Remote state estimation over erasure channels: structure of optimal strategies and fundamental limits

Aditya Mahajan
McGill University

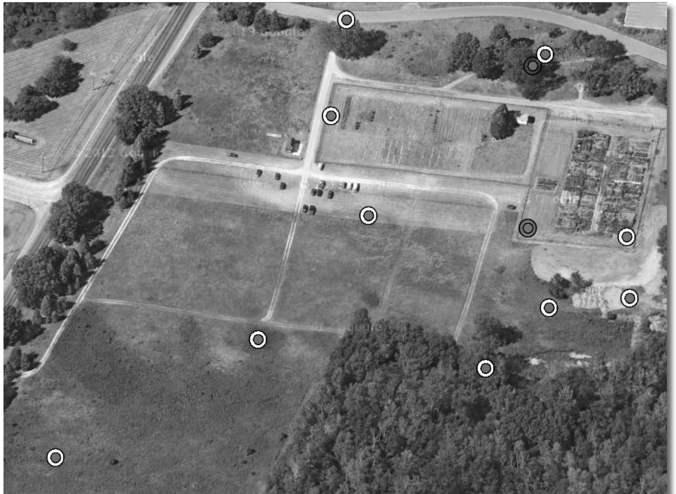
Joint work with Jhelum Chakravorty and Jayakumar Subramanian

Information Theory Forum, Stanford University
4 Nov, 2016

There is a need to revisit rate distortion theory to take network access into account.

Many applications require:

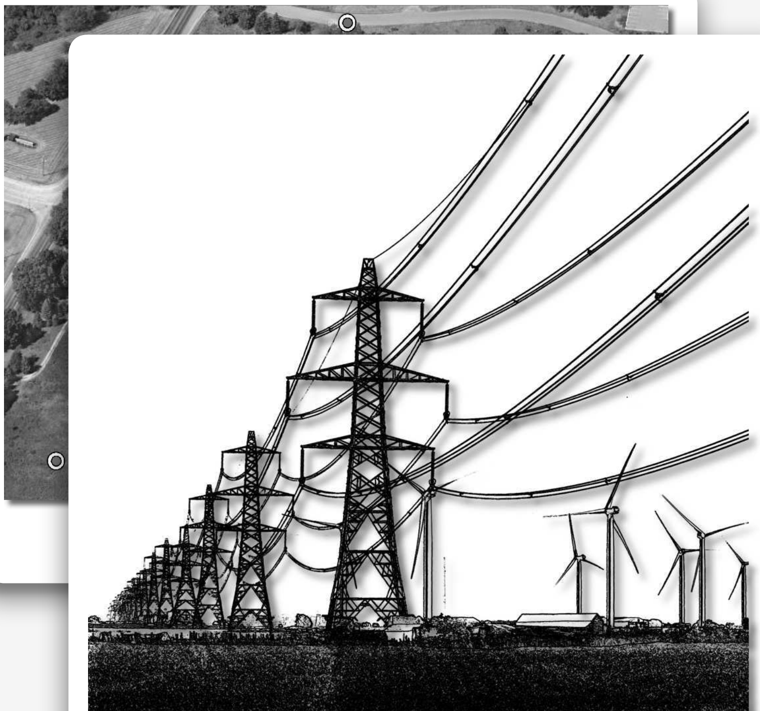
- Sequential transmission of data
- Zero- (or finite-) delay reconstruction



Sensor Networks

Many applications require:

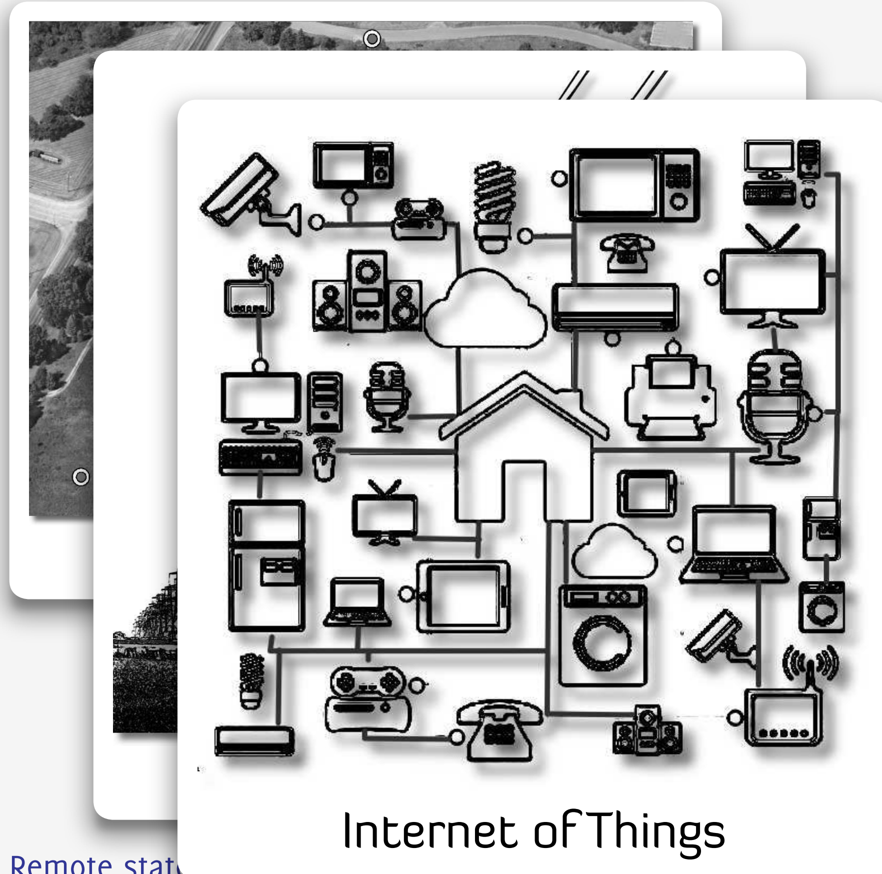
- Sequential transmission of data
- Zero- (or finite-) delay reconstruction



Smart Grids

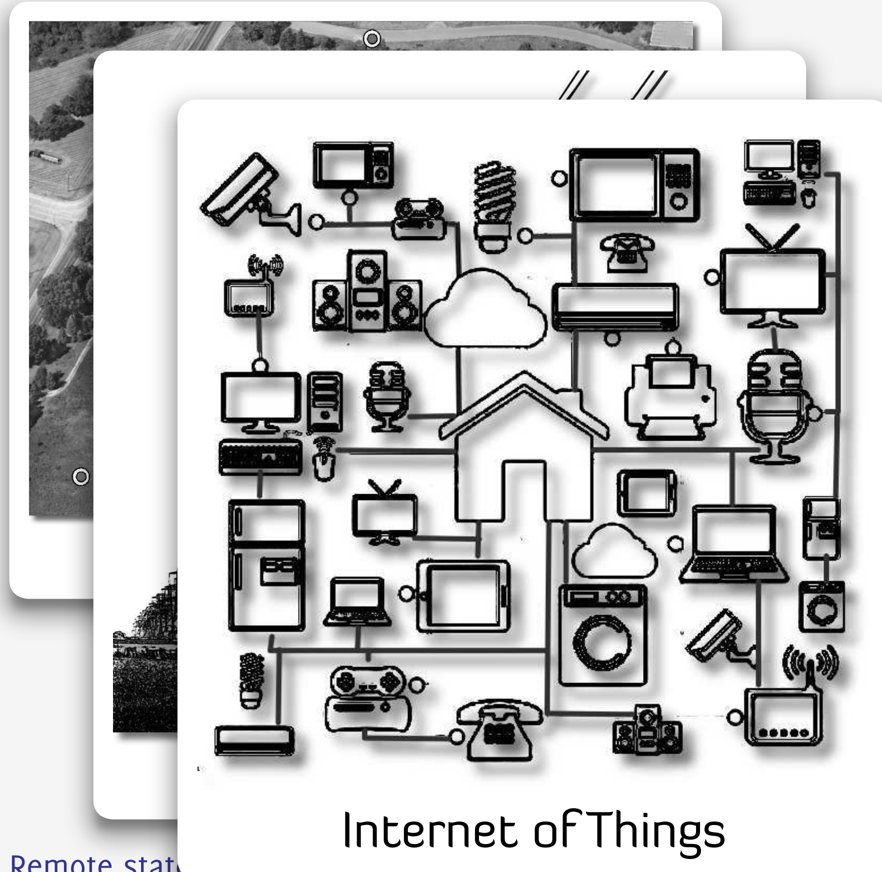
Many applications require:

- Sequential transmission of data
- Zero- (or finite-) delay reconstruction



Many applications require:

- Sequential transmission of data
- Zero- (or finite-) delay reconstruction

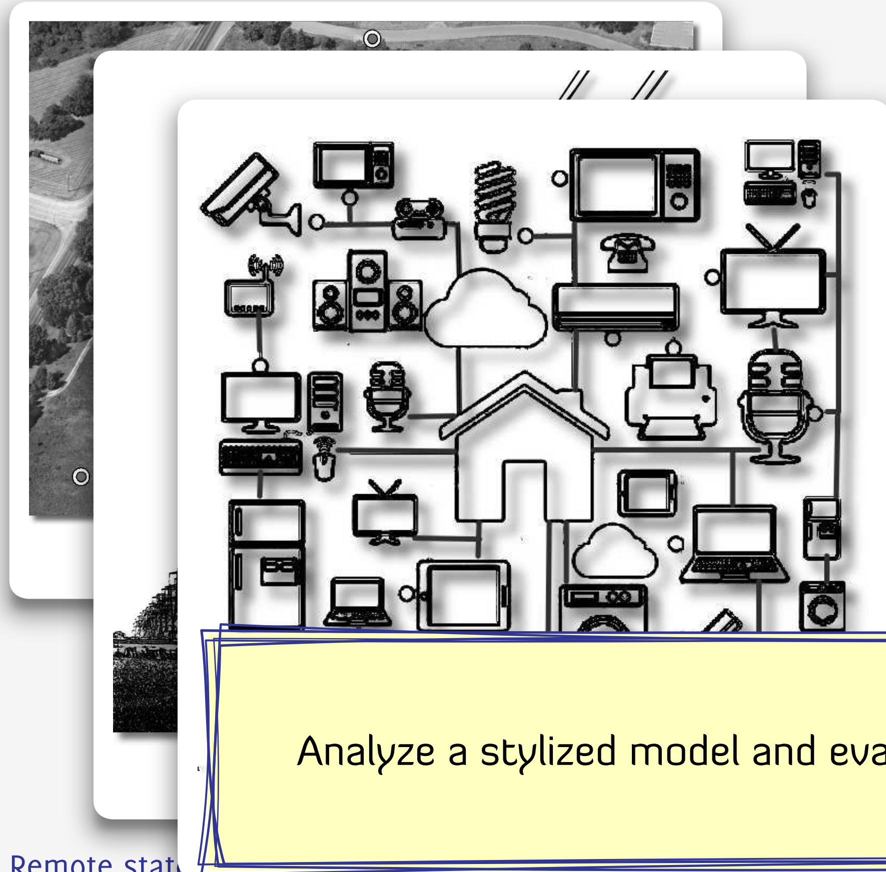


Many applications require:

- Sequential transmission of data
- Zero- (or finite-) delay reconstruction

Salient features

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



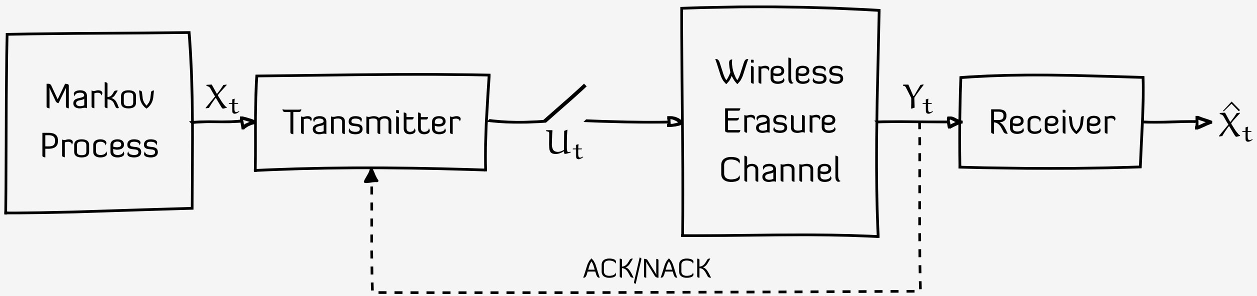
Many applications require:

- Sequential transmission of data
- Zero- (or finite-) delay reconstruction

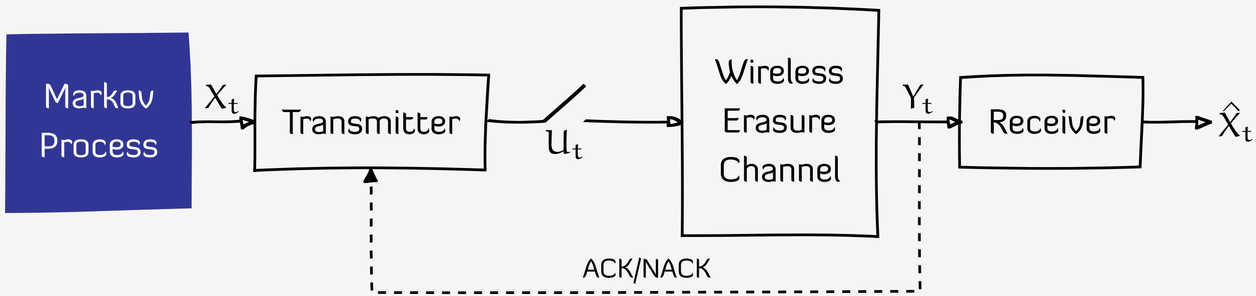
Salient features

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

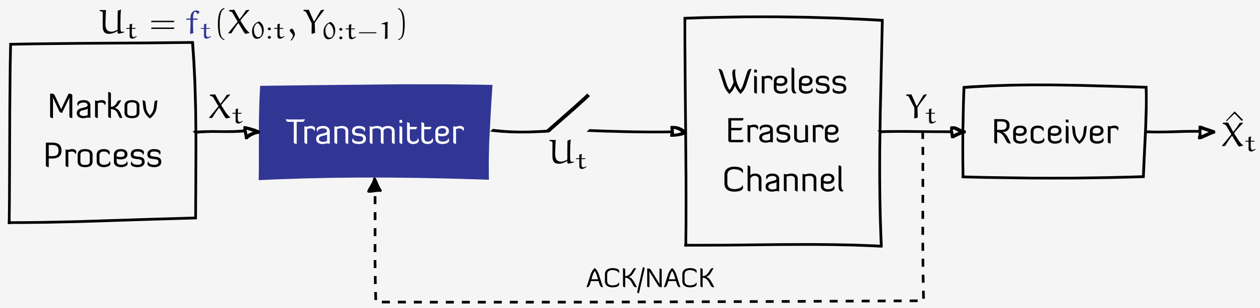
Analyze a stylized model and evaluate fundamental trade-offs



Markov Process
X_t
Transmitter
U_t
Wireless Erasure Channel
Y_t
Receiver
$\hat{X}_t$
ACK/NACK



- First order time-homogeneous Markov process
- The transmitter decides whether or not to transmit the current state
- The transmitted symbol is sent over an erasure channel (with acknowledgments)
- The receiver generates an estimate based on received symbol

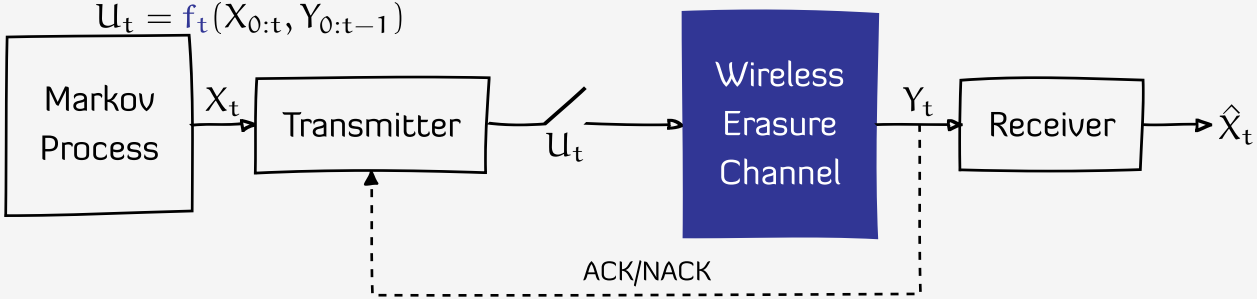


- First order time-homogeneous Markov process
- The transmitter decides whether or not to transmit the current state
- The transmitted symbol is sent over an erasure channel (with acknowledgments)
- The receiver generates an estimate based on received symbol

$$Y_t = \begin{cases} X_t, & \text{if } U_t = 1 \text{ and } S_t = \text{ON} \\ \mathfrak{E}, & \text{if } U_t = 0 \text{ or } S_t = \text{OFF} \end{cases}$$

$$\mathbb{P}(S_t = \text{OFF}) = \varepsilon$$

$$U_t = f_t(X_{0:t}, Y_{0:t-1})$$

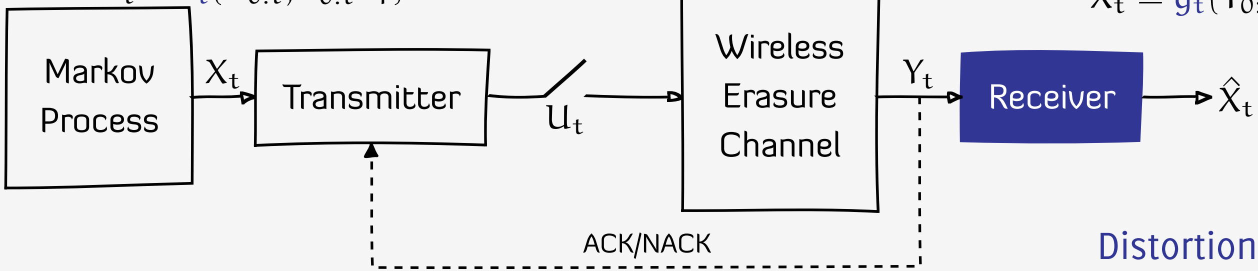


- First order time-homogeneous Markov process
- The transmitter decides whether or not to transmit the current state
- The transmitted symbol is sent over an erasure channel (with acknowledgments)
- The receiver generates an estimate based on received symbol

$$Y_t = \begin{cases} X_t, & \text{if } U_t = 1 \text{ and } S_t = \text{ON} \\ \mathfrak{E}, & \text{if } U_t = 0 \text{ or } S_t = \text{OFF} \end{cases}$$

$$\mathbb{P}(S_t = \text{OFF}) = \varepsilon$$

$$U_t = f_t(X_{0:t}, Y_{0:t-1}) \qquad \hat{X}_t = g_t(Y_{0:t})$$



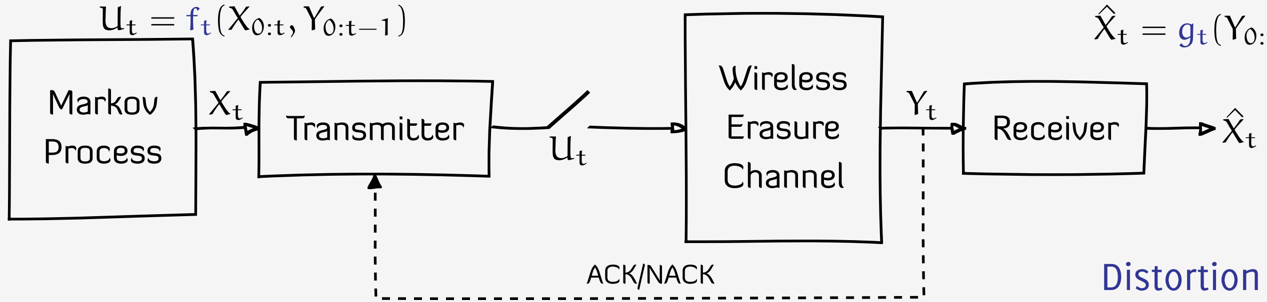
Distortion

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

- First order time-homogeneous Markov process
- The transmitter decides whether or not to transmit the current state
- The transmitted symbol is sent over an erasure channel (with acknowledgments)
- The receiver generates an estimate based on received symbol

$$Y_t = \begin{cases} X_t, & \text{if } U_t = 1 \text{ and } S_t = \text{ON} \\ \mathfrak{E}, & \text{if } U_t = 0 \text{ or } S_t = \text{OFF} \end{cases}$$

$$\mathbb{P}(S_t = \text{OFF}) = \varepsilon$$



Distortion

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

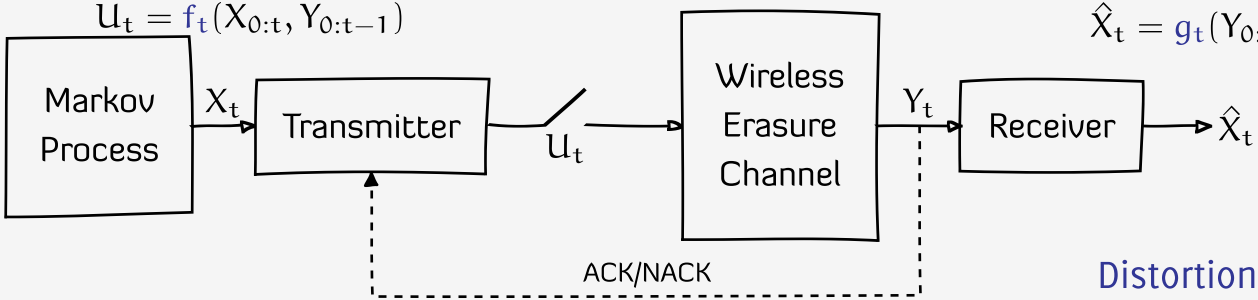
Communication Strategies

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

$$Y_t = \begin{cases} X_t, & \text{if } U_t = 1 \text{ and } S_t = \text{ON} \\ \mathfrak{E}, & \text{if } U_t = 0 \text{ or } S_t = \text{OFF} \end{cases}$$

$$\mathbb{P}(S_t = \text{OFF}) = \varepsilon$$

$$U_t = f_t(X_{0:t}, Y_{0:t-1}) \qquad \hat{X}_t = g_t(Y_{0:t})$$



Distortion

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

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

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

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

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

Optimization problems

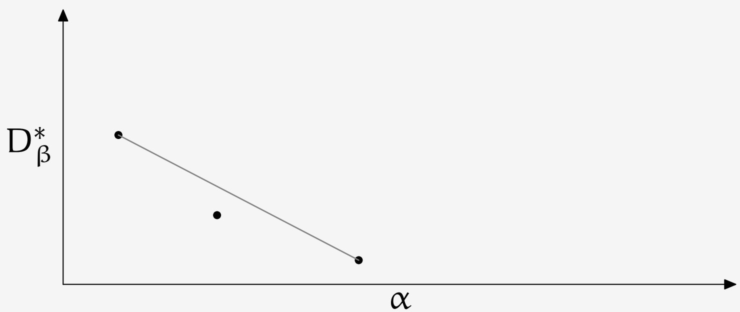
Constrained communication

$$\text{For } \alpha \in (0,1), \quad D^*_\beta(\alpha) \coloneqq \inf_{(f,g)} \left\{ D_\beta(f,g) : N_\beta(f,g) \leqslant \alpha \right\}$$

Optimization problems

Constrained communication

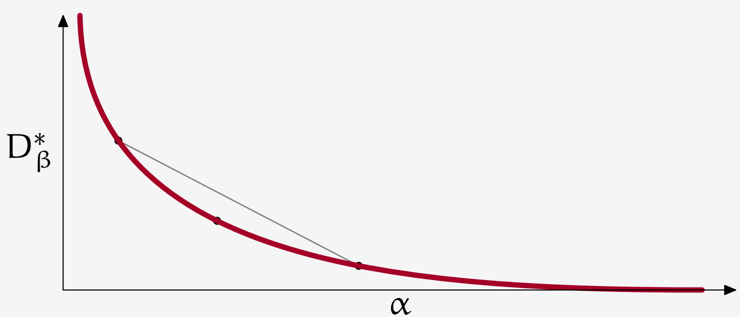
$$\text{For } \alpha \in (0,1), \quad D^*_\beta(\alpha) \coloneqq \inf_{(f,g)} \left\{ D_\beta(f,g) : N_\beta(f,g) \leqslant \alpha \right\}$$



Optimization problems

Constrained communication

$$\text{For } \alpha \in (0,1), \quad D^*_\beta(\alpha) \coloneqq \inf_{(f,g)} \{ D_\beta(f,g) : N_\beta(f,g) \leqslant \alpha \}$$



D^*_β is cts, dec, and convex

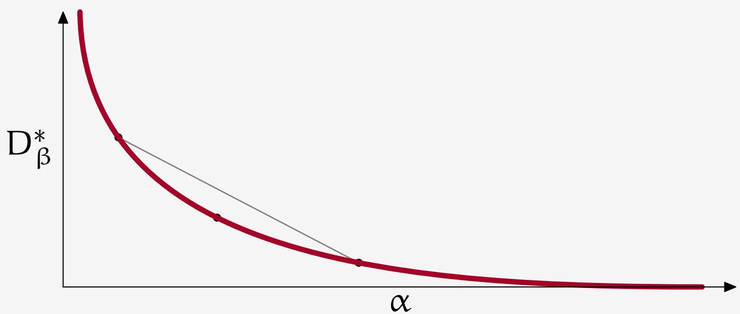
Optimization problems

Constrained communication

$$\text{For } \alpha \in (0,1), \quad D^*_\beta(\alpha) \coloneqq \inf_{(f,g)} \{ D_\beta(f,g) : N_\beta(f,g) \leqslant \alpha \}$$

Costly communication (Lagrange relaxation)

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



D^*_β is cts, dec, and convex

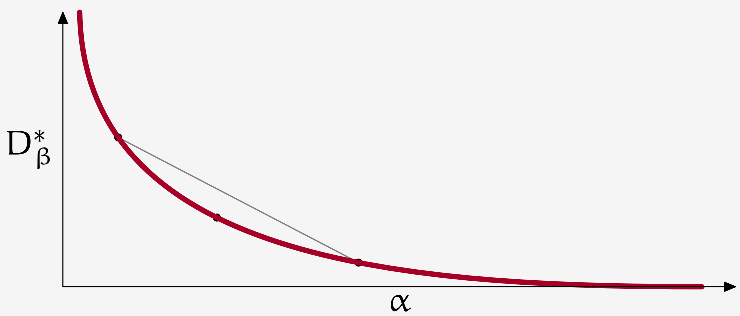
Optimization problems

Constrained communication

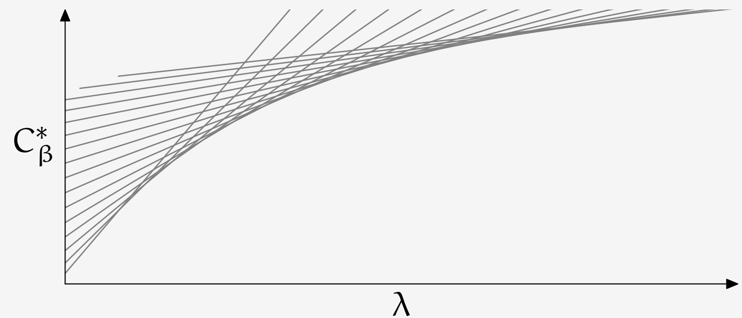
$$\text{For } \alpha \in (0,1), \quad D^*_\beta(\alpha) \coloneqq \inf_{(f,g)} \{ D_\beta(f,g) : N_\beta(f,g) \leqslant \alpha \}$$

Costly communication (Lagrange relaxation)

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



D^*_β is cts, dec, and convex



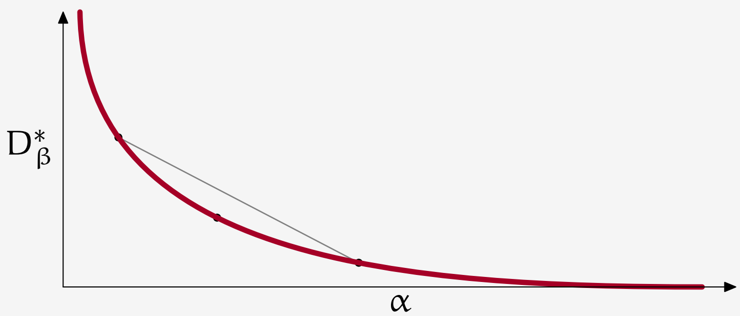
Optimization problems

Constrained communication

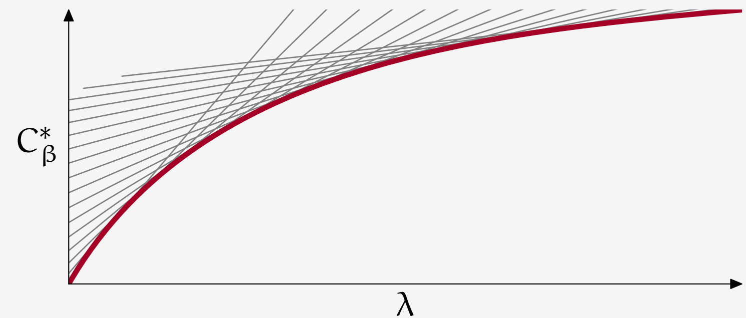
$$\text{For } \alpha \in (0,1), \quad D^*_\beta(\alpha) \coloneqq \inf_{(f,g)} \{ D_\beta(f,g) : N_\beta(f,g) \leqslant \alpha \}$$

Costly communication (Lagrange relaxation)

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



D^*_β is cts, dec, and convex



C^*_β is cts, inc, and concave

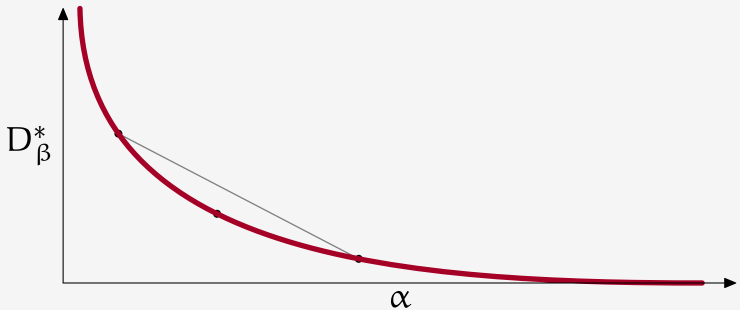
Optimization problems

Constrained communication

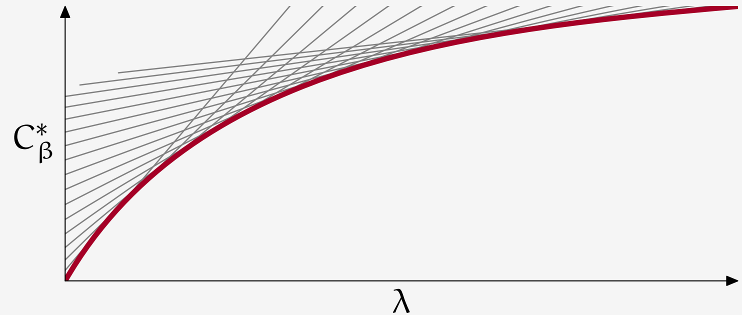
Costly c

Our result: Provide computable expressions for these trade-offs and identify optimal strategies that achieve them.

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



D^*_β is cts, dec, and convex



C^*_β is cts, inc, and concave

Comparison to Information Theory

- Costly communication is analogous to communication under power constraint.
- Constrained communication is analogous to distortion-rate function.
  So, we call it distortion-transmission function.
- Due to zero-delay reconstruction, information theoretic approaches do not apply.

Comparison to Information Theory

- Costly communication is analogous to communication under power constraint.
- Constrained communication is analogous to distortion-rate function.
  So, we call it distortion-transmission function.
- Due to zero-delay reconstruction, information theoretic approaches do not apply.

Previous work on remote-state estimation

- [Marshak 1954] Static (one-shot) problem with arbitrary source distribution
- [Kushner 1964] Off-line choice of measurement times
- [Åstrom Bernhardsson 2002] Lebesque sampling (or event-based sampling)

Comparison to Information Theory

- Costly communication is analogous to communication under power constraint.
- Constrained communication is analogous to distortion-rate function.
  So, we call it distortion-transmission function.
- Due to **zero-delay** reconstruction, information theoretic approaches do not apply.

Previous work on remote-state estimation

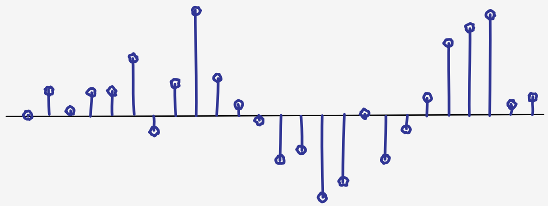
- [Marshak 1954] Static (one-shot) problem with arbitrary source distribution
- [Kushner 1964] Off-line choice of measurement times
- [Åstrom Bernhardsson 2002] Lebesque sampling (or event-based sampling)

Other related work

- Event-based estimation . . .
- Censoring sensors . . .
- Sensor sleep scheduling . . .
- Age of Information . . .

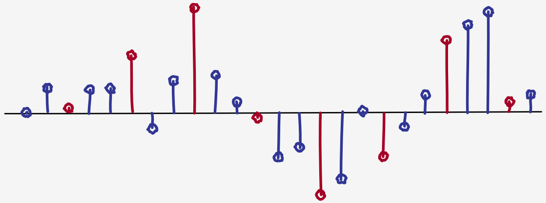
An illustrative example

$X_{t+1} = X_t + W_t$, $W_t \sim \mathcal{N}(0, 1)$. Perfect channel



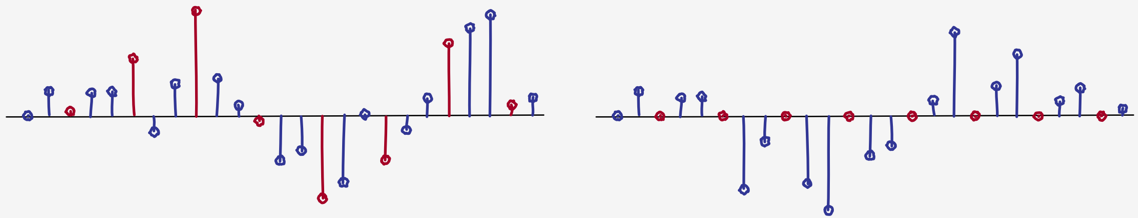
$X_{t+1} = X_t + W_t$, $W_t \sim \mathcal{N}(0, 1)$. Perfect channel

Periodic Transmission Strategy



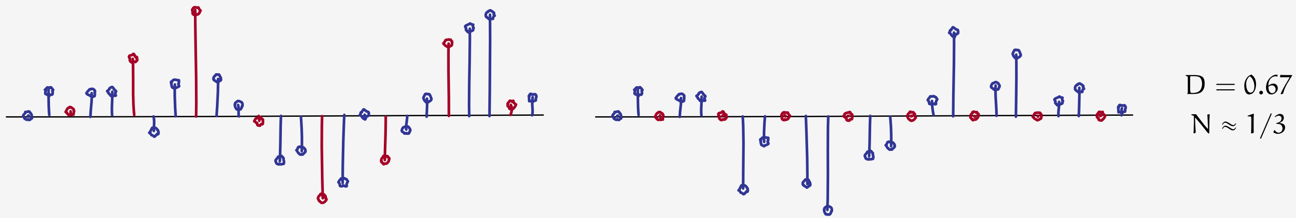
$X_{t+1} = X_t + W_t,\ W_t \sim \mathcal{N}(0, 1)$. Perfect channel

Periodic Transmission Strategy



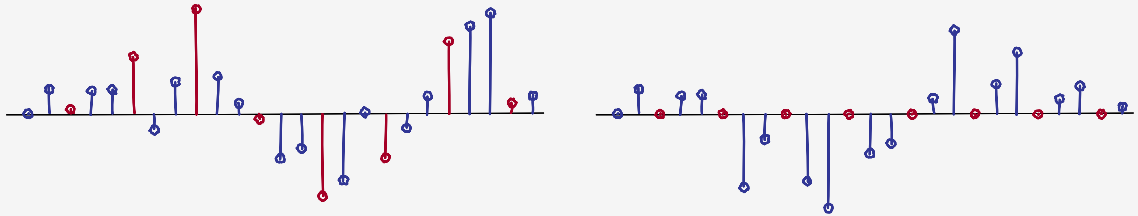
$X_{t+1} = X_t + W_t$, $W_t \sim \mathcal{N}(0, 1)$. Perfect channel

Periodic Transmission Strategy



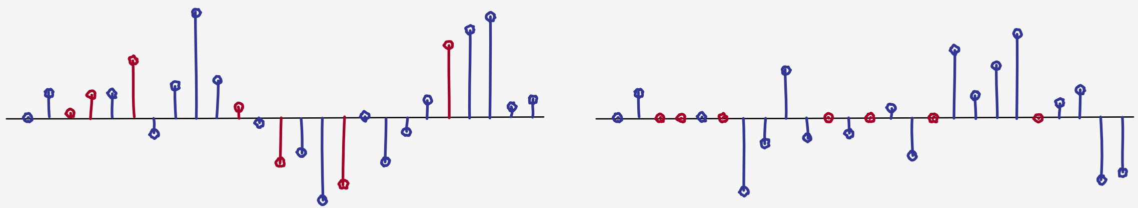
$X_{t+1} = X_t + W_t,\ W_t \sim \mathcal{N}(0,1)$. Perfect channel

Periodic Transmission Strategy



$D = 0.67$
$N \approx 1/3$

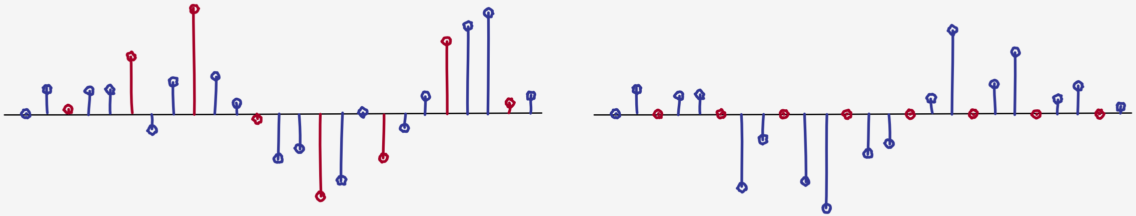
Randomized Transmission Strategy



$D = 2.00$
$N \approx 1/3$

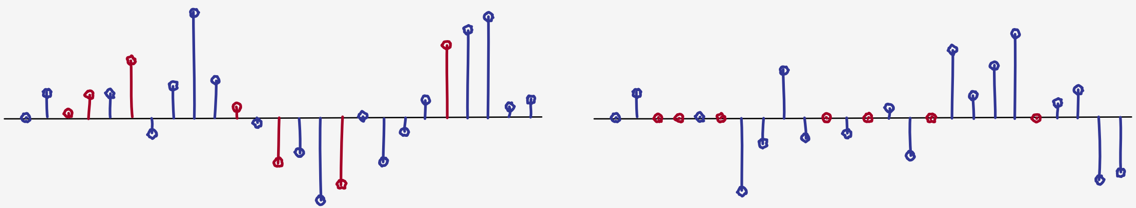
$X_{t+1} = X_t + W_t,\ W_t \sim \mathcal{N}(0,1)$. Perfect channel

Periodic Transmission Strategy



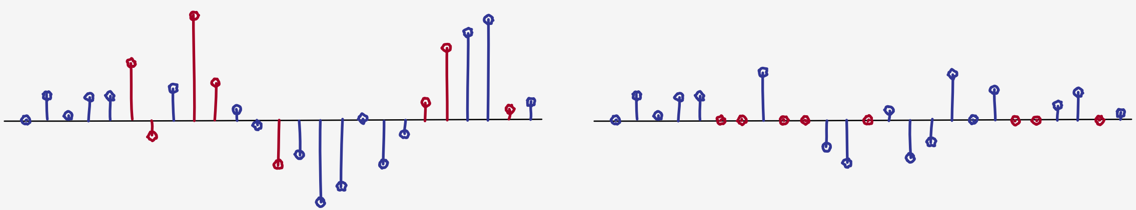
$D = 0.67$
$N \approx 1/3$

Randomized Transmission Strategy



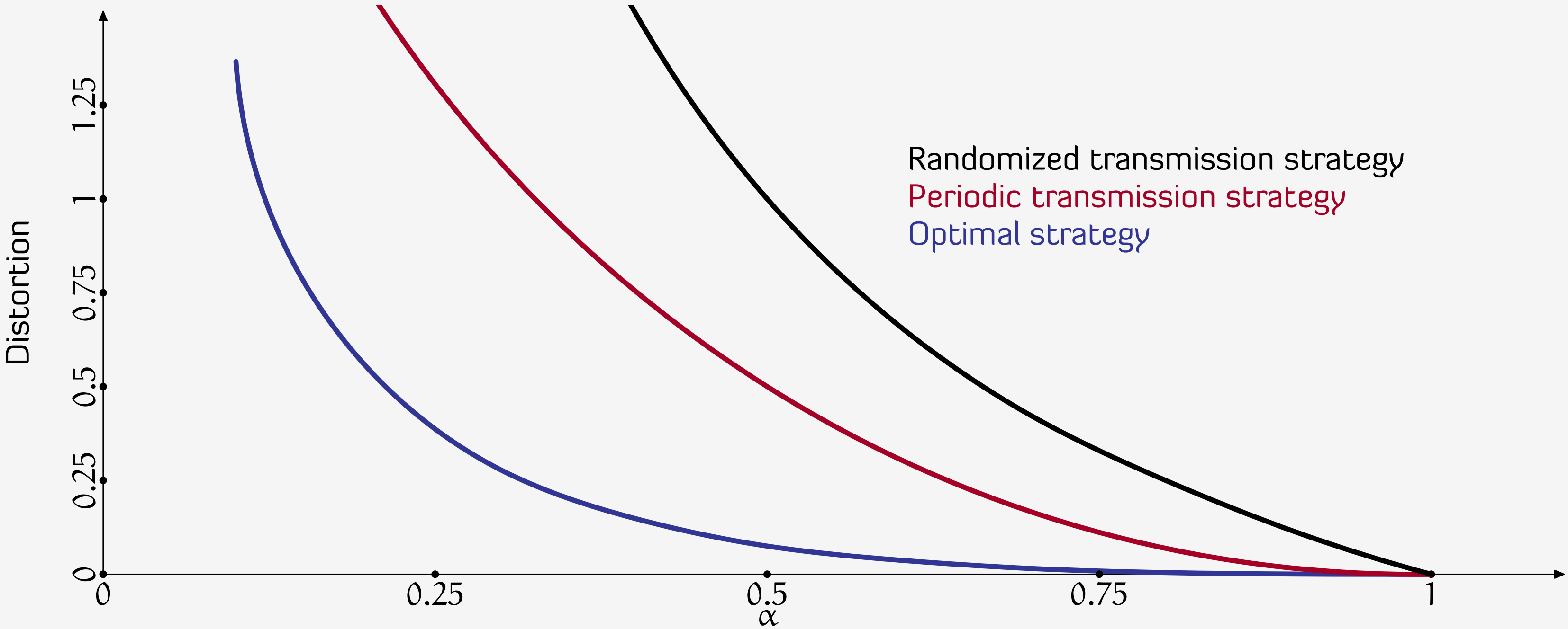
$D = 2.00$
$N \approx 1/3$

Optimal Transmission Strategy



$D = 0.24$
$N \approx 1/3$

Distortion-transmission trade-off: Perfect channel



What’s the conceptual difficulty?

Static (one-shot) problem

Static (one-shot) problem

$\mathcal{S} \subset \mathcal{X}$ is the silence set

Static (one-shot) problem

$\mathcal{X}$

$\mathcal{S} \subset \mathcal{X}$ is the silence set

$\hat{x}$ is the estimate when no packet is received

Static (one-shot) problem

$\mathcal{S} \subset \mathcal{X}$ is the silence set

$\hat{x}$ is the estimate when no packet is received

Cost when $x \in \mathcal{S}$

$d(x - \hat{x})$

Static (one-shot) problem

$\mathcal{S} \subset \mathcal{X}$ is the silence set

$\hat{x}$ is the estimate when no packet is received

Cost when $x \in \mathcal{S}$

$d(x - \hat{x})$

Cost when $x \notin \mathcal{S}$

$\lambda + \varepsilon d(x - \hat{x})$

Static (one-shot) problem

$\mathcal{S} \subset \mathcal{X}$ is the silence set

$\hat{x}$ is the estimate when no packet is received

Cost when $x \in \mathcal{S}$

$d(x - \hat{x})$

Cost when $x \notin \mathcal{S}$

$\lambda + \varepsilon d(x - \hat{x})$

Total expected cost

$$c(\hat{x}, \mathcal{S}) \coloneqq \lambda\, \mathbb{P}(X \notin \mathcal{S}) + \varepsilon \sum_{x \notin \mathcal{S}} \mathbb{P}(X = x) d(x - \hat{x}) + \sum_{x \in \mathcal{S}} \mathbb{P}(X = x) d(x - \hat{x})$$

Static (one-shot) problem

$\mathcal{S} \subset \mathcal{X}$ is the silence set

$\hat{x}$ is the estimate when no packet is received

Cost when $x \in \mathcal{S}$

$d(x - \hat{x})$

Cost when $x \notin \mathcal{S}$

$\lambda + \varepsilon d(x - \hat{x})$

Total expected cost

$$c(\hat{x}, \mathcal{S}) \coloneqq \lambda\, \mathbb{P}(X \notin \mathcal{S}) + \varepsilon \sum_{x \notin \mathcal{S}} \mathbb{P}(X = x) d(x - \hat{x}) + \sum_{x \in \mathcal{S}} \mathbb{P}(X = x) d(x - \hat{x})$$

Choose $(\hat{x}, \mathcal{S})$ to minimize $c(\hat{x}, \mathcal{S})$.

Set-valued (or combinatorial) optimization problem.

Dynamic problem

$\mathcal{S}_1^1 \subset \mathcal{X}$ is the silence set

$\hat{x}_1$ is the estimate when no packet is received

Dynamic problem

$\mathcal{S}^1_1 \subset \mathcal{X}$ is the silence set

$\hat{x}_1$ is the estimate when no packet is received

If a packet is received

$\mathcal{S}^1_2(x_1) \subset \mathcal{X}$ is the silence set

$\hat{x}^1_2$ is the estimate when no packet is received

Dynamic problem

$\mathcal{S}_1^1 \subset \mathcal{X}$ is the silence set

$\hat{x}_1$ is the estimate when no packet is received

If a packet is received

$\mathcal{S}_2^1(x_1) \subset \mathcal{X}$ is the silence set

$\hat{x}_2^1$ is the estimate when no packet is received

If a packet is not received

$\mathcal{S}_2^0(\mathcal{S}_1^1) \subset \mathcal{X}$ is the silence set

$\hat{x}_2^0(\mathcal{S}_1^1)$ is the estimate when no packet is received

Dynamic problem

$\mathcal{S}_1^1 \subset \mathcal{X}$ is the silence set

$\hat{x}_1$ is the estimate when no packet is received

If a packet is received

$\mathcal{S}_2^1(x_1) \subset \mathcal{X}$ is the silence set

$\hat{x}_2^1$ is the estimate when no packet is received

If a packet is not received

$\mathcal{S}_2^0(\mathcal{S}_1^1) \subset \mathcal{X}$ is the silence set

$\hat{x}_2^0(\mathcal{S}_1^1)$ is the estimate when no packet is received

Sequential optimization problem where the optimization problem at each step is a set-valued optimization problem that depends on a history of previously chosen sets!.

Exhaustive search complexity: $\left(|\mathcal{X}|2^{|\mathcal{X}|}\right)^{(2^{|\mathcal{X}|})^T}$

Main results

Distortion transmission function for auto-regressive sources

Source model $X_{t+1} = aX_t + W_t$, where W_t has symmetric and unimodal distribution. $X_t \in \mathbb{Z}/\mathbb{R}$.

Distortion transmission function for auto-regressive sources

Source model $X_{t+1} = aX_t + W_t$, where W_t has symmetric and unimodal distribution. $X_t \in \mathbb{Z}/\mathbb{R}$.

Optimal estimation strategy

$$\hat{X}_t = \begin{cases} a\hat{X}_{t-1}, & \text{if } Y_t = \mathfrak{E} \\ Y_t, & \text{if } Y_t \neq \mathfrak{E} \end{cases}$$

Distortion transmission function for auto-regressive sources

Source model $X_{t+1} = aX_t + W_t$, where W_t has symmetric and unimodal distribution. $X_t \in \mathbb{Z}/\mathbb{R}$.

Optimal transmission strategy

$$U_t = \begin{cases} 1, & \text{if } |X_t - a\hat{X}_{t-1}| \geqslant k \\ 0, & \text{otherwise} \end{cases}$$

Optimal estimation strategy

$$\hat{X}_t = \begin{cases} a\hat{X}_{t-1}, & \text{if } Y_t = \mathfrak{E} \\ Y_t, & \text{if } Y_t \neq \mathfrak{E} \end{cases}$$

Distortion transmission function for auto-regressive sources

Source model $X_{t+1} = aX_t + W_t$, where W_t has symmetric and unimodal distribution. $X_t \in \mathbb{Z}/\mathbb{R}$.

Optimal transmission strategy

$$U_t = \begin{cases} 1, & \text{if } |X_t - a\hat{X}_{t-1}| \geqslant k \\ 0, & \text{otherwise} \end{cases}$$

Optimal estimation strategy

$$\hat{X}_t = \begin{cases} a\hat{X}_{t-1}, & \text{if } Y_t = \mathfrak{E} \\ Y_t, & \text{if } Y_t \neq \mathfrak{E} \end{cases}$$

Salient features

- The transmitter does not try to send information through timing events.
- The estimation strategy is the same to the one for intermittent observations and **does not depend on the choice of the threshold**

Distortion transmission function for auto-regressive sources

Source model $X_{t+1} = aX_t + W_t$, where W_t has symmetric and unimodal distribution. $X_t \in \mathbb{Z}/\mathbb{R}$.

Optimal transmission strategy

$$U_t = \begin{cases} 1, & \text{if } |X_t - a\hat{X}_{t-1}| \geq k \\ 0, & \text{otherwise} \end{cases}$$

Optimal estimation strategy

$$\hat{X}_t = \begin{cases} a\hat{X}_{t-1}, & \text{if } Y_t = \mathfrak{E} \\ Y_t, & \text{if } Y_t \neq \mathfrak{E} \end{cases}$$

Performance of threshold based strategies

- $K_\beta^{(k)}$: Expected discounted number of transmissions until first successful reception.
- $L_\beta^{(k)}$: Expected discounted distortion until first successful reception.
- $M_\beta^{(k)}$: Expected discounted time until first successful reception.

Distortion transmission function for auto-regressive sources

Source model $X_{t+1} = aX_t + W_t$, where W_t has symmetric and unimodal distribution. $X_t \in \mathbb{Z}/\mathbb{R}$.

Optimal transmission strategy

$$U_t = \begin{cases} 1, & \text{if } |X_t - a\hat{X}_{t-1}| \geq k \\ 0, & \text{otherwise} \end{cases}$$

Optimal estimation strategy

$$\hat{X}_t = \begin{cases} a\hat{X}_{t-1}, & \text{if } Y_t = \mathfrak{E} \\ Y_t, & \text{if } Y_t \neq \mathfrak{E} \end{cases}$$

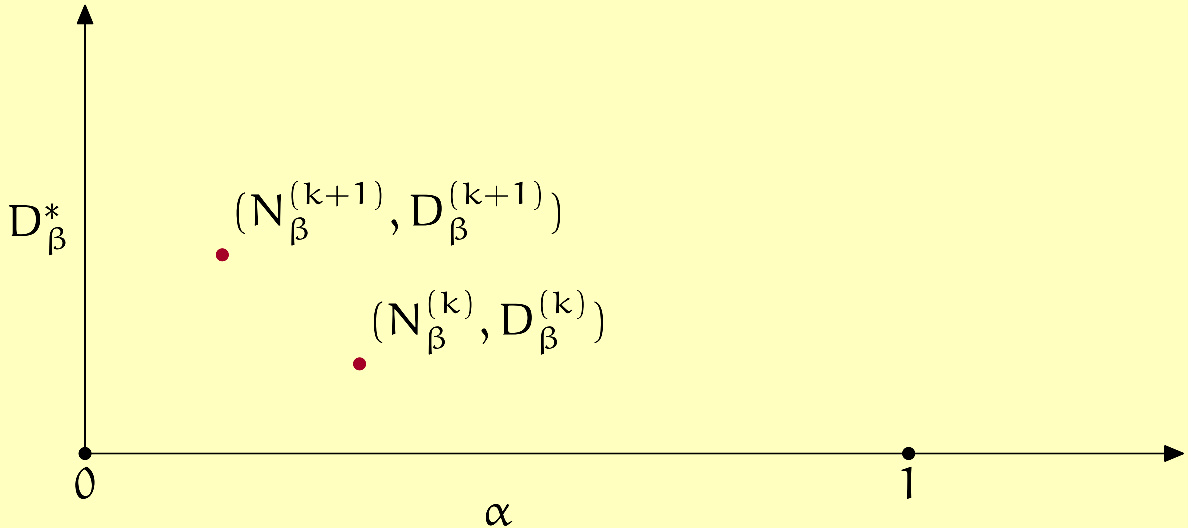
Performance of threshold based strategies

- $K_\beta^{(k)}$: Expected discounted number of transmissions until first successful reception.
- $L_\beta^{(k)}$: Expected discounted distortion until first successful reception.
- $M_\beta^{(k)}$: Expected discounted time until first successful reception.

Then, $D_\beta^{(k)} = \dfrac{L_\beta^{(k)}}{M_\beta^{(k)}}$ and $N_\beta^{(k)} = \dfrac{K_\beta^{(k)}}{M_\beta^{(k)}}$.

Distortion transmission function for auto-regressive sources

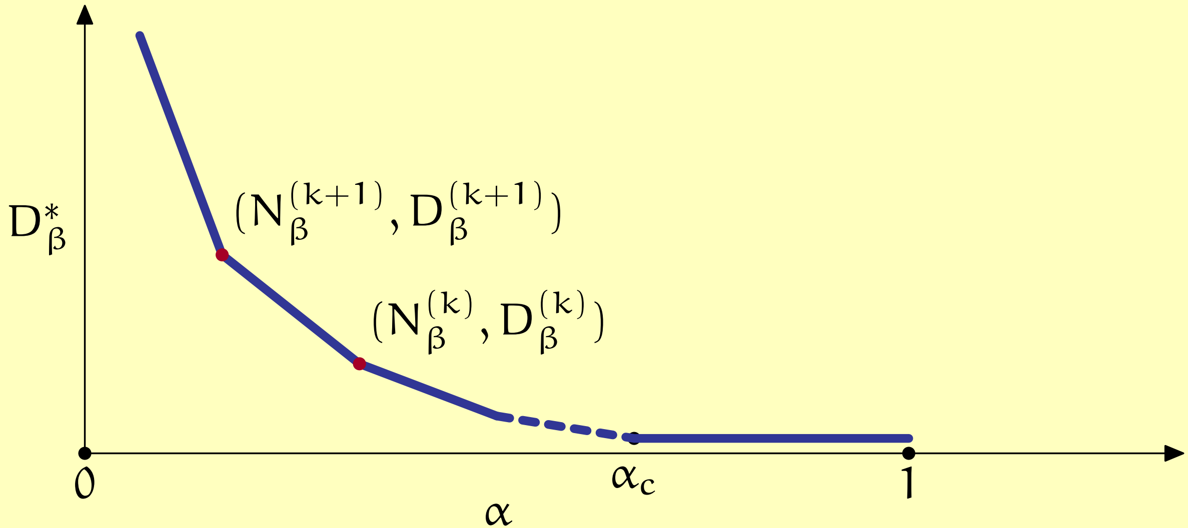
Source model ... on. $X_t \in \mathbb{Z}/\mathbb{R}$.



D^*_β is PWL, dec, and convex

Distortion transmission function for auto-regressive sources

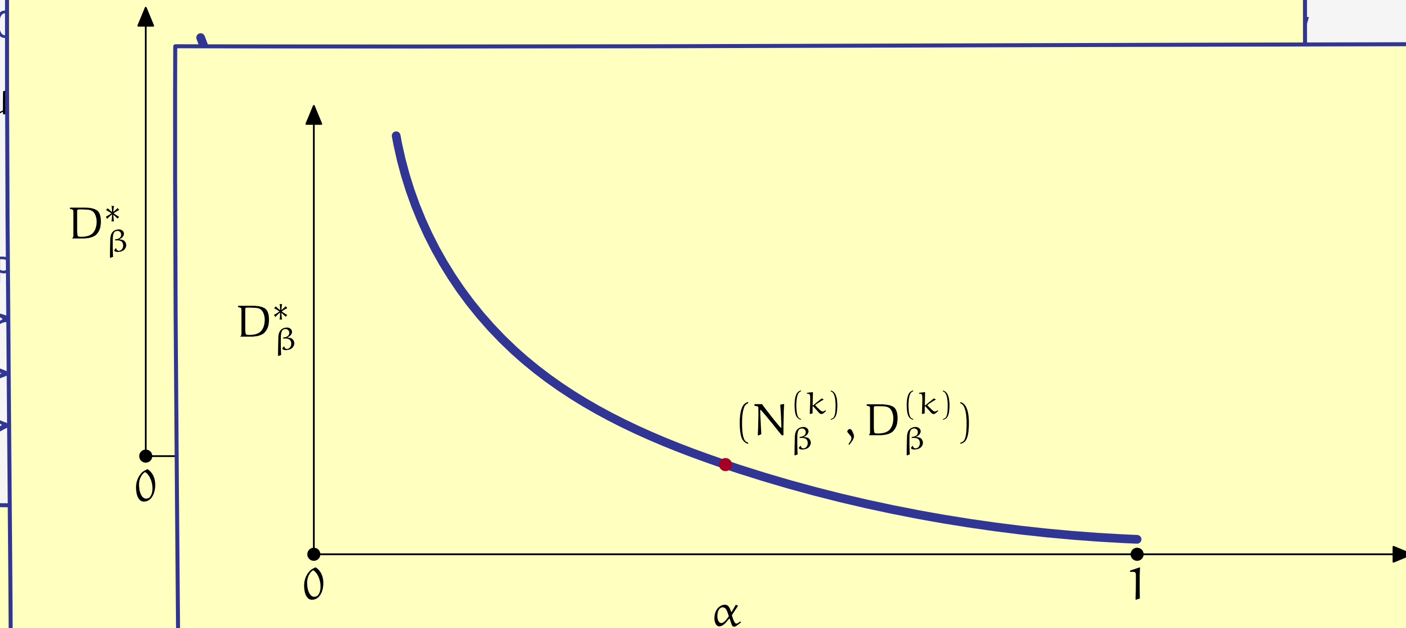
Source model $X_{t+1} = aX_t + W_t$ where W_t has symmetric and unimodal distribution. $X_t \in \mathbb{Z}/\mathbb{R}$.



D^*_β is PWL, dec, and convex

Distortion transmission function for auto-regressive sources

Source model … on. $X_t \in \mathbb{Z}/\mathbb{R}$.



D^*_β is cts, dec, and convex

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 continuous Markov processes

Provide simulation-based algorithms to compute optimal thresholds

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 continuous Markov processes

Provide simulation-based algorithms to compute optimal thresholds

Beautiful example of stochastics and optimization

Decentralized stochastic control (or team theory) and POMDPs

Stochastic orders and majorization

Markov chain analysis, stopping times, and renewal theory

Constrained MDPs and Lagrangian relaxations

Stochastic approximation and simulation based optimization

Solution methodology

Standard technique

- **Achievability**: Identify a good strategy and evaluate its performance.
- **Converse**: Determine a lower bound on distortion.

Solution methodology

Standard technique

- **Achievability**: Identify a good strategy and evaluate its performance.
- **Converse**: Determine a lower bound on distortion.
- **Hope**: The two curves match

Solution methodology

Standard technique

- **Achievability**: Identify a good strategy and evaluate its performance.
- **Converse**: Determine a lower bound on distortion.
- **Hope**: The two curves match

> Converse bounds are hard! Especially for sequential models.

Solution methodology

Standard technique

- **Achievability**: Identify a good strategy and evaluate its performance.
- **Converse**: Determine a lower bound on distortion.
- **Hope**: The two curves match

Converse bounds are hard! Especially for sequential models.

Our approach

- Model the optimization problem as a decentralized stochastic control problem.
  [Witsenhausen 1979, Walrand-Varaiya 1983, Teneketzis 2006, Mahajan-Teneketzis 2009, Kaspi-Merhav 2012, Asnani-Weissman 2013, Yüksel 2013 . . .]
- The system has two decision makers: the transmitter and the estimator, that have access to different information.

Solution methodology

Standard technique

- **Achievability**: Identify a good strategy and evaluate its performance.
- **Converse**: Determine a lower bound on distortion.
- **Hope**: The two curves match

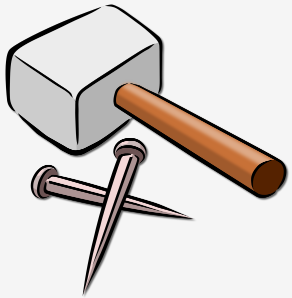
> Converse bounds are hard! Especially for sequential models.

Our approach

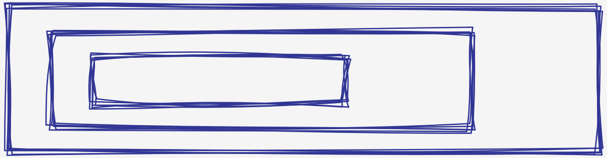
- Model the optimization problem as a decentralized stochastic control problem.
  [Witsenhausen 1979, Walrand-Varaiya 1983, Teneketzis 2006, Mahajan-Teneketzis 2009, Kaspi-Merhav 2012, Asnani-Weissman 2013, Yüksel 2013 . . .]
- The system has two decision makers: the transmitter and the estimator, that have access to different information.
- Identify qualitative properties of optimal strategies
- Identify a dynamic programming decomposition
- Determine optimal strategies based on the dynamic program.

So how do we start?

Decentralized stochastic control

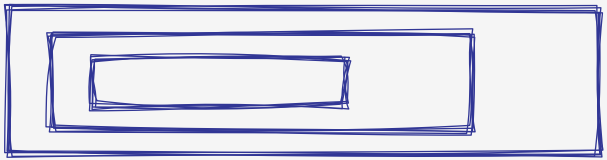


Dealing with non-classical information structure

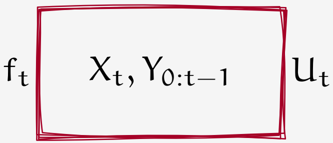


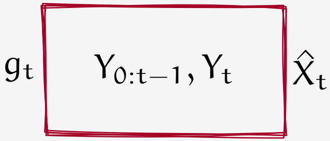
Classical info. struct.

Dealing with non-classical information structure

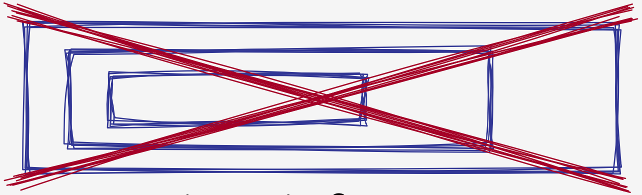


Classical info. struct.





Dealing with non-classical information structure

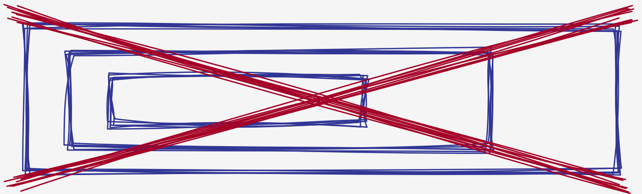


Non-Classical info. struct.

f_t [$X_t, Y_{0:t-1}$] U_t

g_t [$Y_{0:t-1}, Y_t$] $\hat{X}_t$

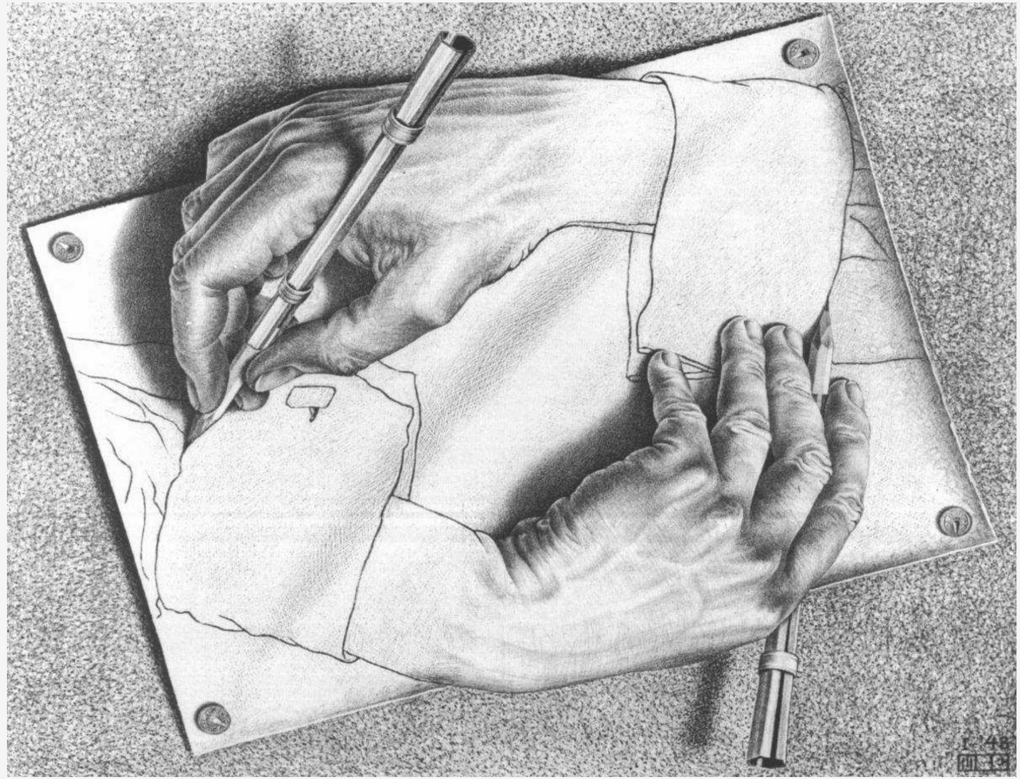
Dealing with non-classical information structure



Non-Classical info. struct.

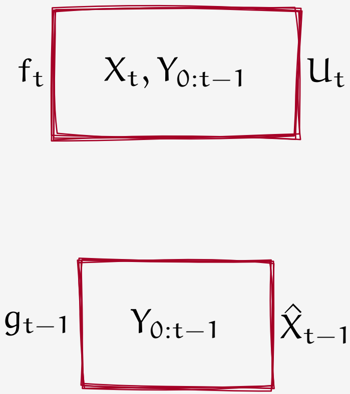
f_t | $X_t, Y_{0:t-1}$ | U_t

g_t | $Y_{0:t-1}, Y_t$ | $\hat{X}_t$



The common information approach (Nayyar, Mahajan, Teneketzis 2013)

Original system

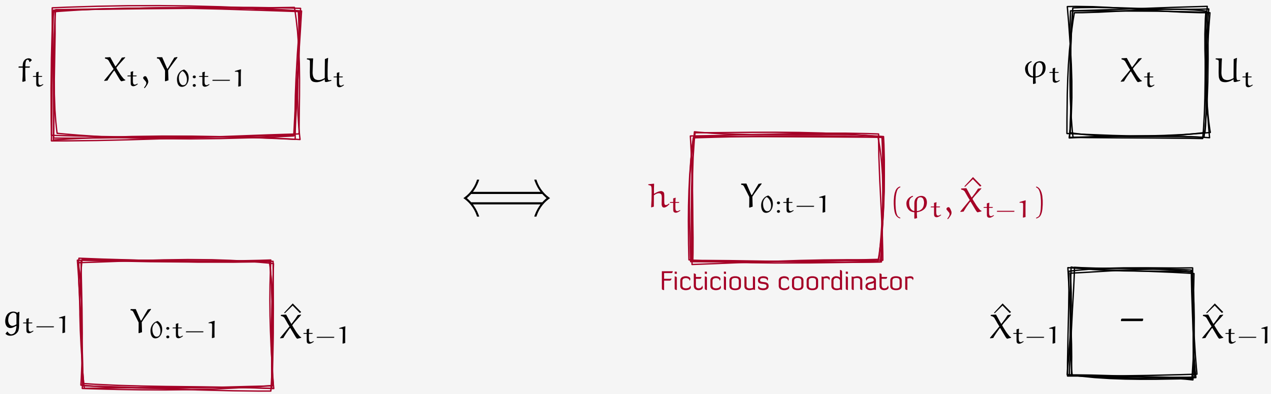


▷ Nayyar, Mahajan and Teneketzis, "Decentralized stochastic control with partial history sharing: A common information approach," IEEE TAC 2013.

The common information approach (Nayyar, Mahajan, Teneketzis 2013)

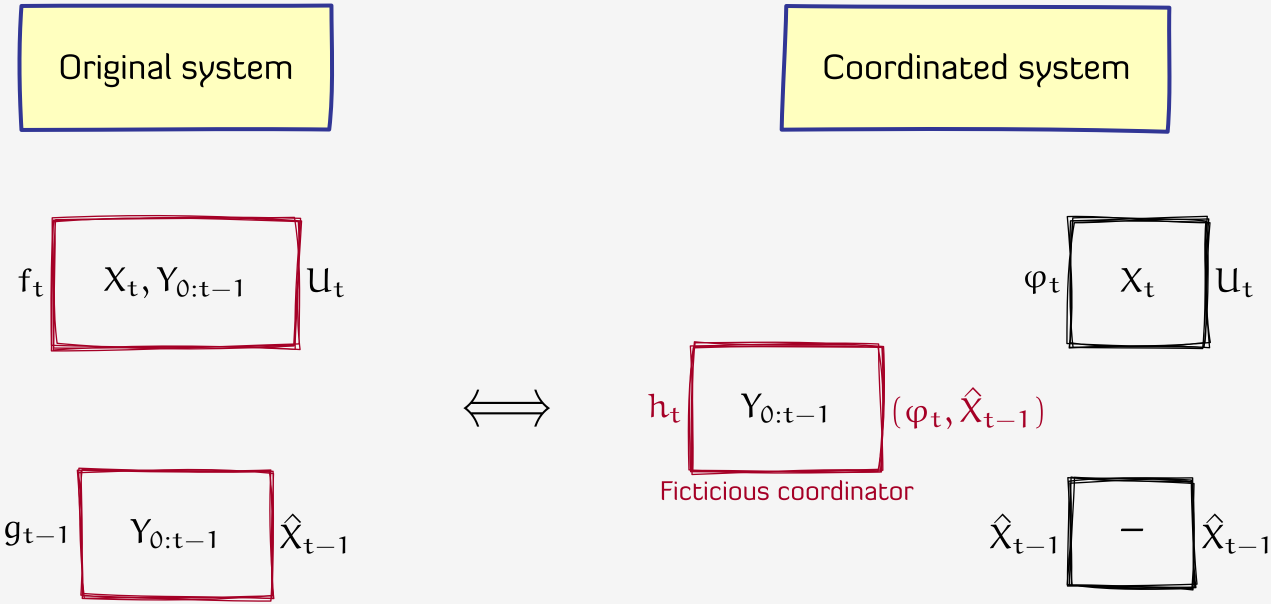
Original system

Coordinated system



▷ Nayyar, Mahajan and Teneketzis, "Decentralized stochastic control with partial history sharing: A common information approach," IEEE TAC 2013.

The common information approach (Nayyar, Mahajan, Teneketzis 2013)



- The coordinated system is equivalent to the original system.

$$f_t(x, y_{0:t-1}) = h_t^1(y_{0:t-1})(x).$$

- The coordinated system is centralized. Belief state $\mathbb{P}(X_t \mid Y_{0:t-1})$.

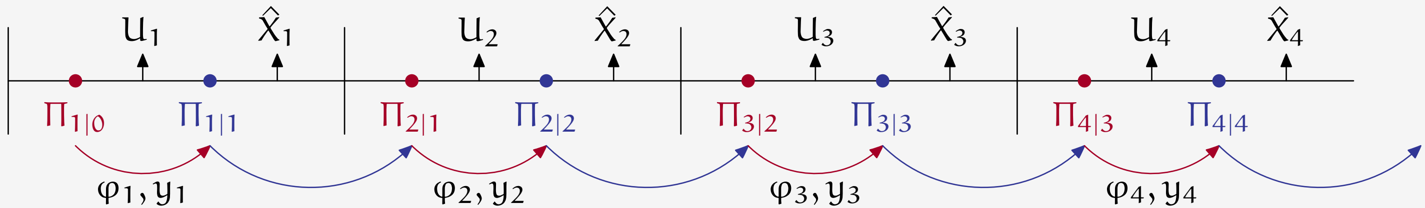
Nayyar, Mahajan and Teneketzis, "Decentralized stochastic control with partial history sharing: A common information approach," IEEE TAC 2013.

Information states and dynamic program

Information states

Pre-transmission belief : $\Pi_{t|t-1}(x) = \mathbb{P}(X_t = x \mid Y_{0:t-1})$.

Post-transmission belief : $\Pi_{t|t}(x) = \mathbb{P}(X_t = x \mid Y_{0:t})$.

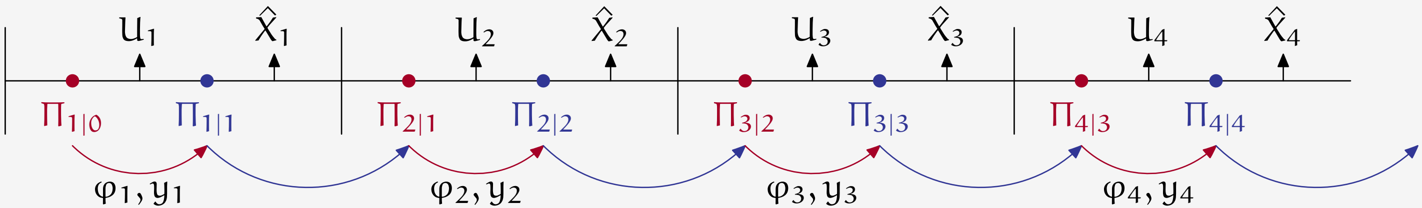


Information states and dynamic program

Information states

Pre-transmission belief : $\Pi_{t|t-1}(x) = \mathbb{P}(X_t = x \mid Y_{0:t-1})$.

Post-transmission belief : $\Pi_{t|t}(x) = \mathbb{P}(X_t = x \mid Y_{0:t})$.



Information state update

- $\pi_{t+1|t} = \tilde{Q}(\pi_{t|t})$.
- $\pi_{t|t} = Q(\pi_{t|t-1}, \varphi_t, y_t)$.

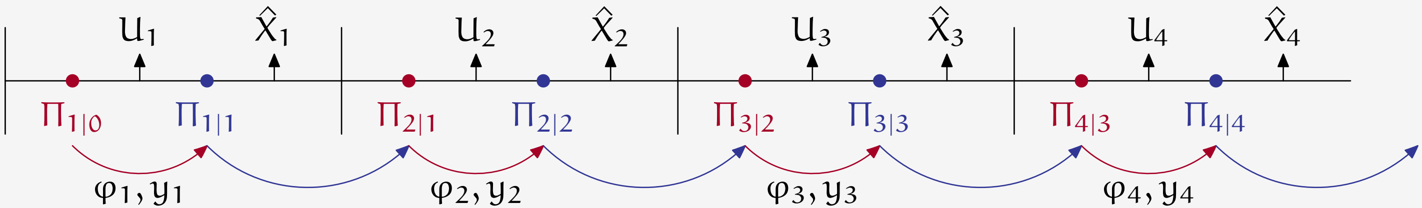
In particular,
$$\pi_{t|t}(x) \coloneqq \begin{cases} \dfrac{\pi_{t|t-1}(x)\big[\varepsilon\varphi_t(x) + (1-\varphi_t(x))\big]}{\sum_{x' \in \mathcal{X}} \pi_{t|t-1}(x')\big[\varepsilon\varphi_t(x') + (1-\varphi_t(x'))\big]}, & \text{if } y_t = \mathfrak{E} \\ \delta_{y_t}, & \text{if } y_t \neq \mathfrak{E} \end{cases}$$

Information states and dynamic program

Information states

Pre-transmission belief : $\Pi_{t|t-1}(x) = \mathbb{P}(X_t = x \mid Y_{0:t-1})$.

Post-transmission belief : $\Pi_{t|t}(x) = \mathbb{P}(X_t = x \mid Y_{0:t})$.



Structural results

There is no loss of optimality in using

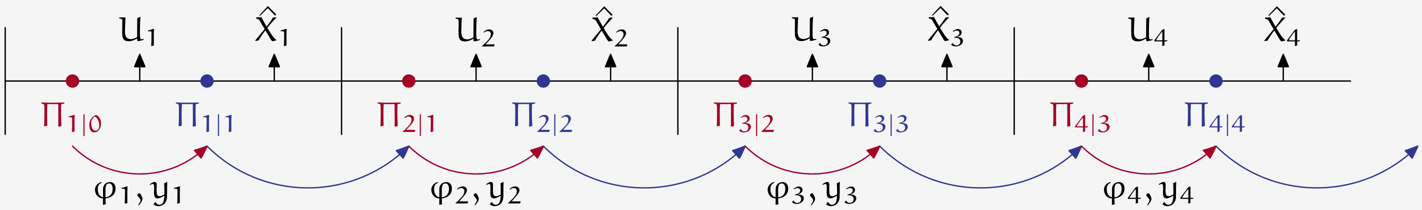
$$U_t = f_t(X_t, \Pi_{t|t-1}) \quad \text{and} \quad \hat{X}_t = g_t(\Pi_{t|t}).$$

Information states and dynamic program

Information states

Pre-transmission belief : $\Pi_{t|t-1}(x) = \mathbb{P}(X_t = x \mid Y_{0:t-1})$.

Post-transmission belief : $\Pi_{t|t}(x) = \mathbb{P}(X_t = x \mid Y_{0:t})$.



Structural results

There is no loss of optimality in using

$$U_t = f_t(X_t, \Pi_{t|t-1}) \quad \text{and} \quad \hat{X}_t = g_t(\Pi_{t|t}).$$

Dynamic Program

$$V_{T+1|T}(\pi) = 0, \quad \text{and for } t = T, \dots, 0$$

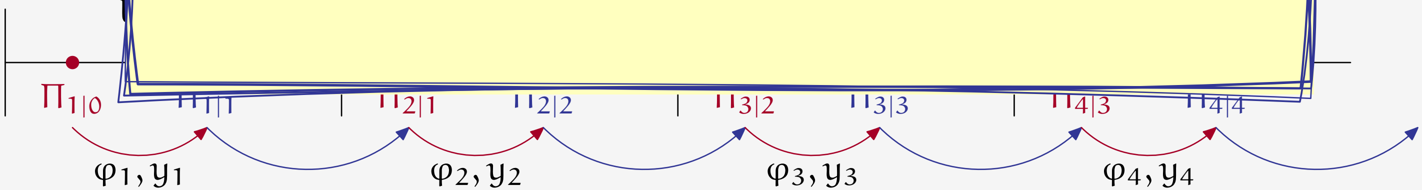
$$V_{t|t}(\pi) = \min_{\hat{x} \in \mathcal{X}} \mathbb{E}[d(X_t - \hat{x}) + V_{t+1|t}(\Pi_{t+1}) \mid \Pi_{t|t} = \pi],$$

$$V_{t|t-1}(\pi) = \min_{\varphi : \mathcal{X} \to \{0,1\}} \mathbb{E}[\lambda \varphi(X_t) + V_{t|t}(\Pi_{t|t}) \mid \Pi_{t|t-1} = \pi, \varphi_t = \varphi].$$

Information states and dynamic program

Information s

"Standard" POMDP. Optimal strategies can be computed numerically (at least, in principle).



Structural results There is no loss of optimality in using

$$U_t = f_t(X_t, \Pi_{t|t-1}) \quad \text{and} \quad \hat{X}_t = g_t(\Pi_{t|t}).$$

Dynamic Program

$$V_{T+1|T}(\pi) = 0, \quad \text{and for } t = T, \dots, 0$$

$$V_{t|t}(\pi) = \min_{\hat{x} \in \mathcal{X}} \mathbb{E}[d(X_t - \hat{x}) + V_{t+1|t}(\Pi_{t+1}) \mid \Pi_{t|t} = \pi],$$

$$V_{t|t-1}(\pi) = \min_{\varphi : \mathcal{X} \to \{0,1\}} \mathbb{E}[\lambda \varphi(X_t) + V_{t|t}(\Pi_{t|t}) \mid \Pi_{t|t-1} = \pi, \varphi_t = \varphi].$$

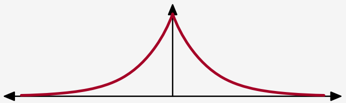
Can we use the DP to say something more about the optimal strategy?

Simplifying modeling assumptions

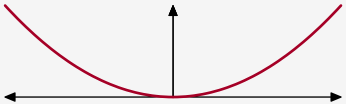
Markov process $X_{t+1} = aX_t + W_t$

- Discrete state process: $X_t, a, W_t \in \mathbb{Z}$
- Continuous state process: $X_t, a, W_t \in \mathbb{R}$

Noise Distribution Unimodal and symmetric



Distortion function Even and increasing

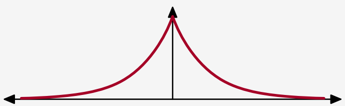


Simplifying modeling assumptions

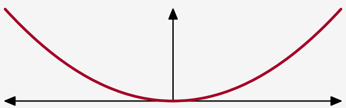
Markov process $X_{t+1} = aX_t + W_t$

- Discrete state process: $X_t, a, W_t \in \mathbb{Z}$
- Continuous state process: $X_t, a, W_t \in \mathbb{R}$

Noise Distribution Unimodal and symmetric



Distortion function Even and increasing



Proof outline

- **Step 1** Show that threshold-based strategies are optimal
- **Step 2** Find performance of arbitrary threshold based strategies
- **Step 3** Solution to the costly communication problem
- **Step 4** Solution to the constrained communication problem

Step 1 Preliminaries: Change of variables

Definition Let σ denote the last time a packet was received successfully. Define

$$E_t = X_t - a^{\sigma-t} X_\sigma$$

$$\hat{E}_t = \hat{X}_t - a^{\sigma-t} X_\sigma$$

Step 1 Preliminaries: Change of variables

Definition Let σ denote the last time a packet was received successfully. Define

$$E_t = X_t - a^{\sigma - t} X_\sigma$$

$$\hat{E}_t = \hat{X}_t - a^{\sigma - t} X_\sigma$$

Note that E_t is a regenerative process:

$$E_{t+1} = \begin{cases} aE_t + W_t, & \text{if } Y_t = \mathfrak{E} \\ W_t, & \text{if } Y_t \neq \mathfrak{E} \end{cases} \qquad \text{and} \qquad d(E_t - \hat{E}_t) = d(X_t - \hat{X}_t)$$

Step 1 Preliminaries: Change of variables

Definition Let σ denote the last time a packet was received successfully. Define

$$E_t = X_t - a^{\sigma - t} X_\sigma$$

$$\hat{E}_t = \hat{X}_t - a^{\sigma - t} X_\sigma$$

Note that E_t is a regenerative process:

$$E_{t+1} = \begin{cases} aE_t + W_t, & \text{if } Y_t = \mathfrak{E} \\ W_t, & \text{if } Y_t \neq \mathfrak{E} \end{cases} \qquad \text{and} \qquad d(E_t - \hat{E}_t) = d(X_t - \hat{X}_t)$$

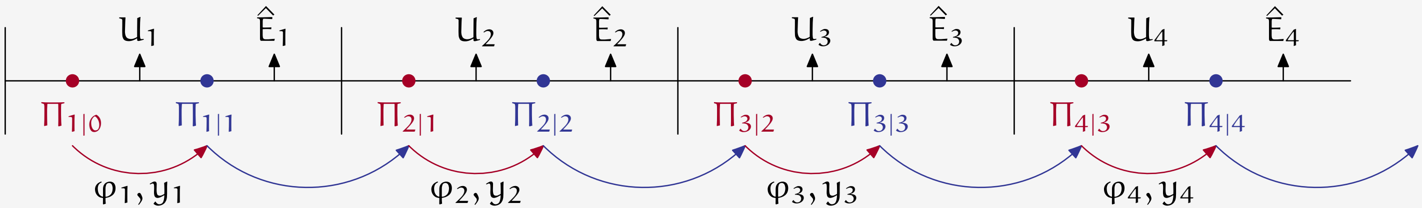
We work with $\{E_t\}_{t \geqslant 0}$ rather than $\{X_t\}_{t \geqslant 0}$

Step 1 Preliminaries: Change of variables

Information states

Pre-transmission belief : $\Pi_{t|t-1}(e) = \mathbb{P}(E_t = x \mid Y_{0:t-1})$.

Post-transmission belief : $\Pi_{t|t}(e) = \mathbb{P}(E_t = e \mid Y_{0:t})$.



Information state update

- $\pi_{t+1|t} = \tilde{Q}(\pi_{t|t})$.
- $\pi_{t|t} = Q(\pi_{t|t-1}, \varphi_t, y_t)$.

In particular, $$\pi_{t|t}(e) \coloneqq \begin{cases} \dfrac{\pi_{t|t-1}(e)\left[\varepsilon\varphi_t(e) + (1-\varphi_t(e))\right]}{\sum_{e' \in \mathcal{X}} \pi_{t|t-1}(e')\left[\varepsilon\varphi_t(e') + (1-\varphi_t(e'))\right]}, & \text{if } y_t = \mathfrak{E} \\ \delta_0, & \text{if } y_t \neq \mathfrak{E} \end{cases}$$

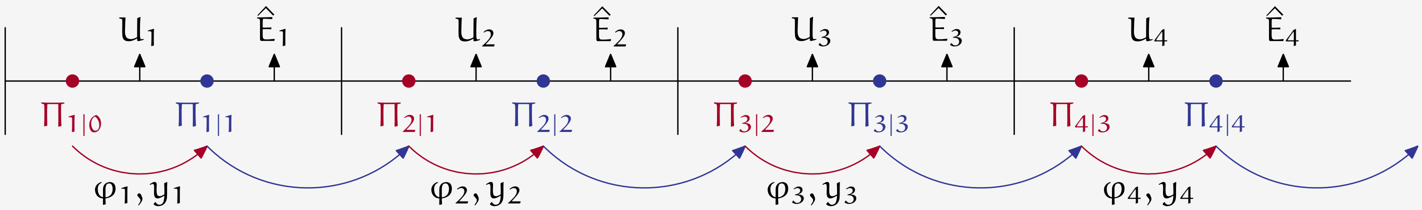
Note that we can write $\pi_{t|t} = Q(\pi_{t|t-1}, \varphi_t, h_t)$, where $h_t = u_t s_t$

Step 1 Preliminaries: Change of variables

Information states

Pre-transmission belief : $\Pi_{t|t-1}(e) = \mathbb{P}(E_t = x \mid Y_{0:t-1})$.

Post-transmission belief : $\Pi_{t|t}(e) = \mathbb{P}(E_t = e \mid Y_{0:t})$.

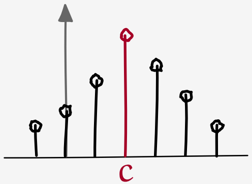


Dynamic Program Remains same as before

Step 1 Preliminaries: Majorization

[Hajek Mitzel Yang 2008, Lipsa Martins 2011, Nayyar et. al. 2013]

Almost uniform and unimodal (ASU) distribution about c

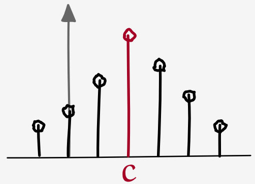


$$\pi_c \geqslant \pi_{c+1} \geqslant \pi_{c-1} \geqslant \pi_{c+2} \geqslant \cdots$$

Step 1 Preliminaries: Majorization

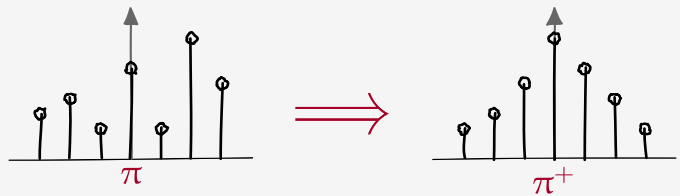
[Hajek Mitzel Yang 2008, Lipsa Martins 2011, Nayyar et. al. 2013]

Almost uniform and unimodal (ASU) distribution about c



$$\pi_c \geqslant \pi_{c+1} \geqslant \pi_{c-1} \geqslant \pi_{c+2} \geqslant \cdots$$

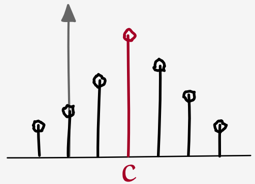
ASU Rearrangement



Step 1 Preliminaries: Majorization

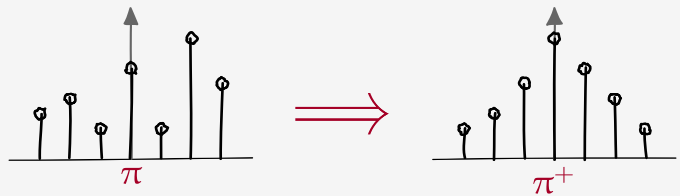
[Hajek Mitzel Yang 2008, Lipsa Martins 2011, Nayyar et. al. 2013]

Almost uniform and unimodal (ASU) distribution about c



$\pi_c \geqslant \pi_{c+1} \geqslant \pi_{c-1} \geqslant \pi_{c+2} \geqslant \cdots$

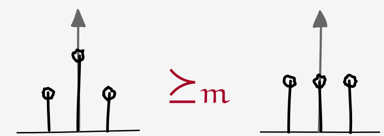
ASU Rearrangement



Majorization

$\xi \succeq_m \pi$ iff

$$\sum_{i=-n}^{n} \xi_i^+ \geqslant \sum_{i=-n}^{n} \pi_i^+ \quad \text{and} \quad \sum_{i=-n}^{n+1} \xi_i^+ \geqslant \sum_{i=-n}^{n+1} \pi_i^+$$

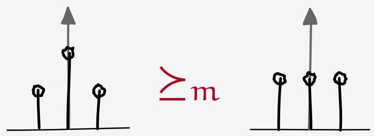


Invariant to permutations.

Step 1 Preliminaries: Majorization

Majorization $\xi \succeq_m \pi$ iff

$$\sum_{i=-n}^{n} \xi_i^+ \geqslant \sum_{i=-n}^{n} \pi_i^+ \quad \text{and} \quad \sum_{i=-n}^{n+1} \xi_i^+ \geqslant \sum_{i=-n}^{n+1} \pi_i^+$$

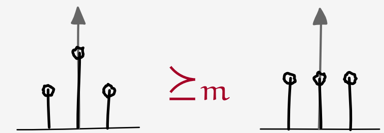


Invariant to permutations.

Step 1 Preliminaries: Majorization

Majorization $\xi \succeq_m \pi$ iff

$$\sum_{i=-n}^{n} \xi_i^+ \geqslant \sum_{i=-n}^{n} \pi_i^+ \quad \text{and} \quad \sum_{i=-n}^{n+1} \xi_i^+ \geqslant \sum_{i=-n}^{n+1} \pi_i^+$$



Invariant to permutations.

ASU Majorization $\xi \succeq_a \pi$ iff ξ is ASU and $\xi \succeq_m \pi$

Step 1 Properties of majorization

Threshold based strategies

Let $\mathcal{F}(c)$ denote the class of all threshold based strategies around c, i.e.,

$$\varphi \in \mathcal{F}(c) \quad \text{if } \exists k \text{ s.t.} \quad \varphi(e) = \begin{cases} 1 & \text{if } |e - ac| \geqslant k \\ 0 & \text{otherwise} \end{cases}$$

Step 1 Properties of majorization

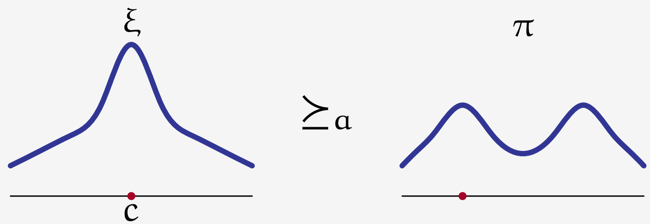
Threshold based strategies

Let $\mathcal{F}(c)$ denote the class of all threshold based strategies around c, i.e.,

$$\varphi \in \mathcal{F}(c) \quad \text{if } \exists k \text{ s.t.} \quad \varphi(e) = \begin{cases} 1 & \text{if } |e - ac| \geqslant k \\ 0 & \text{otherwise} \end{cases}$$

Property 1

For any $\xi \succeq_a \pi$ where ξ is ASU(c),



Step 1 Properties of majorization

Threshold based strategies

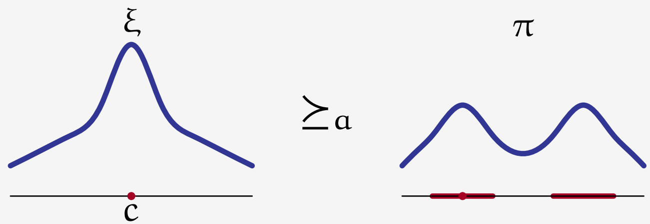
Let $\mathcal{F}(c)$ denote the class of all threshold based strategies around c, i.e.,

$$\varphi \in \mathcal{F}(c) \quad \text{if } \exists k \text{ s.t.} \quad \varphi(e) = \begin{cases} 1 & \text{if } |e - ac| \geqslant k \\ 0 & \text{otherwise} \end{cases}$$

Property 1

For any $\xi \succeq_a \pi$ where ξ is ASU(c),

and any φ,



Step 1 Properties of majorization

Threshold based strategies

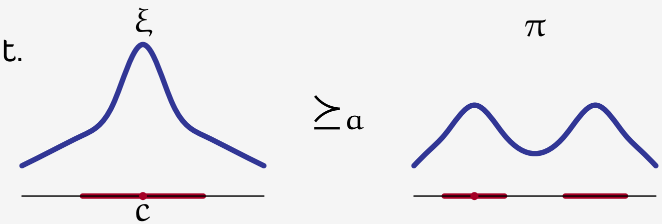
Let $\mathcal{F}(c)$ denote the class of all threshold based strategies around c, i.e.,

$$\varphi \in \mathcal{F}(c) \quad \text{if } \exists k \text{ s.t.} \quad \varphi(e) = \begin{cases} 1 & \text{if } |e - ac| \geq k \\ 0 & \text{otherwise} \end{cases}$$

Property 1

For any $\xi \succeq_a \pi$ where ξ is ASU(c),

and any φ, there exists a $\theta \in \mathcal{F}(c)$ s.t.



Step 1 Properties of majorization

Threshold based strategies

Let $\mathcal{F}(c)$ denote the class of all threshold based strategies around c, i.e.,

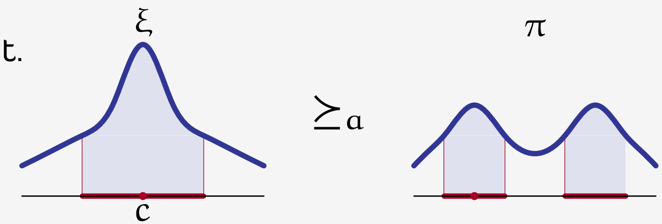
$$\varphi \in \mathcal{F}(c) \quad \text{if } \exists k \text{ s.t.} \quad \varphi(e) = \begin{cases} 1 & \text{if } |e - ac| \geqslant k \\ 0 & \text{otherwise} \end{cases}$$

Property 1

For any $\xi \succeq_a \pi$ where ξ is ASU(c),

and any φ, there exists a $\theta \in \mathcal{F}(c)$ s.t.

$$\sum_{e \in \mathcal{X}} \theta(e)\xi(e) = \sum_{e \in \mathcal{X}} \varphi(e)\pi(e).$$



Properties of majorization

Threshold based strategies

Let $\mathcal{F}(c)$ denote the class of all threshold based strategies around c, i.e.,

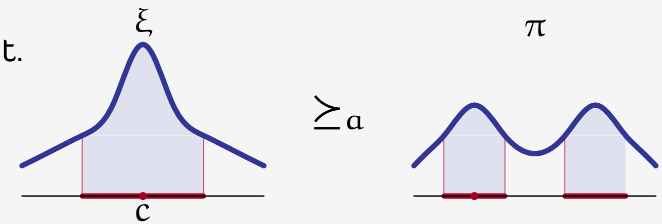
$$\varphi \in \mathcal{F}(c) \quad \text{if } \exists k \text{ s.t.} \quad \varphi(e) = \begin{cases} 1 & \text{if } |e - ac| \geq k \\ 0 & \text{otherwise} \end{cases}$$

Property 1

For any $\xi \succeq_a \pi$ where ξ is ASU(c),

and any φ, there exists a $\theta \in \mathcal{F}(c)$ s.t.

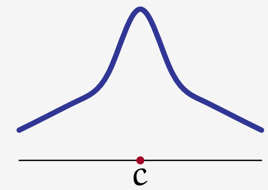
$$\sum_{e \in \mathcal{X}} \theta(e)\xi(e) = \sum_{e \in \mathcal{X}} \varphi(e)\pi(e).$$



Moreover, for $h \in \{0, 1\}$ (recall $h = u \cdot s$), $Q(\xi, \theta, h) \succeq_a Q(\pi, \varphi, h)$.

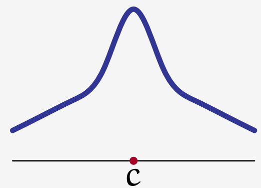
Step 1 Properties of majorization

Property 2 If π is ASU(c), then $c \in \arg\min_{\hat{e} \in \mathcal{X}} \sum_{e \in \mathcal{X}} d(e - \hat{e})\pi(e)$



Step 1 Properties of majorization

Property 2 If π is ASU(c), then $c \in \arg\min_{\hat{e} \in \mathcal{X}} \sum_{e \in \mathcal{X}} d(e - \hat{e}) \pi(e)$

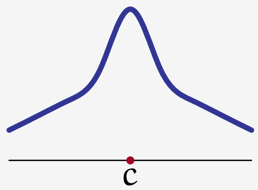


Property 3 if $\xi \succeq_a \pi$, then

$$\min_{\hat{e} \in \mathcal{X}} \sum_{e \in \mathcal{X}} d(e - \hat{e}) \pi(e) \geqslant \min_{\hat{e} \in \mathcal{X}} \sum_{e \in \mathcal{X}} d(e - \hat{e}) \pi^{+}(e) \geqslant \min_{\hat{e} \in \mathcal{X}} \sum_{e \in \mathcal{X}} d(e - \hat{e}) \xi(e)$$

Step 1 Properties of majorization

Property 2 If π is ASU(c), then $c \in \arg\min_{\hat{e} \in \mathcal{X}} \sum_{e \in \mathcal{X}} d(e - \hat{e})\pi(e)$



Property 3 if $\xi \succeq_a \pi$, then

$$\min_{\hat{e} \in \mathcal{X}} \sum_{e \in \mathcal{X}} d(e - \hat{e})\pi(e) \geqslant \min_{\hat{e} \in \mathcal{X}} \sum_{e \in \mathcal{X}} d(e - \hat{e})\pi^{+}(e) \geqslant \min_{\hat{e} \in \mathcal{X}} \sum_{e \in \mathcal{X}} d(e - \hat{e})\xi(e)$$

Property 4 if $\xi \succeq_a \pi$, then $\tilde{Q}(\xi) \succeq_a \tilde{Q}(\pi)$

Step 1 An interchange argument to identify optimal strategies

Main theorem

The optimal estimation strategy is given as follows: $\hat{E}_0 = 0$ and for $t \geqslant 1$

$$\hat{E}_t = \begin{cases} 0, & \text{if } Y_t = \mathfrak{E} \\ E_t, & \text{if } Y_t \neq \mathfrak{E} \end{cases}$$

In addition, there exist thresholds $\{k_t\}_{t \geqslant 0}$ such that the following transmission strategy is optimal

$$U_t = \begin{cases} 1, & \text{if } |E_t| \geqslant k_t \\ 0, & \text{otherwise} \end{cases}$$

Step 1 An interchange argument to identify optimal strategies

Main theorem The optimal estimation strategy is given as follows: $\hat{E}_0 = 0$ and for $t \geqslant 1$

$$\hat{E}_t = \begin{cases} 0, & \text{if } Y_t = \mathfrak{E} \\ E_t, & \text{if } Y_t \neq \mathfrak{E} \end{cases}$$

In addition, there exist thresholds $\{k_t\}_{t \geqslant 0}$ such that the following transmission strategy is optimal

$$U_t = \begin{cases} 1, & \text{if } |E_t| \geqslant k_t \\ 0, & \text{otherwise} \end{cases}$$

Proof by contradiction Suppose t is the first time when $\varphi_t \notin \mathcal{F}(0)$. Note that $\pi_{t|t-1}$ is ASU(0).

Step 1 An interchange argument to identify optimal strategies

Main theorem

The optimal estimation strategy is given as follows: $\hat{E}_0 = 0$ and for $t \geqslant 1$

$$\hat{E}_t = \begin{cases} 0, & \text{if } Y_t = \mathfrak{E} \\ E_t, & \text{if } Y_t \neq \mathfrak{E} \end{cases}$$

In addition, there exist thresholds $\{k_t\}_{t \geqslant 0}$ such that the following transmission strategy is optimal

$$U_t = \begin{cases} 1, & \text{if } |E_t| \geqslant k_t \\ 0, & \text{otherwise} \end{cases}$$

Proof by contradiction

Suppose t is the first time when $\varphi_t \notin \mathcal{F}(0)$. Note that $\pi_{t|t-1}$ is ASU(0).

Since $\pi_{t|t-1} \succeq_a \pi_{t|t-1}$, by Property 1, there is a $\theta_t \in \mathcal{F}(0)$ s.t. that doesn't affect the communication cost at t and . . .

Step 1 An interchange argument to identify optimal strategies

Main theorem

The optimal estimation strategy is given as follows: $\hat{E}_0 = 0$ and for $t \geqslant 1$

$$\hat{E}_t = \begin{cases} 0, & \text{if } Y_t = \mathfrak{E} \\ E_t, & \text{if } Y_t \neq \mathfrak{E} \end{cases}$$

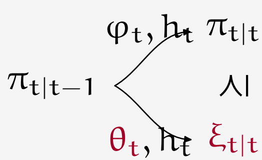
In addition, there exist thresholds $\{k_t\}_{t \geqslant 0}$ such that the following transmission strategy is optimal

$$U_t = \begin{cases} 1, & \text{if } |E_t| \geqslant k_t \\ 0, & \text{otherwise} \end{cases}$$

Proof by contradiction

Suppose t is the first time when $\varphi_t \notin \mathcal{F}(0)$. Note that $\pi_{t|t-1}$ is ASU(0).

Since $\pi_{t|t-1} \succeq_a \pi_{t|t-1}$, by Property 1, there is a $\theta_t \in \mathcal{F}(0)$ s.t. that doesn't affect the communication cost at t and . . .



Step 1 An interchange argument to identify optimal strategies

Main theorem

The optimal estimation strategy is given as follows: $\hat{E}_0 = 0$ and for $t \geqslant 1$

$$\hat{E}_t = \begin{cases} 0, & \text{if } Y_t = \mathfrak{E} \\ E_t, & \text{if } Y_t \neq \mathfrak{E} \end{cases}$$

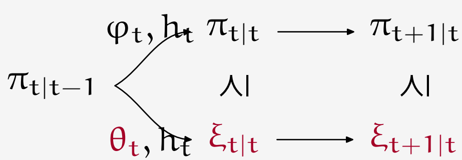
In addition, there exist thresholds $\{k_t\}_{t \geqslant 0}$ such that the following transmission strategy is optimal

$$U_t = \begin{cases} 1, & \text{if } |E_t| \geqslant k_t \\ 0, & \text{otherwise} \end{cases}$$

Proof by contradiction

Suppose t is the first time when $\varphi_t \notin \mathcal{F}(0)$. Note that $\pi_{t|t-1}$ is ASU(0).

Since $\pi_{t|t-1} \succeq_a \pi_{t|t-1}$, by Property 1, there is a $\theta_t \in \mathcal{F}(0)$ s.t. that doesn't affect the communication cost at t and . . .



Step 1 An interchange argument to identify optimal strategies

Main theorem

The optimal estimation strategy is given as follows: $\hat{E}_0 = 0$ and for $t \geqslant 1$

$$\hat{E}_t = \begin{cases} 0, & \text{if } Y_t = \mathfrak{E} \\ E_t, & \text{if } Y_t \neq \mathfrak{E} \end{cases}$$

In addition, there exist thresholds $\{k_t\}_{t \geqslant 0}$ such that the following transmission strategy is optimal

$$U_t = \begin{cases} 1, & \text{if } |E_t| \geqslant k_t \\ 0, & \text{otherwise} \end{cases}$$

Proof by contradiction

Suppose t is the first time when $\varphi_t \notin \mathcal{F}(0)$. Note that $\pi_{t|t-1}$ is ASU(0).

Since $\pi_{t|t-1} \succeq_a \pi_{t|t-1}$, by Property 1, there is a $\theta_t \in \mathcal{F}(0)$ s.t. that doesn't affect the communication cost at t and ...

$$\begin{array}{ccccc} & \varphi_t, h_t \nearrow & \pi_{t|t} & \longrightarrow & \pi_{t+1|t} \longrightarrow \\ \pi_{t|t-1} & & \succeq & & \succeq \\ & \theta_t, h_t \searrow & \xi_{t|t} & \longrightarrow & \xi_{t+1|t} \longrightarrow \end{array}$$

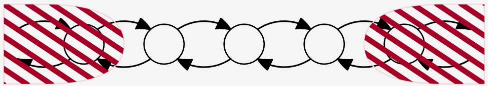
For infinite-horizon setup time-homogeneous **threshold-based strategies are optimal.**

How do we find the optimal threshold-based strategy?

Step 2 Performance of threshold-based strategies

Consider a threshold-based strategy

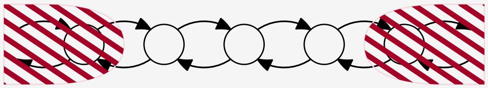
$$f^{(k)}(e) = \begin{cases} 1 & \text{if } |e| \geq k \\ 0 & \text{otherwise} \end{cases}$$



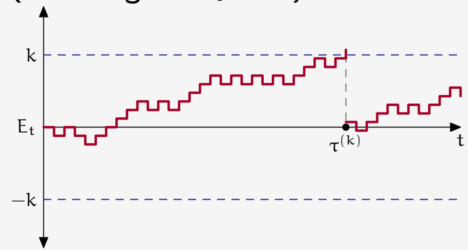
Step 2 Performance of threshold-based strategies

Consider a threshold-based strategy

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



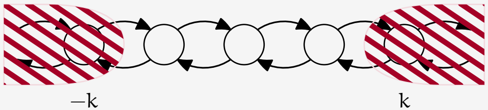
Let $\tau^{(k)}$ denote the stopping time of first reception (starting at $E_0 = 0$).



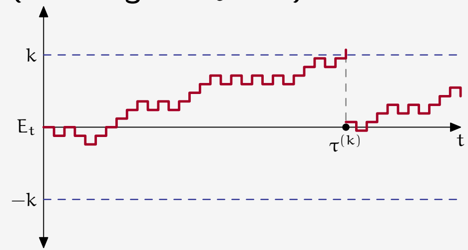
Step 2 Performance of threshold-based strategies

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 reception (starting at $E_0 = 0$).



Define

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

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

$$K_\beta^{(k)}(e) = \mathbb{E}\Big[\sum_{t=0}^{\tau^{(k)}-1} \beta^t U_t \,\Big|\, E_0 = e \Big].$$

Step 2 Performance of threshold-based strategies

Proposition $\{E_t\}_{t=0}^{\infty}$ is a regenerative process. By renewal relationships, we have:

$$D_\beta^{(k)} \coloneqq D_\beta(f^{(k)}, g^*) = \frac{L_\beta^{(k)}(0)}{M_\beta^{(k)}(0)}$$

$$N_\beta^{(k)} \coloneqq N_\beta(f^{(k)}, g^*) = \frac{K_\beta^{(k)}(0)}{M_\beta^{(k)}(0)}$$

Define

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

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

$$K_\beta^{(k)}(e) = \mathbb{E}\Big[\sum_{t=0}^{\tau^{(k)}-1} \beta^t U_t \Big| E_0 = e\Big].$$

Step 2 Performance of threshold-based strategies

Proposition $\{E_t\}_{t=0}^{\infty}$ is a regenerative process. By renewal relationships, we have:

$$D_\beta^{(k)} \coloneqq D_\beta(f^{(k)}, g^*) = \frac{L_\beta^{(k)}(0)}{M_\beta^{(k)}(0)}$$

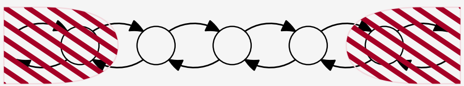
$$N_\beta^{(k)} \coloneqq N_\beta(f^{(k)}, g^*) = \frac{K_\beta^{(k)}(0)}{M_\beta^{(k)}(0)}$$

Define $L_\beta^{(k)}(e) = \mathbb{E}\Big[\sum^{\tau^{(k)}-1} \beta^t d(E_t) \Big| E_0 = e\Big].$

Computing $L_\beta^{(k)}$, $M_\beta^{(k)}$, $K_\beta^{(k)}$ is sufficient to compute the performance of $f^{(k)}$ (i.e., to compute $D_\beta^{(k)}$ and $N_\beta^{(k)}$).

$$K_\beta^{(k)}(e) = \mathbb{E}\Big[\sum_{t=0} \beta^t U_t \Big| E_0 = e\Big].$$

Step 2 Computing $L_\beta^{(k)}$, $M_\beta^{(k)}$, and $K_\beta^{(k)}$: The discrete case

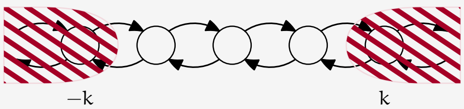


$$L_\beta^{(k)}(e) = \begin{cases} d(e) + \beta \sum_{n \in \mathbb{Z}} p_{n-ae} L_\beta^{(k)}(n), & \text{if } |e| < k \\ \varepsilon\Big[d(e) + \beta \sum_{n \in \mathbb{Z}} p_{n-ae} L_\beta^{(k)}(n)\Big], & \text{if } |e| \geqslant k \end{cases}$$

$$M_\beta^{(k)}(e) = \begin{cases} 1 + \beta \sum_{n \in \mathbb{Z}} p_{n-ae} M_\beta^{(k)}(n), & \text{if } |e| < k \\ \varepsilon\Big[1 + \beta \sum_{n \in \mathbb{Z}} p_{n-ae} M_\beta^{(k)}(n)\Big], & \text{if } |e| \geqslant k \end{cases}$$

$$K_\beta^{(k)}(e) = \begin{cases} \beta \sum_{n \in \mathbb{Z}} p_{n-ae} K_\beta^{(k)}(n), & \text{if } |e| < k \\ 1 + \varepsilon\beta \sum_{n \in \mathbb{Z}} p_{n-ae} M_\beta^{(k)}(n), & \text{if } |e| \geqslant k \end{cases}$$

Step 2 Computing $L_\beta^{(k)}$, $M_\beta^{(k)}$, and $K_\beta^{(k)}$: The discrete case



$$L_\beta^{(k)}(e) = \begin{cases} d(e) + \beta \sum_{n \in \mathbb{Z}} p_{n-ae} L_\beta^{(k)}(n), & \text{if } |e| < k \\ \varepsilon\left[d(e) + \beta \sum_{n \in \mathbb{Z}} p_{n-ae} L_\beta^{(k)}(n)\right], & \text{if } |e| \geqslant k \end{cases}$$

$$M_\beta^{(k)}(e) = \begin{cases} 1 + \beta \sum_{n \in \mathbb{Z}} p_{n-ae} M_\beta^{(k)}(n), & \text{if } |e| < k \\ \varepsilon\left[1 + \beta \sum p_{n-ae} M_\beta^{(k)}(n)\right] & \text{if } |e| \geqslant k \end{cases}$$

$$\begin{bmatrix} \vdots \\ L_\beta^{(k)}(-2) \\ L_\beta^{(k)}(-1) \\ L_\beta^{(k)}(0) \\ L_\beta^{(k)}(1) \\ L_\beta^{(k)}(2) \\ \vdots \end{bmatrix} = \begin{bmatrix} \vdots \\ \varepsilon d(-2) \\ d(-1) \\ d(0) \\ d(1) \\ \varepsilon d(2) \\ \vdots \end{bmatrix} + \beta \begin{bmatrix} \vdots & \cdots & \cdots & \cdots & \cdots & \cdots & \vdots \\ \cdots & \varepsilon p_1 & \varepsilon p_2 & \varepsilon p_3 & \varepsilon p_4 & \varepsilon p_5 & \cdots \\ \cdots & p_0 & p_1 & p_2 & p_3 & p_4 & \cdots \\ \cdots & p_{-1} & p_0 & p_1 & p_2 & p_3 & \cdots \\ \cdots & p_{-2} & p_{-1} & p_0 & p_1 & p_2 & \cdots \\ \cdots & \varepsilon p_{-3} & \varepsilon p_{-2} & \varepsilon p_{-1} & \varepsilon p_0 & \varepsilon p_1 & \cdots \\ \vdots & \cdots & \cdots & \cdots & \cdots & \cdots & \vdots \end{bmatrix} \begin{bmatrix} \vdots \\ L_\beta^{(k)}(-2) \\ L_\beta^{(k)}(-1) \\ L_\beta^{(k)}(0) \\ L_\beta^{(k)}(1) \\ L_\beta^{(k)}(2) \\ \vdots \end{bmatrix}$$

Step 2 Computing $L_\beta^{(k)}$, $M_\beta^{(k)}$, and $K_\beta^{(k)}$: The discrete case

$L_\beta^{(k)} = [I - \beta h^{(k)} \odot P]^{-1} h^{(k)} \odot d$
where $h^{(k)} \odot P$ is substochastic.

$$L_\beta^{(k)}(e) = \begin{cases} d(e) + \beta \sum_{n \in \mathbb{Z}} p_{n-ae} L_\beta^{(k)}(n), & \text{if } |e| < k \\ \varepsilon \Big[d(e) + \beta \sum_{n \in \mathbb{Z}} p_{n-ae} L_\beta^{(k)}(n) \Big], & \text{if } |e| \geqslant k \end{cases}$$

$$M_\beta^{(k)}(e) = \begin{cases} 1 + \beta \sum_{n \in \mathbb{Z}} p_{n-ae} M_\beta^{(k)}(n), & \text{if } |e| < k \\ \varepsilon \Big[1 + \beta \sum_{n \in \mathbb{Z}} p_{n-ae} M_\beta^{(k)}(n) \Big], & \text{if } |e| \geqslant k \end{cases}$$

$$K_\beta^{(k)}(e) = \begin{cases} \beta \sum_{n \in \mathbb{Z}} p_{n-ae} K_\beta^{(k)}(n), & \text{if } |e| < k \\ 1 + \varepsilon\beta \sum_{n \in \mathbb{Z}} p_{n-ae} M_\beta^{(k)}(n), & \text{if } |e| \geqslant k \end{cases}$$

Step 2 Computing $L_\beta^{(k)}$, $M_\beta^{(k)}$, and $K_\beta^{(k)}$: The discrete case

$L_\beta^{(k)} = [I - \beta h^{(k)} \odot P]^{-1} h^{(k)} \odot d$
where $h^{(k)} \odot P$ is substochastic.

$$L_\beta^{(k)}(e) = \begin{cases} d(e) + \beta \sum_{n \in \mathbb{Z}} p_{n-ae} L_\beta^{(k)}(n), & \text{if } |e| < k \\ \varepsilon\left[d(e) + \beta \sum_{n \in \mathbb{Z}} p_{n-ae} L_\beta^{(k)}(n)\right], & \text{if } |e| \geqslant k \end{cases}$$

$M_\beta^{(k)} = [I - \beta h^{(k)} \odot P]^{-1} h^{(k)}$

$$M_\beta^{(k)}(e) = \begin{cases} 1 + \beta \sum_{n \in \mathbb{Z}} p_{n-ae} M_\beta^{(k)}(n), & \text{if } |e| < k \\ \varepsilon\left[1 + \beta \sum_{n \in \mathbb{Z}} p_{n-ae} M_\beta^{(k)}(n)\right], & \text{if } |e| \geqslant k \end{cases}$$

$$K_\beta^{(k)}(e) = \begin{cases} \beta \sum_{n \in \mathbb{Z}} p_{n-ae} K_\beta^{(k)}(n), & \text{if } |e| < k \\ 1 + \varepsilon\beta \sum_{n \in \mathbb{Z}} p_{n-ae} M_\beta^{(k)}(n), & \text{if } |e| \geqslant k \end{cases}$$

Step 2 Computing $L_\beta^{(k)}$, $M_\beta^{(k)}$, and $K_\beta^{(k)}$: The discrete case

$$L_\beta^{(k)} = [I - \beta h^{(k)} \odot P]^{-1} h^{(k)} \odot d$$

where $h^{(k)} \odot P$ is substochastic.

$$L_\beta^{(k)}(e) = \begin{cases} d(e) + \beta \sum_{n \in \mathbb{Z}} p_{n-ae} L_\beta^{(k)}(n), & \text{if } |e| < k \\ \varepsilon\Big[d(e) + \beta \sum_{n \in \mathbb{Z}} p_{n-ae} L_\beta^{(k)}(n)\Big], & \text{if } |e| \geq k \end{cases}$$

$$M_\beta^{(k)} = [I - \beta h^{(k)} \odot P]^{-1} h^{(k)}$$

$$M_\beta^{(k)}(e) = \begin{cases} 1 + \beta \sum_{n \in \mathbb{Z}} p_{n-ae} M_\beta^{(k)}(n), & \text{if } |e| < k \\ \varepsilon\Big[1 + \beta \sum_{n \in \mathbb{Z}} p_{n-ae} M_\beta^{(k)}(n)\Big], & \text{if } |e| \geq k \end{cases}$$

$$K_\beta^{(k)} = [I - \beta h^{(k)} \odot P]^{-1} \bar{h}^{(k)}$$

$$K_\beta^{(k)}(e) = \begin{cases} \beta \sum_{n \in \mathbb{Z}} p_{n-ae} K_\beta^{(k)}(n), & \text{if } |e| < k \\ 1 + \varepsilon\beta \sum_{n \in \mathbb{Z}} p_{n-ae} M_\beta^{(k)}(n), & \text{if } |e| \geq k \end{cases}$$

Step 2 Computing $L_\beta^{(k)}$, $M_\beta^{(k)}$, and $K_\beta^{(k)}$: The discrete case

$L_\beta^{(k)} = [I - \beta h^{(k)} \odot P]^{-1} h^{(k)} \odot d$
where $h^{(k)} \odot P$ is substochastic.

$$L_\beta^{(k)}(e) = \begin{cases} d(e) + \beta \sum_{n \in \mathbb{Z}} p_{n-ae} L_\beta^{(k)}(n), & \text{if } |e| < k \\ \ldots \end{cases}$$

$M_\beta^{(k)} = [I - \beta h^{(k)} \odot P]^{-1} h^{(k)}$

$D_\beta^{(k)}$ and $N_\beta^{(k)}$ can be computed using these expressions.

$K_\beta^{(k)} = [I - \beta h^{(k)} \odot P]^{-1} \bar{h}^{(k)}$

$$K_\beta^{(k)}(e) = \begin{cases} \beta \sum_{n \in \mathbb{Z}} p_{n-ae} K_\beta^{(k)}(n), & \text{if } |e| < k \\ 1 + \varepsilon\beta \sum_{n \in \mathbb{Z}} p_{n-ae} M_\beta^{(k)}(n), & \text{if } |e| \geq k \end{cases}$$

Step 2 Computing $L_\beta^{(k)}$, $M_\beta^{(k)}$, and $K_\beta^{(k)}$: The continuous case

- The expressions are similar to the discrete case.
- $h^{(k)} \odot P$ is a contraction operator
- The equations for $L_\beta^{(k)}$, etc. are Fredholm integral equations of the second kind. Numerical solution can be obtained by using Picard's iteration and Nystrom interpolation.

We will later provide a simulation based approach to compute $C_\beta^*(\lambda)$ and $D_\beta^*(\alpha)$ that does not need an exact computation of $L_\beta^{(k)}$, etc.

Optimal trade-offs for costly and constrained communication for discrete sources

Step 3 Solution to costly optimization problem

Proposition

- $C_\beta^{(k)}(\lambda) \coloneqq D_\beta^{(k)} + \lambda N_\beta^{(k)}$ is submodular in (k, λ).
- Hence, $k_\beta^*(\lambda) \coloneqq \arg\min_{k \geq 0} C_\beta^{(k)}(\lambda)$ is increasing in λ

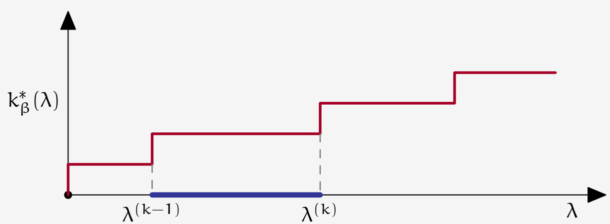
Step 3 Solution to costly optimization problem

Proposition

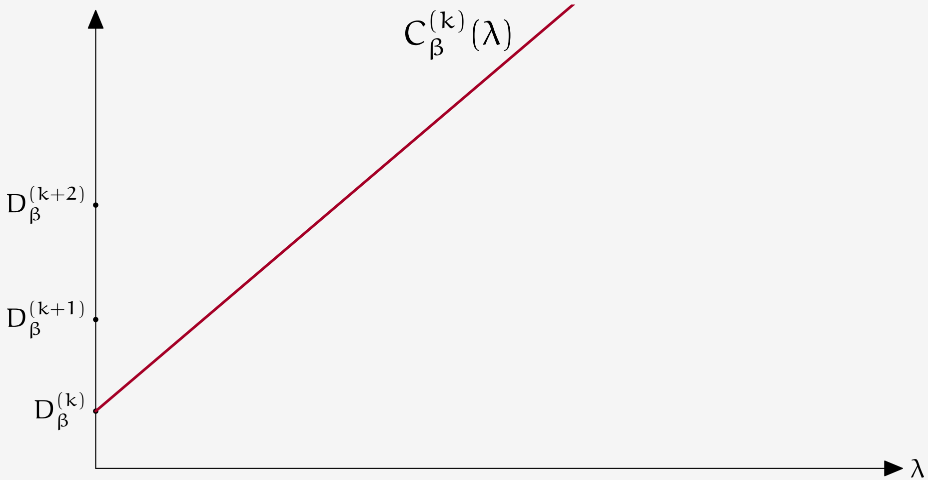
- $C^{(k)}_\beta(\lambda) \coloneqq D^{(k)}_\beta + \lambda N^{(k)}_\beta$ is submodular in (k, λ).
- Hence, $k^*_\beta(\lambda) \coloneqq \arg\min_{k \geqslant 0} C^{(k)}_\beta(\lambda)$ is increasing in λ

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

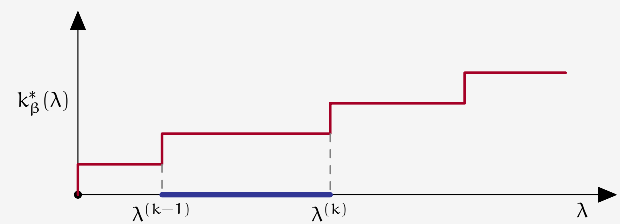
$$C^{(k)}_\beta(\lambda^{(k)}_\beta) = C^{(k+1)}_\beta(\lambda^{(k)}_\beta)$$



Step 3 Solution to costly optimization problem

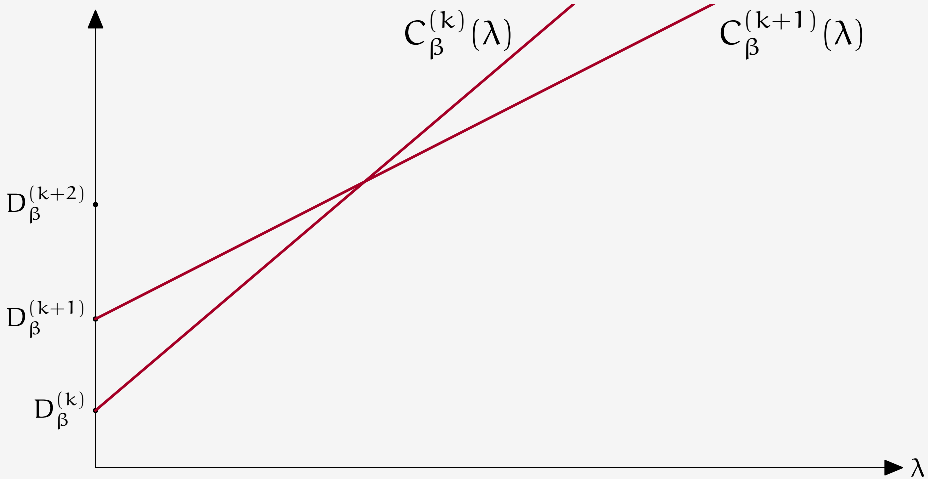


Define $\Lambda_\beta^{(k)} := \{\lambda \in \mathbb{R}_{\geq 0} : k_\beta^*(\lambda) = k\}$
$= [\lambda_\beta^{(k-1)}, \lambda_\beta^{(k)}]$.



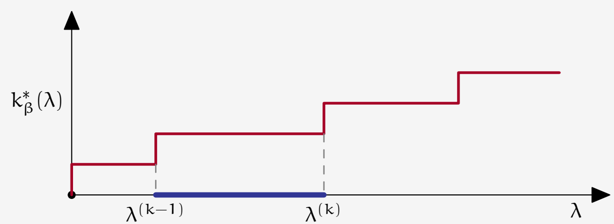
$$C_\beta^{(k)}(\lambda_\beta^{(k)}) = C_\beta^{(k+1)}(\lambda_\beta^{(k)})$$

Step 3 Solution to costly optimization problem

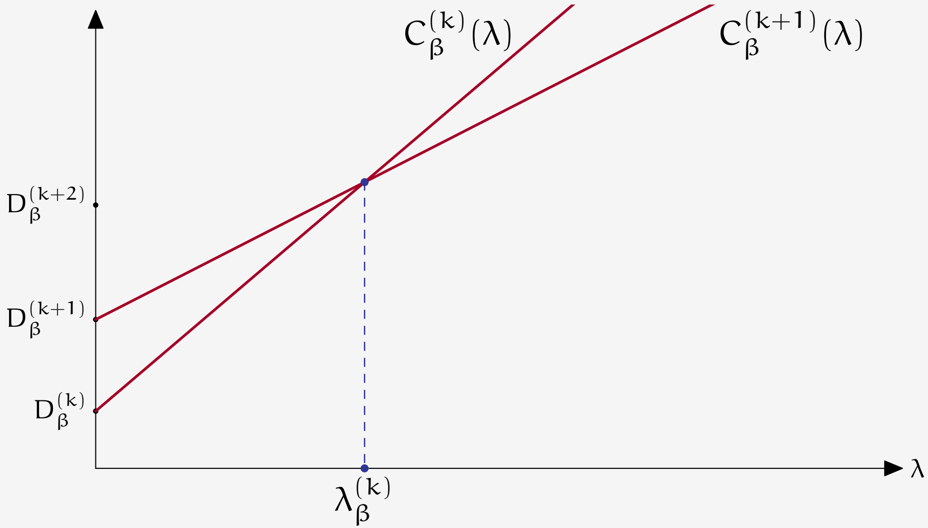


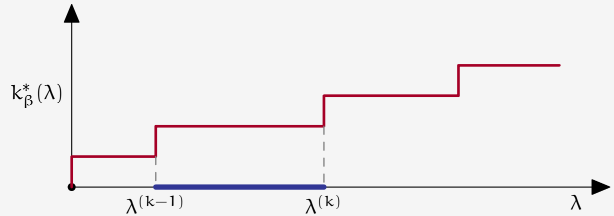
Define $\Lambda_\beta^{(k)} := \{\lambda \in \mathbb{R}_{\geqslant 0} : k^*_\beta(\lambda) = k\}$
$= [\lambda_\beta^{(k-1)}, \lambda_\beta^{(k)}].$

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



Solution to costly optimization problem



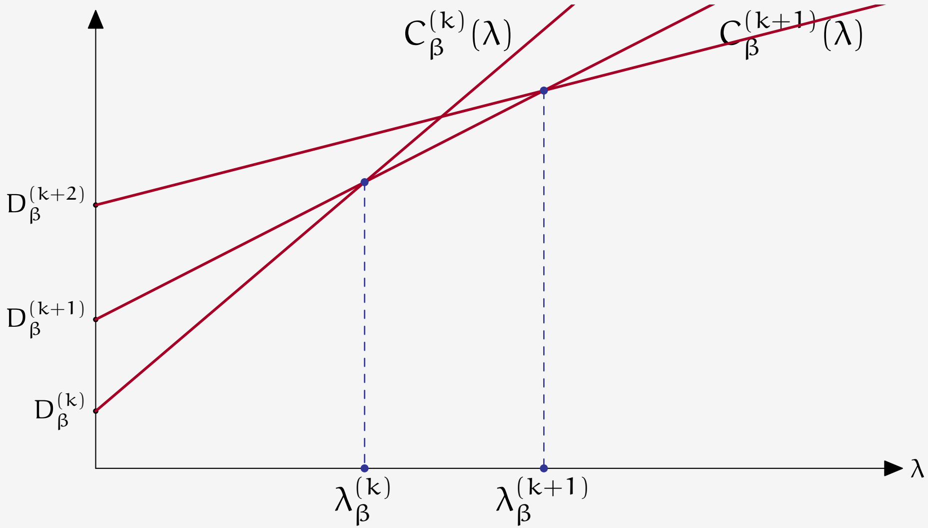


Define $\Lambda_\beta^{(k)} \coloneqq \{\lambda \in \mathbb{R}_{\geqslant 0} : k^*_\beta(\lambda) = k\}$

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

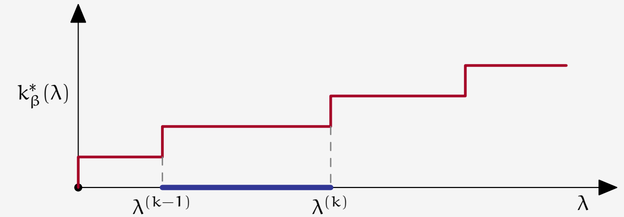
$$C_\beta^{(k)}(\lambda_\beta^{(k)}) = C_\beta^{(k+1)}(\lambda_\beta^{(k)})$$

Step 3 Solution to costly optimization problem

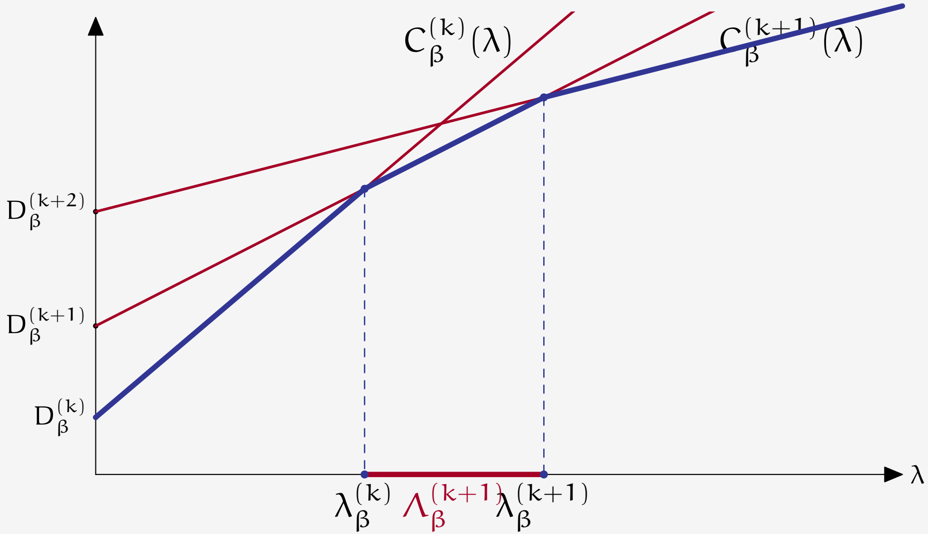


Define $\Lambda_\beta^{(k)} \coloneqq \{\lambda \in \mathbb{R}_{\geqslant 0} : k_\beta^*(\lambda) = k\}$
$= [\lambda_\beta^{(k-1)}, \lambda_\beta^{(k)}]$.

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

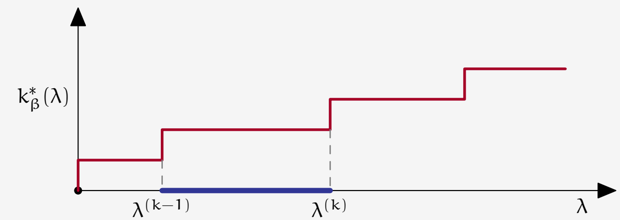


Step 3 Solution to costly optimization problem



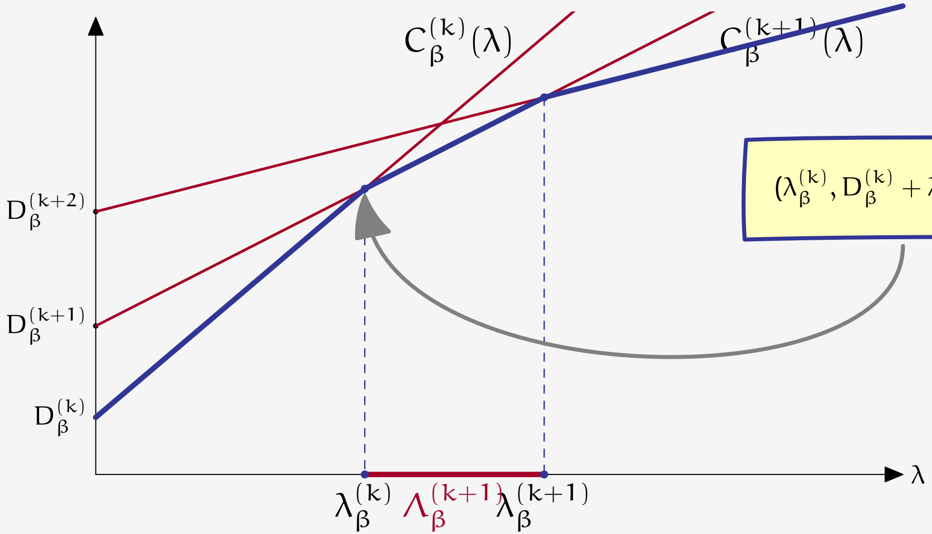
Define $\Lambda^{(k)}_\beta \coloneqq \{\lambda \in \mathbb{R}_{\geqslant 0} : k^*_\beta(\lambda) = k\}$

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

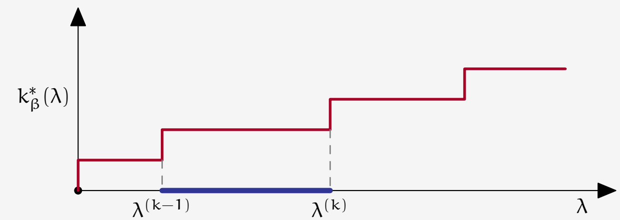


$$C^{(k)}_\beta(\lambda^{(k)}_\beta) = C^{(k+1)}_\beta(\lambda^{(k)}_\beta)$$

Step 3 Solution to costly optimization problem

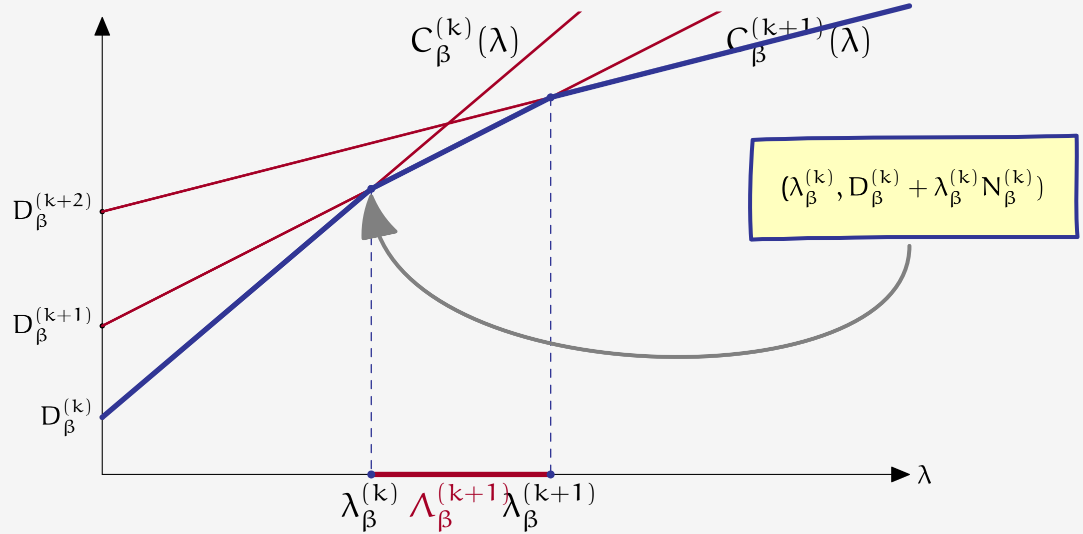


Define $\Lambda_\beta^{(k)} \coloneqq \{\lambda \in \mathbb{R}_{\geqslant 0} : k_\beta^*(\lambda) = k\}$
$= [\lambda_\beta^{(k-1)}, \lambda_\beta^{(k)}]$.



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

Step 3 Solution to costly optimization problem



Theorem

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

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

Step 4 Solution to constrained communication problem

Sufficient condition for optimality

A strategy (f°, g°) is optimal for the constrained problem if

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

(C2) There exists $\lambda^\circ \geqslant 0$ such that (f°, g°) is optimal for the Lagrange relaxation with parameter λ°.

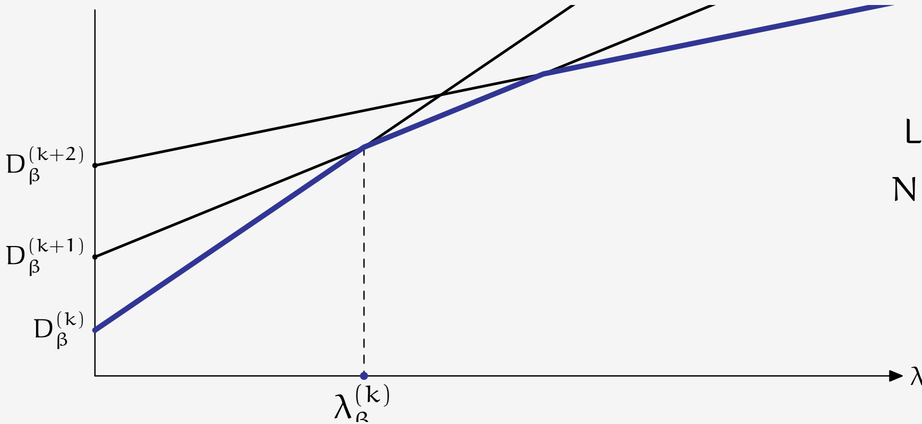
Solution to constrained communication problem

Sufficient condition for optimality

A strategy (f°, g°) is optimal for the constrained problem if

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

(C2) There exists $\lambda^\circ \geqslant 0$ such that (f°, g°) is optimal for the Lagrange relaxation with parameter λ°.



Let k_β^* be such that

$$N_\beta^{(k_\beta^*)} > \alpha > N_\beta^{(k_\beta^*+1)}$$

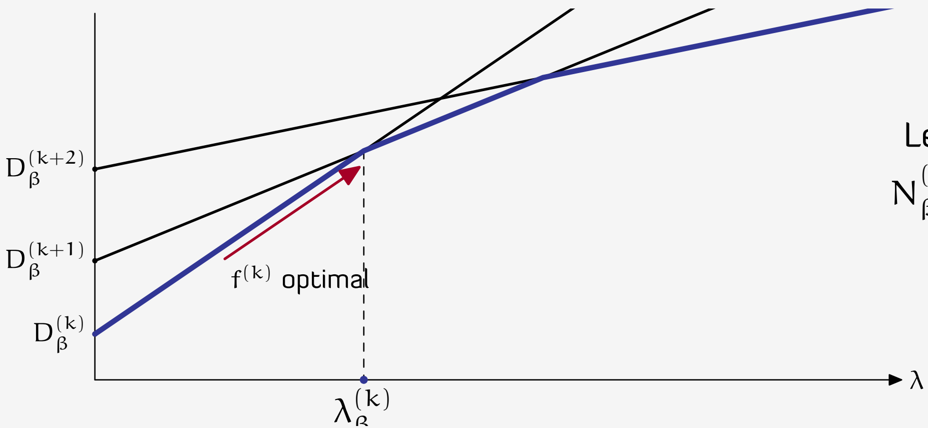
Step 4 Solution to constrained communication problem

Sufficient condition for optimality

A strategy (f°, g°) is optimal for the constrained problem if

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

(C2) There exists $\lambda^\circ \geqslant 0$ such that (f°, g°) is optimal for the Lagrange relaxation with parameter λ°.



Let k_β^* be such that

$N_\beta^{(k_\beta^*)} > \alpha > N_\beta^{(k_\beta^*+1)}$

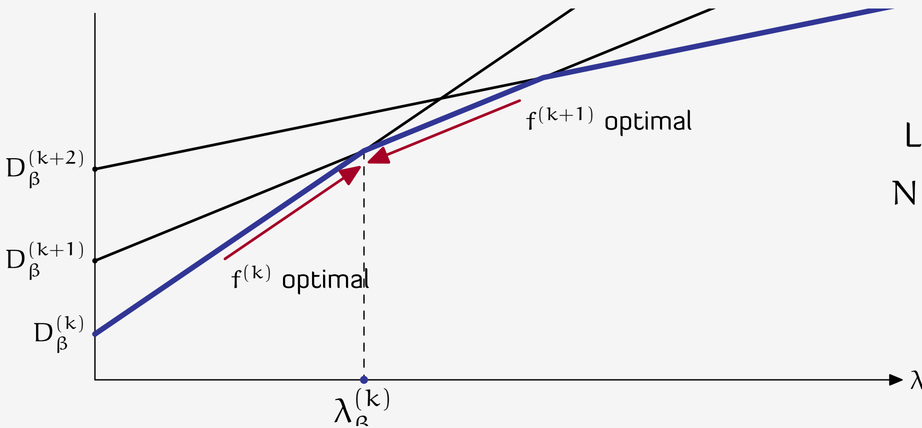
Solution to constrained communication problem

Sufficient condition for optimality

A strategy (f°, g°) is optimal for the constrained problem if

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

(C2) There exists $\lambda^\circ \geqslant 0$ such that (f°, g°) is optimal for the Lagrange relaxation with parameter λ°.



Let k_β^* be such that
$N_\beta^{(k_\beta^*)} > \alpha > N_\beta^{(k_\beta^*+1)}$

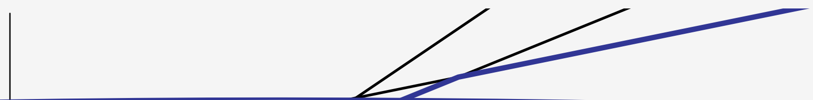
Step 4 Solution to constrained communication problem

Sufficient condition for optimality

A strategy (f°, g°) is optimal for the constrained problem if

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

(C2) There exists $\lambda^\circ \geqslant 0$ such that (f°, g°) is optimal for the Lagrange relaxation with parameter λ°.



such that

$> N_\beta^{(k^*_\beta+1)}$

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

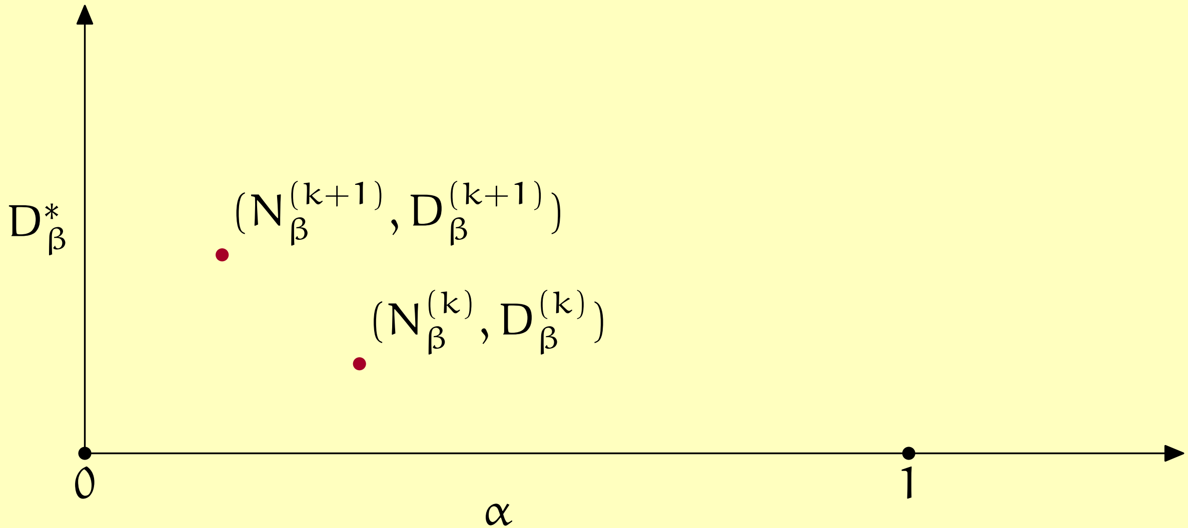
$$\theta^* N_\beta^{(k)} + (1-\theta^*) N_\beta^{(k+1)} = \alpha$$

Sufficient condition for optimality

A strate

(C1

(C2 … rameter λ°.



D^*_β is PWL, dec, and convex

R

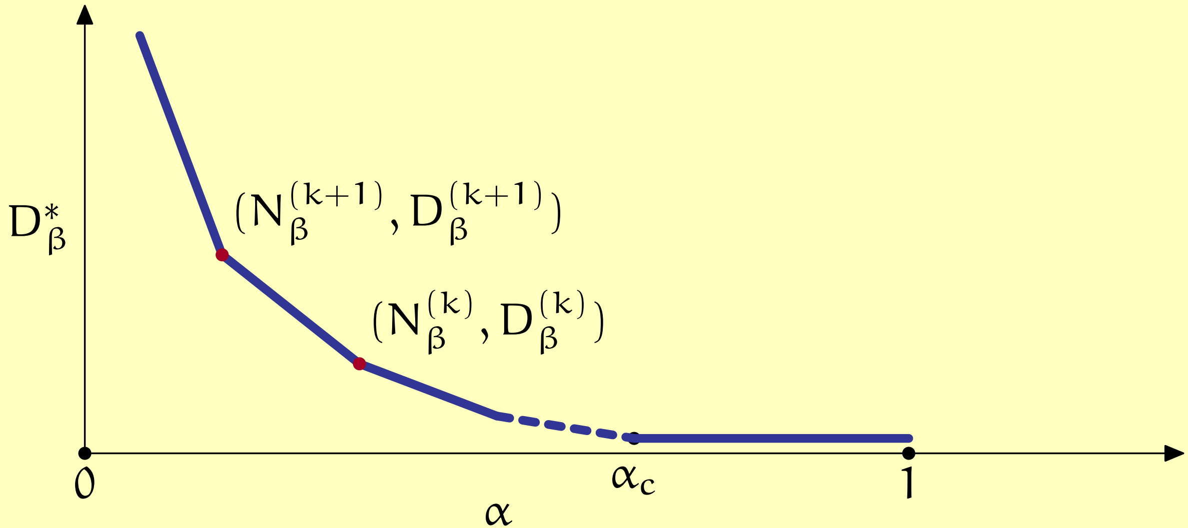
Sufficient condition for optimality

A strate

(C

(C2 rameter λ°.

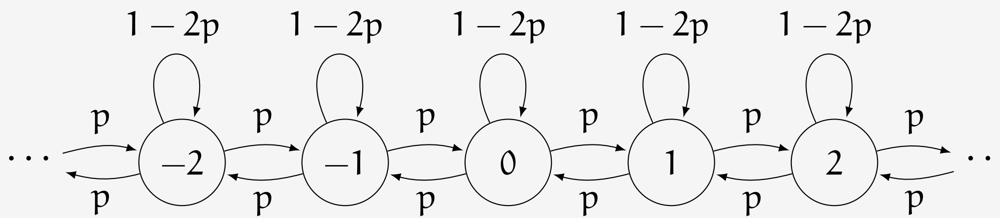
R



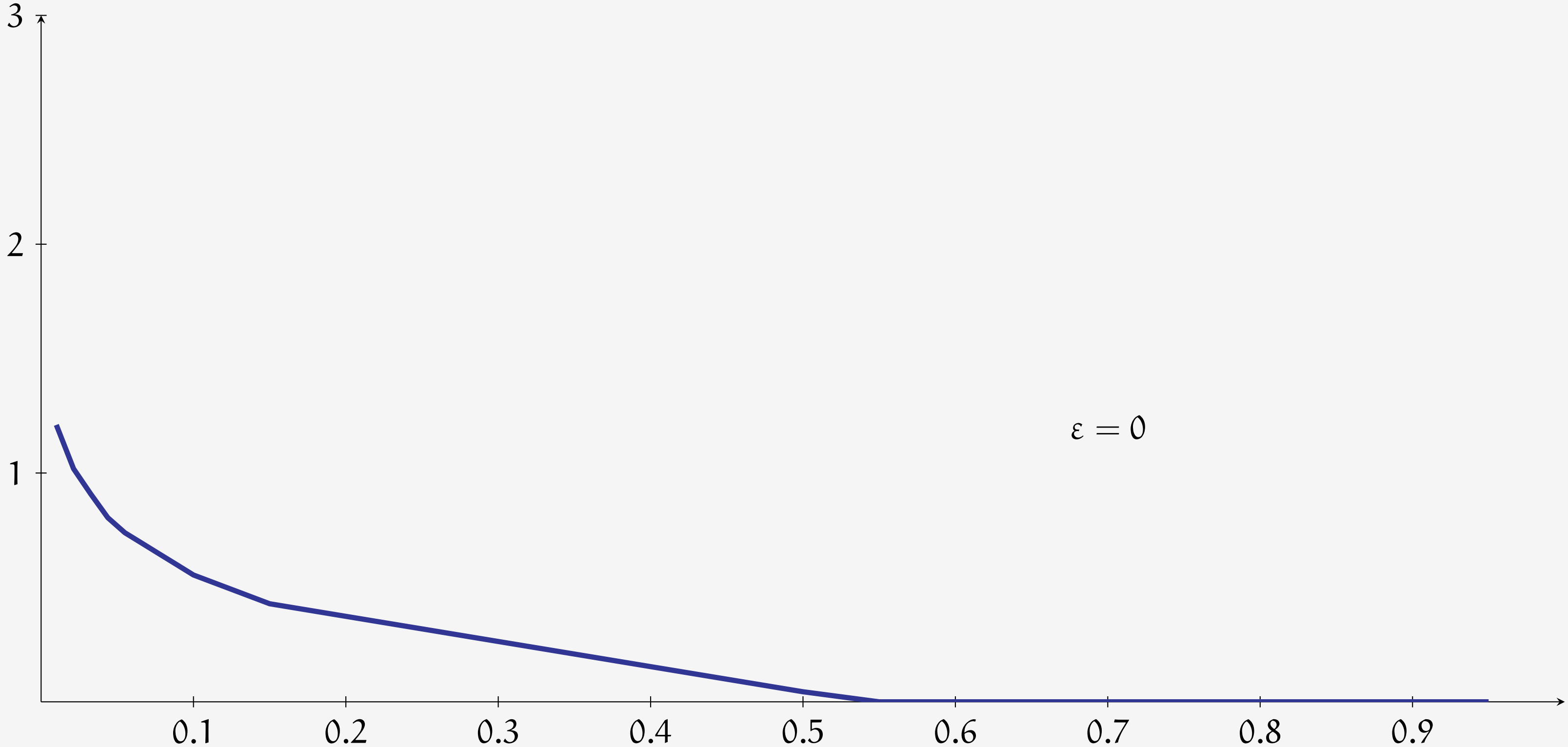
D^*_β is PWL, dec, and convex

Example Symmetric birth-death Markov chain

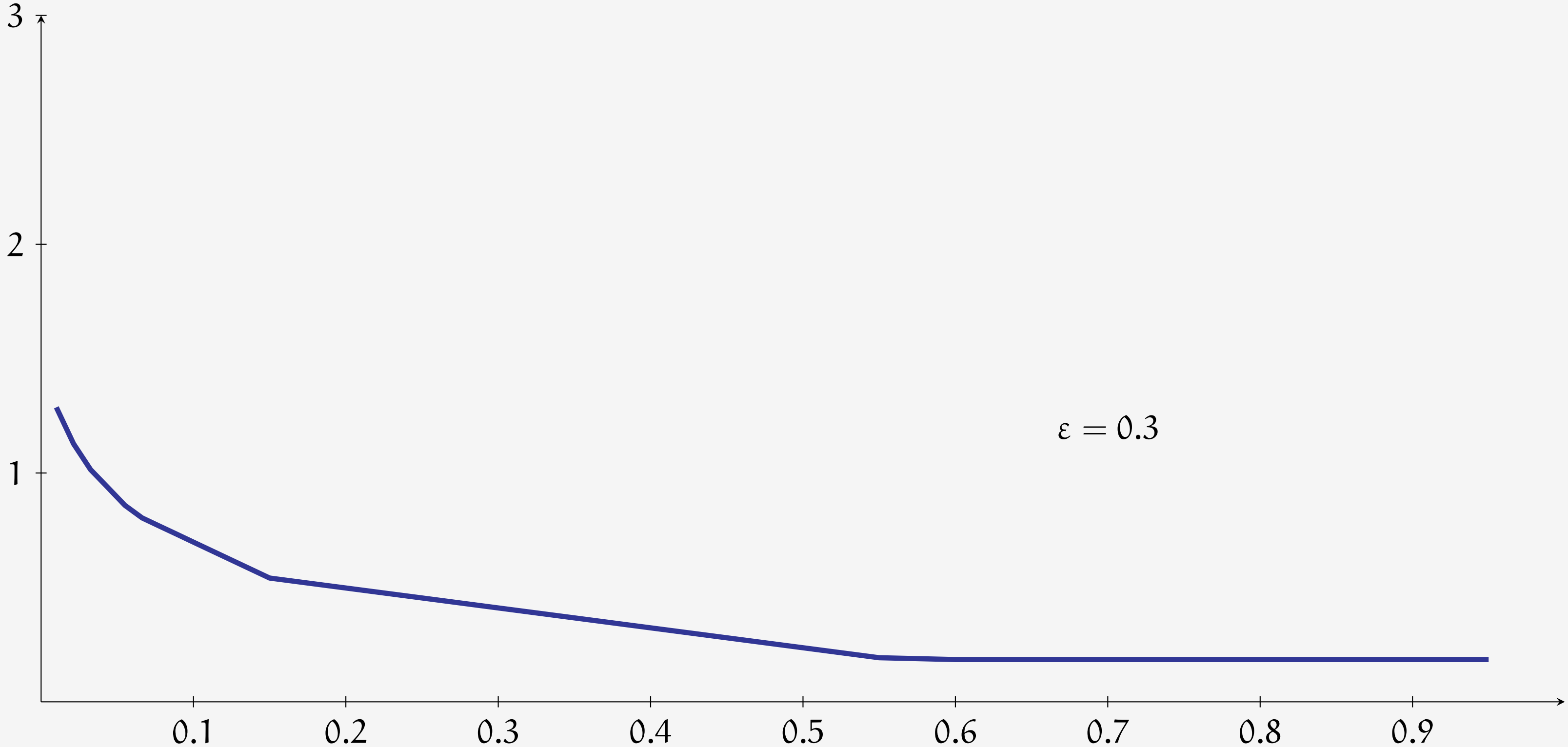
$$p_n = \begin{cases} p, & \text{if } |n| = 1; \\ 1-2p, & \text{if } n = 0; \\ 0, & \text{otherwise,} \end{cases} \qquad \text{where } p \in (0, \tfrac{1}{3}), \qquad d(e) = |e|$$



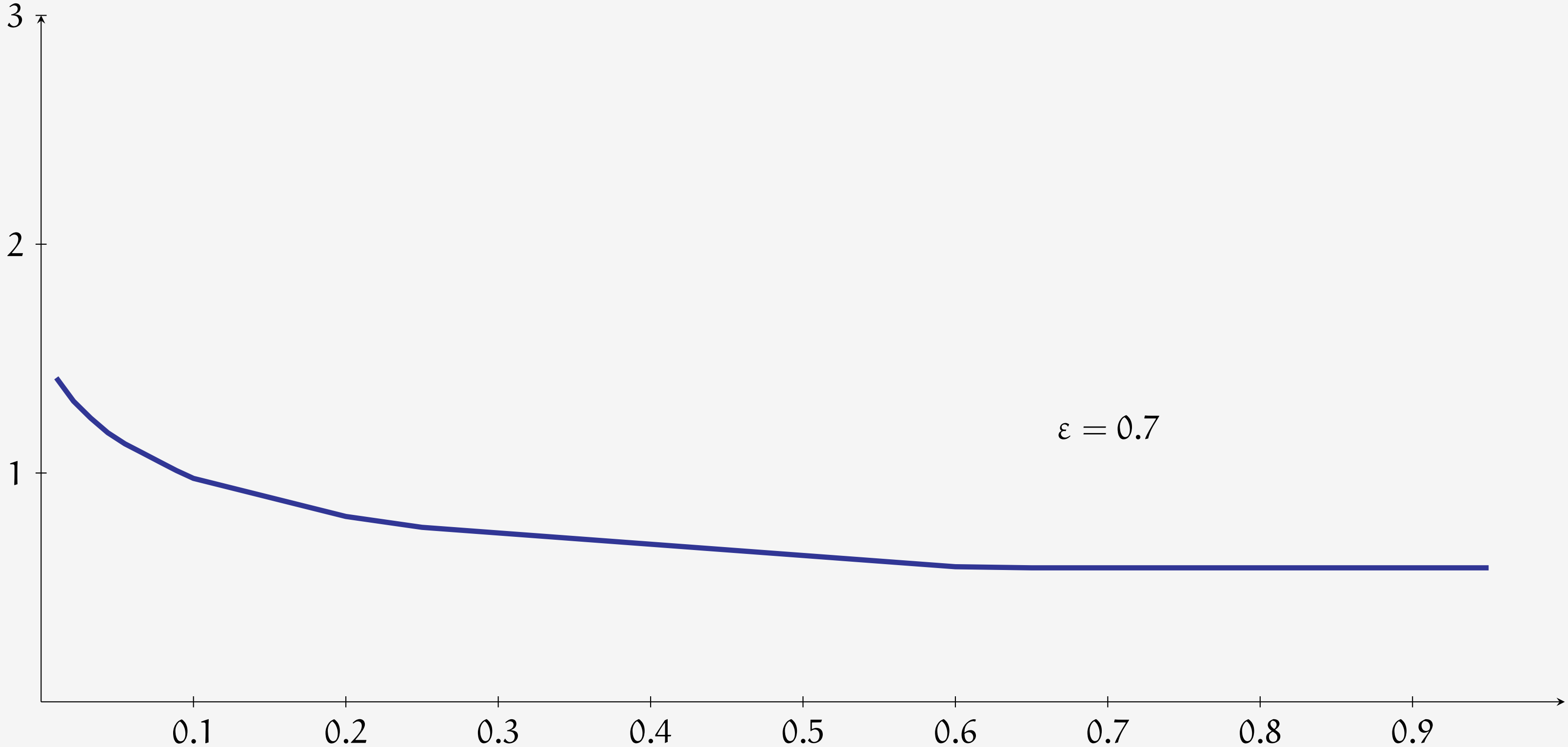
Example Symmetric birth-death Markov chain $(p = 0.3, \beta = 0.9)$



Example Symmetric birth-death Markov chain $(p = 0.3, \beta = 0.9)$



Example Symmetric birth-death Markov chain ($p = 0.3$, $\beta = 0.9$)



Optimal trade-offs for costly and constrained communication for continuous sources

Step 3 Solution to costly optimization problem

Proposition As in the case of discrete sources:

- $C^{(k)}_\beta(\lambda) \coloneqq D^{(k)}_\beta + \lambda N^{(k)}_\beta$ is submodular in (k, λ).
- Hence, $k^*_\beta(\lambda) \coloneqq \arg\min_{k \geqslant 0} C^{(k)}_\beta(\lambda)$ is increasing in λ

Step 3 Solution to costly optimization problem

Proposition As in the case of discrete sources:

- $C_\beta^{(k)}(\lambda) \coloneqq D_\beta^{(k)} + \lambda N_\beta^{(k)}$ is submodular in (k, λ).
- Hence, $k_\beta^*(\lambda) \coloneqq \arg\min_{k \geqslant 0} C_\beta^{(k)}(\lambda)$ is increasing in λ

Theorem If the pair (λ, k) satisfies

$$\lambda = -\frac{\partial_k D_\beta^{(k)}}{\partial_k N_\beta^{(k)}} \quad (\text{i.e., } \partial_k D_\beta^{(k)} + \lambda \partial_k N_\beta^{(k)} = 0)$$

then the strategy $(f^{(k)}, g^*)$ is optimal for the costly communication with cost λ.

The optimal performance $C_\beta^*(\lambda)$ is continuous, concave and increasing function of λ.

Step 3 Solution to costly optimization problem

Proposition As in the case of discrete sources:

- $C^{(k)}_\beta(\lambda) \coloneqq D^{(k)}_\beta + \lambda N^{(k)}_\beta$ is submodular in (k, λ).
- Hence, $k^*_\beta(\lambda) \coloneqq \arg\min_{k \geqslant 0} C^{(k)}_\beta(\lambda)$ is increasing in λ

Theorem If the pair (λ, k) satisfies

$$\lambda = -\frac{\partial_k D^{(k)}_\beta}{\partial_k N^{(k)}_\beta} \quad (\text{i.e., } \partial_k D^{(k)}_\beta + \lambda \partial_k N^{(k)}_\beta = 0)$$

then the strategy $(f^{(k)}, g^*)$ is optimal for the costly communication with cost λ.

The optimal performance $C^*_\beta(\lambda)$ is continuous, concave and increasing function of λ.

Scaling with variance for Gaussian noise

$$C^*_{\beta,\sigma}(\lambda) = \sigma^2 C^*_{\beta,1}\left(\frac{\lambda}{\sigma^2}\right).$$

Step 4 Solution to constrained optimization problem

Theorem For any $\beta \in (0,1]$ and $\alpha \in (0,1)$, let $k^*_\beta(\alpha)$ be such that

$$N_\beta^{(k^*_\beta(\alpha))} = \alpha.$$

Such a $k^*_\beta(\alpha)$ always exists and we have the following:

- The strategy $(f^{(k^*_\beta(\alpha))}, g^*)$ is optimal for the constrained optimization problem with constraint α
- The distortion transmission function $D^*_\beta(\alpha)$ is continuous, convex, and decreasing in α and is given by

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

Step 4 Solution to constrained optimization problem

Theorem For any $\beta \in (0, 1]$ and $\alpha \in (0, 1)$, let $k^*_\beta(\alpha)$ be such that

$$N_\beta^{(k^*_\beta(\alpha))} = \alpha.$$

Such a $k^*_\beta(\alpha)$ always exists and we have the following:

- The strategy $(f^{(k^*_\beta(\alpha))}, g^*)$ is optimal for the constrained optimization problem with constraint α
- The distortion transmission function $D^*_\beta(\alpha)$ is continuous, convex, and decreasing in α and is given by

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

Scaling with variance for Gaussian noise

$$k^*_{\beta,\sigma}(\alpha) = k^*_{\beta,1}(\alpha) \text{ and } D^*_{\beta,\sigma}(\alpha) = \sigma^2 D^*_{\beta,1}(\alpha).$$

Computation of optimal thresholds

Costly communication Given λ, find k such that $\partial_k(D_\beta^{(k)} + \lambda N_\beta^{(k)}) = 0$.

Constrained communication Given α, find k such that $N_\beta^{(k)} = \alpha$.

Computation of optimal thresholds

Costly communication Given λ, find k such that $\partial_k(D^{(k)}_\beta + \lambda N^{(k)}_\beta) = 0$.

Constrained communication Given α, find k such that $N^{(k)}_\beta = \alpha$.

Main idea

- Pick a threshold k and use strategy $f^{(k)}$ until first successful reception.
- The sample path values of L, M, and K may be viewed as a "noisy" observation of true $L^{(k)}_\beta$, $M^{(k)}_\beta$, and $K^{(k)}_\beta$.
- Use stochastic approximation to find optimal thresholds.

Computation of optimal thresholds

Kiefer-Wolfowitz Algorithm

Robbins-Monro Algorithm

Costly communication Given λ, find k such that $\partial_k(D_\beta^{(k)} + \lambda N_\beta^{(k)}) = 0$.

Constrained communication Given α, find k such that $N_\beta^{(k)} = \alpha$.

Main idea

- Pick a threshold k and use strategy $f^{(k)}$ until first successful reception.
- The sample path values of L, M, and K may be viewed as a "noisy" observation of true $L_\beta^{(k)}$, $M_\beta^{(k)}$, and $K_\beta^{(k)}$.
- Use stochastic approximation to find optimal thresholds.

Computation of optimal thresholds

Costly communic

Constrai
communicat

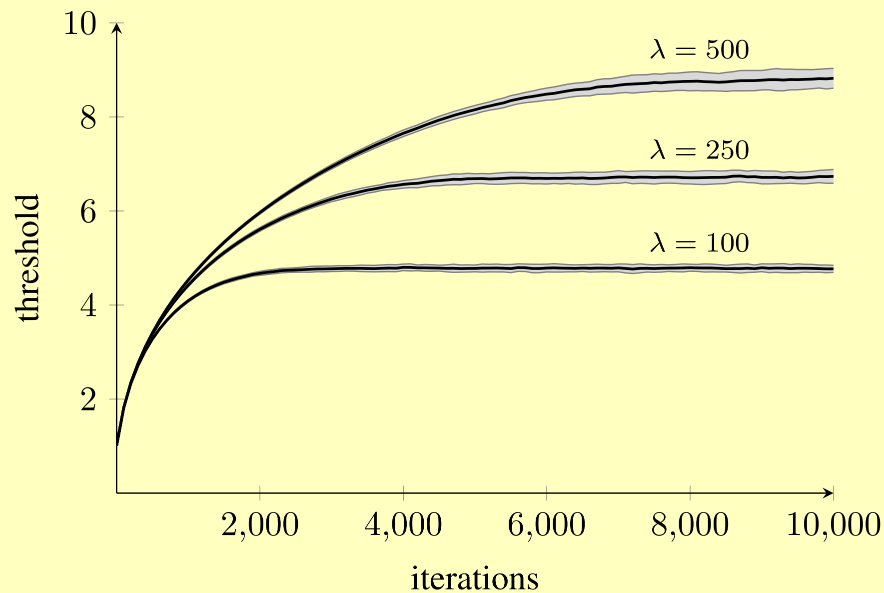
Main i

…lfowitz Algorithm

…Monro Algorithm

…reception.

…oisy” observation



Costly communication ($\beta = 0.9$, $\varepsilon = 0.3$)

Computation of optimal thresholds

Costly communic

Constrai
communicat

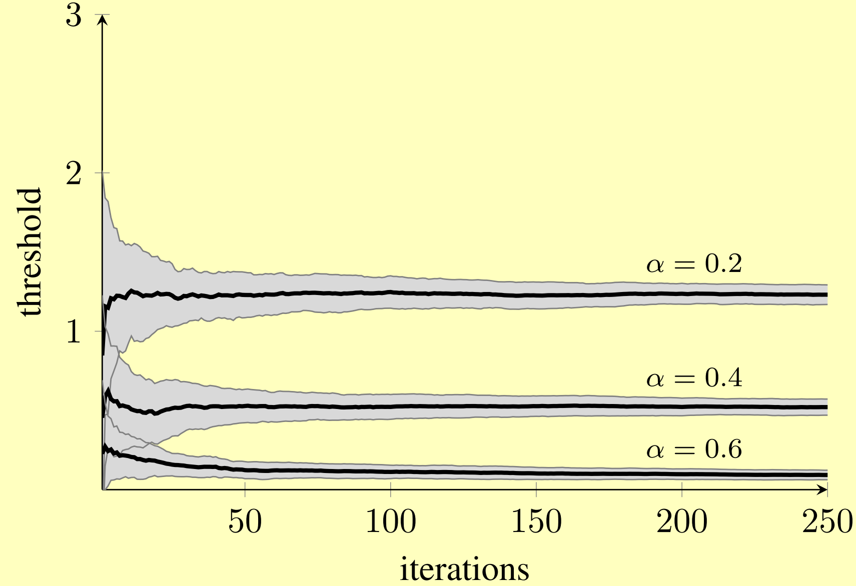
Main i

ılfowitz Algorithm

nro Algorithm

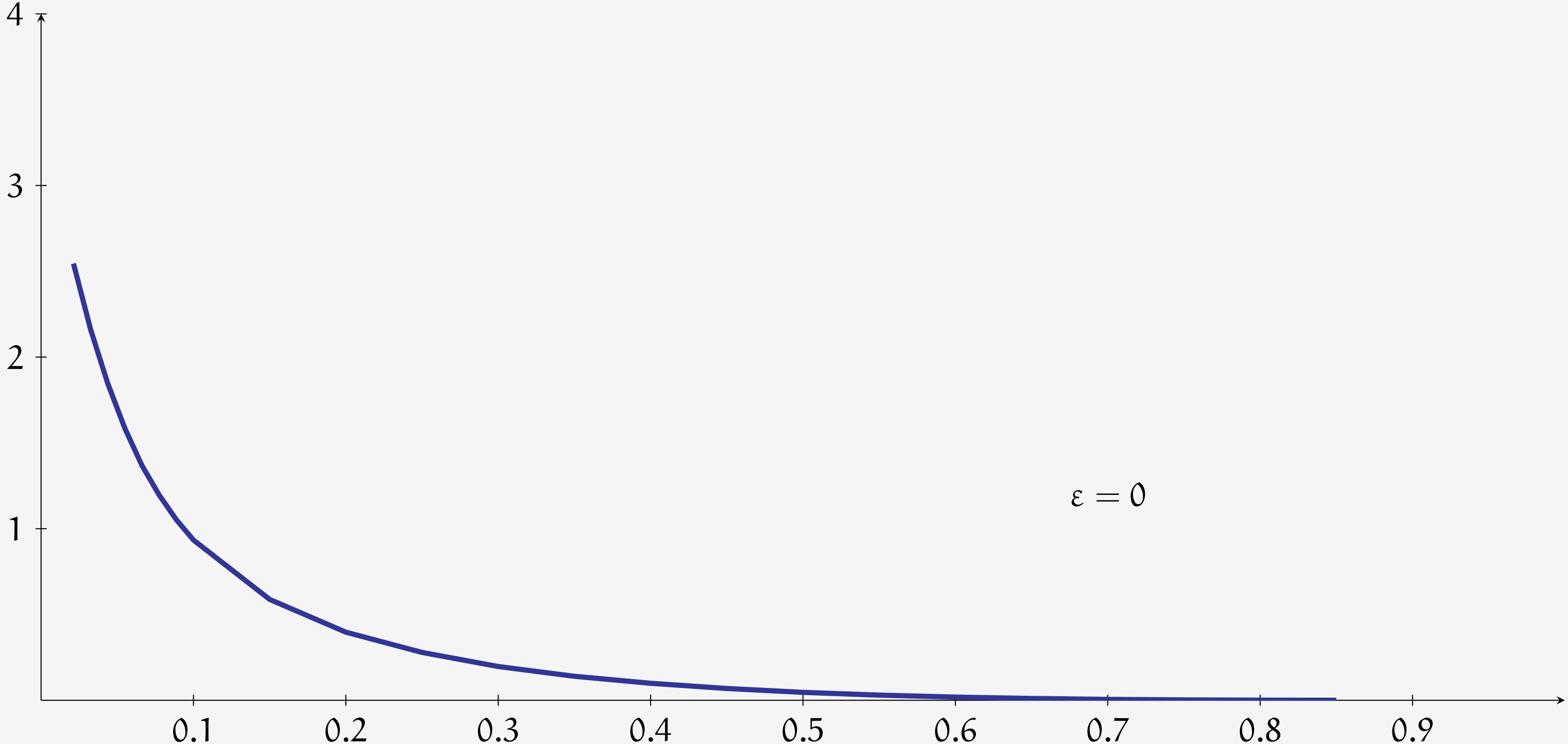
ception.

sy" observation

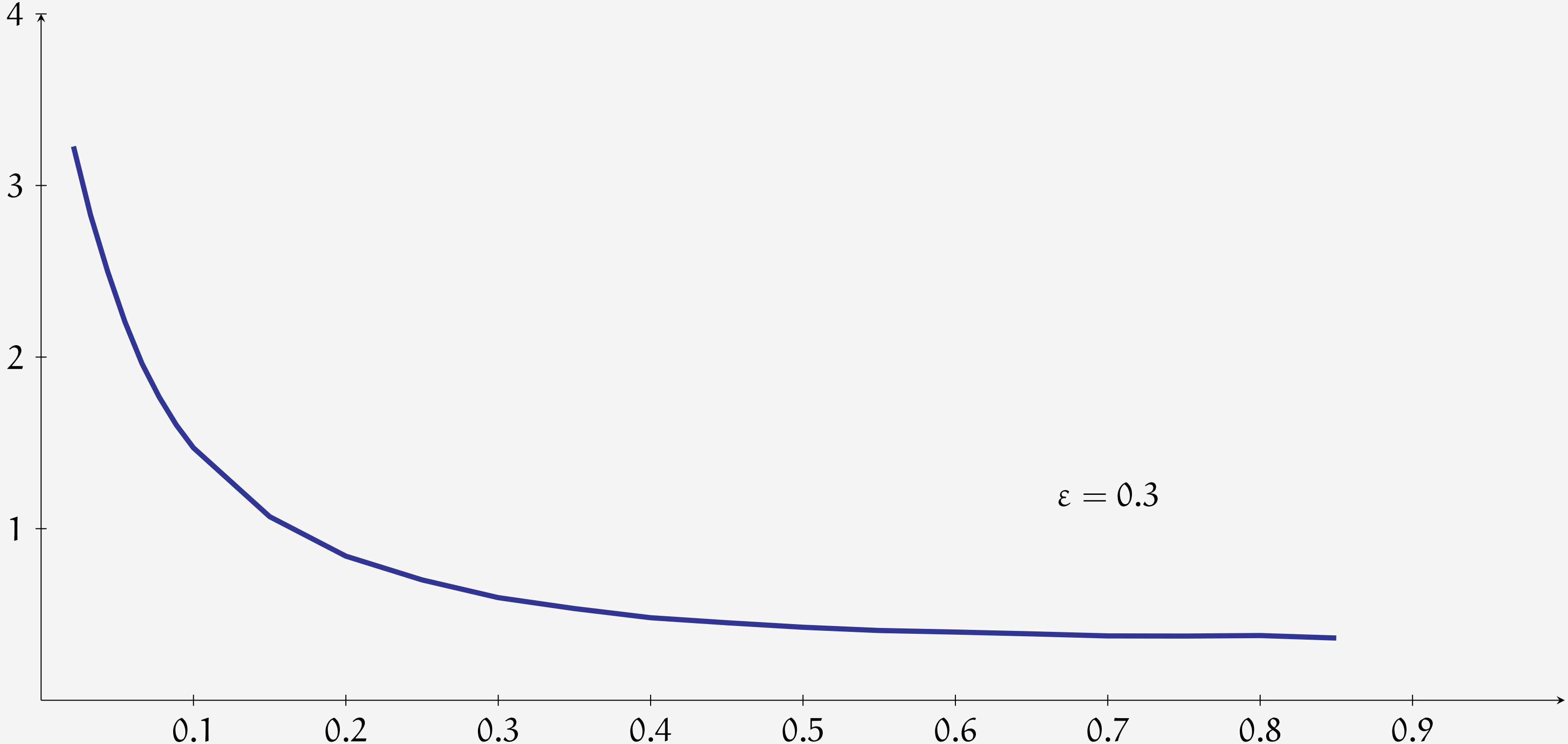


Constrained communication ($\beta = 0.9$, $\varepsilon = 0.3$)

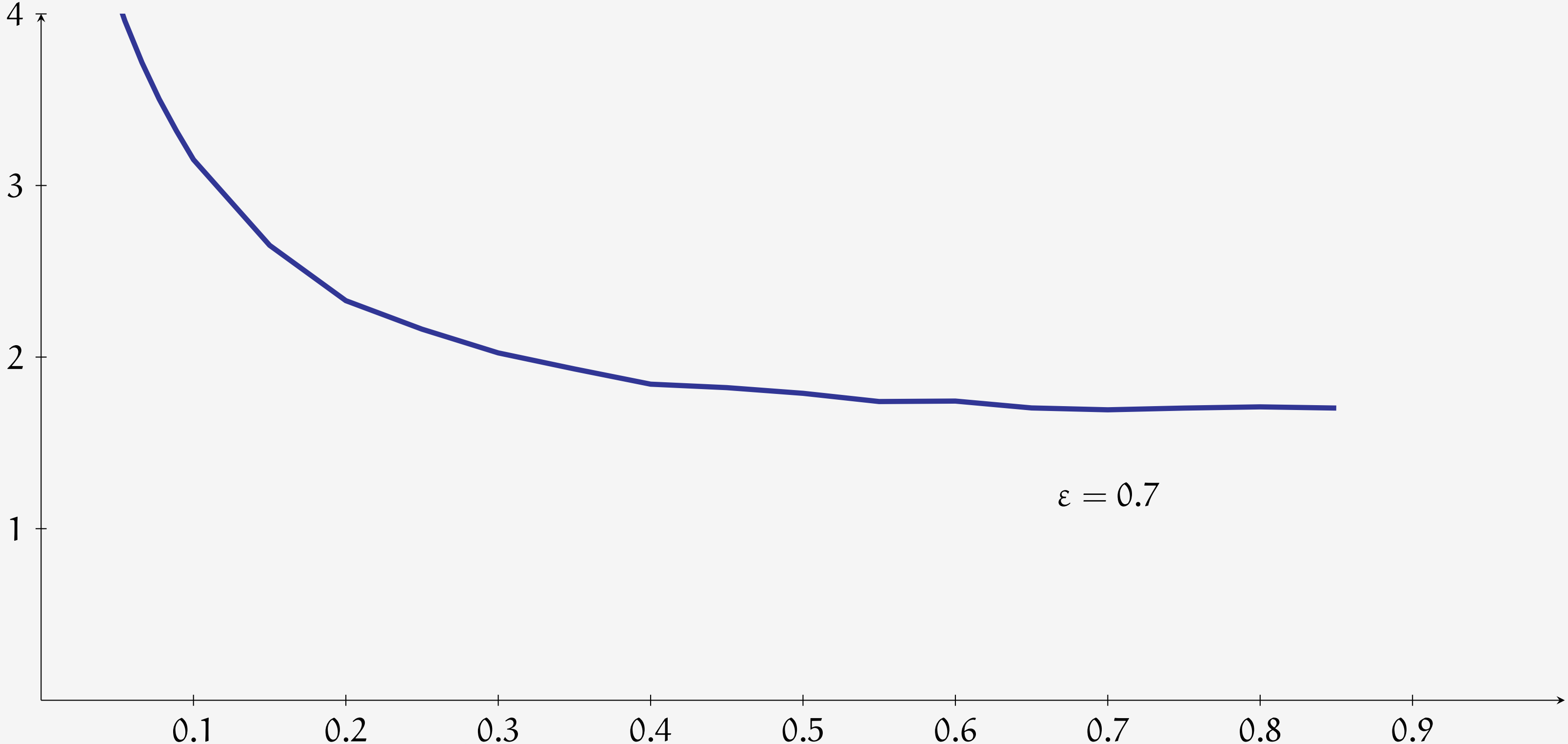
Distortion transmission fn for Gauss-Markov process $(\sigma^2 = 1, \beta = 0.9)$



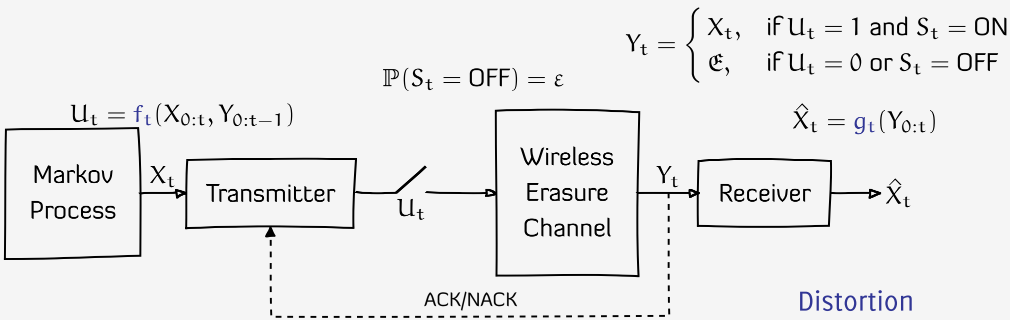
Distortion transmission fn for Gauss-Markov process $(\sigma^2 = 1, \beta = 0.9)$



Distortion transmission fn for Gauss-Markov process $(\sigma^2 = 1, \beta = 0.9)$



Summary



Distortion

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

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

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

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

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

Summary

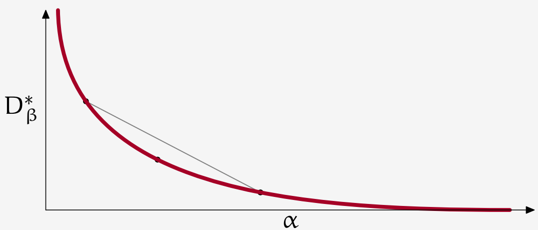
Optimization problems

Constrained communication

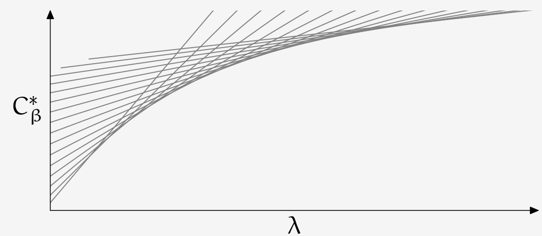
$$\text{For } \alpha \in (0,1), \quad D^*_\beta(\alpha) \coloneqq \inf_{(f,g)} \{ D_\beta(f,g) : N_\beta(f,g) \leqslant \alpha \}$$

Costly communication (Lagrange relaxation)

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



D^*_β is cts, dec, and convex



Summary

Optimization problems

Distortion transmission function for auto-regressive sources

Source model $X_{t+1} = aX_t + W_t$, where W_t has symmetric and unimodal distribution. $X_t \in \mathbb{Z}/\mathbb{R}$.

Optimal transmission strategy

$$U_t = \begin{cases} 1, & \text{if } |X_t - a\hat{X}_{t-1}| \geq k \\ 0, & \text{otherwise} \end{cases}$$

Optimal estimation strategy

$$\hat{X}_t = \begin{cases} a\hat{X}_{t-1}, & \text{if } Y_t = \mathfrak{E} \\ Y_t, & \text{if } Y_t \neq \mathfrak{E} \end{cases}$$

Performance of threshold based strategies

- $K_\beta^{(k)}$: Expected discounted number of transmissions until first successful reception.
- $L_\beta^{(k)}$: Expected discounted distortion until first successful reception.
- $M_\beta^{(k)}$: Expected discounted time until first successful reception.

Then, $D_\beta^{(k)} = \dfrac{L_\beta^{(k)}}{M_\beta^{(k)}}$ and $N_\beta^{(k)} = \dfrac{K_\beta^{(k)}}{M_\beta^{(k)}}$.

1

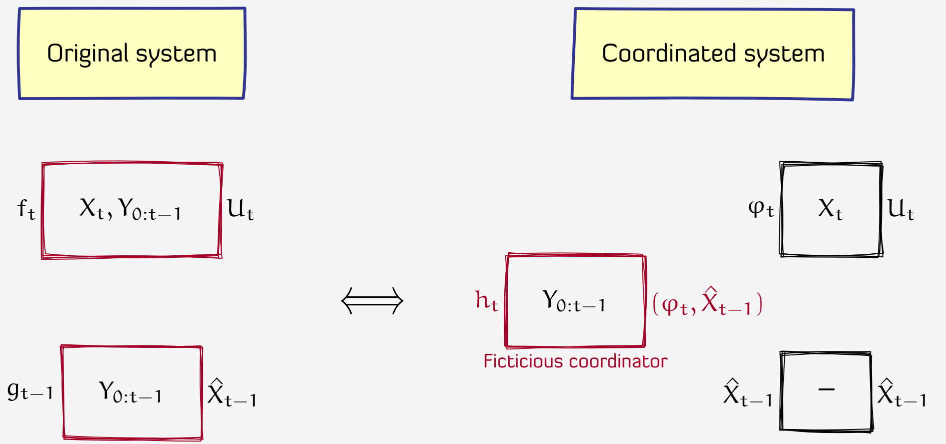
2

Summary

Optimization problems

Distortion transmission function for auto-regressive sources

The common information approach (Nayyar, Mahajan, Teneketzis 2013)



- The coordinated system is equivalent to the original system.

  $$f_t(x, y_{0:t-1}) = h_t^1(y_{0:t-1})(x).$$

- The coordinated system is centralized. Belief state $\mathbb{P}(X_t \mid Y_{0:t-1})$.

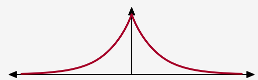
- Nayyar, Mahajan and Teneketzis, "Decentralized stochastic control with partial history sharing: A common information approach," IEEE TAC 2013.

Summary

Simplifying modeling assumptions

Markov process: $X_{t+1} = aX_t + W_t$

- Discrete state process: $X_t,\ a,\ W_t \in \mathbb{Z}$
- Continuous state process: $X_t,\ a,\ W_t \in \mathbb{R}$



Noise Distribution: Unimodal and symmetric

Distortion function: Even and increasing

Proof outline

- **Step 1** Show that threshold-based strategies are optimal
- **Step 2** Find performance of arbitrary threshold based strategies
- **Step 3** Solution to the costly communication problem
- **Step 4** Solution to the constrained communication problem

Summary

Computation of optimal thresholds

Costly communication: Given λ, find k such that $\partial_k(D^{(k)}_\beta + \lambda N^{(k)}_\beta) = 0$.

Kiefer-Wolfowitz Algorithm

Constrained communication: Given α, find k such that $N^{(k)}_\beta = \alpha$.

Robbins-Monro Algorithm

Main idea

- Pick a threshold k and use strategy $f^{(k)}$ until first successful reception.
- The sample path values of L, M, and K may be viewed as a "noisy" observation of true $L^{(k)}_\beta$, $M^{(k)}_\beta$, and $K^{(k)}_\beta$.
- Use stochastic approximation to find optimal thresholds.

Concluding Remarks

Generalization to vector sources

- Difficulty: If X_t is ASU, is $AX_t + W_t$ also ASU?
- Even if threshold policies are not optimal, the tools developed may be useful to identify best threshold-based strategy.

Concluding Remarks

Generalization to vector sources

- Difficulty: If X_t is ASU, is $AX_t + W_t$ also ASU?
- Even if threshold policies are not optimal, the tools developed may be useful to identify best threshold-based strategy.

Results are derived under idealized assumptions

Future directions

- Power or rate control ...
- Markovian or burst erasures ...
- Scheduling multiple sources ...
- Model network delays ...

References

Jhelum Chakravorty and Aditya Mahajan, "Fundamental limits of remote estimation of autoregressive Markov processes under communication constraints," IEEE TAC, Sep 2017 (to appear).

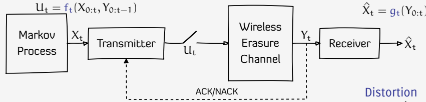
Jhelum Chakravorty and Aditya Mahajan, "Remote-state estimation with packet drop," IFAC Conference on Networked Control Systems (NecSys), Aug 2016. (Best Student Paper Award)

Jhelum Chakravorty, Jayakumar Subramanian, and Aditya Mahajan, "Stochastic approximation based methods for computing the optimal thresholds in remote-state estimation with packet drops," ACC 2017 (submitted)

$$Y_t = \begin{cases} X_t, & \text{if } U_t = 1 \text{ and } S_t = \text{ON} \\ \mathfrak{E}, & \text{if } U_t = 0 \text{ or } S_t = \text{OFF} \end{cases}$$

$\mathbb{P}(S_t = \text{OFF}) = \varepsilon$

$U_t = f_t(X_{0:t}, Y_{0:t-1})$ $\hat{X}_t = g_t(Y_{0:t})$



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

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

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

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

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

Distortion transmission function for auto-regressive sources

Source model $X_{t+1} = aX_t + W_t$, where W_t has symmetric and unimodal distribution. $X_t \in \mathbb{Z}/\mathbb{R}$.

Optimal transmission strategy

$$U_t = \begin{cases} 1, & \text{if } |X_t - a\hat{X}_{t-1}| \geq k \\ 0, & \text{otherwise} \end{cases}$$

Optimal estimation strategy

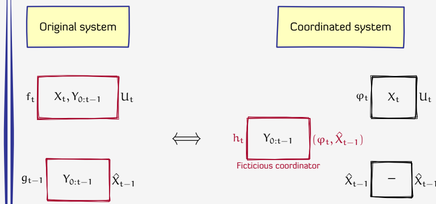
$$\hat{X}_t = \begin{cases} a\hat{X}_{t-1}, & \text{if } Y_t = \mathfrak{E} \\ Y_t, & \text{if } Y_t \neq \mathfrak{E} \end{cases}$$

Performance of threshold based strategies

- $K_\beta^{(k)}$: Expected discounted number of transmissions until first successful reception.
- $L_\beta^{(k)}$: Expected discounted distortion until first successful reception.
- $M_\beta^{(k)}$: Expected discounted time until first successful reception.

Then, $D_\beta^{(k)} = \dfrac{L_\beta^{(k)}}{M_\beta^{(k)}}$ and $N_\beta^{(k)} = \dfrac{K_\beta^{(k)}}{M_\beta^{(k)}}$.

The common information approach (Nayyar, Mahajan, Teneketzis 2013)



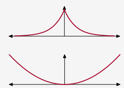
- The coordinated system is equivalent to the original system.
  $f_t(x, y_{0:t-1}) = h_t^1(y_{0:t-1})(x)$.
- The coordinated system is centralized. Belief state $\mathbb{P}(X_t \mid Y_{0:t-1})$.
- Nayyar, Mahajan and Teneketzis, "Decentralized stochastic control with partial history sharing: A common information approach," IEEE TAC 2013.

Simplifying modeling assumptions

Markov process $X_{t+1} = aX_t + W_t$
- Discrete state process: $X_t, a, W_t \in \mathbb{Z}$
- Continuous state process: $X_t, a, W_t \in \mathbb{R}$

Noise Distribution Unimodal and symmetric

Distortion function Even and increasing



Proof outline

Step 1 Show that threshold-based strategies are optimal

Step 2 Find performance of arbitrary threshold based strategies

Step 3 Solution to the costly communication problem

Step 4 Solution to the constrained communication problem

Step 1 Properties of majorization

Threshold based strategies

Let $\mathcal{F}(c)$ denote the class of all threshold based strategies around c, i.e.,

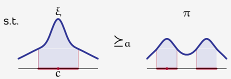
$$\varphi \in \mathcal{F}(c) \quad \text{if } \exists k \text{ s.t.} \quad \varphi(e) = \begin{cases} 1 & \text{if } |e - ac| \geq k \\ 0 & \text{otherwise} \end{cases}$$

Property 1

For any $\xi \succeq_a \pi$ where ξ is ASU(c),

and any φ, there exists a $\theta \in \mathcal{F}(c)$ s.t.

$$\sum_{e\in\mathcal{X}} \theta(e)\xi(e) = \sum_{e\in\mathcal{X}} \varphi(e)\pi(e).$$

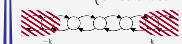


Moreover, for $h \in \{0,1\}$ (recall $h = u \cdot s$), $Q(\xi, \theta, h) \succeq_a Q(\pi, \varphi, h)$.

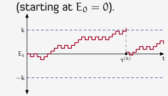
Step 2 Performance of threshold-based strategies

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 reception (starting at $E_0 = 0$).



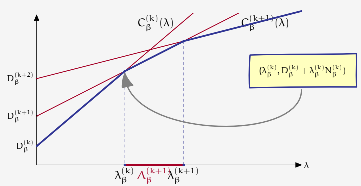
Define

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

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

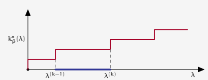
$$K_\beta^{(k)}(e) = \mathbb{E}\Big[\sum_{t=0}^{\tau^{(k)}-1} \beta^t U_t \Big| E_0 = e\Big].$$

Step 3 Solution to costly optimization problem



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

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



Step 4 Solution to constrained communication problem

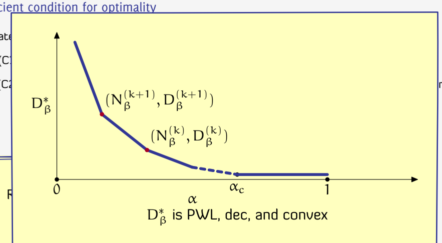
Sufficient condition for optimality

A strate

(C

(C

rameter λ°.



Computation of optimal thresholds

Costly communication Given λ, find k such that $\partial_k(D_\beta^{(k)} + \lambda N_\beta^{(k)}) = 0$.

Kiefer-Wolfowitz Algorithm

Constrained communication Given α, find k such that $N_\beta^{(k)} = \alpha$.

Robbins-Monro Algorithm

Main idea
- Pick a threshold k and use strategy $f^{(k)}$ until first successful reception.
- The sample path values of L, M, and K may be viewed as a "noisy" observation of true $L_\beta^{(k)}$, $M_\beta^{(k)}$, and $K_\beta^{(k)}$.
- Use stochastic approximation to find optimal thresholds.