$$C^{(k)}_\beta(\lambda^{(k)}_\beta) = C^{(k+1)}_\beta(\lambda^{(k)}_\beta)$$
Step 3 Optimal costly communication: Markov chain

Define $\Lambda^{(k)}_\beta := \{\lambda \in \mathbb{R}_{\geq 0} : k^*_\beta(\lambda) = k\} = [\lambda^{(k-1)}_\beta, \lambda^{(k)}_\beta]$.

$C^{(k)}_{\beta}(\lambda^{(k)}_\beta) = C^{(k+1)}_{\beta}(\lambda^{(k)}_\beta)$
Step 3 Optimal costly communication: Markov chain

\[ \lambda_{β}^{(k)} = \frac{D_{β}^{(k+1)} - D_{β}^{(k)}}{N_{β}^{(k)} - N_{β}^{(k+1)}} \]

**Theorem**

Strategy \( f^{(k+1)} \) is optimal for \( \lambda \in (\lambda_{β}^{(k)}, \lambda_{β}^{(k+1)}) \).

\[ C_{β}^{*}(λ) = \min_{k \in \mathbb{Z}_{≥0}} C_{β}^{(k)} \] is piecewise linear, continuous, concave, and increasing function of \( λ \).
Step 3 Optimal costly communication: Markov chain

Strategy $f^{(k+1)}$ is optimal for $\lambda \in (\lambda^{(k)}_\beta, \lambda^{(k+1)}_\beta]$. 

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

Estimation under communication constraints–(Mahajan and Chakravorty)
Example Symmetric birth-death Markov chain ($p = 0.3$)

\[ \beta = 1.0 \]
\[ \beta = 0.95 \]
\[ \beta = 0.9 \]
Step 3 Optimal costly communication: Gauss-Markov

Lemma

$D^{(k)}_{\beta}$ is increasing in $k$ and $N^{(k)}_{\beta}$ is decreasing in $k$.

$D^{(k)}_{\beta}$ and $N^{(k)}_{\beta}$ are differentiable in $k$. 
**Step 3** Optimal costly communication: Gauss-Markov

**Lemma**

\( D_{\beta}^{(k)} \) is increasing in \( k \) and \( N_{\beta}^{(k)} \) is decreasing in \( k \).

\( D_{\beta}^{(k)} \) and \( N_{\beta}^{(k)} \) are differentiable in \( k \).

**Theorem**

Strategy \( f^{(k)} \) is optimal for
\[
\lambda = -\frac{\partial_k D_{\beta}^{(k)}}{\partial_k N_{\beta}^{(k)}}
\]

\( C_{\beta}^*(\lambda) = \min_{k \in \mathbb{R}_{\geq 0}} C_{\beta}^{(k)} \) is continuous, concave, and increasing function of \( \lambda \).
Step 3  Optimal costly communication: Gauss-Markov

Lemma  \( D^{(k)}_\beta \) is increasing in \( k \) and \( N^{(k)}_\beta \) is decreasing in \( k \).
\( D^{(k)}_\beta \) and \( N^{(k)}_\beta \) are differentiable in \( k \).

Theorem  Strategy \( f^{(k)} \) is optimal for \( \lambda = -\frac{\partial_k D^{(k)}_\beta}{\partial_k N^{(k)}_\beta} \).
\( C^*_\beta (\lambda) = \min_{k \in \mathbb{R} \geq 0} C^{(k)}_\beta \) is continuous, concave, and increasing function of \( \lambda \).

Scaling with variance \( \sigma^2 \)  \( C^*_{\beta, \sigma}(\lambda) = \sigma^2 C^*_{\beta, 1} \left( \frac{\lambda}{\sigma^2} \right) \).
**Step 3** Optimal costly communication: Gauss-Markov

**Lemma**
- $D^{(k)}_\beta$ is increasing in $k$ and $N^{(k)}_\beta$ is decreasing in $k$.
- $D^{(k)}_\beta$ and $N^{(k)}_\beta$ are differentiable in $k$.

**Theorem**
- Strategy $f^{(k)}$ is optimal for $\lambda = -\frac{\partial_k D^{(k)}_\beta}{\partial_k N^{(k)}_\beta}$

$$C^*_\beta(\lambda) = \min_{k \in \mathbb{R}_{\geq 0}} C^{(k)}_\beta$$ is continuous, concave, and increasing function of $\lambda$.

**Scaling with variance $\sigma^2$**
- $C^*_{\beta, \sigma}(\lambda) = \sigma^2 C^*_\beta \left( \frac{\lambda}{\sigma^2} \right)$

**Computation**
- Use bisection search to find $k$ such that $\lambda = -\frac{\partial_k D^{(k)}_\beta}{\partial_k N^{(k)}_\beta}$
Example \textbf{Gauss-Markov with } $\sigma^2 = 1$

\begin{align*}
\beta &= 1.0 \\
\beta &= 0.95 \\
\beta &= 0.9
\end{align*}

![Graph showing $C^*_\beta(\lambda)$ for different values of $\beta$.](image)
**Step 1** Structure of optimal strategies

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

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

**Step 3** Optimal costly comm.

**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

\begin{enumerate}
\item[(C1)] \(N_\beta(f^\circ, g^\circ) = \alpha\)
\item[(C2)] There exists \(\lambda^\circ \geq 0\) such that \((f^\circ, g^\circ)\) is optimal for \(C_\beta(f, g; \lambda^\circ)\).
\end{enumerate}
Step 4: Distortion-transmission trade-off: Markov chain

Sufficient conditions for constrained optimality

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

\((C1)\) \(N_{\beta}(f^\circ, g^\circ) = \alpha\)

\((C2)\) 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

\[ N_{\beta}(k^*_\beta) > \alpha > N_{\beta}(k^*_\beta + 1) \]
Step 4  Distortion-transmission trade-off: Markov chain

Sufficient conditions for constrained optimality

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

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

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

Let \(k^*_\beta\) be such that

\[N_\beta^{(k^*_\beta)} > \alpha > N_\beta^{(k^*_\beta + 1)}\]
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

$N_\beta^{(k_\beta^*)} > \alpha > N_\beta^{(k_\beta^*+1)}$
Step 4 Distortion-transmission trade-off: Markov chain

Sufficient conditions for constrained optimality

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

1. \(N_\beta(f^o, g^o) = \alpha\) \hspace{2cm} \text{(C1)}

2. There exists \(\lambda^o \geq 0\) such that \((f^o, g^o)\) is optimal for \(C_\beta(f, g; \lambda^o)\). \hspace{2cm} \text{(C2)}

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

\[
\theta^* N_\beta^{(k)} + (1 - \theta^*) N_\beta^{(k+1)} = \alpha
\]
Sufficient conditions for constrained optimality

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

1. \(N^{(k)}(f, g) = \alpha\)
2. There exists \(\lambda \circ \geq 0\) such that \((f, g)\) is optimal for \(C^{(k)}(f, g; \lambda \circ)\).

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

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

Sufficient conditions for constrained optimality

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

(C1) \(N^{(k)} (f \circ, g \circ) = \alpha\)

(C2) There exists \(\lambda \circ \geq 0\) such that \((f \circ, g \circ)\) is optimal for \(C^{(k)}(f, g; \lambda \circ)\).

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

\[ \theta^* N_{(k)}^{(k)} + (1 - \theta^*) N_{(k+1)}^{(k+1)} = \alpha \]
Example Symmetric birth-death Markov chain ($p = 0.3$)

$$D^*_\beta(\alpha)$$

- $\beta = 1.0$
- $\beta = 0.95$
- $\beta = 0.9$
Step 4 Distortion-transmission trade-off: Gauss-Markov

Sufficient conditions for constrained optimality

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

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

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

Theorem

There exists a \(k^*_\beta(\alpha)\) such that \(N_{\beta}^{(k^*_\beta(\alpha))} = \alpha\). Therefore,

\[D^*_\beta(\alpha) = D_{\beta}^{(k^*_\beta(\alpha))}\]
Sufficient conditions for constrained optimality

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

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

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

Theorem

There exists a \(k^*_\beta(\alpha)\) such that \(N_{\beta}^{(k^*_\beta(\alpha))} = \alpha\). Therefore,

\[D^*_\beta(\alpha) = D_{\beta}^{(k^*_\beta(\alpha))}\]

Scaling with variance \(\sigma^2\)

\[D_{\beta, \sigma}^*(\alpha) = \sigma^2 D_{\beta, 1}^*(\alpha).\]
Sufficient conditions for constrained optimality

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

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

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

**Theorem**

There exists a \(k^*_\beta(\alpha)\) such that \(N_{\beta}^{(k^*_\beta(\alpha))} = \alpha\). Therefore,

\[D^*_\beta(\alpha) = D_{\beta}^{(k^*_\beta(\alpha))}\]

**Scaling with variance \(\sigma^2\)**

\[D^*_{\beta, \sigma}(\alpha) = \sigma^2 D^*_{\beta, 1}(\alpha).\]

**Computation**

Use bisection search to find \(k\) such that \(N_{\beta}^{(k)} = \alpha\).
Example: Gauss-Markov with $\sigma^2 = 1$

Graph showing $D^*_\beta(\alpha)$ for different values of $\beta$: $\beta = 1.0$, $\beta = 0.95$, $\beta = 0.9$. The graph plots $\alpha$ on the x-axis and $D^*_\beta(\alpha)$ on the y-axis.
Conclusion

Analyze fundamental limits of estimation under communication constraints
Conclusion

Analyze fundamental limits of estimation under communication constraints

Possible generalizations to more realistic models
- Packet drops
- Rate constraints (effect of quantization)
- Network delays
Conclusion

Analyze fundamental limits of estimation under communication constraints

Possible generalizations to more realistic models

- Packet drops
- Rate constraints (effect of quantization)
- Network delays

A simple non-trivial “toy-problem” for decentralized control

- Decentralized control is full of difficult problems and negative results.
- It is important to identify “easy” problems and positive results.
Conclusion

Analyze fundamental limits of estimation under communication constraints

Possible generalizations to more realistic models
- Packet drops
- Rate constraints (effect of quantization)
- Network delays

A simple non-trivial “toy-problem” for decentralized control
- Decentralized control is full of difficult problems and negative results.
- It is important to identify “easy” problems and positive results.

The system model

Markov Process

Distortion

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

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

\[ D_\beta(f, g) = (1 - \beta) \mathbb{E}_t \left[ \sum_{i=0}^{\infty} \beta^i d(X_t - \hat{X}_t) \right] \]

2. Average cost setup, \( \beta = 1 \)

\[ D_\beta(f, g) = \mathbb{E}_t \left[ \sum_{i=0}^{\infty} d(X_t - \hat{X}_t) \right] \]

\[ N_\beta(f, g) = (1 - \beta) \mathbb{E}_t \left[ \sum_{i=0}^{\infty} \beta^i U_i \right] \]

Structural results

Let \( f \) and \( g \) be as above. By renewal theory, for any \( \beta \in (0, 1) \):

\[ \mathbb{E}_t \left[ d(X_t - \hat{X}_t) \right] = \frac{1}{1 - \beta} \mathbb{E}_t \left[ \sum_{i=0}^{\infty} \beta^i d(X_t - \hat{X}_t) \right] \]

Step 2

Consider a \( D \)-average cost setup, \( 1 - k(f, g) = (k) \beta(\Pi) \) and for \( e \in \{1, \ldots, N \} \),

\[ \Pi_t = \{ X_t = e \} \]

\[ \mathbb{E}_t \left[ d(E_t) \right] = k \beta \]

\[ \mathbb{E}_t \left[ d(x_t) \right] = k \beta \]

Step 3

Optimal costly communication: Markov chain

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

Define \( L^{(k)}(e) \) and \( M^{(k)}(e) \)

\[ L^{(k)}(e) = (1 - \beta) \mathbb{E}_t \left[ \sum_{i=0}^{\tau^{(k)}-1} \beta^i d(E_i) | E_0 = e \right] \]

\[ M^{(k)}(e) = (1 - \beta) \mathbb{E}_t \left[ \sum_{i=0}^{\tau^{(k)}-1} \beta^i e_i | E_0 = e \right] \]

Proposition

\( \{ E_i \}_{i=0}^{\infty} \) is a regenerative process. By renewal theory,

\[ D^{(k)}_\beta = D_\beta(f^{(k)}, g^{(k)}) = \frac{\mathbb{E}_t \left[ \tau^{(k)} \right]}{M^{(k)}(0)} \]

\[ N^{(k)}_\beta = N_\beta(f^{(k)}, g^{(k)}) = \frac{1}{M^{(k)}(0)} \]

Define \( A^{(k)}_\beta \)

\[ A^{(k)}_\beta = \inf_{\lambda \in \mathbb{R}_+} \left\{ \lambda \mathbb{E}_{\beta} \left[ \sum_{i=0}^{\tau^{(k)}-1} \beta^i d(E_i) | E_0 = 0 \right] \right\} \]

Sufficient conditions for constrained optimality

A strategy \( \theta^{(k)} \) is ASU about \( z_0 \) if

\[ \lambda^{(k)} \mathbb{E}_t \left[ \sum_{i=0}^{\tau^{(k)}-1} \beta^i d(E_i) | E_0 = 0 \right] \]

Dealing with non-classical information structure

Non-Classical info. struct.

\[ g(C, L) = \psi(C) \]

Belongs to the class of tractable non-classical information structures (called partial-history-sharing) identified in [Mahajan-Nayar-Teneketzis 2013]

Optimization problems

Costly communication

For \( \lambda \in \mathbb{R}_+ \),

\[ C^{(k)}_\beta = C^{(k)}_\beta(f^{(k)}, g^{(k)}) = \inf_{\lambda \in \mathbb{R}_+} \left\{ D_\beta(f^{(k)}, g^{(k)}) + \lambda N_\beta(f^{(k)}, g^{(k)}) \right\} \]

Constrained communication

For \( \alpha \in [0, 1] \),

\[ D^{(k)}_\beta = D_\beta(f^{(k)}; \alpha) = \inf_{\lambda \in \mathbb{R}_+} \left\{ D_\beta(f^{(k)}; \alpha) + \lambda N_\beta(f^{(k)}; \alpha) \right\} \]

Estimation under communication constraints

Step 1: Structure of optimal constraints

Step 2: Performance of threshold strategies

Step 3: Optimal costly communication: Markov chain

Step 4: Distortion-transmission tradeoff: Markov chain

Estimation under communication constraints

Sufficient conditions for constrained optimality

A strategy \( \theta^{(k)} \) is ASU about \( z_0 \) if

\[ \lambda^{(k)} \mathbb{E}_t \left[ \sum_{i=0}^{\tau^{(k)}-1} \beta^i d(E_i) | E_0 = 0 \right] \]

Randomized strategy \( \theta^{(k)} \) is optimal where

\[ \theta^{(k)} N_\beta^{(k)} + (1 - \theta^{(k)} N_\beta^{(k)} \alpha = \alpha \]

Estimation under communication constraints

Non-Classical info. struct.

\[ g(C, L) = \psi(C) | L \]

Belongs to the class of tractable non-classical information structures (called partial-history-sharing) identified in [Mahajan-Nayar-Teneketzis 2013]