Fundamental limits of remote state estimation

Aditya Mahajan
McGill University

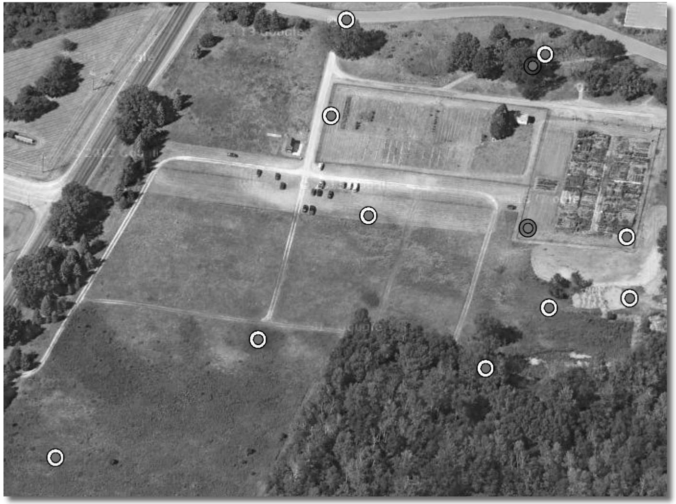
Joint work with Jhelum Chakravorty and Jayakumar Subramanian

BLISS Seminar, UC Berkeley
20 March, 2017

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

Many applications require:

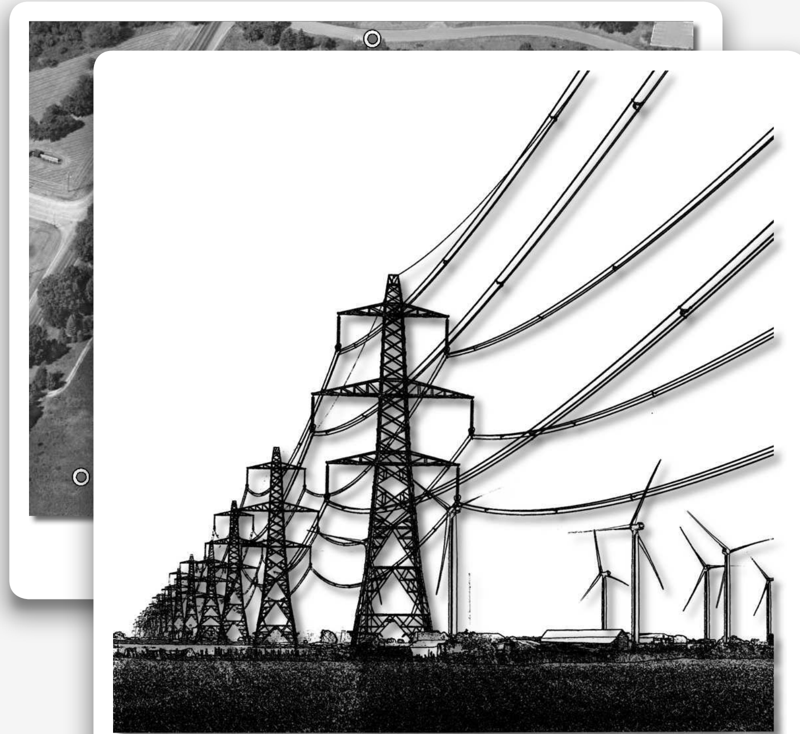
- Sequential transmission of data
- Zero- (or finite-) delay reconstruction



Sensor Networks

Many applications require:

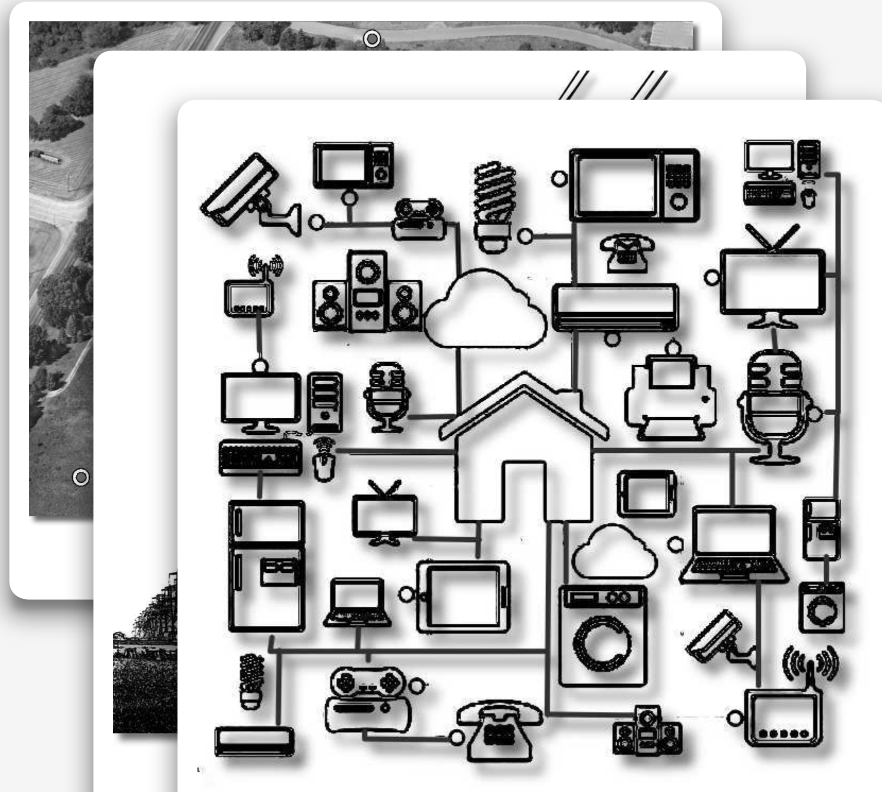
- Sequential transmission of data
- Zero- (or finite-) delay reconstruction



Smart Grids

Many applications require:

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

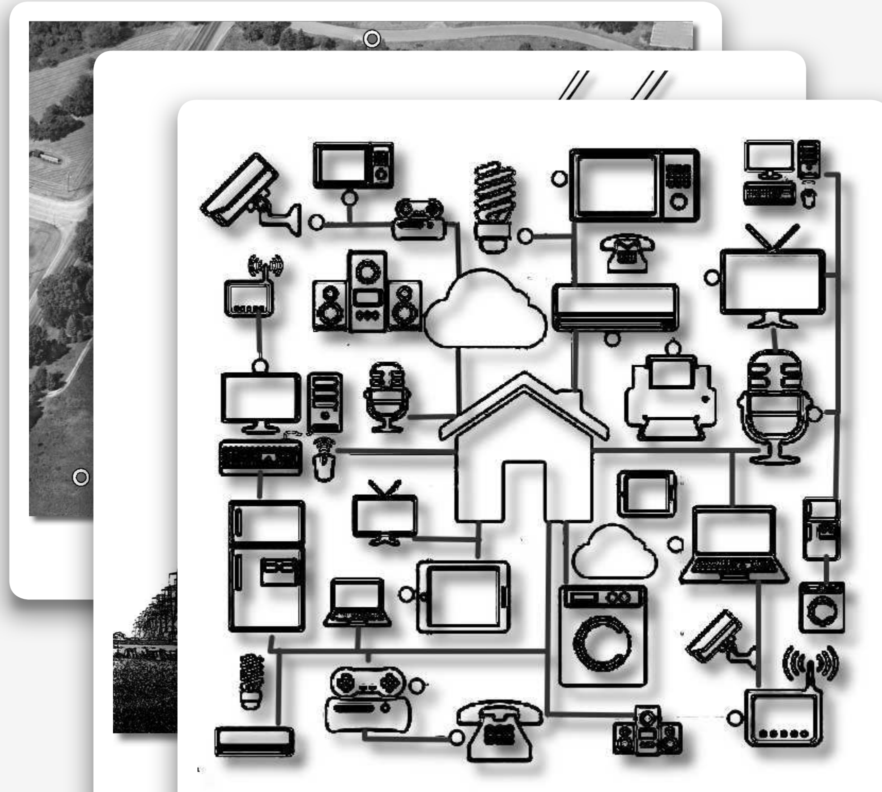


Internet of Things

Remote stat

Many applications require:

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



Internet of Things

Remote stat

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

Remote stat

Communication system

Source model

$\{X_t\}_{t \geqslant 0}$, $X_t \in \mathcal{X}$, is a first-order Markov process.

For some results, we restrict to **autoregressive model**: $X_{t+1} = aX_t + W_t$, $X_t \in \mathbb{Z}/\mathbb{R}$.

Communication system

Source model $\{X_t\}_{t \geq 0}$, $X_t \in \mathcal{X}$, is a first-order Markov process.

For some results, we restrict to **autoregressive model**: $X_{t+1} = aX_t + W_t$, $X_t \in \mathbb{Z}/\mathbb{R}$.

Channel model Gilbert-Elliot channel (at the packet level). Transition matrix Q.

Communication system

Source model

$\{X_t\}_{t \geqslant 0}$, $X_t \in \mathcal{X}$, is a first-order Markov process.

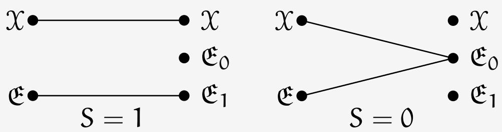
For some results, we restrict to **autoregressive model**: $X_{t+1} = aX_t + W_t$, $X_t \in \mathbb{Z}/\mathbb{R}$.

Channel model

Gilbert-Elliot channel (at the packet level). Transition matrix Q.

When $S_t = 1$ (Channel is ON)
channel output = channel input

When $S_t = 0$ (Channel is OFF)
channel output = noise



Communication system

Source model

$\{X_t\}_{t \geqslant 0}$, $X_t \in \mathcal{X}$, is a first-order Markov process.

For some results, we restrict to **autoregressive model**: $X_{t+1} = aX_t + W_t$, $X_t \in \mathbb{Z}/\mathbb{R}$.

Channel model

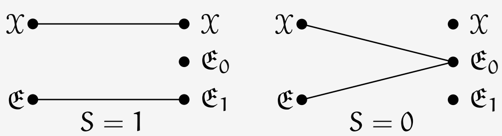
Gilbert-Elliot channel (at the packet level). Transition matrix Q.

When $S_t = 1$ (Channel is ON)
channel output = channel input

When $S_t = 0$ (Channel is OFF)
channel output = noise

Formal definition

Input alphabet $\bar{\mathcal{X}} = \mathcal{X} \cup \{\mathfrak{E}\}$ Output alphabet $\mathcal{Y} = \mathcal{X} \cup \{\mathfrak{E}_0, \mathfrak{E}_1\}$.



Communication system

Source model

$\{X_t\}_{t \geqslant 0}$, $X_t \in \mathcal{X}$, is a first-order Markov process.

For some results, we restrict to **autoregressive model**: $X_{t+1} = aX_t + W_t$, $X_t \in \mathbb{Z}/\mathbb{R}$.

Channel model

Gilbert-Elliot channel (at the packet level). Transition matrix Q.

When $S_t = 1$ (Channel is ON)
channel output = channel input

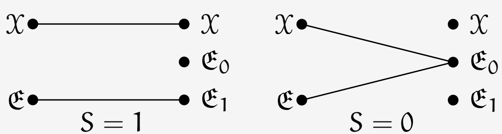
When $S_t = 0$ (Channel is OFF)
channel output = noise

Formal definition

Input alphabet $\bar{\mathcal{X}} = \mathcal{X} \cup \{\mathfrak{E}\}$ Output alphabet $\mathcal{Y} = \mathcal{X} \cup \{\mathfrak{E}_0, \mathfrak{E}_1\}$.

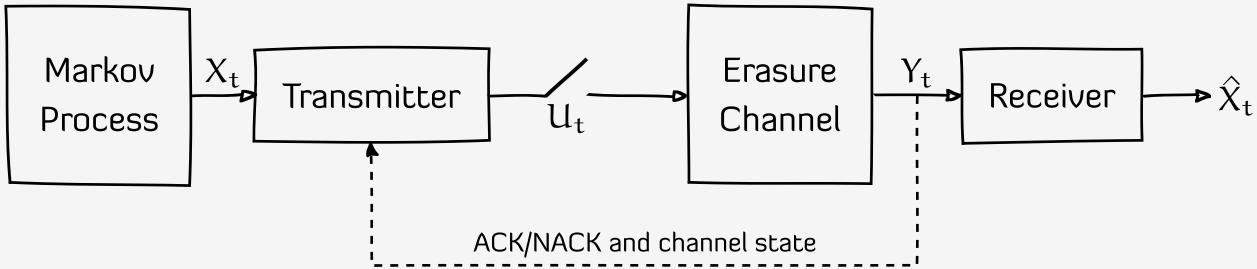
Channel input/output relationship

$$\mathbb{P}(Y_t \mid \bar{X}_{0:t}, S_{0:t}) = \mathbb{P}(Y_t \mid \bar{X}_t, S_t).$$



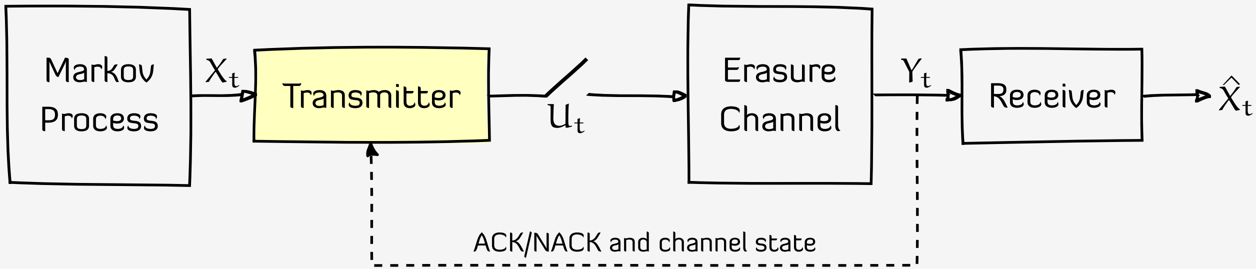
$$= \begin{cases} \mathfrak{E}_1, & \text{if } \bar{X}_t = \mathfrak{E} \text{ and } S_t = 1 \quad \text{(No received energy)} \\ \mathfrak{E}_0, & \text{if } S_t = 0 \quad \text{(Received energy)} \\ \bar{X}_t, & \text{if } \bar{X}_t \in \mathcal{X} \text{ and } S_t = 1 \quad \text{(Packet can be decoded)} \end{cases}$$

Communication system (cont.)



Feedback The receiver sends two bits of feedback: ACK/NACK and channel state.

Communication system (cont.)



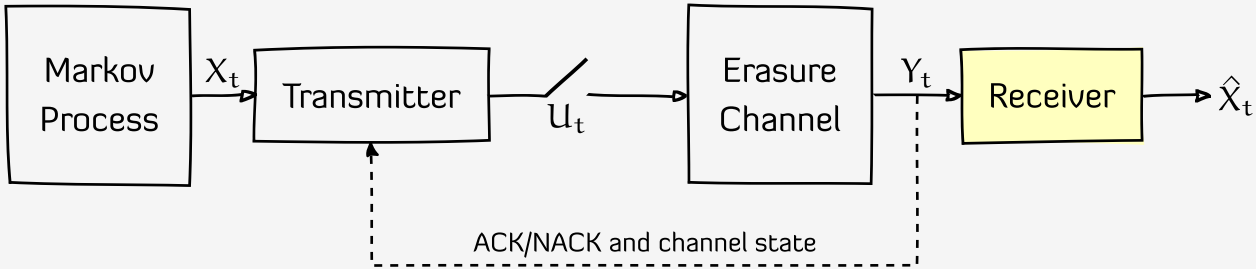
Feedback The receiver sends two bits of feedback: ACK/NACK and channel state.

Transmitter Decides whether to transmit or not. Denoted by $U_t \in \{0, 1\}$.

If $U_t = 0$, $\bar{X}_t = \mathfrak{E}$. If $U_t = 1$, $\bar{X}_t = X_t$.

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

Communication system (cont.)



Feedback The receiver sends two bits of feedback: ACK/NACK and channel state.

Transmitter Decides whether to transmit or not. Denoted by $U_t \in \{0, 1\}$.

If $U_t = 0$, $\bar{X}_t = \mathfrak{E}$. If $U_t = 1$, $\bar{X}_t = X_t$.

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

Receiver Chooses an estimate $\hat{X}_t \in \mathcal{X}$

$$\hat{X}_t = g_t(Y_{1:t}, S_{1:t})$$

An illustration

Transmitter	Receiver
$U_1 = f_1(X_1)$	

An illustration

Transmitter	Receiver
$U_1 = f_1(X_1)$	$\hat{X}_1 = g_1(Y_1, S_1)$

An illustration

Transmitter	Receiver
$U_1 = f_1(X_1)$	$\hat{X}_1 = g_1(Y_1, S_1)$
$U_2 = f_2(X_{1:2}, Y_1, S_1)$	

An illustration

Transmitter	Receiver
$U_1 = f_1(X_1)$	$\hat{X}_1 = g_1(Y_1, S_1)$
$U_2 = f_2(X_{1:2}, Y_1, S_1)$	$\hat{X}_2 = g_2(Y_{1:2}, S_{1:2})$

An illustration

Transmitter	Receiver
$U_1 = f_1(X_1)$	$\hat{X}_1 = g_1(Y_1, S_1)$
$U_2 = f_2(X_{1:2}, Y_1, S_1)$	$\hat{X}_2 = g_2(Y_{1:2}, S_{1:2})$
$U_3 = f_3(X_{1:3}, Y_{1:2}, S_{1:2})$	

An illustration

Transmitter	Receiver
$U_1 = f_1(X_1)$	$\hat{X}_1 = g_1(Y_1, S_1)$
$U_2 = f_2(X_{1:2}, Y_1, S_1)$	$\hat{X}_2 = g_2(Y_{1:2}, S_{1:2})$
$U_3 = f_3(X_{1:3}, Y_{1:2}, S_{1:2})$	$\hat{X}_3 = g_3(Y_{1:3}, S_{1:3})$

An illustration

Transmitter	Receiver
$U_1 = f_1(X_1)$	$\hat{X}_1 = g_1(Y_1, S_1)$
$U_2 = f_2(X_{1:2}, Y_1, S_1)$	$\hat{X}_2 = g_2(Y_{1:2}, S_{1:2})$
$U_3 = f_3(X_{1:3}, Y_{1:2}, S_{1:2})$	$\hat{X}_3 = g_3(Y_{1:3}, S_{1:3})$
$\vdots$	$\vdots$

Performance metrics

Performance metrics   Distortion D and Number of transmissions N

Performance metrics

Performance metrics Distortion D and Number of transmissions N

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, $\beta = 1$

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

Optimization problems

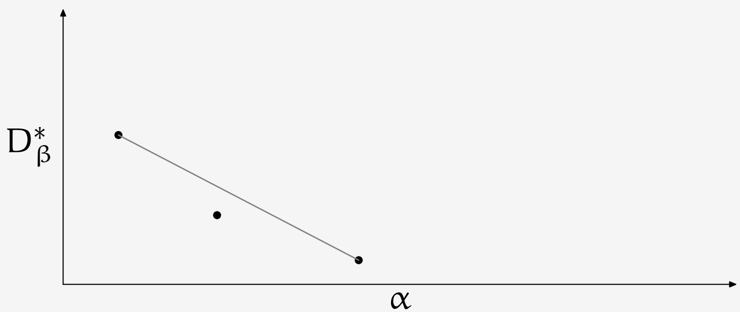
Constrained communication

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

Optimization problems

Constrained communication

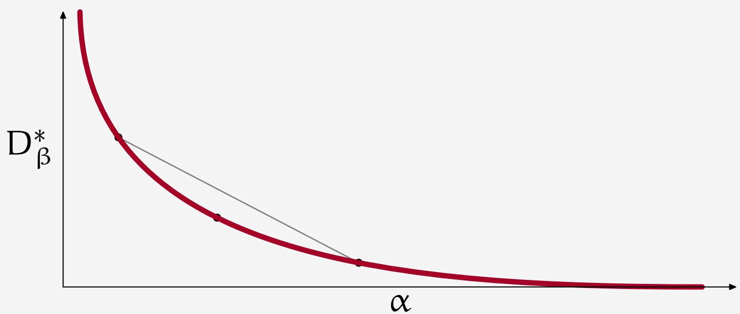
$$\text{For } \alpha \in (0,1), \quad D^*_\beta(\alpha) \coloneqq \inf_{(f,g)} \left\{ D_\beta(f,g) : N_\beta(f,g) \leqslant \alpha \right\}$$



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

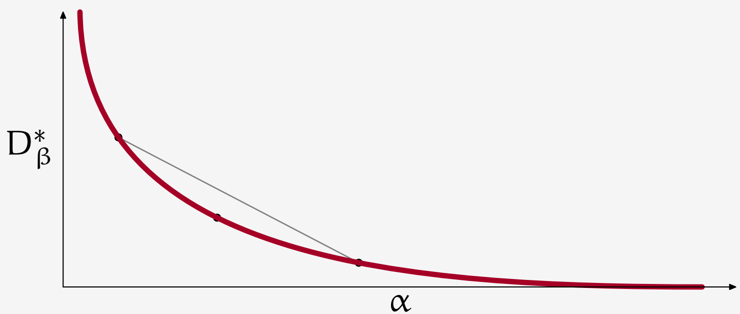
Optimization problems

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

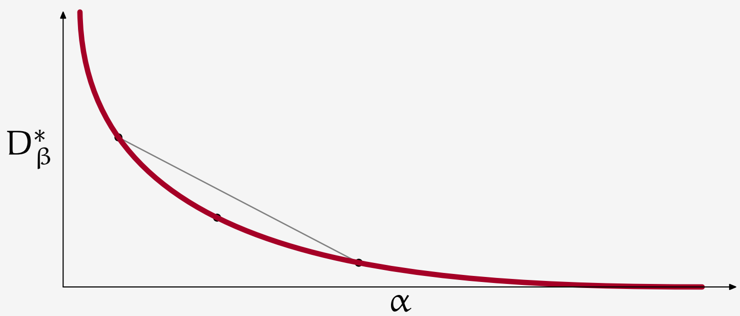
Optimization problems

Constrained communication

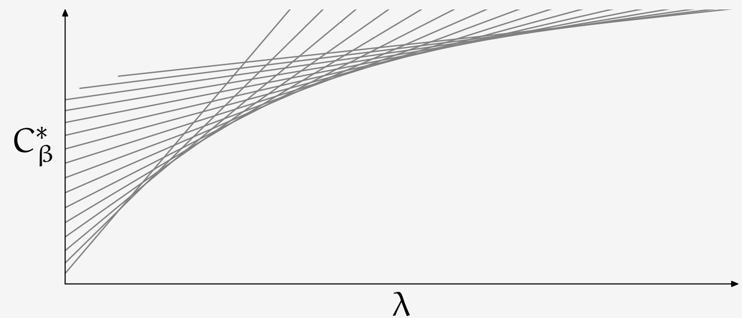
$$\text{For } \alpha \in (0,1), \quad D^*_\beta(\alpha) \coloneqq \inf_{(f,g)} \{ D_\beta(f,g) : N_\beta(f,g) \leqslant \alpha \}$$

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



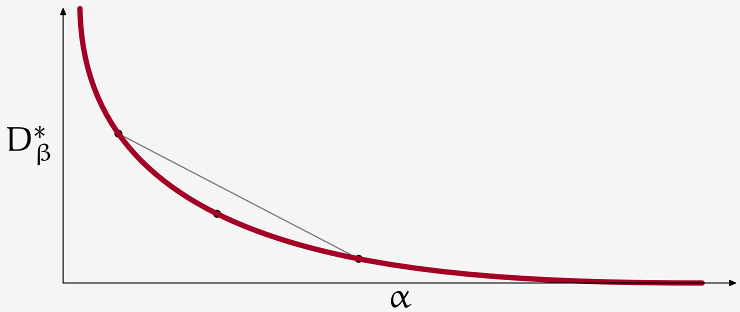
Optimization problems

Constrained communication

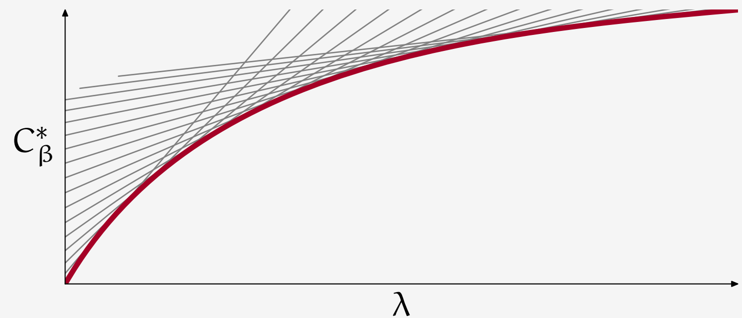
$$\text{For } \alpha \in (0,1), \quad D^*_\beta(\alpha) \coloneqq \inf_{(f,g)} \{ D_\beta(f,g) : N_\beta(f,g) \leqslant \alpha \}$$

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

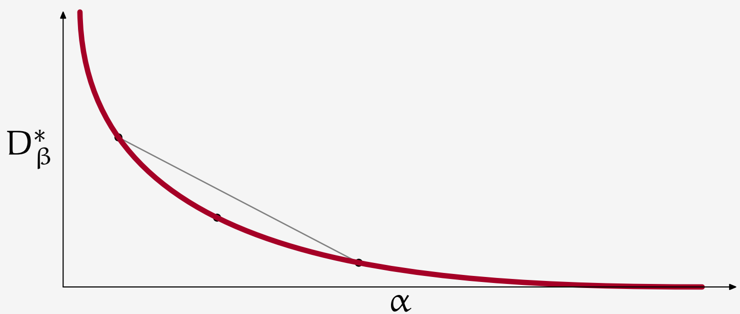
Optimization problems

Constrained communication

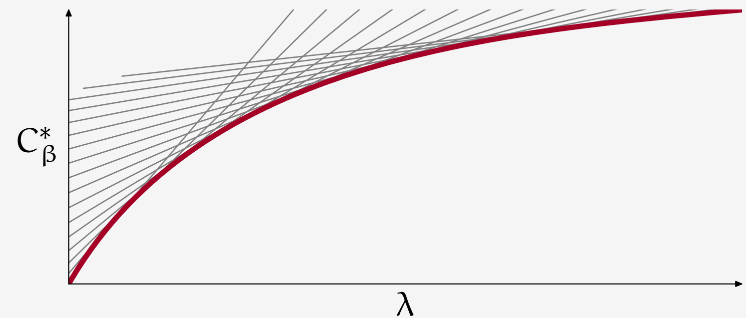
Our result: Provide computable expressions for these trade-offs and identify optimal strategies that achieve them.

Costly c

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



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)
- many others . . .

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

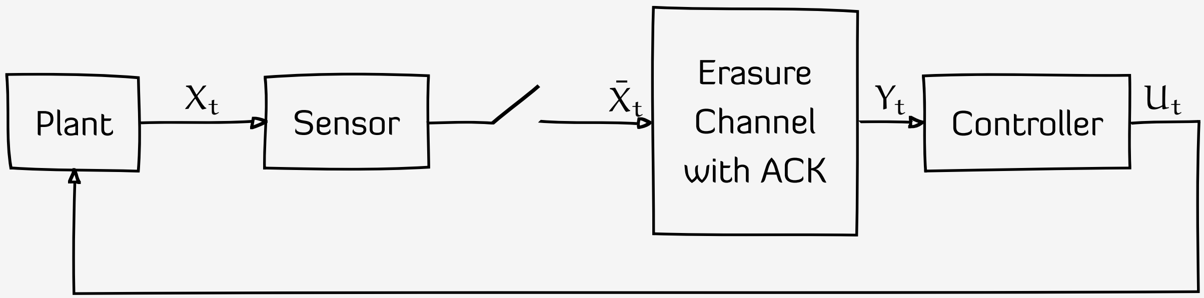
- [Marshak 1954] Static (one-shot) problem with arbitrary source distribution
- [Kushner 1964] Off-line choice of measurement times
- [Åstrom Bernhardsson 2002] Lebesque sampling (or event-based sampling)
- many others ...

Other related work

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

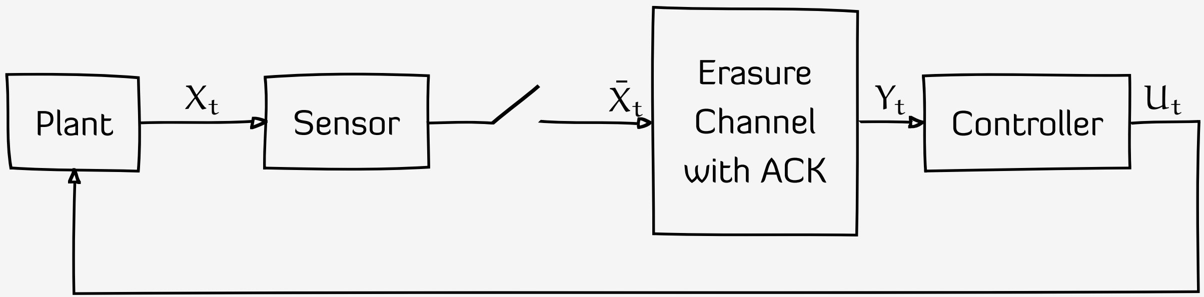
A networked control motivation

Networked control system



Model $\quad X_{t+1} = AX_t + BU_t + W_t, \quad \bar{X}_t \in \{X_t, \mathfrak{E}\}, \quad U_t = g_t(Y_{1:t}). \quad$ Min. quadratic cost

Networked control system



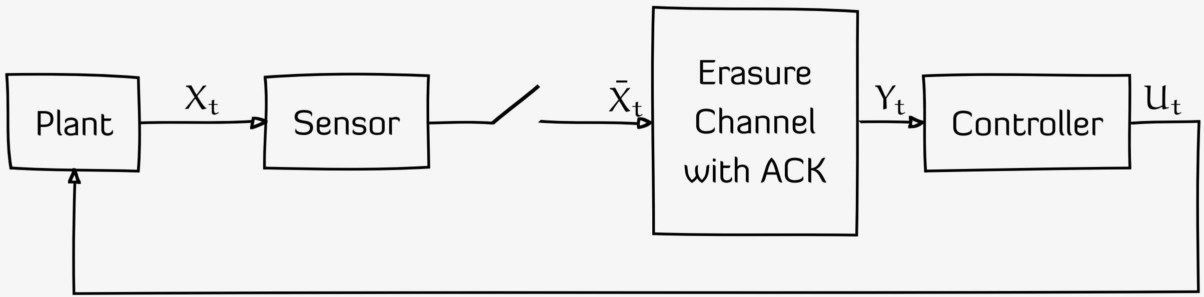
Model $\quad X_{t+1} = AX_t + BU_t + W_t, \quad \bar{X}_t \in \{X_t, \mathfrak{E}\}, \quad U_t = g_t(Y_{1:t}).$ Min. quadratic cost

Separation of estimation and control

- Consider the innovation process: $Z_t = X_t - \tilde{X}_t$, where $\tilde{X}_t = \sum_{s=0}^{t-1} A^{t-s-1} BU_s$
- There is no loss of optimality in deciding to transmit based on Z_t.
- Certainty equivalent controller is optimal: $U_t = K_t(\hat{Z}_t + \tilde{X}_t)$

- Yüksel, "Jointly Optimal LQG Quantization and Control Policies for Multi-Dimensional Systems," TAC 2014
- Rabi, Ramesh, and Johansson, "Separated design of encoder and controller for networked linear quadratic optimal control," SICON 2016

Networked control system



Model $\quad X_{t+1} = AX_t + BU_t + W_t, \quad \bar{X}_t \in \{X_t, \mathfrak{E}\}, \quad U_t = g_t(Y_{1:t}).$ Min. quadratic cost

Separation of estimation and control

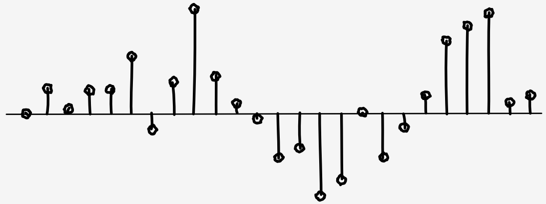
- Consider the innovation process: $Z_t = X_t - \tilde{X}_t$, where $\tilde{X}_t = \sum_{s=0}^{t-1} A^{t-s-1} BU_s$
- There is no loss of optimality in deciding to transmit based on Z_t.
- Certainty equivalent controller is optimal: $U_t = K_t(\hat{Z}_t + \tilde{X}_t)$
- Innovations do not depend on control $Z_{t+1} = AZ_t + W_t$

- Yüksel, "Jointly Optimal LQG Quantization and Control Policies for Multi-Dimensional Systems," TAC 2014
- Rabi, Ramesh, and Johansson, "Separated design of encoder and controller for networked linear quadratic optimal control," SICON 2016

Why bother?

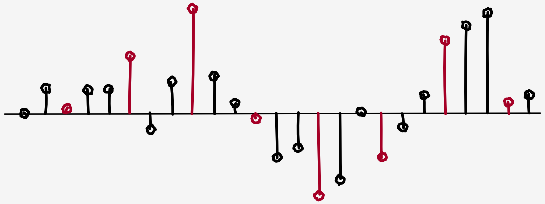
How much do we gain compared to simple strategies?

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



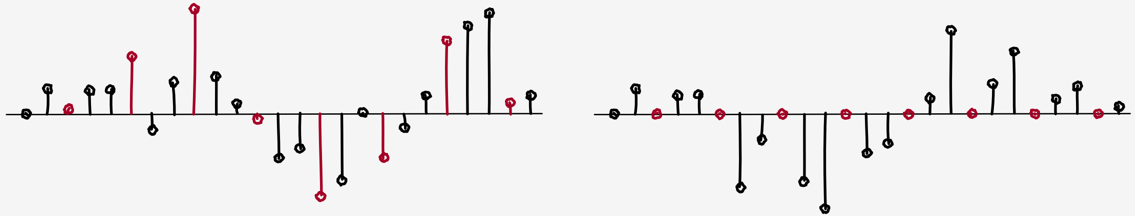
$X_{t+1} = X_t + W_t$, $W_t \sim \mathcal{N}(0,1)$. Perfect channel

Periodic Transmission Strategy

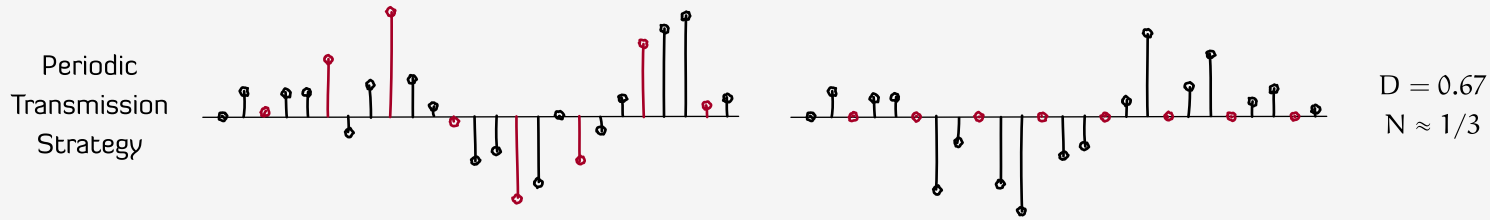


$X_{t+1} = X_t + W_t$, $W_t \sim \mathcal{N}(0,1)$. Perfect channel

Periodic Transmission Strategy

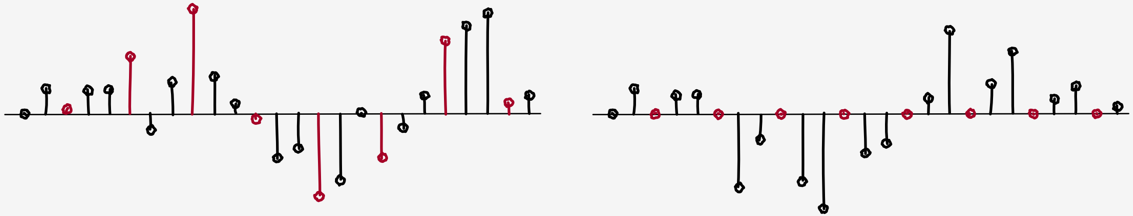


$X_{t+1} = X_t + W_t$, $W_t \sim \mathcal{N}(0,1)$. Perfect channel



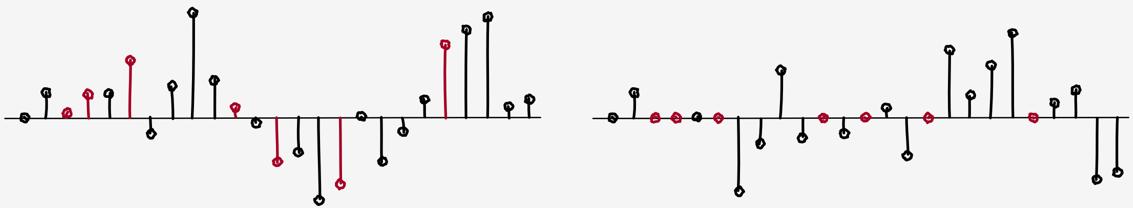
$X_{t+1} = X_t + W_t$, $W_t \sim \mathcal{N}(0, 1)$. Perfect channel

Periodic Transmission Strategy



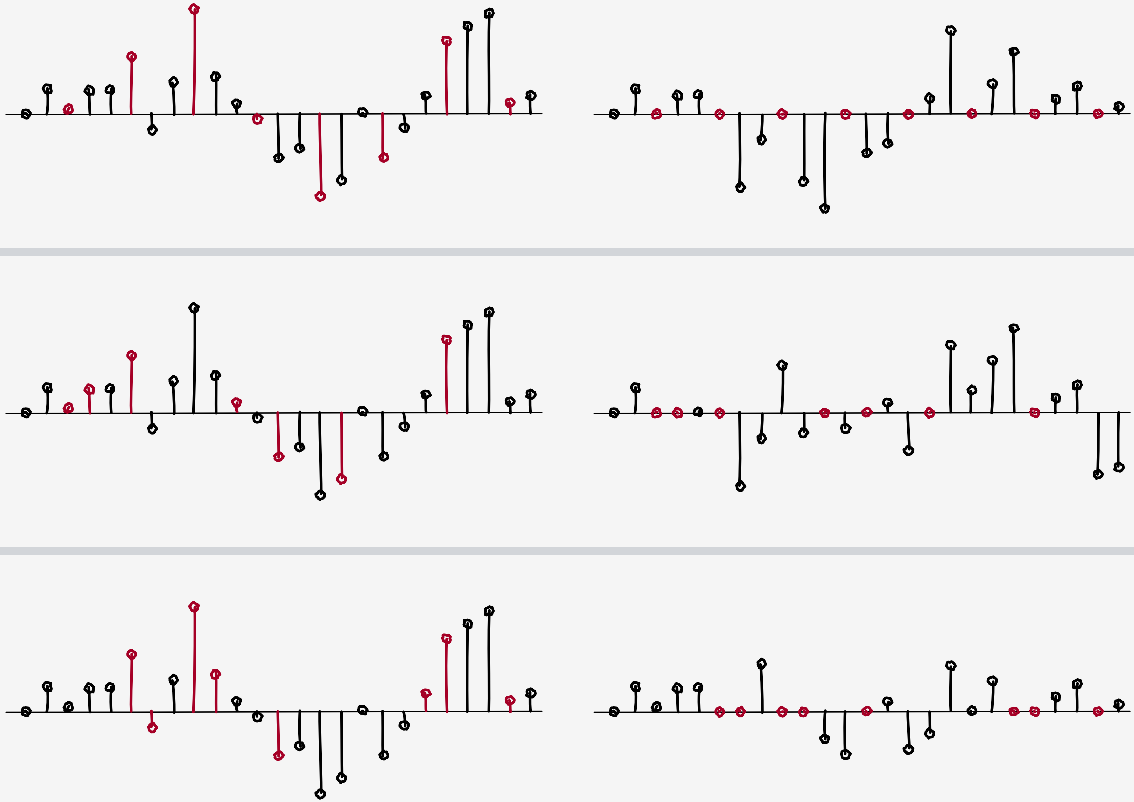
$D = 0.67$
$N \approx 1/3$

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

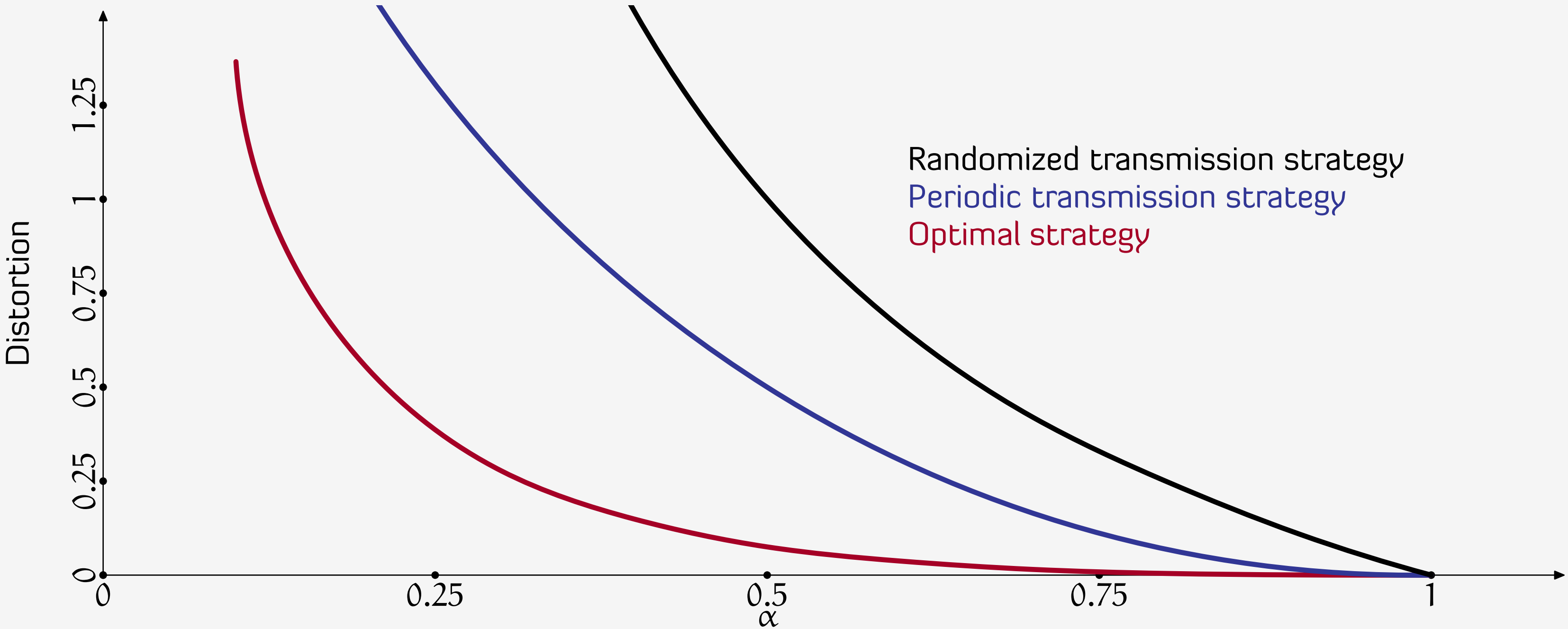


Periodic Transmission Strategy: $D = 0.67$, $N \approx 1/3$

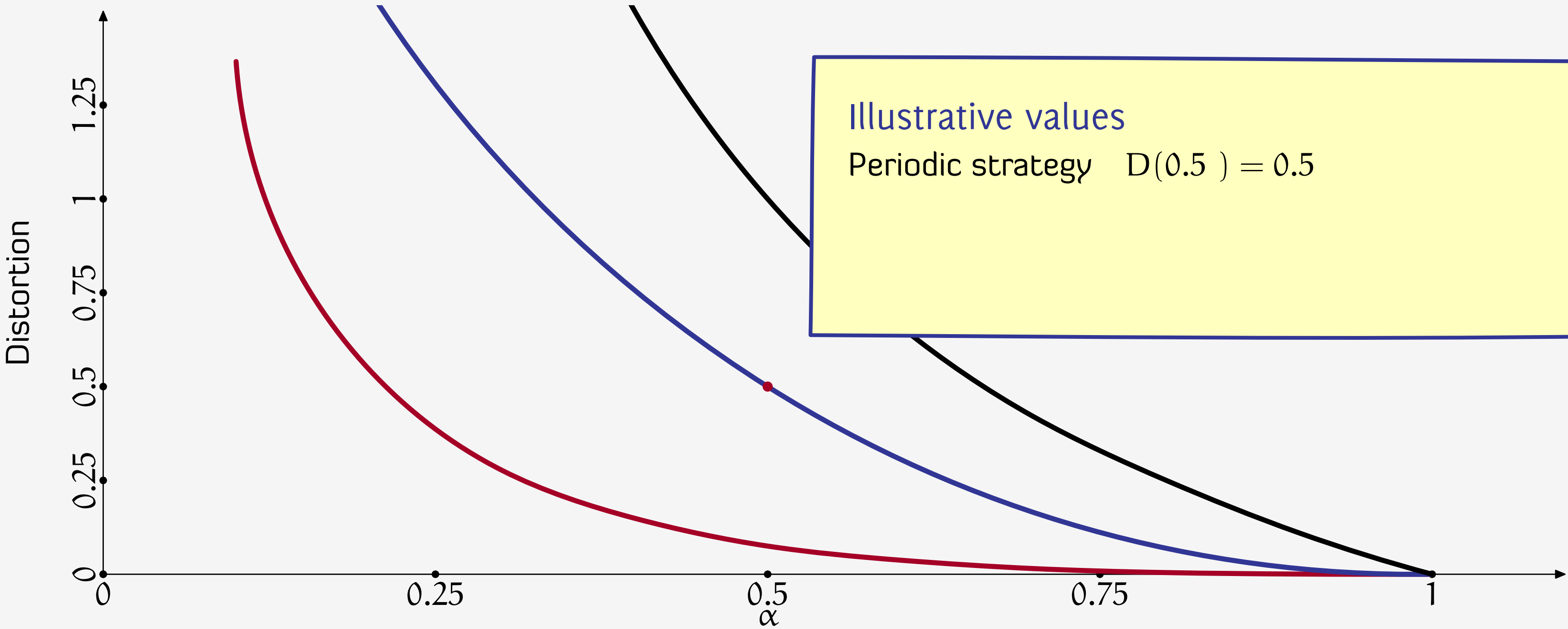
Randomized Transmission Strategy: $D = 2.00$, $N \approx 1/3$

Optimal Transmission Strategy: $D = 0.24$, $N \approx 1/3$

Distortion-transmission trade-off: Perfect channel



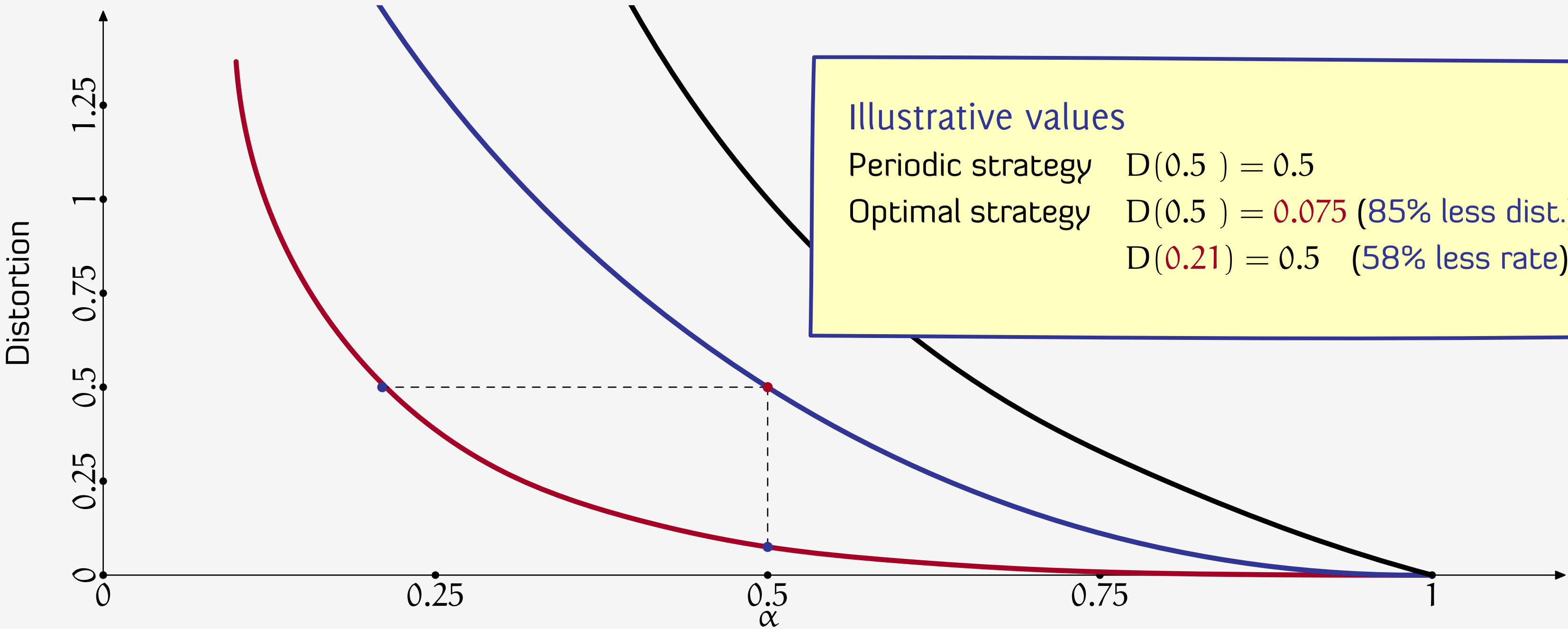
Distortion-transmission trade-off: Perfect channel



Illustrative values

Periodic strategy $D(0.5\ ) = 0.5$

Distortion-transmission trade-off: Perfect channel



Illustrative values

Periodic strategy $D(0.5) = 0.5$

Optimal strategy $D(0.5) = 0.075$ (85% less dist.)

$D(0.21) = 0.5$ (58% less rate)

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

$\mathcal{S} \subset \mathcal{X}$ 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 $x \in \mathcal{S}$	Cost when $x \notin \mathcal{S}$
$d(x - \hat{x})$	$\lambda + \varepsilon d(x - \hat{x})$

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

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

$\hat{x}_2^1$ 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

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

Optimal strategies and their performance

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 $d(x, \hat{x}) = d(x - \hat{x})$ where $d(\cdot)$ is symmetric and quasi-convex.

Optimal strategies and their performance

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 $d(x, \hat{x}) = d(x - \hat{x})$ where $d(\cdot)$ is symmetric and quasi-convex.

Optimal estimation strategy

$$\hat{X}_t = \begin{cases} a\hat{X}_{t-1}, & \text{if } Y_t \in \{\mathfrak{E}_0, \mathfrak{E}_1\} \\ Y_t, & \text{if } Y_t \in \mathcal{X} \end{cases}$$

Optimal strategies and their performance

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 $d(x, \hat{x}) = d(x - \hat{x})$ where $d(\cdot)$ is symmetric and quasi-convex.

Optimal transmission strategy

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

Optimal estimation strategy

$$\hat{X}_t = \begin{cases} a\hat{X}_{t-1}, & \text{if } Y_t \in \{\mathfrak{E}_0, \mathfrak{E}_1\} \\ Y_t, & \text{if } Y_t \in \mathcal{X} \end{cases}$$

Optimal strategies and their performance

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 $d(x, \hat{x}) = d(x - \hat{x})$ where $d(\cdot)$ is symmetric and quasi-convex.

Optimal transmission strategy

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

Optimal estimation strategy

$$\hat{X}_t = \begin{cases} a\hat{X}_{t-1}, & \text{if } Y_t \in \{\mathfrak{E}_0, \mathfrak{E}_1\} \\ Y_t, & \text{if } Y_t \in \mathcal{X} \end{cases}$$

Salient features

- Optimal strategies are simple and intuitive

Optimal strategies and their performance

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 $d(x, \hat{x}) = d(x - \hat{x})$ where $d(\cdot)$ is symmetric and quasi-convex.

Optimal transmission strategy

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

Optimal estimation strategy

$$\hat{X}_t = \begin{cases} a\hat{X}_{t-1}, & \text{if } Y_t \in \{\mathfrak{E}_0, \mathfrak{E}_1\} \\ Y_t, & \text{if } Y_t \in \mathcal{X} \end{cases}$$

Salient features

- Optimal strategies are simple and intuitive
- The transmitter does not try to send information through timing events (or length of silence intervals).
- The estimation strategy does not depend on the value of the threshold
- When the estimator does not receive a packet, it behaves as if the packet was dropped by the channel, **even when the channel is perfect!**

Optimal strategies and their performance

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 $d(x, \hat{x}) = d(x - \hat{x})$ where $d(\cdot)$ is symmetric and quasi-convex.

Optimal transmission strategy

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

Optimal estimation strategy

$$\hat{X}_t = \begin{cases} a\hat{X}_{t-1}, & \text{if } Y_t \in \{\mathfrak{E}_0, \mathfrak{E}_1\} \\ Y_t, & \text{if } Y_t \in \mathcal{X} \end{cases}$$

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.

Optimal strategies and their performance

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 $d(x, \hat{x}) = d(x - \hat{x})$ where $d(\cdot)$ is symmetric and quasi-convex.

Optimal transmission strategy

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

Optimal estimation strategy

$$\hat{X}_t = \begin{cases} a\hat{X}_{t-1}, & \text{if } Y_t \in \{\mathfrak{E}_0, \mathfrak{E}_1\} \\ Y_t, & \text{if } Y_t \in \mathcal{X} \end{cases}$$

Performance of threshold based strategies

- $K^{(k)}_\beta$: Expected discounted number of transmissions until first successful reception.
- $L^{(k)}_\beta$: Expected discounted distortion until first successful reception.
- $M^{(k)}_\beta$: Expected discounted time until first successful reception.

Then, $D^{(k)}_\beta = \dfrac{L^{(k)}_\beta}{M^{(k)}_\beta}$ and $N^{(k)}_\beta = \dfrac{K^{(k)}_\beta}{M^{(k)}_\beta}$. **(Renewal Relationships)**

Optimal trade-offs for discrete sources

Assume i.i.d. packet drops (i.e., transition matrix $\begin{bmatrix} \varepsilon & 1-\varepsilon \\ \varepsilon & 1-\varepsilon \end{bmatrix}$).

Optimal trade-offs for discrete sources

Assume i.i.d. packet drops (i.e., transition matrix $\begin{bmatrix} \varepsilon & 1-\varepsilon \\ \varepsilon & 1-\varepsilon \end{bmatrix}$).

- For i.i.d. packet drops $k^*(0) = k^*(1)$.
- For every $k \in \mathbb{Z}_{>0}$, compute $D_\beta^{(k)}$ and $N_\beta^{(k)}$. Define $\lambda_\beta^{(k)} = \big(D_\beta^{(k+1)} - D_\beta^{(k)}\big) / \big(N_\beta^{(k)} - N_\beta^{(k+1)}\big)$.

Optimal trade-offs for discrete sources

Assume i.i.d. packet drops (i.e., transition matrix $\begin{bmatrix} \varepsilon & 1-\varepsilon \\ \varepsilon & 1-\varepsilon \end{bmatrix}$).

- For i.i.d. packet drops $k^*(0) = k^*(1)$.
- For every $k \in \mathbb{Z}_{>0}$, compute $D_\beta^{(k)}$ and $N_\beta^{(k)}$. Define $\lambda_\beta^{(k)} = \big(D_\beta^{(k+1)} - D_\beta^{(k)}\big) / \big(N_\beta^{(k)} - N_\beta^{(k+1)}\big)$.

Costly communication

$$C_\beta^*(\lambda) \coloneqq \inf_{(f,g)} \big\{ D_\beta(f,g) + \lambda N_\beta(f,g) \big\}$$

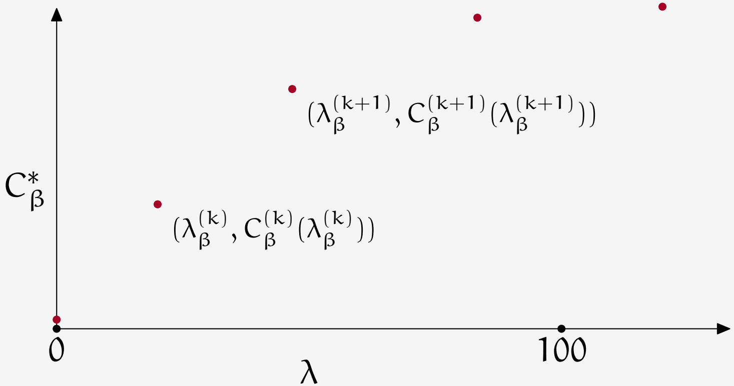
Optimal trade-offs for discrete sources

Assume i.i.d. packet drops (i.e., transition matrix $\begin{bmatrix} \varepsilon & 1-\varepsilon \\ \varepsilon & 1-\varepsilon \end{bmatrix}$).

- For i.i.d. packet drops $k^*(0) = k^*(1)$.
- For every $k \in \mathbb{Z}_{>0}$, compute $D_\beta^{(k)}$ and $N_\beta^{(k)}$. Define $\lambda_\beta^{(k)} = (D_\beta^{(k+1)} - D_\beta^{(k)})/(N_\beta^{(k)} - N_\beta^{(k+1)})$.

Costly communication

$$C^*_\beta(\lambda) \coloneqq \inf_{(f,g)} \{D_\beta(f,g) + \lambda N_\beta(f,g)\}$$



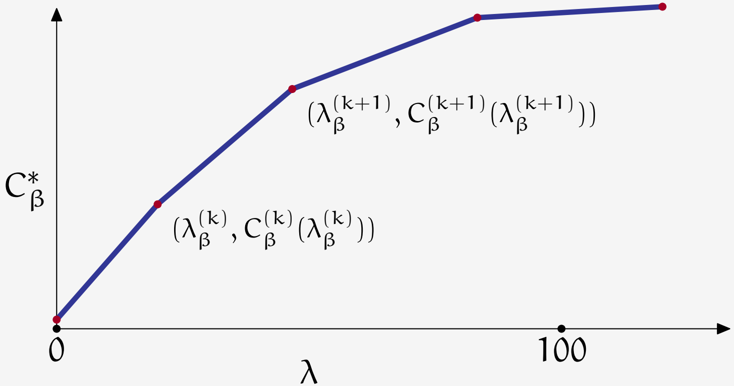
Optimal trade-offs for discrete sources

Assume i.i.d. packet drops (i.e., transition matrix $\begin{bmatrix} \varepsilon & 1-\varepsilon \\ \varepsilon & 1-\varepsilon \end{bmatrix}$).

- For i.i.d. packet drops $k^*(0) = k^*(1)$.
- For every $k \in \mathbb{Z}_{>0}$, compute $D_\beta^{(k)}$ and $N_\beta^{(k)}$. Define $\lambda_\beta^{(k)} = (D_\beta^{(k+1)} - D_\beta^{(k)})/(N_\beta^{(k)} - N_\beta^{(k+1)})$.

Costly communication

$$C^*_\beta(\lambda) \coloneqq \inf_{(f,g)} \{D_\beta(f,g) + \lambda N_\beta(f,g)\}$$



Optimal trade-offs for discrete sources

Assume i.i.d. packet drops (i.e., transition matrix $\begin{bmatrix} \varepsilon & 1-\varepsilon \\ \varepsilon & 1-\varepsilon \end{bmatrix}$).

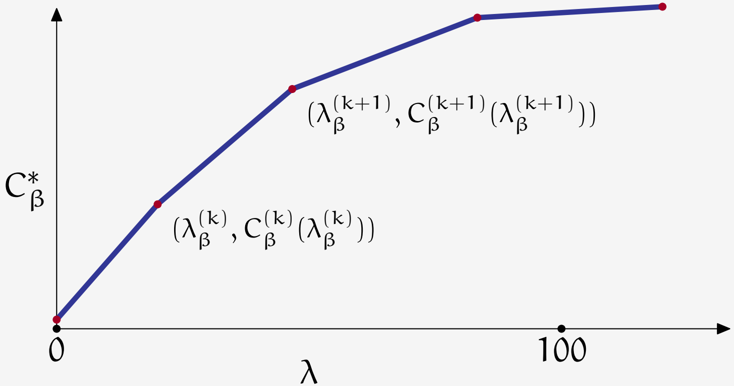
- For i.i.d. packet drops $k^*(0) = k^*(1)$.
- For every $k \in \mathbb{Z}_{>0}$, compute $D_\beta^{(k)}$ and $N_\beta^{(k)}$. Define $\lambda_\beta^{(k)} = \big(D_\beta^{(k+1)} - D_\beta^{(k)}\big)/\big(N_\beta^{(k)} - N_\beta^{(k+1)}\big)$.

Costly communication

$$C^*_\beta(\lambda) \coloneqq \inf_{(f,g)} \{D_\beta(f,g) + \lambda N_\beta(f,g)\}$$

Constrained communication

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



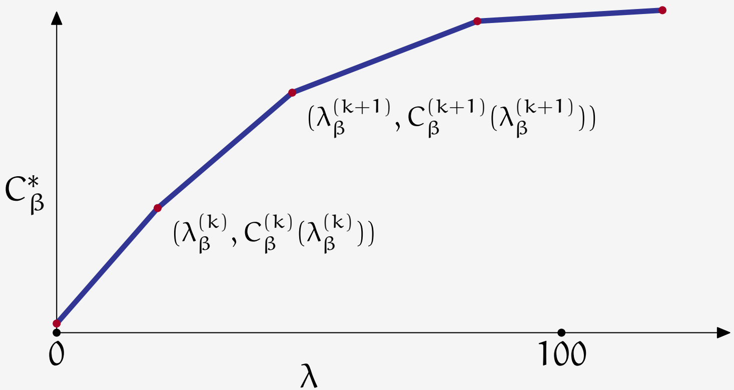
Optimal trade-offs for discrete sources

Assume i.i.d. packet drops (i.e., transition matrix $\begin{bmatrix} \varepsilon & 1-\varepsilon \\ \varepsilon & 1-\varepsilon \end{bmatrix}$).

- For i.i.d. packet drops $k^*(0) = k^*(1)$.
- For every $k \in \mathbb{Z}_{>0}$, compute $D_\beta^{(k)}$ and $N_\beta^{(k)}$. Define $\lambda_\beta^{(k)} = \big(D_\beta^{(k+1)} - D_\beta^{(k)}\big)/\big(N_\beta^{(k)} - N_\beta^{(k+1)}\big)$.

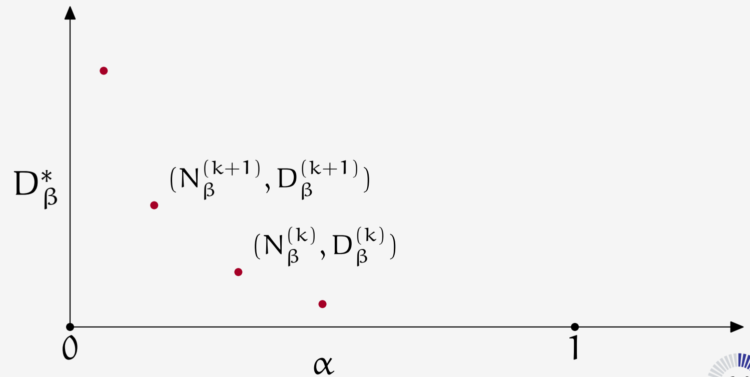
Costly communication

$$C^*_\beta(\lambda) \coloneqq \inf_{(f,g)} \big\{D_\beta(f,g) + \lambda N_\beta(f,g)\big\}$$



Constrained communication

$$D^*_\beta(\alpha) \coloneqq \inf_{(f,g)} \big\{D_\beta(f,g) : N_\beta(f,g) \leqslant \alpha\big\}$$



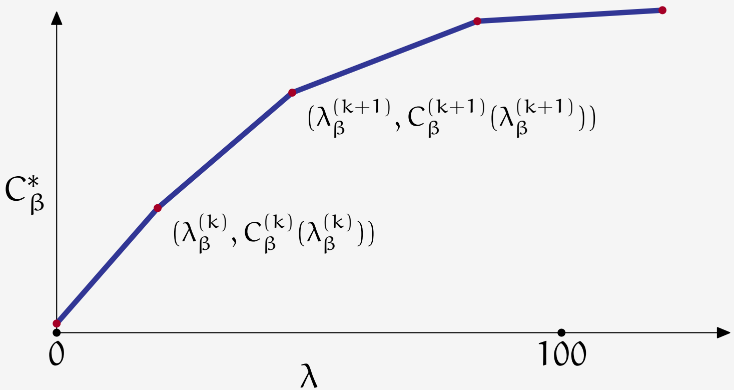
Optimal trade-offs for discrete sources

Assume i.i.d. packet drops (i.e., transition matrix $\begin{bmatrix} \varepsilon & 1-\varepsilon \\ \varepsilon & 1-\varepsilon \end{bmatrix}$).

- For i.i.d. packet drops $k^*(0) = k^*(1)$.
- For every $k \in \mathbb{Z}_{>0}$, compute $D_\beta^{(k)}$ and $N_\beta^{(k)}$. Define $\lambda_\beta^{(k)} = \big(D_\beta^{(k+1)} - D_\beta^{(k)}\big)/\big(N_\beta^{(k)} - N_\beta^{(k+1)}\big)$.

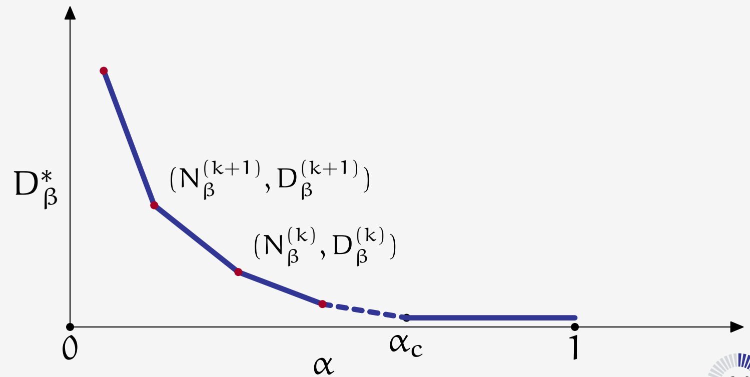
Costly communication

$$C^*_\beta(\lambda) \coloneqq \inf_{(f,g)} \{D_\beta(f,g) + \lambda N_\beta(f,g)\}$$



Constrained communication

$$D^*_\beta(\alpha) \coloneqq \inf_{(f,g)} \{D_\beta(f,g) : N_\beta(f,g) \leqslant \alpha\}$$



Optimal trade-offs for continuous sources

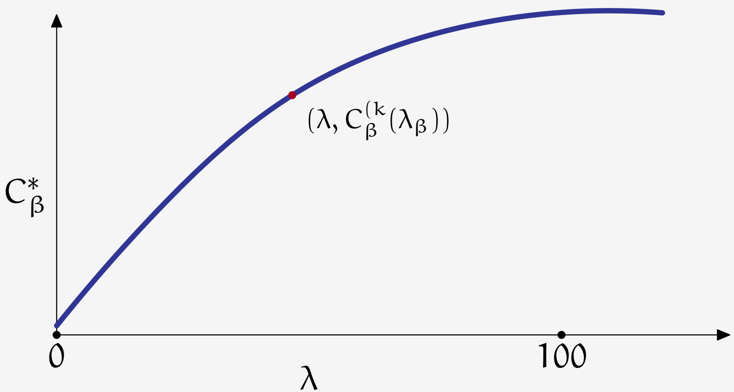
Again assume i.i.d. packet drops.

Optimal trade-offs for continuous sources

Again assume i.i.d. packet drops.

Costly communication

$$C^*_\beta(\lambda) \coloneqq \inf_{(f,g)} \left\{ D_\beta(f,g) + \lambda N_\beta(f,g) \right\}$$



where (λ, k) satisfy $\partial_k C_\beta^{(k)}(\lambda) = 0$

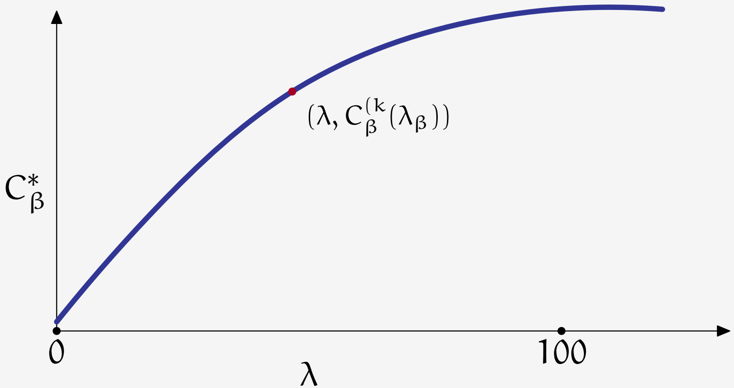
or, equivalently, $\lambda = -\dfrac{\partial_k D_\beta^{(k)}}{\partial_k N_\beta^{(k)}}$

Optimal trade-offs for continuous sources

Again assume i.i.d. packet drops.

Costly communication

$$C^*_\beta(\lambda) \coloneqq \inf_{(f,g)} \left\{ D_\beta(f,g) + \lambda N_\beta(f,g) \right\}$$

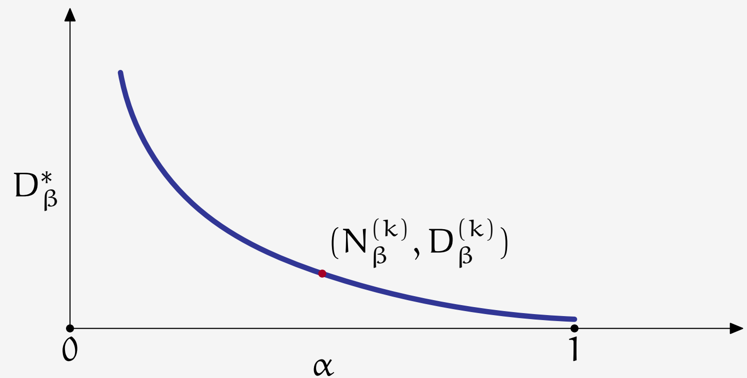


where (λ, k) satisfy $\partial_k C^{(k)}_\beta(\lambda) = 0$

or, equivalently, $\lambda = -\dfrac{\partial_k D^{(k)}_\beta}{\partial_k N^{(k)}_\beta}$

Constrained communication

$$D^*_\beta(\alpha) \coloneqq \inf_{(f,g)} \left\{ D_\beta(f,g) : N_\beta(f,g) \leqslant \alpha \right\}$$



Proof outline

How to prove the optimality of a coding scheme?

Information theory approach

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

How to prove the optimality of a coding scheme?

Information theory approach

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

How to prove the optimality of a coding scheme?

Information theory approach

- 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**.

How to prove the optimality of a coding scheme?

Information theory approach

- 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**.

Stochastic control approach

- Dynamic program: Identify sufficient statistics dynamic program
- Structural results: Determine qualitative properties of optimal solutions

How to prove the optimality of a coding scheme?

Information theory approach

- 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**.

Stochastic control approach

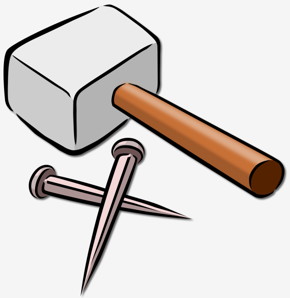
- Dynamic program: Identify sufficient statistics dynamic program
- Structural results: Determine qualitative properties of optimal solutions

> Structural results are hard! Especially for **multi-agent systems**.

- Related results (real-time comm.): [Witsenhausen 1979, Walrand-Varaiya 1983, Teneketzis 2006, Mahajan-Teneketzis 2009, Kaspi-Merhav 2012, Asnani-Weissman 2013, Yüksel 2013 . . .]

So how do we start?

Decentralized stochastic control

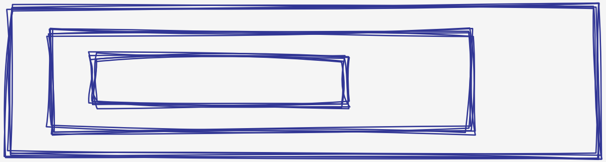


Dealing with non-classical information structure

- **Structure of optimal strategies**
  Instead of $f(\text{history of obs})$ use $f(\text{info state})$.

- **Compute optimal strategy using DP**
  $V(\text{info state}) = \min_{\text{action}} \left[\mathcal{B}_{\text{action}} V\right](\text{info state})$

Dealing with non-classical information structure



Classical info. struct.

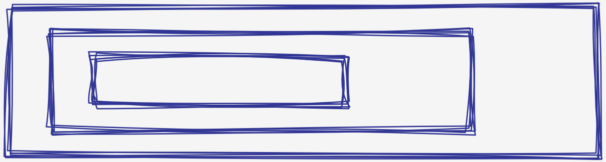
- Structure of optimal strategies

  Instead of $f(\text{history of obs})$ use $f(\text{info state})$.

- Compute optimal strategy using DP

  $$V(\text{info state}) = \min_{\text{action}} \left[\mathcal{B}_{\text{action}} V \right] (\text{info state})$$

Dealing with non-classical information structure



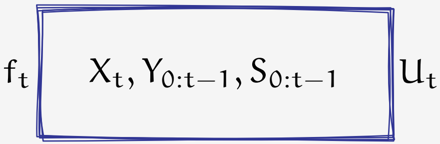
Non-Classical info. struct.

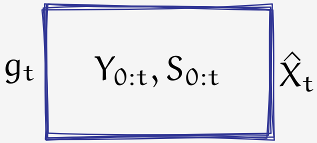
- **Structure of optimal strategies**

  Instead of f(history of obs) use f(info state).

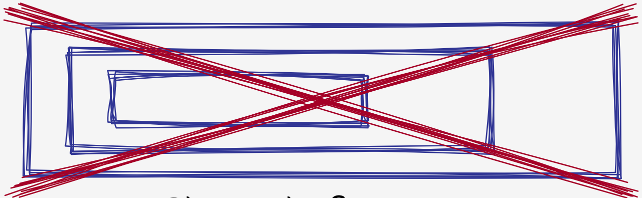
- **Compute optimal strategy using DP**

  $V(\text{info state}) = \min_{\text{action}} \left[\mathcal{B}_{\text{action}} V\right](\text{info state})$





Dealing with non-classical information structure



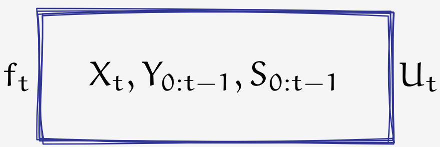
Non-Classical info. struct.

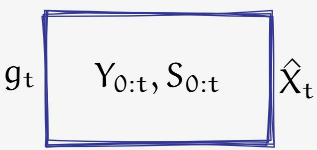
- **Structure of optimal strategies**

  Instead of f(history of obs) use f(info state).

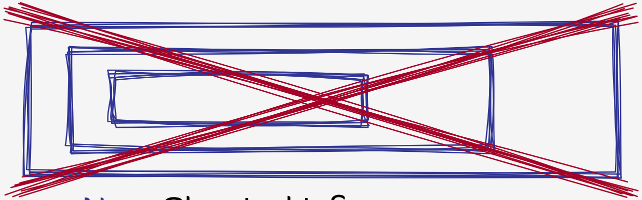
- **Compute optimal strategy using DP**

  $$V(\text{info state}) = \min_{\text{action}} \left[\mathcal{B}_{\text{action}} V\right](\text{info state})$$





Dealing with non-classical information structure



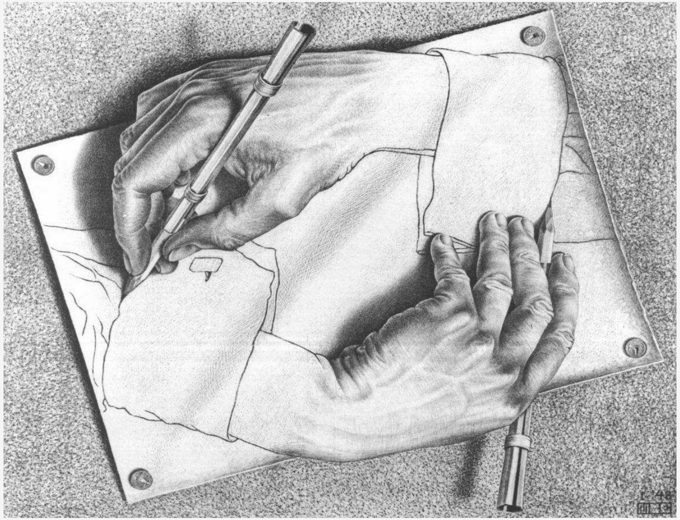
Non-Classical info. struct.

f_t $\boxed{X_t, Y_{0:t-1}, S_{0:t-1}}$ U_t

g_t $\boxed{Y_{0:t}, S_{0:t}}$ $\hat{X}_t$

- **Structure of optimal strategies**
  Instead of f(history of obs) use f(info state).

- **Compute optimal strategy using DP**
  $$V(\text{info state}) = \min_{\text{action}} \left[\mathcal{B}_{\text{action}} V\right](\text{info state})$$



The common information approach

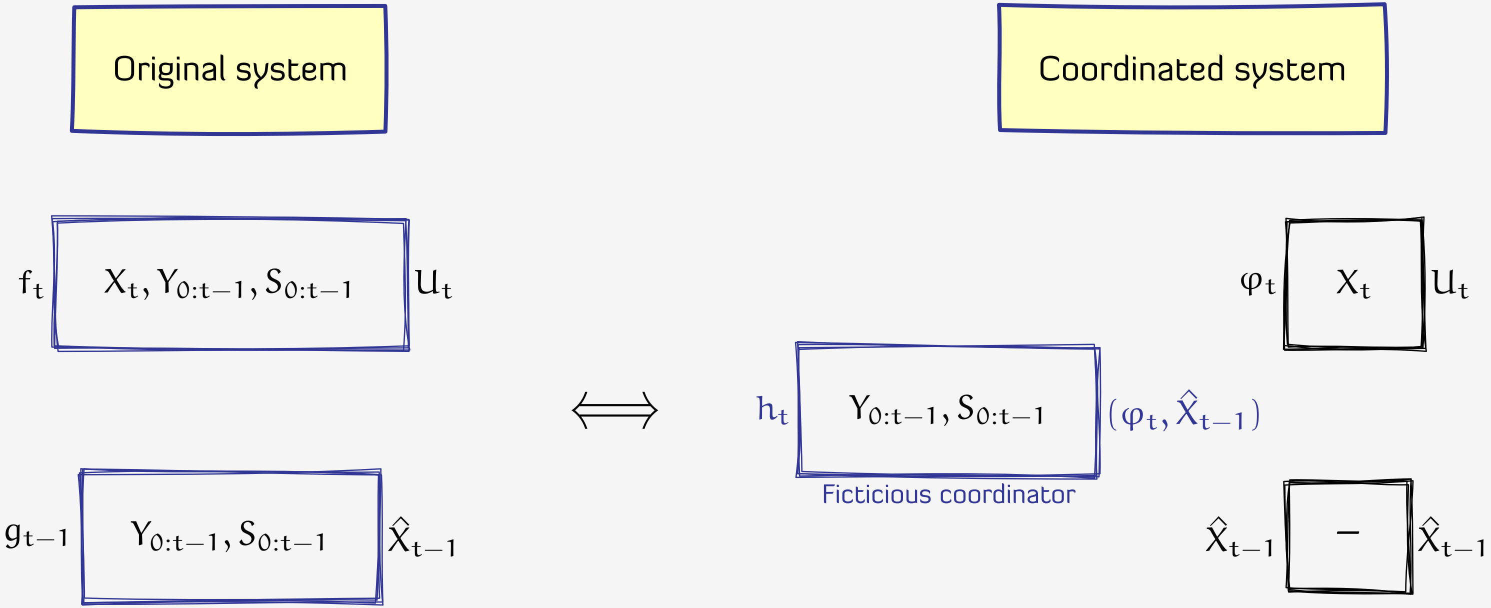
Original system

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

g_{t-1} $Y_{0:t-1}, S_{0:t-1}$ $\hat{X}_{t-1}$

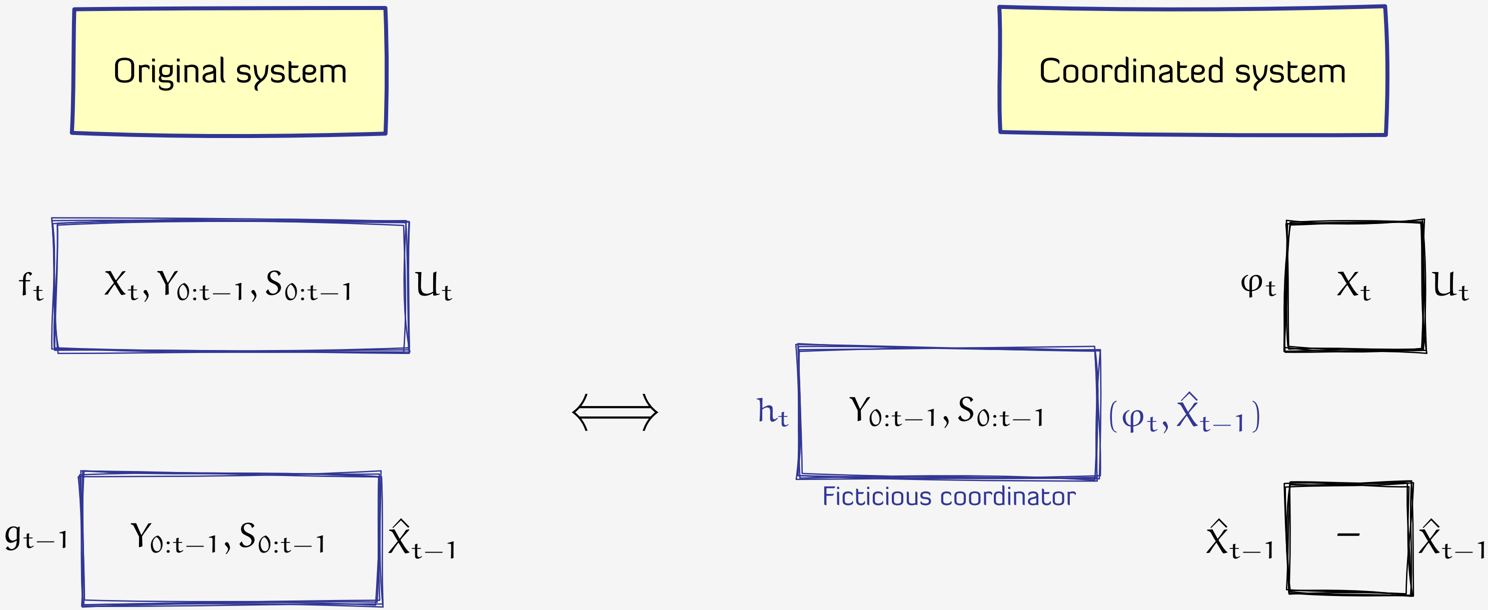
▷ Nayyar, Mahajan and Teneketzis, "Decentralized stochastic control with partial history sharing: A common information approach," IEEE TAC 2013.

The common information approach



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

The common information approach



- The coordinated system is equivalent to the original system.

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

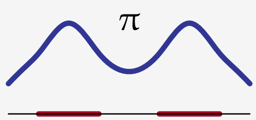
- The coordinated system is centralized. Belief state $\mathbb{P}(X_t \mid Y_{0:t-1}, S_{0:t-1})$.

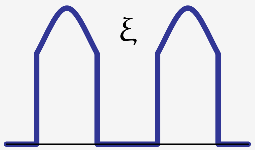
- Nayyar, Mahajan and Teneketzis, "Decentralized stochastic control with partial history sharing: A common information approach," IEEE TAC 2013.

Information states or sufficient statistics

Notation For any $\pi \in \Delta(\mathcal{X})$ and $\varphi \colon \mathcal{X} \to \{0, 1\}$

- $B_i(\varphi) = \{x \in \mathcal{X} : \varphi(x) = i\}, \; i \in \{0, 1\}$
- $\xi = \pi|_\varphi$ means $\xi(x) = \dfrac{\mathbb{1}_0\{\varphi(x)\}\pi(x)}{\pi(B_0(\varphi))}$

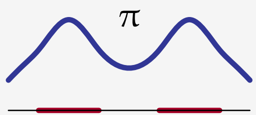


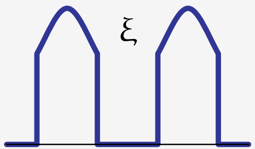


Information states or sufficient statistics

Notation For any $\pi \in \Delta(\mathcal{X})$ and $\varphi\colon \mathcal{X} \to \{0,1\}$

- $B_i(\varphi) = \{x \in \mathcal{X} : \varphi(x) = i\}, \quad i \in \{0,1\}$
- $\xi = \pi|_\varphi$ means $\xi(x) = \dfrac{\mathbb{1}_0\{\varphi(x)\}\pi(x)}{\pi(B_0(\varphi))}$





Pre-transmission belief $\pi^1_t(x) = \mathbb{P}(X_t = x | S_{0:t-1} = s_{0:t-1}, Y_{0:t-1} = y_{0:t-1}).$

Post-transmission belief $\pi^2_t(x) = \mathbb{P}(X_t = x | S_{0:t} = s_{0:t}, Y_{0:t} = y_{0:t}).$

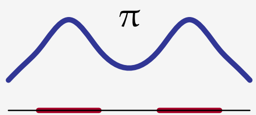
Information states or sufficient statistics

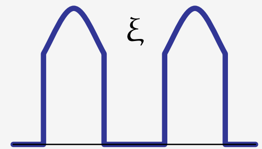
Notation

For any $\pi \in \Delta(\mathcal{X})$ and $\varphi \colon \mathcal{X} \to \{0, 1\}$

▷ $B_i(\varphi) = \{x \in \mathcal{X} : \varphi(x) = i\}, \ \ i \in \{0, 1\}$

▷ $\xi = \pi|_\varphi$ means $\xi(x) = \dfrac{\mathbb{1}_0\{\varphi(x)\}\pi(x)}{\pi(B_0(\varphi))}$





Pre-transmission belief

$$\pi_t^1(x) = \mathbb{P}(X_t = x | S_{0:t-1} = s_{0:t-1}, Y_{0:t-1} = y_{0:t-1}).$$

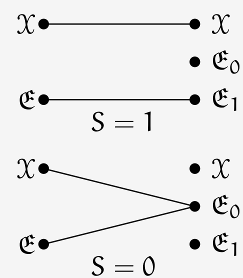
Post-transmission belief

$$\pi_t^2(x) = \mathbb{P}(X_t = x | S_{0:t} = s_{0:t}, Y_{0:t} = y_{0:t}).$$

Belief update

$$\pi_{t+1}^1 = \pi_t^2 P$$

$$\pi_t^2 = F^2(\pi_t^1, \varphi_t, y_t) = \begin{cases} \delta_{y_t}, & \text{if } y_t \in \mathcal{X} \\ \pi_t^1|_{\varphi_t}, & \text{if } y_t = \mathfrak{E}_1 \\ \pi_t^1, & \text{if } y_t = \mathfrak{E}_0 \end{cases}$$



Dynamic program

$$V^1_{T+1}(s, \pi^1) = 0$$

and for $t \in \{T, \dots, 0\}$

$$V^1_t(s, \pi^1) = \min_{\varphi : \mathcal{X} \to \{0,1\}} \Big\{ \lambda \pi^1(B_1(\varphi)) + \pi^1(B_0(\varphi)) W^0_t(\pi^1, \varphi) + \sum_{x \in B_1(\varphi)} \pi^1(x) W^1_t(\pi^1, \varphi, x) \Big\}$$

$$V^2_t(s, \pi^2) = \min_{\hat{x} \in \mathcal{X}} \sum_{x \in \mathcal{X}} \pi^2(x) d(x, \hat{x}) + V^1_{t+1}(s, \pi^2 P)$$

where $W^0_t(\pi^1, \varphi) = Q_{s0} V^2_t(0, \pi^1) + Q_{s1} V^2_t(1, \pi^1|_{\varphi})$

$$W^1_t(\pi^2, \varphi, x) = Q_{s0} V^2_t(0, \pi^1) + Q_{s1} V^2_t(1, \delta_x)$$

Dynamic program

$$V^1_{T+1}(s, \pi^1) = 0$$

and for $t \in \{T, \dots, 0\}$

$$V^1_t(s, \pi^1) = \min_{\varphi : \mathcal{X} \to \{0,1\}} \Big\{ \lambda \pi^1(B_1(\varphi)) + \pi^1(B_0(\varphi)) W^0_t(\pi^1, \varphi) + \sum_{x \in B_1(\varphi)} \pi^1(x) W^1_t(\pi^1, \varphi, x) \Big\}$$

$$V^2_t(s, \pi^2) = \min_{\hat{x} \in \mathcal{X}} \sum_{x \in \mathcal{X}} \pi^2(x) d(x, \hat{x}) + V^1_{t+1}(s, \pi^2 P)$$

where $W^0_t(\pi^1, \varphi) = Q_{s0} V^2_t(0, \pi^1) + Q_{s1} V^2_t(1, \pi^1|_\varphi)$

$$W^1_t(\pi^2, \varphi, x) = Q_{s0} V^2_t(0, \pi^1) + Q_{s1} V^2_t(1, \delta_x)$$

Dynamic program

$$V^1_{T+1}(s, \pi^1) = 0$$

and for $t \in \{T, \dots, 0\}$

$$V^1_t(s, \pi^1) = \min_{\varphi : \mathcal{X} \to \{0,1\}} \Big\{ \lambda \pi^1(B_1(\varphi)) + \pi^1(B_0(\varphi)) W^0_t(\pi^1, \varphi) + \sum_{x \in B_1(\varphi)} \pi^1(x) W^1_t(\pi^1, \varphi, x) \Big\}$$

$$V^2_t(s, \pi^2) = \min_{\hat{x} \in \mathcal{X}} \sum_{x \in \mathcal{X}} \pi^2(x) d(x, \hat{x}) + V^1_{t+1}(s, \pi^2 P)$$

where $W^0_t(\pi^1, \varphi) = Q_{s0} V^2_t(0, \pi^1) + Q_{s1} V^2_t(1, \pi^1|_\varphi)$

$$W^1_t(\pi^2, \varphi, x) = Q_{s0} V^2_t(0, \pi^1) + Q_{s1} V^2_t(1, \delta_x)$$

Dynamic program

$$V^1_{T+1}(s, \pi^1) = 0$$

and for $t \in \{T, \dots, 0\}$

$$V^1_t(s, \pi^1) = \min_{\varphi : \mathcal{X} \to \{0,1\}} \Big\{ \lambda \pi^1(B_1(\varphi)) + \pi^1(B_0(\varphi)) W^0_t(\pi^1, \varphi) + \sum_{x \in B_1(\varphi)} \pi^1(x) W^1_t(\pi^1, \varphi, x) \Big\}$$

$$V^2_t(s, \pi^2) = \min_{\hat{x} \in \mathcal{X}} \sum_{x \in \mathcal{X}} \pi^2(x) d(x, \hat{x}) + V^1_{t+1}(s, \pi^2 P)$$

where $W^0_t(\pi^1, \varphi) = Q_{s0} V^2_t(0, \pi^1) + Q_{s1} V^2_t(1, \pi^1|_{\varphi})$

$$W^1_t(\pi^2, \varphi, x) = Q_{s0} V^2_t(0, \pi^1) + Q_{s1} V^2_t(1, \delta_x)$$

Dynamic program

$$V^1_{T+1}(s, \pi^1) = 0$$

and for $t \in \{T, \dots, 0\}$

$$V^1_t(s, \pi^1) = \min_{\varphi : \mathcal{X} \to \{0,1\}} \Big\{ \lambda \pi^1(B_1(\varphi)) + \pi^1(B_0(\varphi)) W^0_t(\pi^1, \varphi) + \sum_{x \in B_1(\varphi)} \pi^1(x) W^1_t(\pi^1, \varphi, x) \Big\}$$

$$V^2_t(s, \pi^2) = \min_{\hat{x} \in \mathcal{X}} \sum_{x \in \mathcal{X}} \pi^2(x) d(x, \hat{x}) + V^1_{t+1}(s, \pi^2 P)$$

where $W^0_t(\pi^1, \varphi) = Q_{s0} V^2_t(0, \pi^1) + Q_{s1} V^2_t(1, \pi^1|_\varphi)$

$$W^1_t(\pi^2, \varphi, x) = Q_{s0} V^2_t(0, \pi^1) + Q_{s1} V^2_t(1, \delta_x)$$

Dynamic program

$$V^1_{T+1}(s, \pi^1) = 0$$

and for $t \in \{T, \dots, 0\}$

$$V^1_t(s, \pi^1) = \min_{\varphi : \mathcal{X} \to \{0,1\}} \Big\{ \lambda \pi^1(B_1(\varphi)) + \pi^1(B_0(\varphi)) W^0_t(\pi^1, \varphi) + \sum_{x \in B_1(\varphi)} \pi^1(x) W^1_t(\pi^1, \varphi, x) \Big\}$$

$$V^2_t(s, \pi^2) = \min_{\hat{x} \in \mathcal{X}} \sum_{x \in \mathcal{X}} \pi^2(x) d(x, \hat{x}) + V^1_{t+1}(s, \pi^2 P)$$

where $W^0_t(\pi^1, \varphi) = Q_{s0} V^2_t(0, \pi^1) + Q_{s1} V^2_t(1, \pi^1|_{\varphi})$

$$W^1_t(\pi^2, \varphi, x) = Q_{s0} V^2_t(0, \pi^1) + Q_{s1} V^2_t(1, \delta_x)$$

Salient features

- Minimization over functions φ
- Similar to DP for POMDPs. Can be solved using similar numerical techniques.

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

Simplifying modeling assumptions

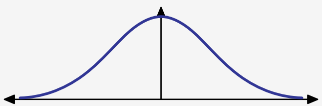
Markov process

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

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

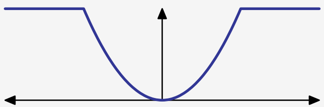
Noise Distribution

Unimodal and symmetric

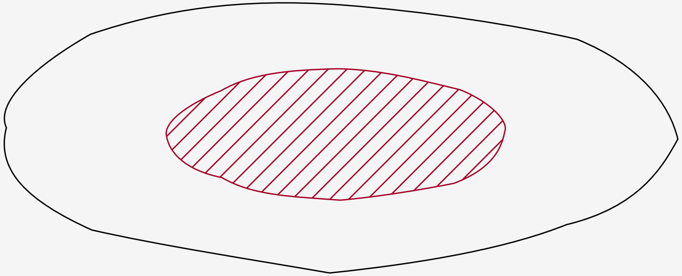


Distortion function

Symmetric and quasi-convex

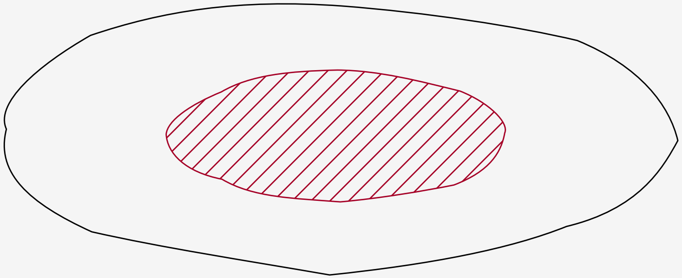


Step 1 Threshold strategies are optimal



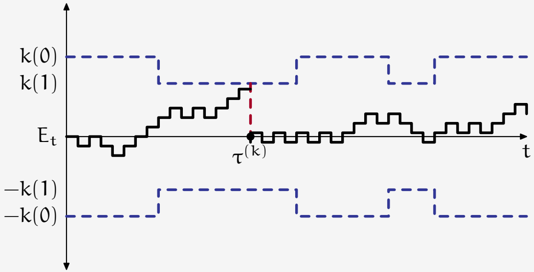
Search space of strategies (f, g)

Step 1 Threshold strategies are optimal

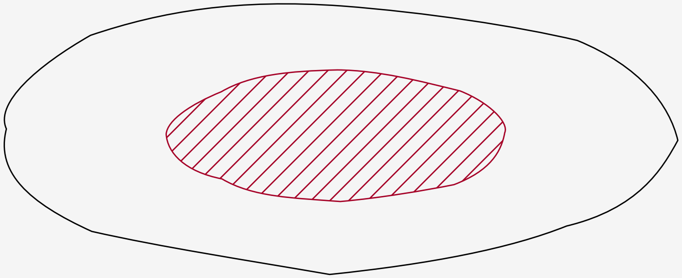


Search space of strategies (f, g)

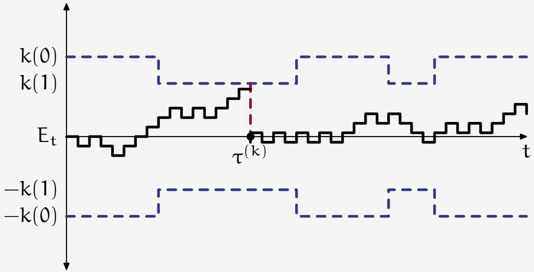
Step 2 Performance of threshold strategies



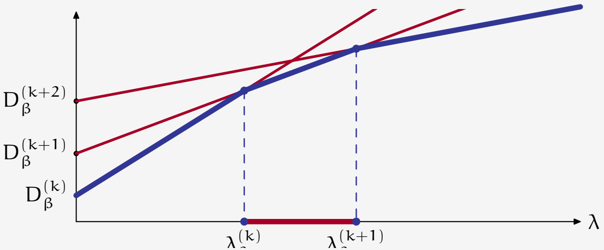
Step 1 Threshold strategies are optimal



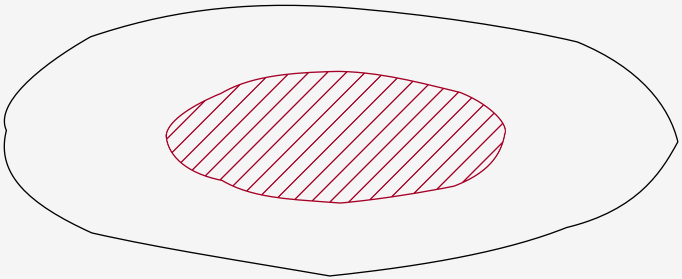
Step 2 Performance of threshold strategies



Step 3 Optimal costly communication

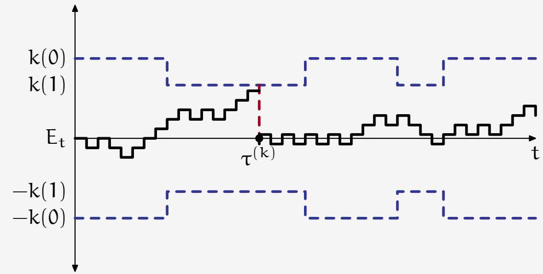


Step 1 Threshold strategies are optimal

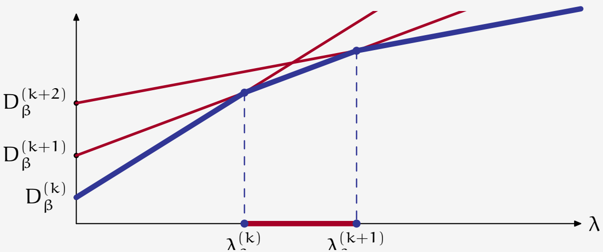


Search space of strategies (f, g)

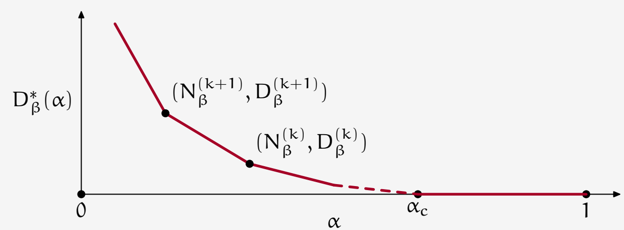
Step 2 Performance of threshold strategies



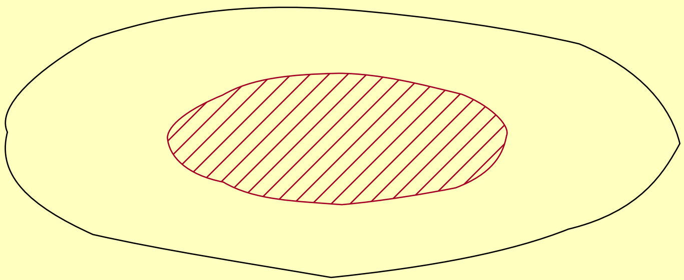
Step 3 Optimal costly communication



Step 4 Optimal constrained communication

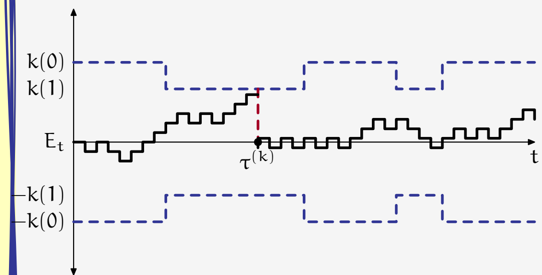


Step 1 Threshold strategies are optimal

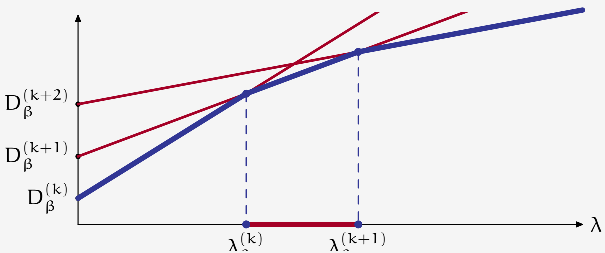


Search space of strategies (f, g)

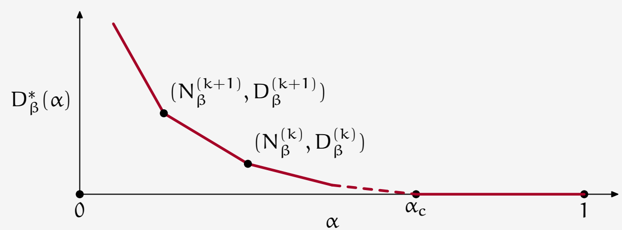
Step 2 Performance of threshold strategies



Step 3 Optimal costly communication



Step 4 Optimal constrained communication



Step 1 A change of variables

Define $$Z_0 = 0 \text{ and } Z_t = \begin{cases} aZ_{t-1}, & \text{if } Y_t \in \{\mathfrak{E}_0, \mathfrak{E}_1\} \\ Y_t, & \text{if } Y_t \in \mathcal{X} \end{cases}$$ (Observable at both Tx and Rx)

Step 1 A change of variables

Define $$Z_0 = 0 \text{ and } Z_t = \begin{cases} aZ_{t-1}, & \text{if } Y_t \in \{\mathfrak{E}_0, \mathfrak{E}_1\} \\ Y_t, & \text{if } Y_t \in \mathcal{X} \end{cases}$$ (Observable at both Tx and Rx)

$$E_t = X_t - aZ_{t-1}, \quad E_t^+ = X_t - Z_t, \quad \hat{E}_t = \hat{X}_t - Z_t$$

Step 1 A change of variables

Define $$Z_0 = 0 \text{ and } Z_t = \begin{cases} aZ_{t-1}, & \text{if } Y_t \in \{\mathfrak{E}_0, \mathfrak{E}_1\} \\ Y_t, & \text{if } Y_t \in \mathcal{X} \end{cases}$$ (Observable at both Tx and Rx)

$$E_t = X_t - aZ_{t-1}, \quad E_t^+ = X_t - Z_t, \quad \hat{E}_t = \hat{X}_t - Z_t$$

Thus, these are related as

$$E_t^+ = \begin{cases} E_t, & \text{if } Y_t \in \{\mathfrak{E}_0, \mathfrak{E}_1\} \\ 0, & \text{if } Y_t \in \mathcal{X} \end{cases} \qquad \text{and} \qquad E_{t+1} = aE_t^+ + W_t$$

Step 1 A change of variables

Define $$Z_0 = 0 \text{ and } Z_t = \begin{cases} aZ_{t-1}, & \text{if } Y_t \in \{\mathfrak{E}_0, \mathfrak{E}_1\} \\ Y_t, & \text{if } Y_t \in \mathcal{X} \end{cases}$$ (Observable at both Tx and Rx)

$$E_t = X_t - aZ_{t-1}, \quad E_t^+ = X_t - Z_t, \quad \hat{E}_t = \hat{X}_t - Z_t$$

Thus, these are related as

$$E_t^+ = \begin{cases} E_t, & \text{if } Y_t \in \{\mathfrak{E}_0, \mathfrak{E}_1\} \\ 0, & \text{if } Y_t \in \mathcal{X} \end{cases} \quad \text{and} \quad E_{t+1} = aE_t^+ + W_t$$

Note $X_t - \hat{X}_t = E_t^+ - \hat{E}_t$ and hence $d(X_t - \hat{X}_t) = d(E_t^+ - \hat{E}_t)$.

Implication of change of variables

Pre-transmission belief $\pi_t^1(e) = \mathbb{P}(E_t = e | S_{0:t-1} = s_{0:t-1}, Y_{0:t-1} = y_{0:t-1}).$

Post-transmission belief $\pi_t^2(e) = \mathbb{P}(E_t^+ = e | S_{0:t} = s_{0:t}, Y_{0:t} = y_{0:t}).$

Implication of change of variables

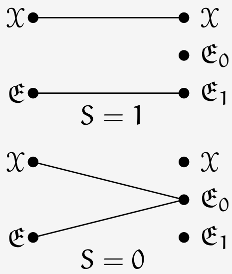
Pre-transmission belief $\pi^1_t(e) = \mathbb{P}(E_t = e | S_{0:t-1} = s_{0:t-1}, Y_{0:t-1} = y_{0:t-1})$.

Post-transmission belief $\pi^2_t(e) = \mathbb{P}(E^+_t = e | S_{0:t} = s_{0:t}, Y_{0:t} = y_{0:t})$.

Belief update

$$\pi^1_{t+1} = \pi^2_t P$$

$$\pi^2_t = F^2(\pi^1_t, \varphi_t) = \begin{cases} \delta_0, & \text{if } y_t \in \mathcal{X} \\ \pi^1_t|_{\varphi_t}, & \text{if } y_t = \mathfrak{E}_1 \\ \pi^1_t, & \text{if } y_t = \mathfrak{E}_0 \end{cases}$$



Dynamic program

$$V^1_{T+1}(s, \pi^1) = 0$$

and for $t \in \{T, \dots, 0\}$

$$V^1_t(s, \pi^1) = \min_{\varphi : \mathcal{X} \to \{0,1\}} \Big\{ \lambda \pi^1(B_1(\varphi)) + \pi^1(B_0(\varphi)) W^0_t(\pi^1, \varphi) + \sum_{x \in B_1(\varphi)} \pi^1(x) W^1_t(\pi^1, \varphi, x) \Big\}$$

$$V^2_t(s, \pi^2) = \min_{\hat{x} \in \mathcal{X}} \sum_{x \in \mathcal{X}} \pi^2(x) d(x, \hat{x}) + V^1_{t+1}(s, \pi^2 P)$$

where $W^0_t(\pi^1, \varphi) = Q_{s0} V^2_t(0, \pi^1) + Q_{s1} V^2_t(1, \pi^1|_\varphi)$

$$W^1_t(\pi^2, \varphi, x) = Q_{s0} V^2_t(0, \pi^1) + Q_{s1} V^2_t(1, \delta_x)$$

Dynamic program

$$V^1_{T+1}(s, \pi^1) = 0$$

and for $t \in \{T, \dots, 0\}$

$$\pi^1(B_1(\varphi))W^1_t(\pi^1, \varphi)$$

$$V^1_t(s, \pi^1) = \min_{\varphi : \mathcal{X} \to \{0,1\}} \Big\{ \lambda \pi^1(B_1(\varphi)) + \pi^1(B_0(\varphi))W^0_t(\pi^1, \varphi) + \sum_{x \in B_1(\varphi)} \pi^1(x) W^1_t(\pi^1, \varphi, x) \Big\}$$

$$V^2_t(s, \pi^2) = \min_{\hat{x} \in \mathcal{X}} \sum_{x \in \mathcal{X}} \pi^2(x) d(x, \hat{x}) + V^1_{t+1}(s, \pi^2 P)$$

where $W^0_t(\pi^1, \varphi) = Q_{s0} V^2_t(0, \pi^1) + Q_{s1} V^2_t(1, \pi^1|_\varphi)$

$$W^1_t(\pi^2, \varphi, x) = Q_{s0} V^2_t(0, \pi^1) + Q_{s1} V^2_t(1, \delta_x)$$

Dynamic program

$$V^1_{T+1}(s, \pi^1) = 0$$

and for $t \in \{T, \dots, 0\}$

$$\pi^1(B_1(\varphi))W^1_t(\pi^1, \varphi)$$

$$V^1_t(s, \pi^1) = \min_{\varphi : \mathcal{X} \to \{0,1\}} \Big\{ \lambda \pi^1(B_1(\varphi)) + \pi^1(B_0(\varphi)) W^0_t(\pi^1, \varphi) + \sum_{x \in B_1(\varphi)} \pi^1(x) W^1_t(\pi^1, \varphi, x) \Big\}$$

$$V^2_t(s, \pi^2) = \min_{\hat{x} \in \mathcal{X}} \sum_{x \in \mathcal{X}} \pi^2(x) d(x, \hat{x}) + V^1_{t+1}(s, \pi^2 P)$$

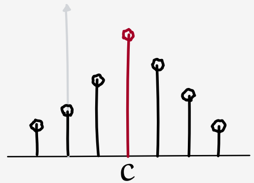
where $W^0_t(\pi^1, \varphi) = Q_{s0} V^2_t(0, \pi^1) + Q_{s1} V^2_t(1, \pi^1|_\varphi)$

$$W^1_t(\pi^2, \varphi, x) = Q_{s0} V^2_t(0, \pi^1) + Q_{s1} V^2_t(1, \delta_x) \quad Q_{s1} V^2_t(1, \delta_0)$$

Step 1 Preliminaries: Majorization

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

Almost uniform and unimodal (ASU) distribution about c

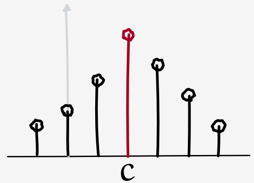


$$\pi_c \geqslant \pi_{c+1} \geqslant \pi_{c-1} \geqslant \pi_{c+2} \geqslant \cdots$$

Step 1 Preliminaries: Majorization

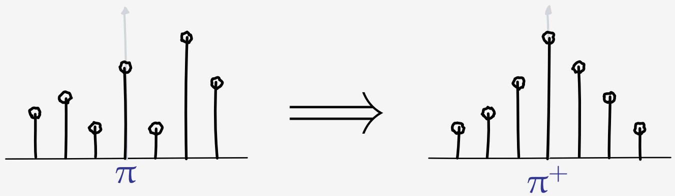
[Hajek Mitzel Yang 2008, Lipsa Martins 2011, Nayyar et. al. 2013]

Almost uniform and unimodal (ASU) distribution about c



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

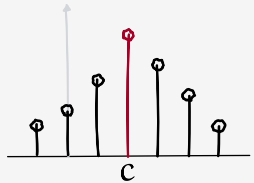
ASU Rearrangement



Step 1 Preliminaries: Majorization

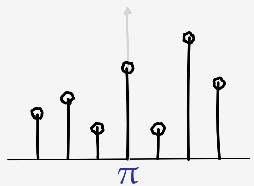
[Hajek Mitzel Yang 2008, Lipsa Martins 2011, Nayyar et. al. 2013]

Almost uniform and unimodal (ASU) distribution about c

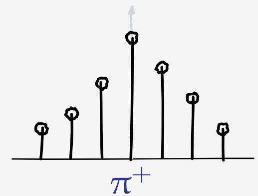


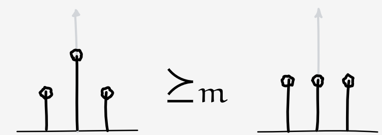
$$\pi_c \geqslant \pi_{c+1} \geqslant \pi_{c-1} \geqslant \pi_{c+2} \geqslant \cdots$$

ASU Rearrangement



$\Longrightarrow$





Majorization

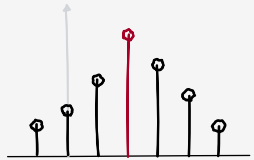
$$\xi \succeq_m \pi \text{ iff } \sum_{i=-n}^{n} \xi_i^+ \geqslant \sum_{i=-n}^{n} \pi_i^+ \quad \text{and} \quad \sum_{i=-n}^{n+1} \xi_i^+ \geqslant \sum_{i=-n}^{n+1} \pi_i^+$$

Invariant to permutations.

Step 1 Preliminaries: Majorization

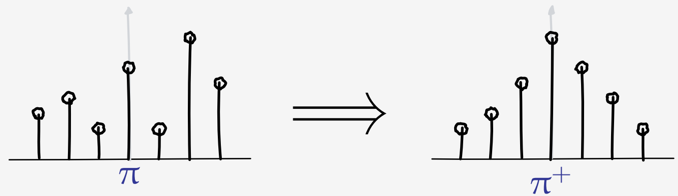
[Hajek Mitzel Yang 2008, Lipsa Martins 2011, Nayyar et. al. 2013]

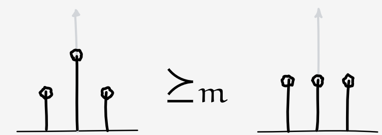
Almost uniform and unimodal (ASU) distribution about c



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

ASU Rearrangement





Majorization

$\xi \succeq_m \pi$ iff $\sum_{i=-n}^{n} \xi_i^+ \geqslant \sum_{i=-n}^{n} \pi_i^+$ and $\sum_{i=-n}^{n+1} \xi_i^+ \geqslant \sum_{i=-n}^{n+1} \pi_i^+$

Invariant to permutations.

ASU Majorization

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

Step 1 Implication of Majorization

Recall DP

$$V_t^1(s, \pi^1) = \min_{\varphi : \mathcal{X} \to \{0,1\}} \{\lambda \pi^1(B_1(\varphi)) + \pi^1(B_0(\varphi)) W_t^0(\pi^1, \varphi) + \pi^1(B_1(\varphi)) W_t^1(\pi^1, \varphi)\}$$

$$V_t^2(s, \pi^2) = \min_{\hat{x} \in \mathcal{X}} \sum_{x \in \mathcal{X}} \pi^2(x) d(x, \hat{x}) + V_{t+1}^1(s, \pi^2 P)$$

Proposition

V_t^1 and V_t^2 satisfy the following property:

- For any $s \in \{0, 1\}$ and $\pi \succeq_a \xi$, then $V_t^i(s, \pi) \geqslant V_t^i(s, \xi)$
  (Similar to Schur convexity, so we call it ASU Schur convexity)

Step 1 Implication of Majorization

Recall DP

$$V_t^1(s,\pi^1) = \min_{\varphi:\mathcal{X}\to\{0,1\}} \left\{\lambda\pi^1(B_1(\varphi)) + \pi^1(B_0(\varphi))W_t^0(\pi^1,\varphi) + \pi^1(B_1(\varphi))W_t^1(\pi^1,\varphi)\right\}$$

$$V_t^2(s,\pi^2) = \min_{\hat{x}\in\mathcal{X}} \sum_{x\in\mathcal{X}} \pi^2(x)d(x,\hat{x}) + V_{t+1}^1(s,\pi^2 P)$$

Proposition

V_t^1 and V_t^2 satisfy the following property:

- For any $s \in \{0,1\}$ and $\pi \succeq_a \xi$, then $V_t^i(s,\pi) \geqslant V_t^i(s,\xi)$
  (Similar to Schur convexity, so we call it ASU Schur convexity)

Step 1 Implication of Majorization

Recall DP

$$V_t^1(s,\pi^1) = \min_{\varphi:\mathcal{X}\to\{0,1\}} \{\lambda\pi^1(B_1(\varphi)) + \pi^1(B_0(\varphi))W_t^0(\pi^1,\varphi) + \pi^1(B_1(\varphi))W_t^1(\pi^1,\varphi)\}$$

$$V_t^2(s,\pi^2) = \min_{\hat{x}\in\mathcal{X}} \sum_{x\in\mathcal{X}} \pi^2(x)d(x,\hat{x}) + V_{t+1}^1(s,\pi^2 P)$$

Proposition

V_t^1 and V_t^2 satisfy the following property:

- For any $s \in \{0,1\}$ and $\pi \succeq_a \xi$, then $V_t^i(s,\pi) \geqslant V_t^i(s,\xi)$
  (Similar to Schur convexity, so we call it ASU Schur convexity)

Definition

A prescription φ is called **threshold based** if there exists a $k \in \mathcal{X}$ such that $\varphi(e) = 1$ if $|e| > 1$ and 0 otherwise.

Step 1 Implication of Majorization

Recall DP

$$V_t^1(s,\pi^1) = \min_{\varphi:\mathcal{X}\to\{0,1\}} \{\lambda\pi^1(B_1(\varphi)) + \pi^1(B_0(\varphi))W_t^0(\pi^1,\varphi) + \pi^1(B_1(\varphi))W_t^1(\pi^1,\varphi)\}$$

$$V_t^2(s,\pi^2) = \min_{\hat{x}\in\mathcal{X}} \sum_{x\in\mathcal{X}} \pi^2(x)d(x,\hat{x}) + V_{t+1}^1(s,\pi^2 P)$$

Proposition

V_t^1 and V_t^2 satisfy the following property:

- For any $s\in\{0,1\}$ and $\pi \succeq_a \xi$, then $V_t^i(s,\pi) \geqslant V_t^i(s,\xi)$ (Similar to Schur convexity, so we call it ASU Schur convexity)

Definition

A prescription φ is called **threshold based** if there exists a $k\in\mathcal{X}$ such that $\varphi(e)=1$ if $|e|>1$ and 0 otherwise.

Theorem

There is no loss optimality in restricting attention to threshold based transmission strategies and using estimation strategies of form

$$\hat{E}_t = \begin{cases} 0, & \text{if } Y_t\in\mathcal{X} \\ aE_{t-1}, & \text{if } Y_t\in\{\mathfrak{E}_0,\mathfrak{E}_1\} \end{cases}$$

Structure of optimal strategies

Theorem For the infinite horizon costly communication problem, we have the following:

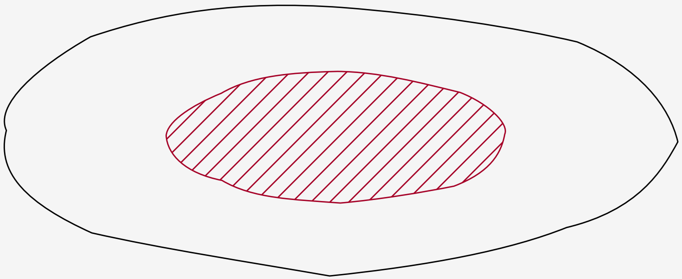
- **Structure of optimal estimation strategies**: The optimal estimation strategy is $\hat{X}_0 = 0$ and for $t > 0$

$$\hat{X}_t = \begin{cases} Y_t, & \text{if } Y_t \in \mathcal{X} \\ a\hat{X}_{t-1}, & \text{if } Y_t \in \{\mathfrak{E}_0, \mathfrak{E}_1\} \end{cases}$$

- **Structure of optimal transmission strategy**: There exist time-invariant thresholds $k(0), k(1) \in \mathcal{X}$ such that the strategy

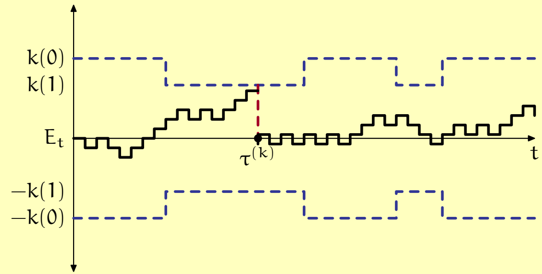
$$U_t = \begin{cases} 1, & \text{if } |X_t - aX_{t-1}| \geqslant k(S_{t-1}) \\ 0, & \text{otherwise} \end{cases}$$

Step 1 Threshold strategies are optimal

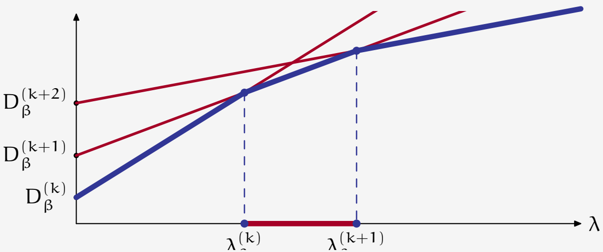


Search space of strategies (f, g)

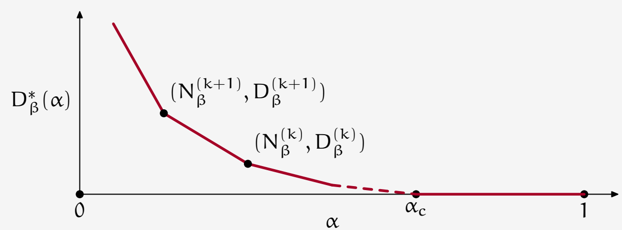
Step 2 Performance of threshold strategies



Step 3 Optimal costly communication



Step 4 Optimal constrained communication



Step 2 Performance of threshold-based strategies

Consider a threshold-based strategy

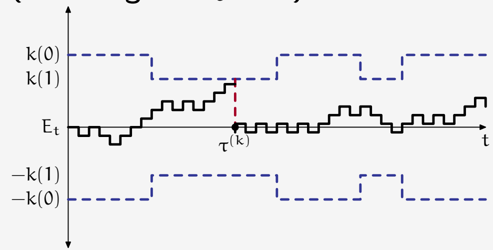
$$f^{(k)}(e, s) = \begin{cases} 1 & \text{if } |e| \geqslant k(s) \\ 0 & \text{otherwise} \end{cases}$$

Performance of threshold-based strategies

Consider a threshold-based strategy

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

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

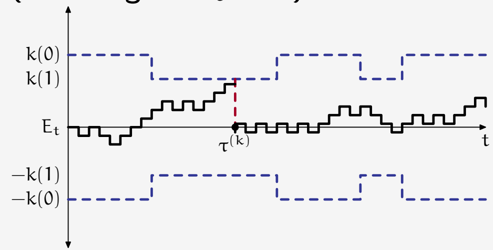


Step 2 Performance of threshold-based strategies

Consider a threshold-based strategy

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

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



Define

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

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

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

n

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

$$L_\beta^{(k)}(e) = \mathbb{E}\left[\sum_{t=0}^{\tau^{(k)}-1} \beta^t d(E_t) \,\middle|\, E_0 = e\right]. \qquad \text{(Distortion until first reception)}$$

$$M_\beta^{(k)}(e) = \mathbb{E}\left[\sum_{t=0}^{\tau^{(k)}-1} \beta^t \,\middle|\, E_0 = e\right]. \qquad \text{(Time until the first reception)}$$

$$K_\beta^{(k)}(e) = \mathbb{E}\left[\sum_{t=0}^{\tau^{(k)}} \beta^t U_t \,\middle|\, E_0 = e\right]. \qquad \text{(Transmissions until the first reception)}$$

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

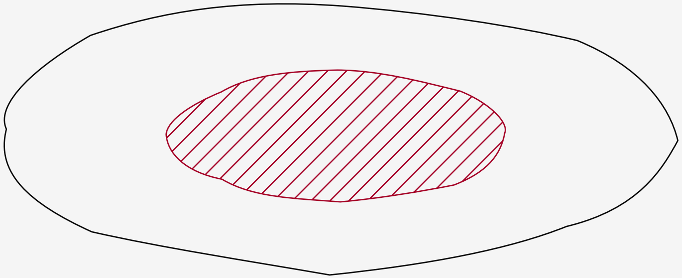
$$D_\beta^{(k)} \coloneqq D_\beta(f^{(k)}, g^*) = \frac{L_\beta^{(k)}(0)}{M_\beta^{(k)}(0)}$$

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

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 $N_\beta^{(k)}$).

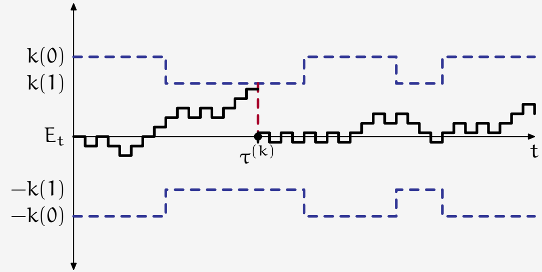
These can be computed using standard Markov chain formulas.

Step 1 Threshold strategies are optimal

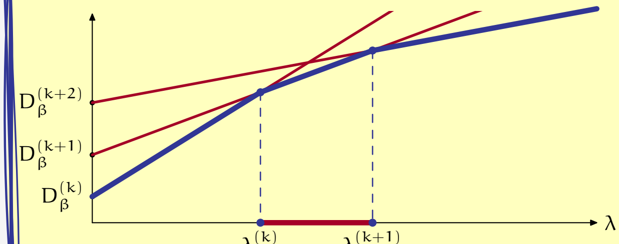


Search space of strategies (f, g)

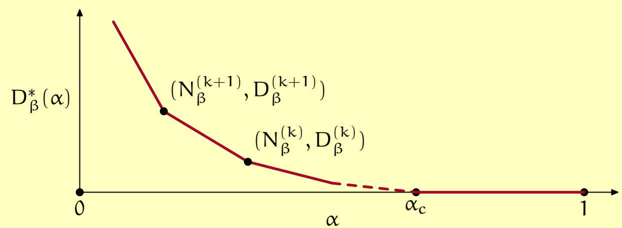
Step 2 Performance of threshold strategies



Step 3 Optimal costly communication



Step 4 Optimal constrained communication

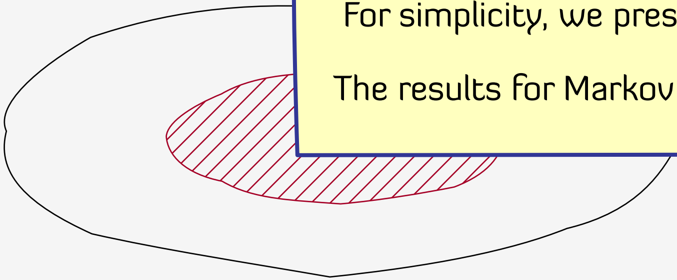


Step 1 Threshold strategies are optimal

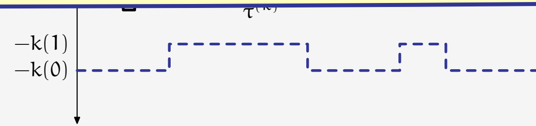
Step 2 Performance of threshold strategies

For simplicity, we present the results for i.i.d. packet drops.

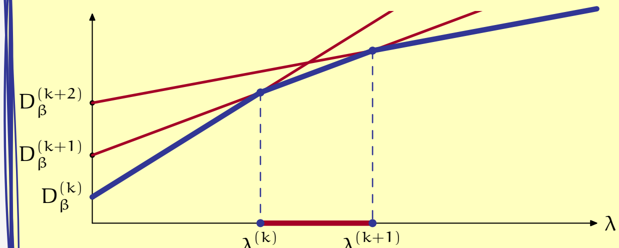
The results for Markov packet drops are similar (but harder to describe).



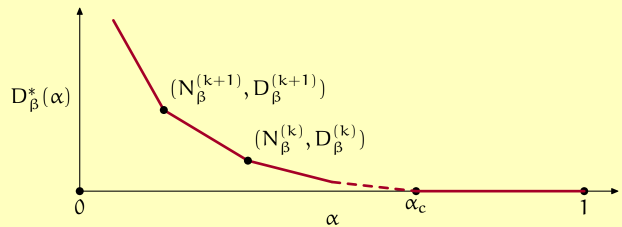
Search space of strategies (f, g)



Step 3 Optimal costly communication



Step 4 Optimal constrained communication



Step 3 Solution to costly comm. for discrete sources

Proposition

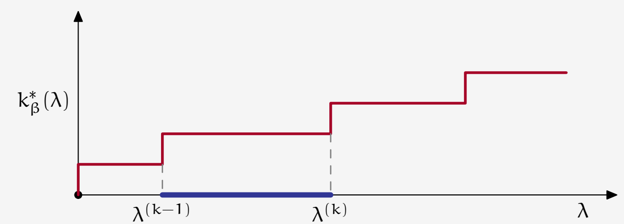
- $C^{(k)}_\beta(\lambda) \coloneqq D^{(k)}_\beta + \lambda N^{(k)}_\beta$ is submodular in (k, λ).
- Hence, $k^*_\beta(\lambda) \coloneqq \arg\min_{k \geq 0} C^{(k)}_\beta(\lambda)$ is increasing in λ

Step 3 Solution to costly comm. for discrete sources

Proposition

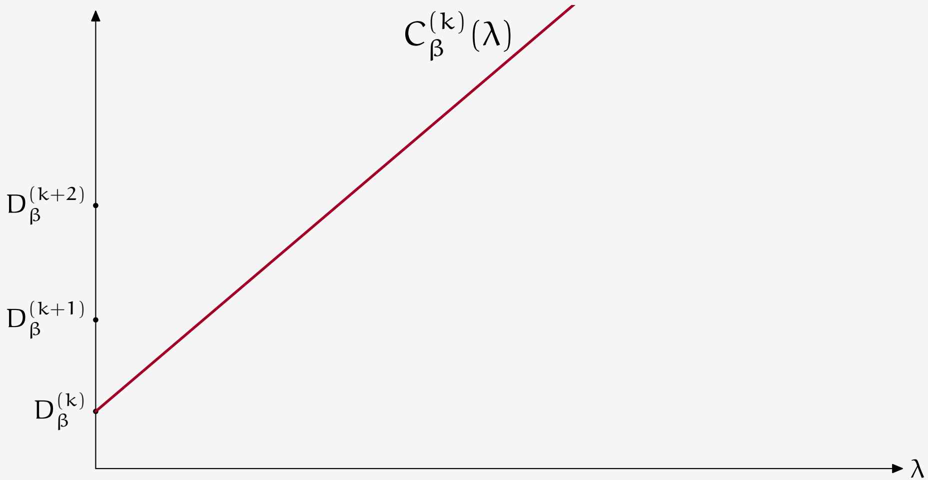
- $C_\beta^{(k)}(\lambda) \coloneqq D_\beta^{(k)} + \lambda N_\beta^{(k)}$ is submodular in (k, λ).
- Hence, $k_\beta^*(\lambda) \coloneqq \arg\min_{k \geqslant 0} C_\beta^{(k)}(\lambda)$ is increasing in λ

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

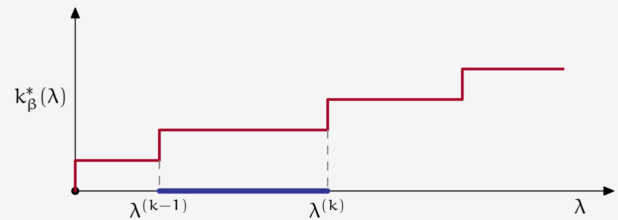


$$C_\beta^{(k)}(\lambda_\beta^{(k)}) = C_\beta^{(k+1)}(\lambda_\beta^{(k)})$$
$$\implies \lambda_\beta^{(k)} = \big(D_\beta^{(k+1)} - D_\beta^{(k)}\big) / \big(N_\beta^{(k)} - N_\beta^{(k+1)}\big).$$

Solution to costly comm. for discrete sources

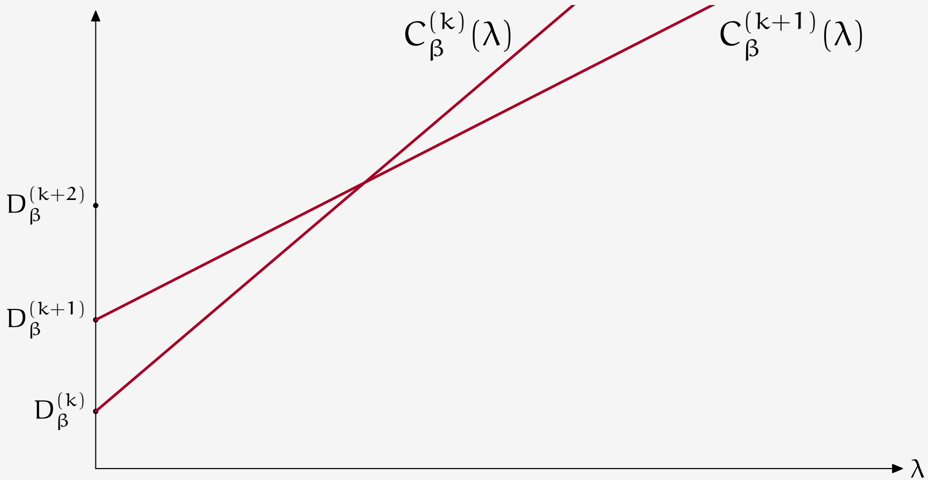


Define $\Lambda_\beta^{(k)} \coloneqq \{\lambda \in \mathbb{R}_{\geqslant 0} : k_\beta^*(\lambda) = k\}$
$= [\lambda_\beta^{(k-1)}, \lambda_\beta^{(k)}]$.

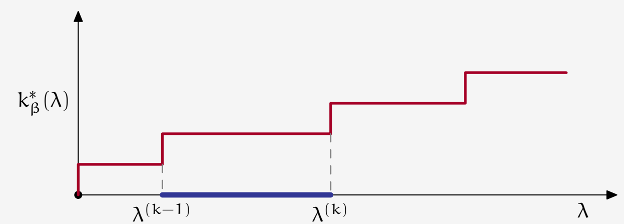


$$C_\beta^{(k)}(\lambda_\beta^{(k)}) = C_\beta^{(k+1)}(\lambda_\beta^{(k)})$$
$$\Longrightarrow \lambda_\beta^{(k)} = \big(D_\beta^{(k+1)} - D_\beta^{(k)}\big) / \big(N_\beta^{(k)} - N_\beta^{(k+1)}\big).$$

Solution to costly comm. for discrete sources

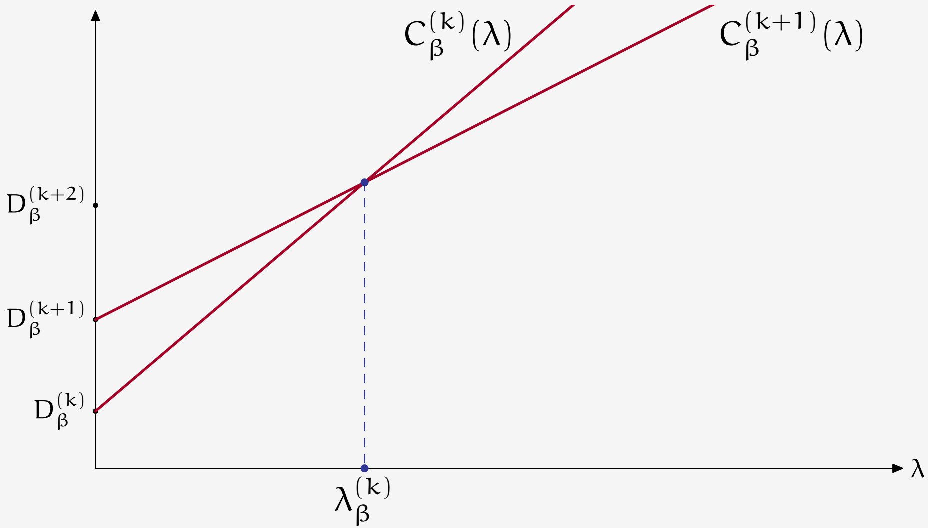


Define $\Lambda_\beta^{(k)} \coloneqq \{\lambda \in \mathbb{R}_{\geqslant 0} : k_\beta^*(\lambda) = k\}$
$= [\lambda_\beta^{(k-1)}, \lambda_\beta^{(k)}]$.

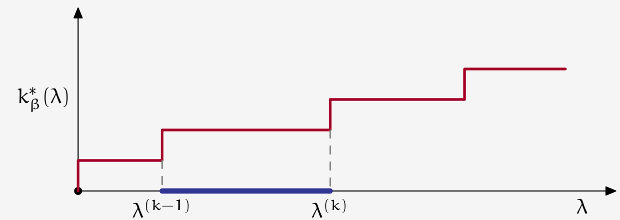


$$C_\beta^{(k)}(\lambda_\beta^{(k)}) = C_\beta^{(k+1)}(\lambda_\beta^{(k)})$$
$$\implies \lambda_\beta^{(k)} = \big(D_\beta^{(k+1)} - D_\beta^{(k)}\big) / \big(N_\beta^{(k)} - N_\beta^{(k+1)}\big).$$

Step 3 Solution to costly comm. for discrete sources

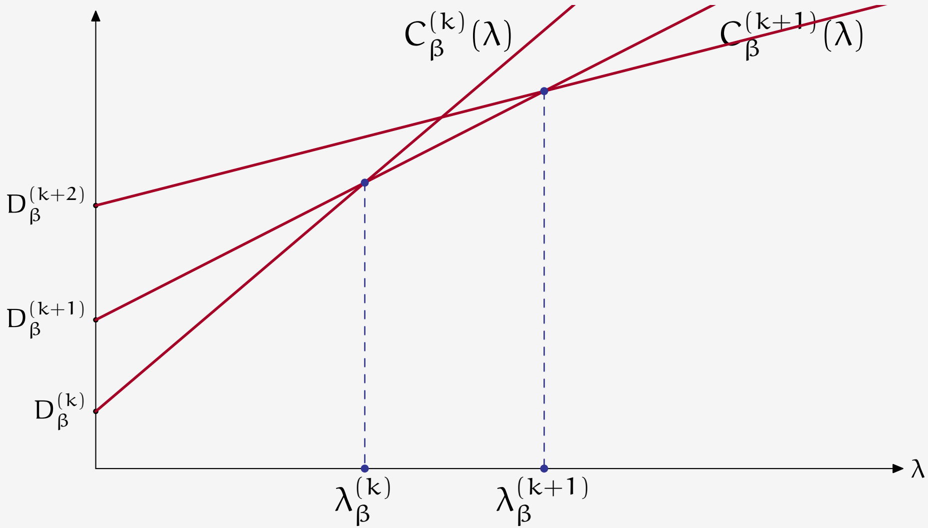


Define $\Lambda_\beta^{(k)} := \{\lambda \in \mathbb{R}_{\geqslant 0} : k^*_\beta(\lambda) = k\}$
$= [\lambda_\beta^{(k-1)}, \lambda_\beta^{(k)}]$.

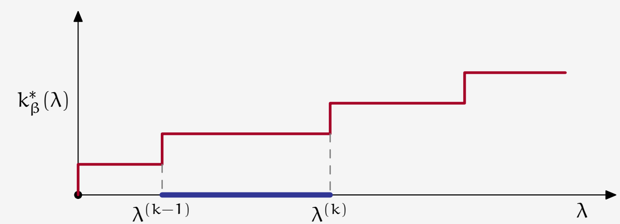


$$C_\beta^{(k)}(\lambda_\beta^{(k)}) = C_\beta^{(k+1)}(\lambda_\beta^{(k)})$$
$$\implies \lambda_\beta^{(k)} = \big(D_\beta^{(k+1)} - D_\beta^{(k)}\big)/\big(N_\beta^{(k)} - N_\beta^{(k+1)}\big).$$

Step 3 Solution to costly comm. for discrete sources

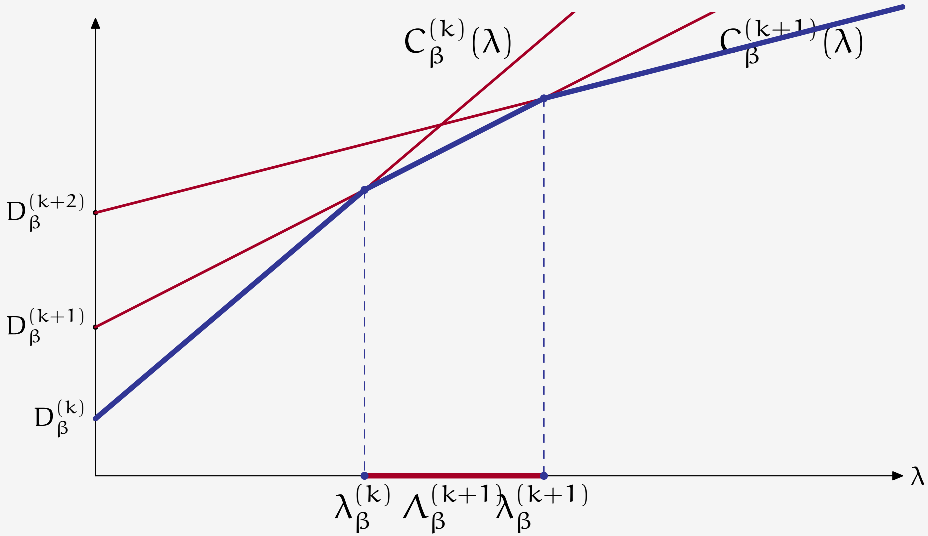


Define $\Lambda_\beta^{(k)} \coloneqq \{\lambda \in \mathbb{R}_{\geqslant 0} : k_\beta^*(\lambda) = k\}$
$= [\lambda_\beta^{(k-1)}, \lambda_\beta^{(k)}].$

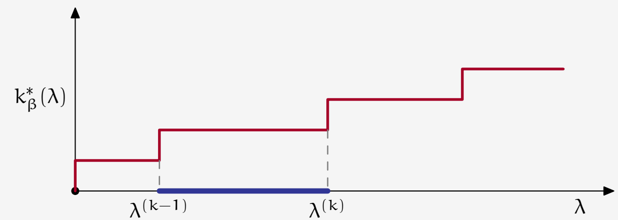


$$C_\beta^{(k)}(\lambda_\beta^{(k)}) = C_\beta^{(k+1)}(\lambda_\beta^{(k)})$$
$$\implies \lambda_\beta^{(k)} = \big(D_\beta^{(k+1)} - D_\beta^{(k)}\big)\big/\big(N_\beta^{(k)} - N_\beta^{(k+1)}\big).$$

Solution to costly comm. for discrete sources

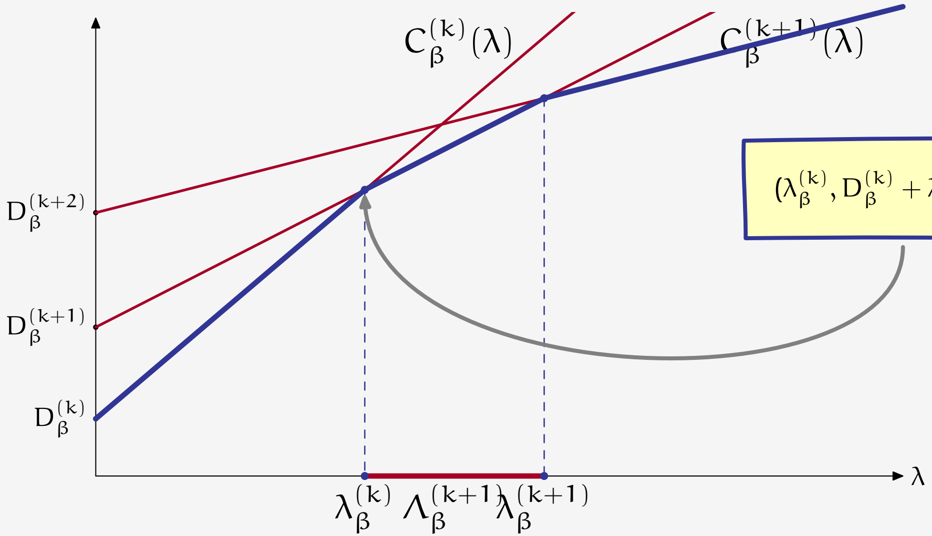


Define $\Lambda_\beta^{(k)} \coloneqq \{\lambda \in \mathbb{R}_{\geqslant 0} : k_\beta^*(\lambda) = k\}$
$= [\lambda_\beta^{(k-1)}, \lambda_\beta^{(k)}].$

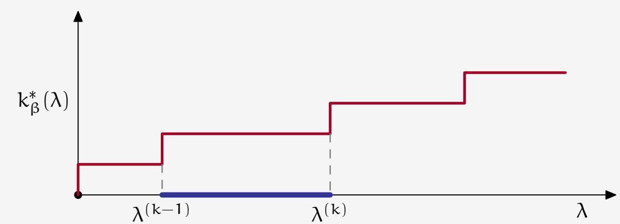


$$C_\beta^{(k)}(\lambda_\beta^{(k)}) = C_\beta^{(k+1)}(\lambda_\beta^{(k)})$$
$$\implies \lambda_\beta^{(k)} = \big(D_\beta^{(k+1)} - D_\beta^{(k)}\big) / \big(N_\beta^{(k)} - N_\beta^{(k+1)}\big).$$

Step 3 Solution to costly comm. for discrete sources

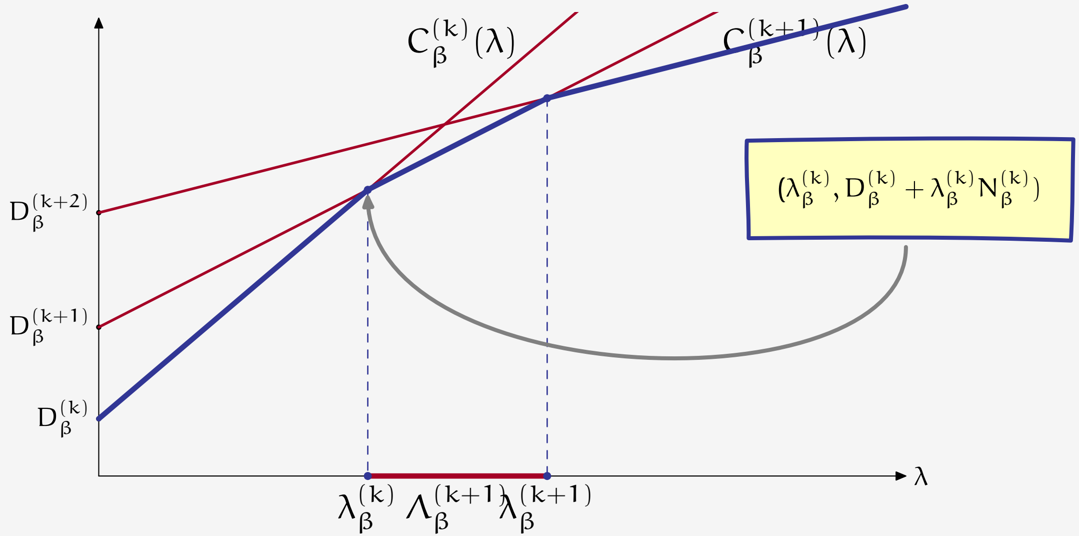


Define $\Lambda_\beta^{(k)} := \{\lambda \in \mathbb{R}_{\geqslant 0} : k_\beta^*(\lambda) = k\}$
$= [\lambda_\beta^{(k-1)}, \lambda_\beta^{(k)}].$



$$C_\beta^{(k)}(\lambda_\beta^{(k)}) = C_\beta^{(k+1)}(\lambda_\beta^{(k)})$$
$$\implies \lambda_\beta^{(k)} = \big(D_\beta^{(k+1)} - D_\beta^{(k)}\big) / \big(N_\beta^{(k)} - N_\beta^{(k+1)}\big).$$

Step 3 Solution to costly comm. for discrete sources



Theorem

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

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

Step 4 Solution to constrained comm. for discrete sources

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°.

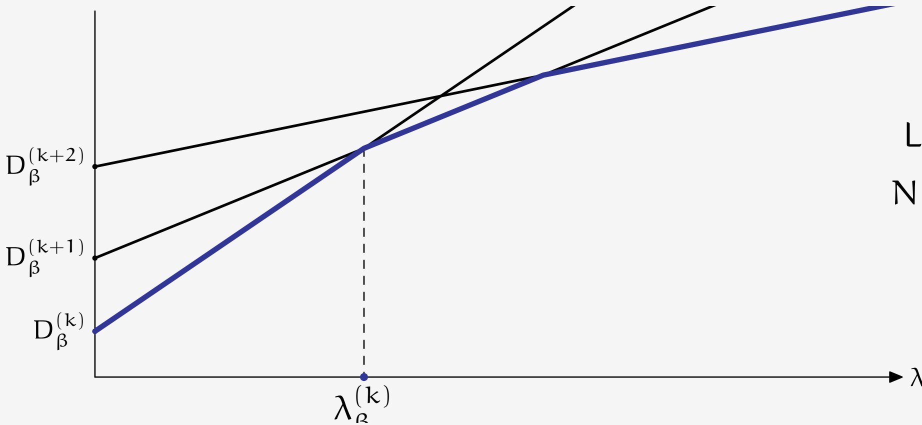
Step 4 Solution to constrained comm. for discrete sources

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

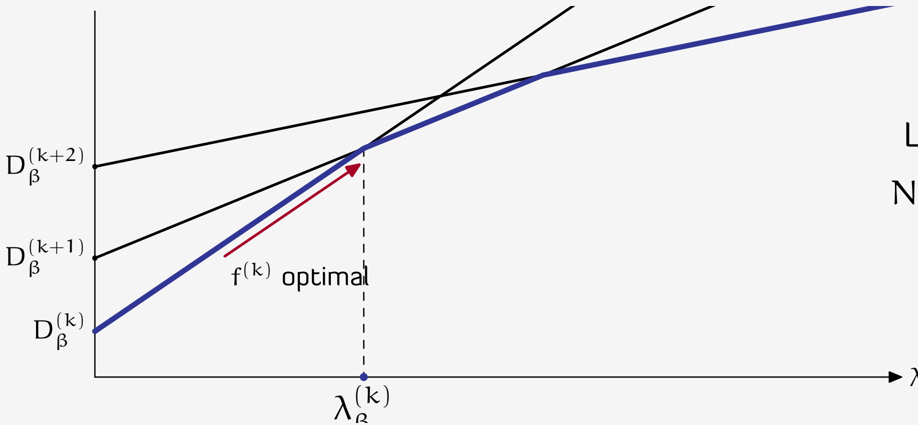
Solution to constrained comm. for discrete sources

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

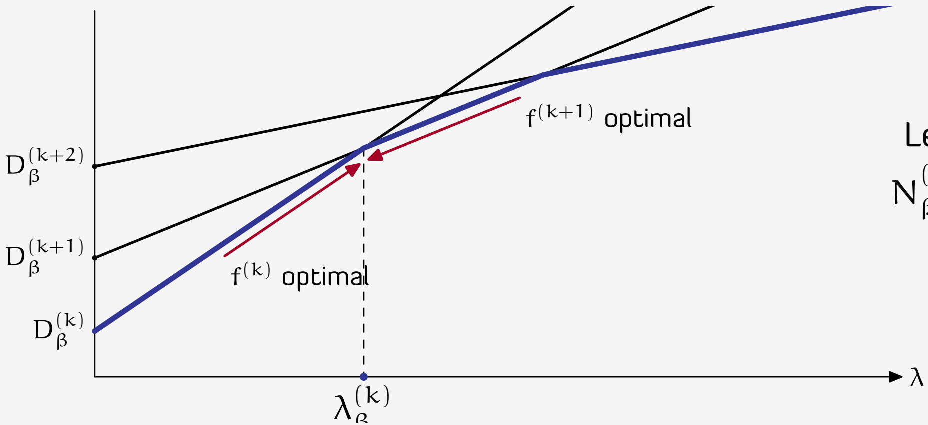
Step 4 Solution to constrained comm. for discrete sources

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 comm. for discrete sources

Sufficient conditio

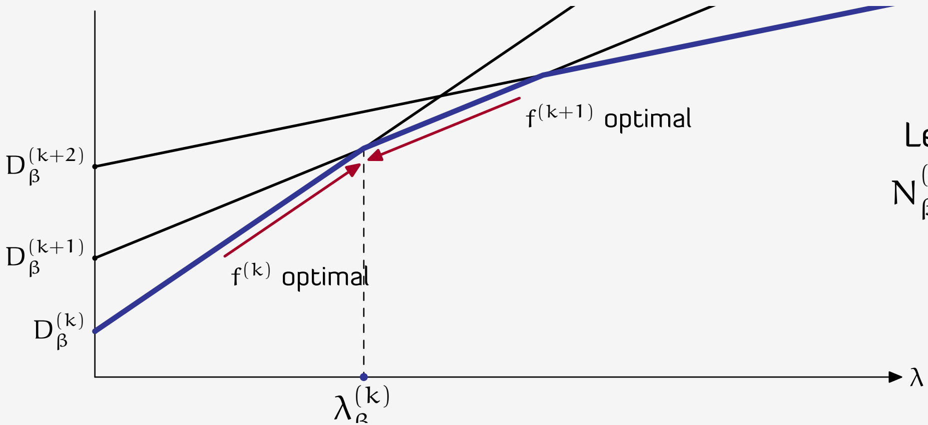
A strategy (f°, g°) i

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

(C2) There exi ameter λ°.

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

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



Let k_β^* be such that

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

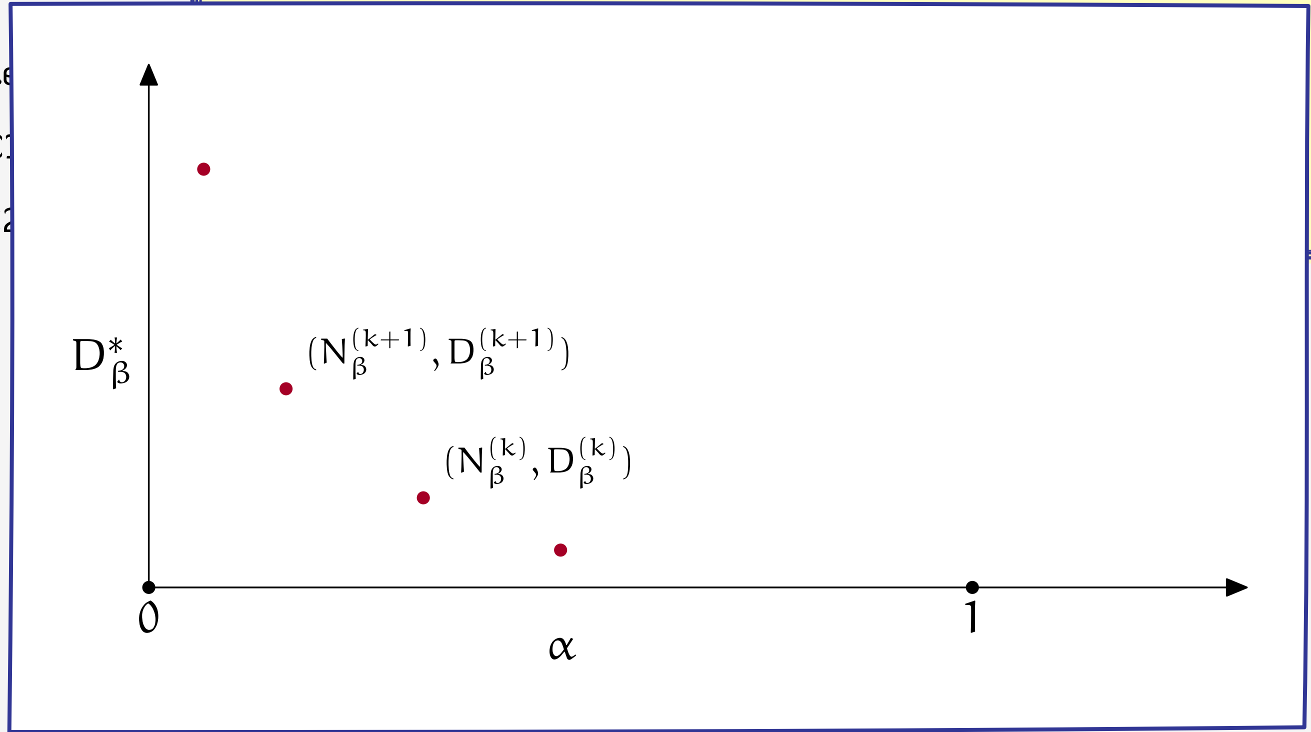
Step 4 Solution to constrained comm. for discrete sources

Sufficient conditio

A strate

(C1

(C2 ameter λ°.



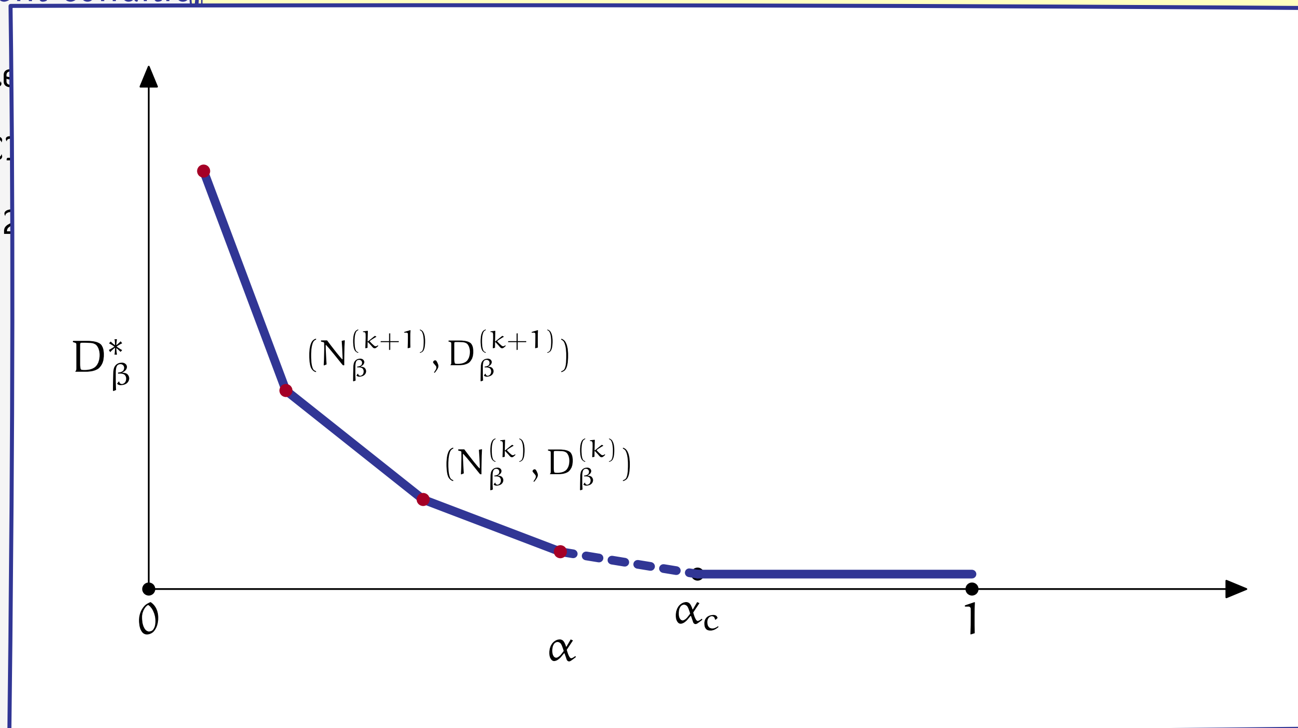
$\lambda_\beta^{(k)}$

Sufficient conditio

A strate

(C1

(C2 ameter λ°.



$\lambda_\beta^{(k)}$

Step 3 Solution to costly communication for continuous sources

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 communication for continuous sources

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.

Step 3 Solution to costly communication for continuous sources

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, $k^*_{\beta}(\lambda) \coloneqq \arg\min_{k \geq 0} C^{(k)}_{\beta}(\lambda)$ is increasing in λ

Theorem If the pair (λ, k) satisfies

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

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 communication for continuous sources

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

  (For the Markov packet drop case, we need to check additional KKT conditions)

- 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 communication for continuous sources

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

  (For the Markov packet drop case, we need to check additional KKT conditions)

- 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

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

Computation of optimal thresholds

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

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

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

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

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

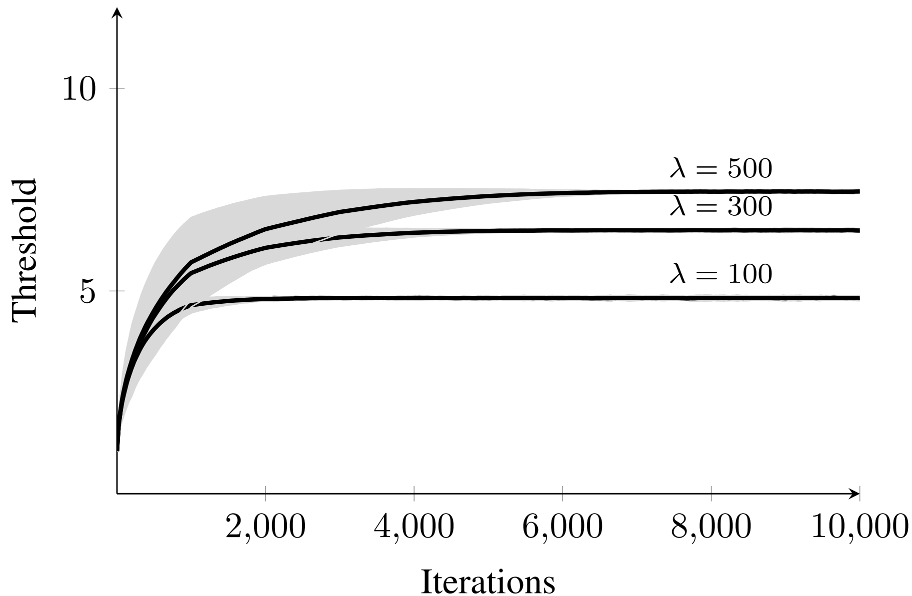
Main i

-Wolfowitz Algorithm

ns-Monro Algorithm

sful reception.

a "noisy" observation



Costly communication ($\beta = 1.0$, i.i.d. packet drop $\varepsilon = 0.3$)

Computation of optimal thresholds

Costly communic

-Wolfowitz Algorithm

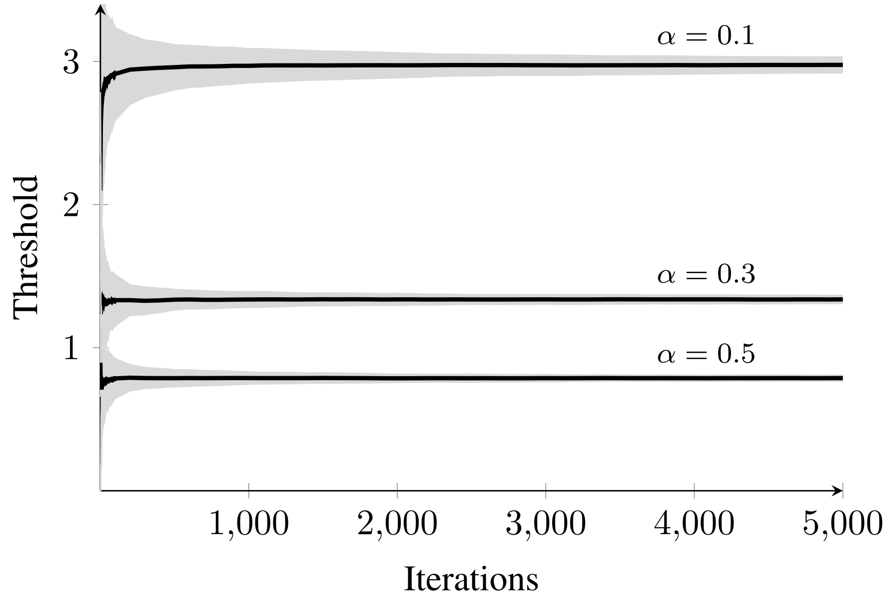
Constrai
communicat

s-Monro Algorithm

Main i

ful reception.

"noisy" observation

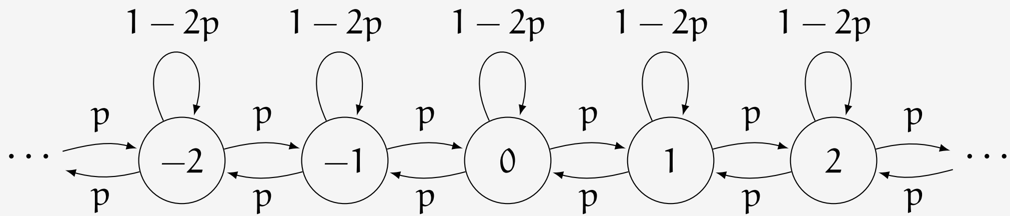


Constrained communication ($\beta = 1.0$, i.i.d. packet drop $\varepsilon = 0.3$)

Examples: Birth-death Markov chain and Gauss-Markov process

Example: Symmetric birth-death Markov chain (perfect channel)

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



Example Symmetric birth-death Markov chain (perfect channel)

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

Discounted cost

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

$$D_\beta^{(k)} = \frac{\sinh(k m_\beta) - k \sinh(m_\beta)}{2 \sinh^2(k m_\beta/2) \sinh(m_\beta)}$$

$$N_\beta^{(k)} = \frac{2\beta p \sinh^2(m_\beta/2) \cosh(k m_\beta)}{\sinh^2(k m_\beta/2)} - (1-\beta)$$

Average cost

$$D_1^{(k)} = \frac{k^2-1}{3k} \quad \text{and} \quad N_1^{(k)} = \frac{2p}{k^2}$$

Example Symmetric birth-death Markov chain (perfect channel)

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

Discounted cost

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

$$D_\beta^{(k)} = \frac{\sinh(k m_\beta) - k \sinh(m_\beta)}{2 \sinh^2(k m_\beta/2) \sinh(m_\beta)}$$

$$N_\beta^{(k)} = \frac{2\beta p \sinh^2(m_\beta/2) \cosh(k m_\beta)}{\sinh^2(k m_\beta/2)} - (1-\beta)$$

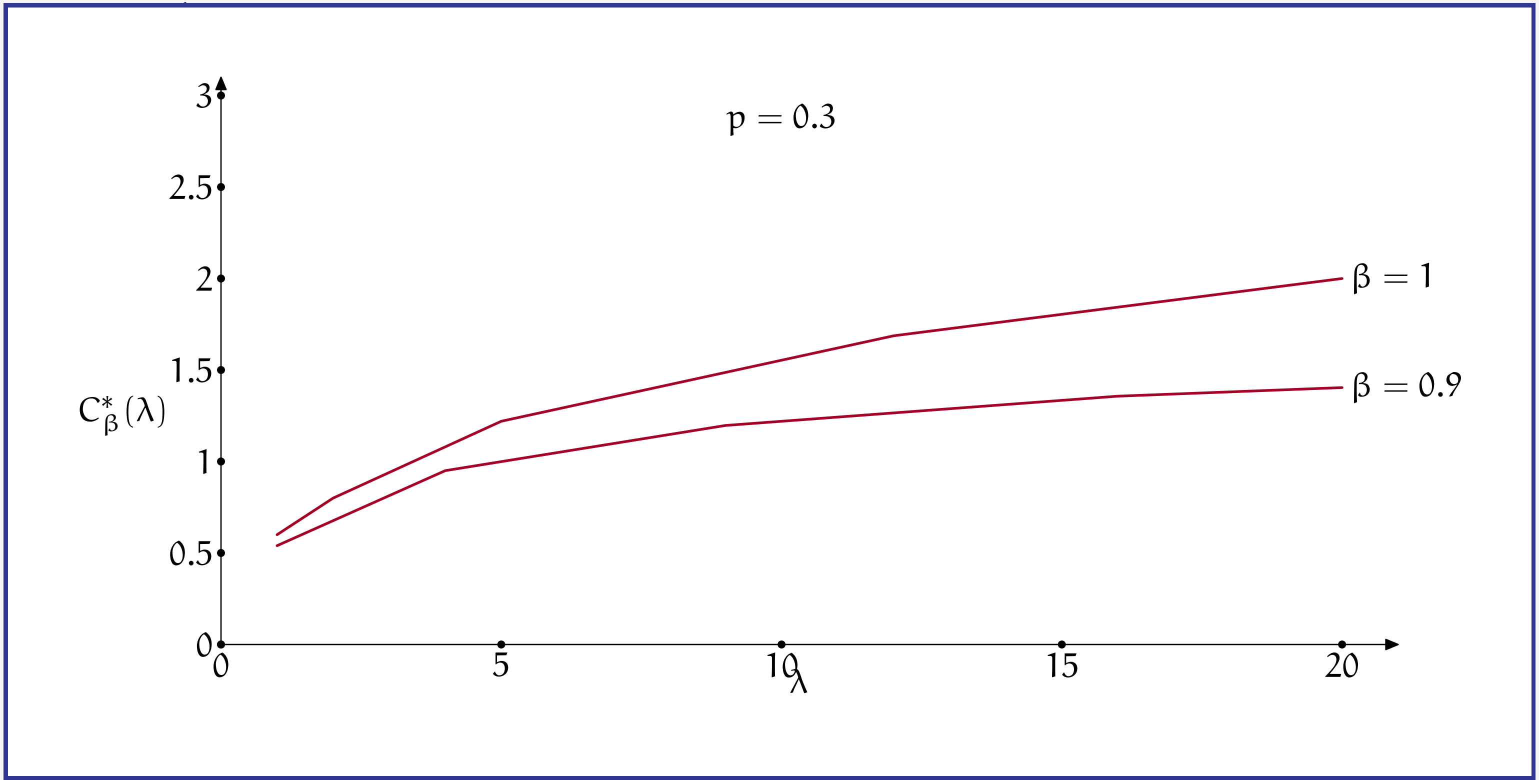
$\lambda_\beta^{(k)}$ can be computed in terms of $D_\beta^{(k)}$ and $N_\beta^{(k)}$.

Average cost

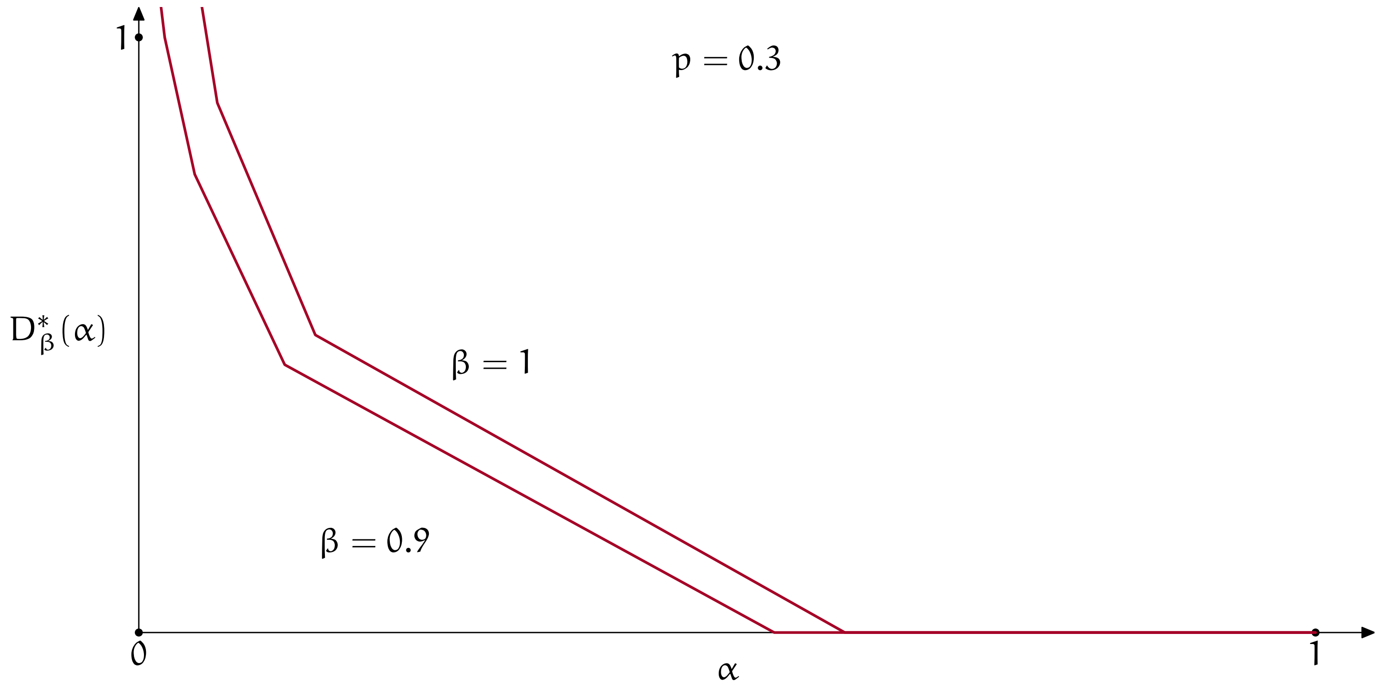
$D_1^{(k)} = \dfrac{k^2-1}{3k}$ and $N_1^{(k)} = \dfrac{2p}{k^2}$

$$\lambda_1^{(k)} = \frac{k(k+1)(k^2+k+1)}{6p(2k+1)}$$

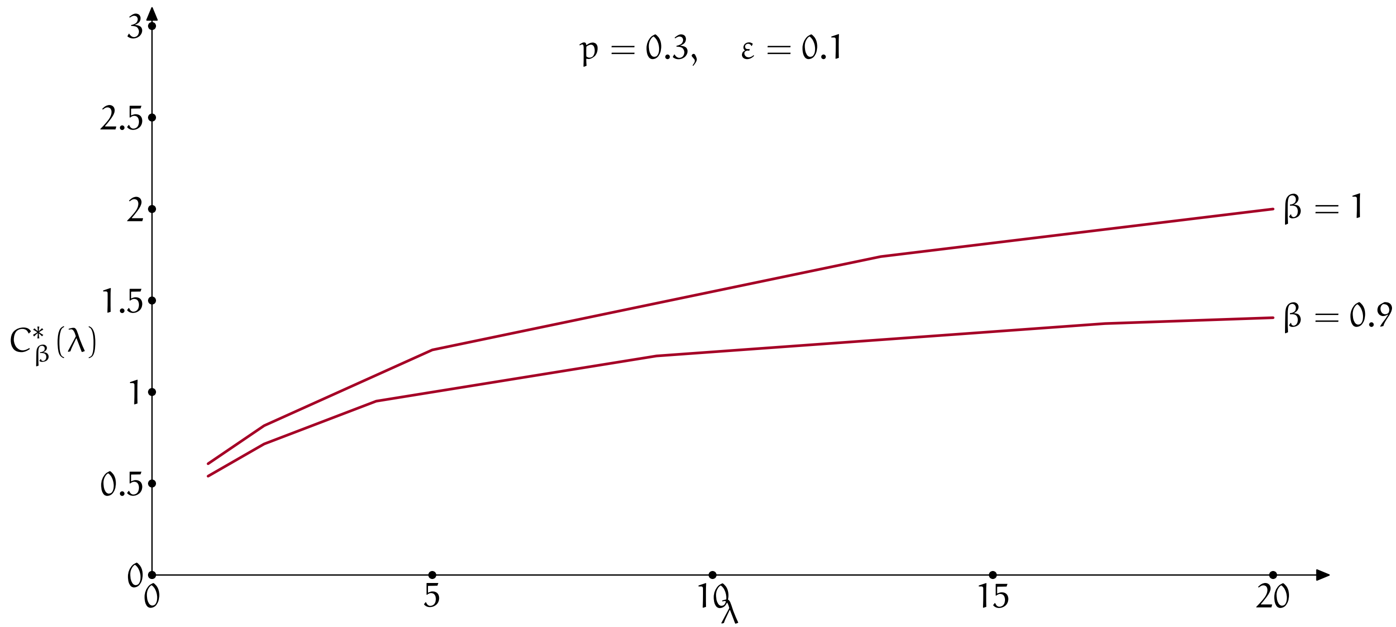
Example Symmetric birth-death Markov chain (perfect channel)



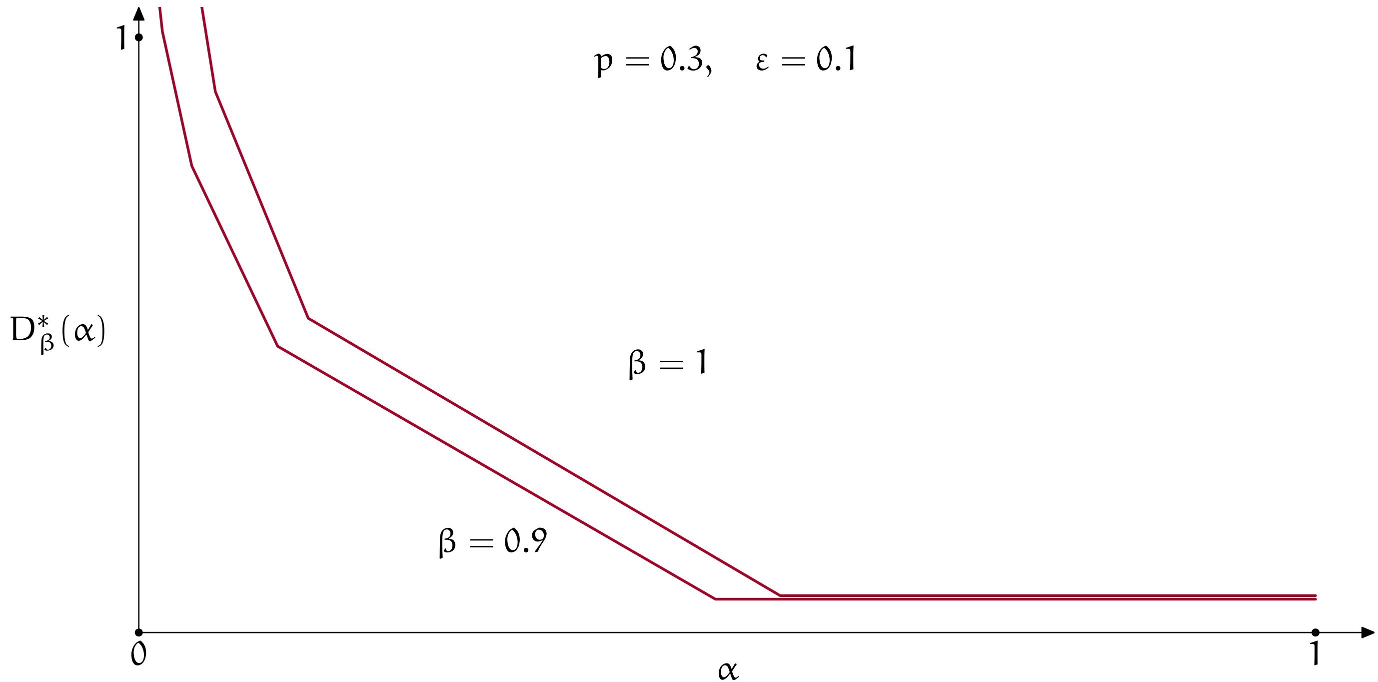
Example Symmetric birth-death Markov chain (perfect channel)



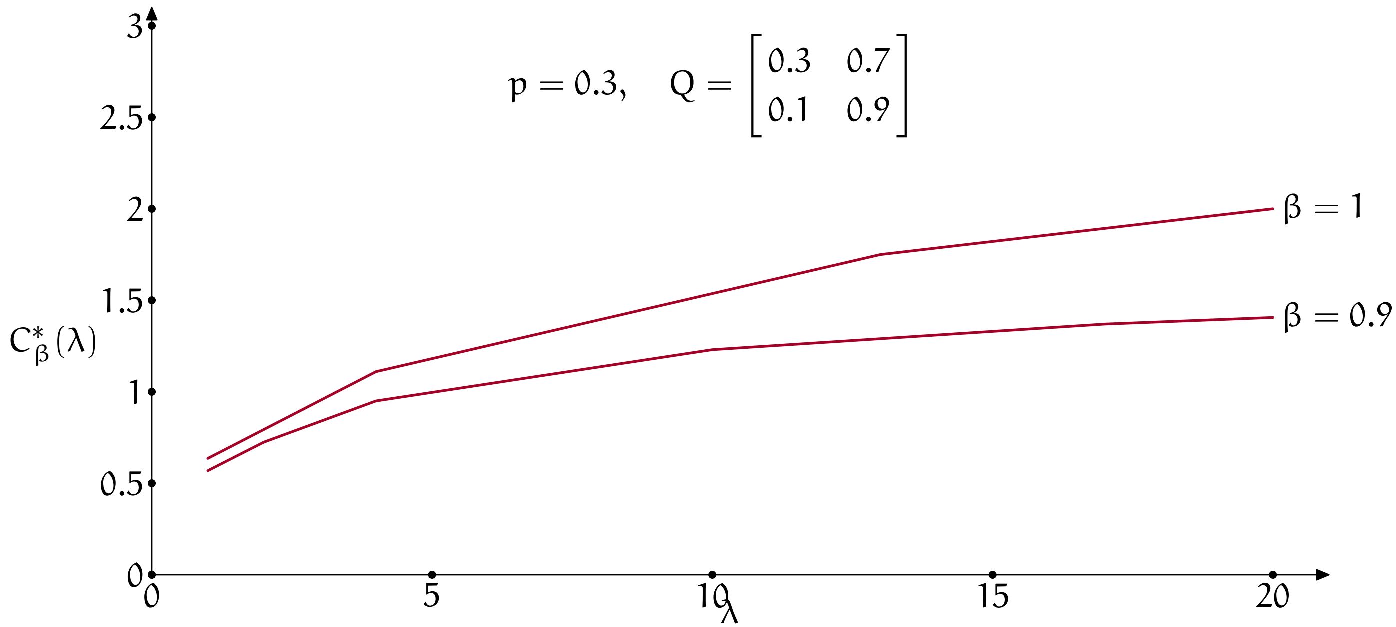
Example Symmetric birth-death Markov chain (i.i.d. packet drops)



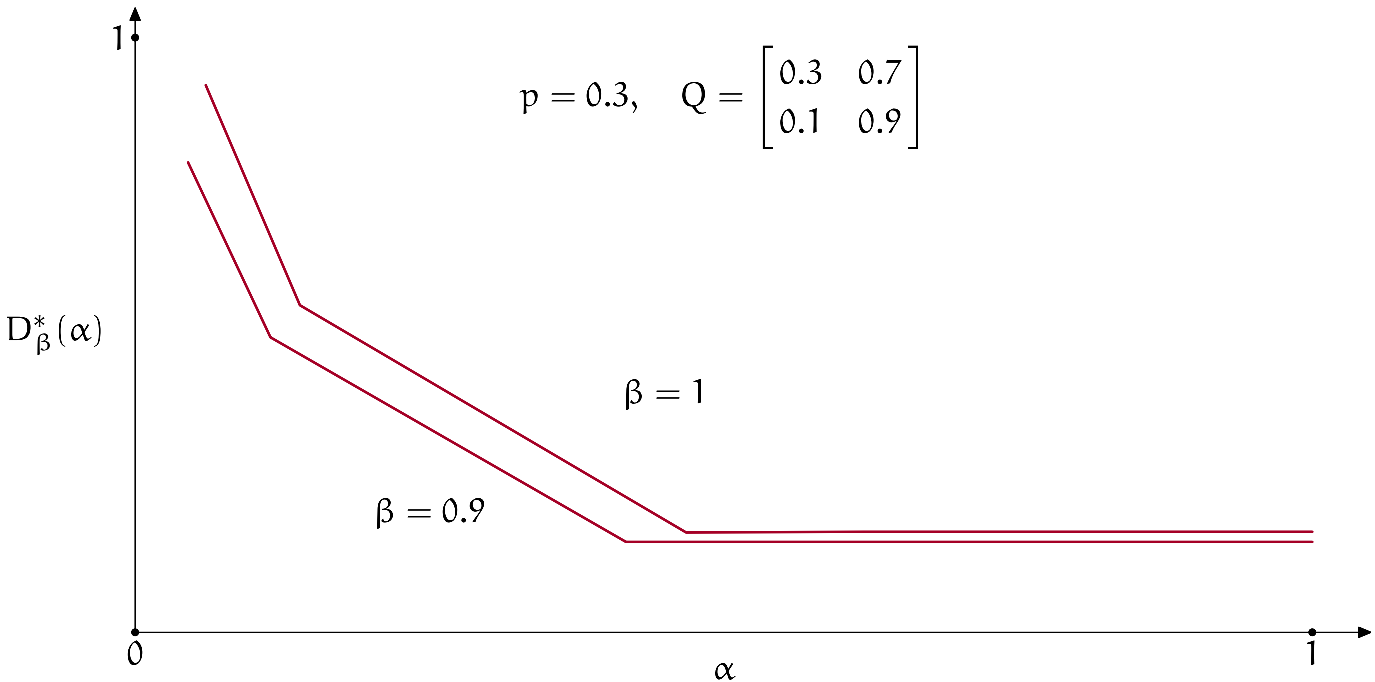
Example Symmetric birth-death Markov chain (i.i.d. packet drops)



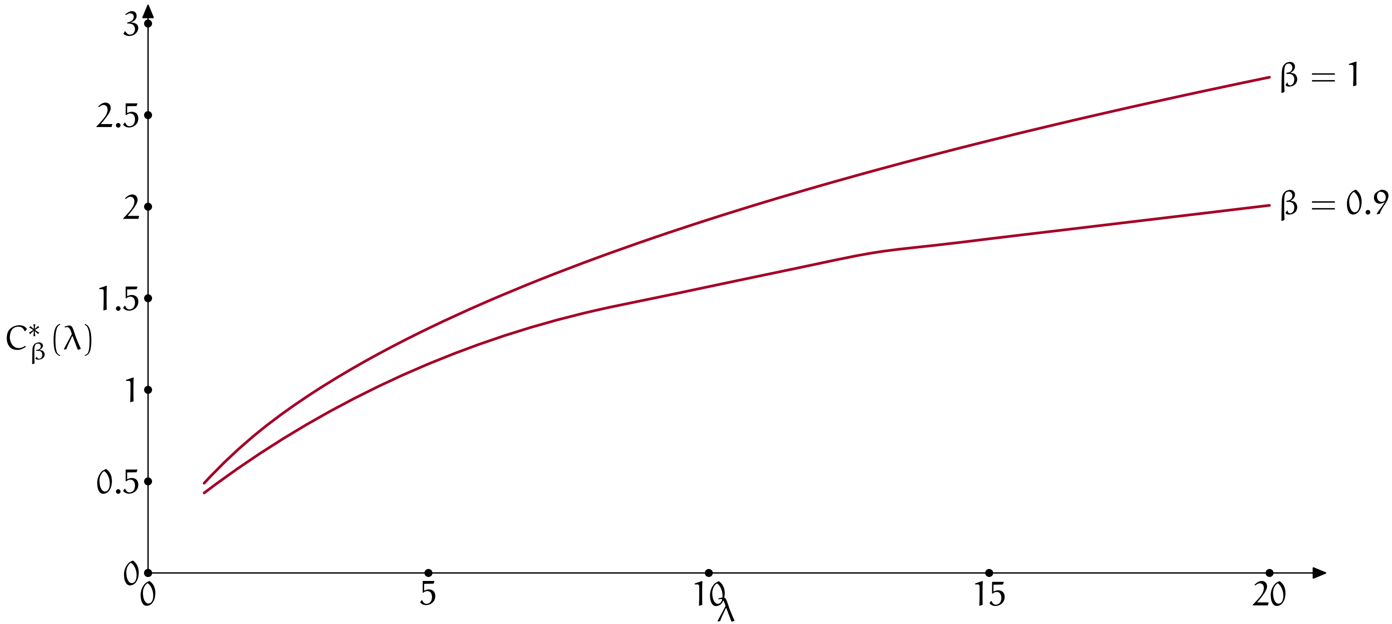
Example Symmetric birth-death Markov chain (Markov packet drops)



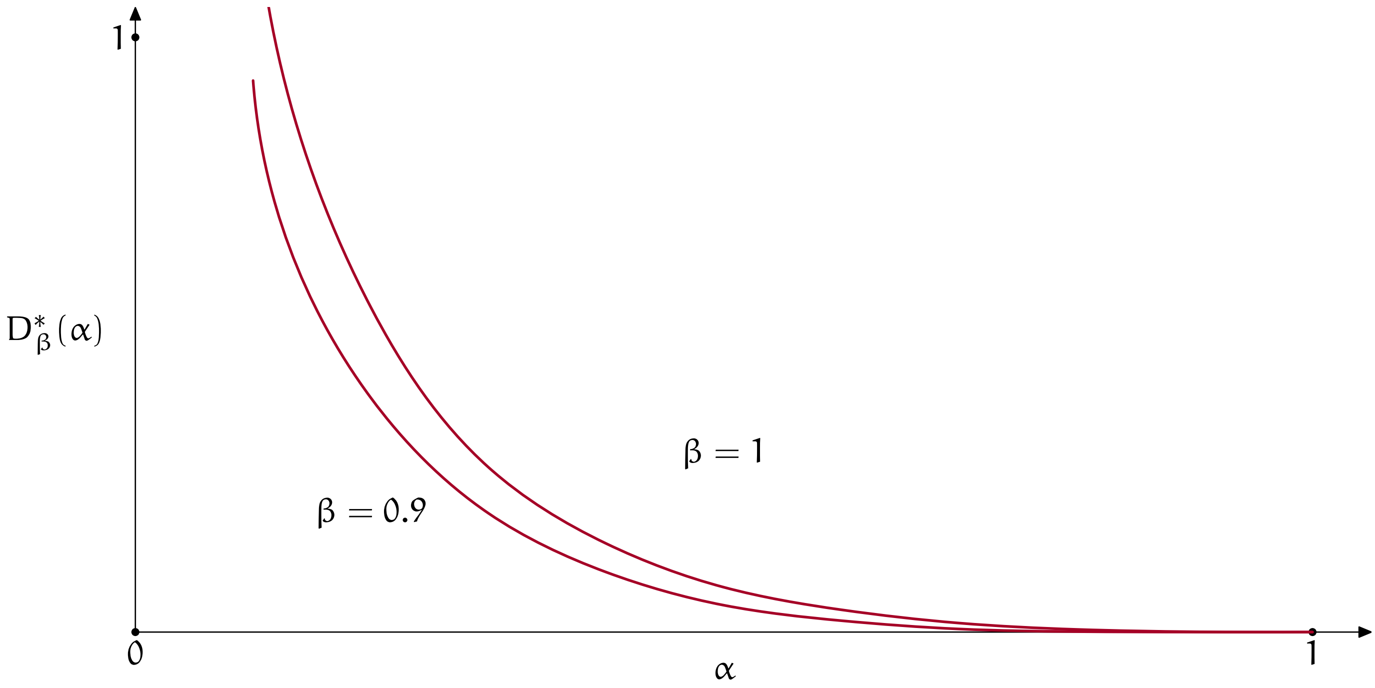
Example Symmetric birth-death Markov chain (Markov packet drops)



Gauss-Markov process ($\alpha = 1$, $\sigma^2 = 1$)

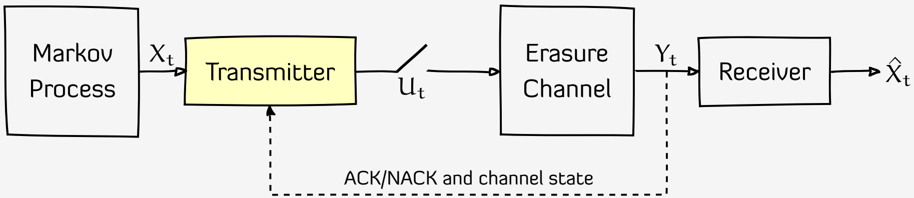


Gauss-Markov process ($a = 1$, $\sigma^2 = 1$)



Summary

Communication system (cont.)



Feedback The receiver sends two bits of feedback: ACK/NACK and channel state.

Transmitter Decides whether to transmit or not. Denoted by $U_t \in \{0, 1\}$.
If $U_t = 0$, $\bar{X}_t = \mathfrak{E}$. If $U_t = 1$, $\bar{X}_t = X_t$.

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

Summary

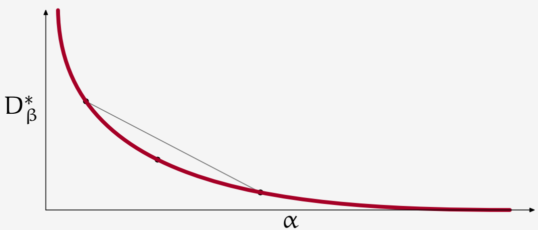
Optimization problems

Constrained communication

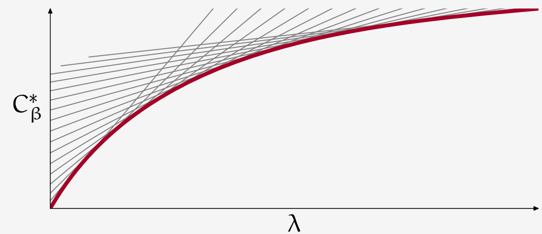
$$\text{For } \alpha \in (0,1), \quad D^*_\beta(\alpha) \coloneqq \inf_{(f,g)} \{ D_\beta(f,g) : N_\beta(f,g) \leqslant \alpha \}$$

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



C^*_β is cts, inc, and concave

Summary

Communication system (cont.)

Optimization problems

Optimal strategies and their performance

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 $d(x, \hat{x}) = d(x - \hat{x})$ where $d(\cdot)$ is symmetric and quasi-convex.

Optimal transmission strategy

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

Optimal estimation strategy

$$\hat{X}_t = \begin{cases} a\hat{X}_{t-1}, & \text{if } Y_t \in \{\mathfrak{E}_0, \mathfrak{E}_1\} \\ Y_t, & \text{if } Y_t \in \mathcal{X} \end{cases}$$

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 $N_\beta^{(k)} = \dfrac{K_\beta^{(k)}}{M_\beta^{(k)}}$. **(Renewal Relationships)**

Summary

Communication system (cont.)

Optimization problems

Optimal strategies and their performance

Dynamic program

$$V^1_{T+1}(s, \pi^1) = 0$$

and for $t \in \{T, \dots, 0\}$

$$V^1_t(s, \pi^1) = \min_{\varphi: \mathcal{X} \to \{0,1\}} \{\lambda \pi^1(B_1(\varphi)) + \pi^1(B_0(\varphi)) W^0_t(\pi^1, \varphi) + \sum_{x \in B_1(\varphi)} \pi^1(x) W^1_t(\pi^1, \varphi, x)\}$$

$$V^2_t(s, \pi^2) = \min_{\hat{x} \in \mathcal{X}} \sum_{x \in \mathcal{X}} \pi^2(x) d(x, \hat{x}) + V^1_{t+1}(s, \pi^2 P)$$

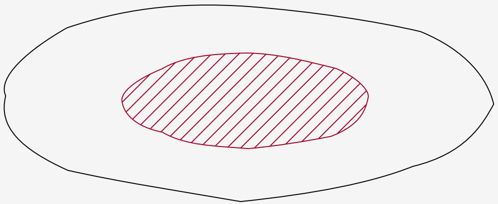
where $W^0_t(\pi^1, \varphi) = Q_{s0} V^2_t(0, \pi^1) + Q_{s1} V^2_t(1, \pi^1|_\varphi)$

$$W^1_t(\pi^2, \varphi, x) = Q_{s0} V^2_t(0, \pi^1) + Q_{s1} V^2_t(1, \delta_x)$$

Summary

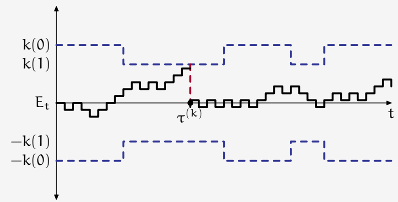
Communication system (cont.)
Optimization problems
Optimal strategies and their performance
Dynamic program

Step 1 Threshold strategies are optimal

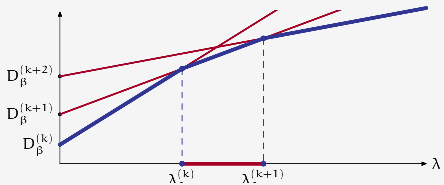


Search space of strategies (f, g)

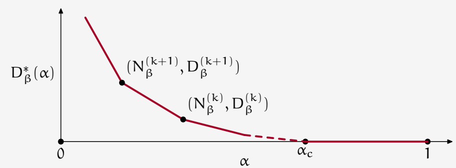
Step 2 Performance of threshold strategies



Step 3 Optimal costly communication



Step 4 Optimal constrained communication



Summary

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.

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

- Quantization . . .
- Power control . . .
- Scheduling multiple sources . . .
- Model network delays . . .

Beautiful example of stochastics and optimization

Decentralized control, POMDP, stochastic orders, majorization, Markov chains, constrained optimization, stochastic approximation

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

- Quantization . . .
- Power control . . .
- Scheduling multiple sources . . .
- Model network delays . . .

Beautiful example of stochastics and optimization

Decentralized control, POMDP, stochastic orders, majorization, Markov chains, constrained optimization, stochastic approximation

References

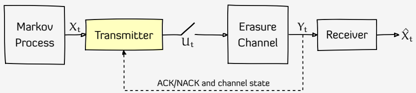
Jhelum Chakravorty and Aditya Mahajan, “Fundamental limits of remote estimation of autoregressive Markov processes under communication constraints,” IEEE TAC, March 2017.

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

Jhelum Chakravorty and Aditya Mahajan, “Structure of optimal strategies for remote estimation over Gilbert-Elliott channel with feedback,” ISIT 2017 (submitted)

Communication system (cont.)



Feedback The receiver sends two bits of feedback: ACK/NACK and channel state.

Transmitter Decides whether to transmit or not. Denoted by $U_t \in \{0,1\}$.
If $U_t = 0$, $\bar{X}_t = \mathfrak{E}$. If $U_t = 1$, $\bar{X}_t = X_t$.
$U_t = f_t(X_{1:t}, Y_{1:t-1}, S_{1:t-1})$

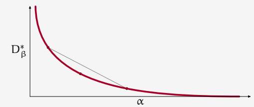
Optimization problems

Constrained communication

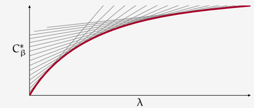
For $\alpha \in (0,1)$, $D^*_\beta(\alpha) := \inf_{(f,g)} \{D_\beta(f,g) : N_\beta(f,g) \leqslant \alpha\}$

Costly communication (Lagrange relaxation)

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



D^*_β is cts, dec, and convex



C^*_β is cts, inc, and concave

Optimal strategies and their performance

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 $d(x,\hat{x}) = d(x - \hat{x})$ where $d(\cdot)$ is symmetric and quasi-convex.

Optimal transmission strategy

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

Optimal estimation strategy

$$\hat{X}_t = \begin{cases} a\hat{X}_{t-1}, & \text{if } Y_t \in \{\mathfrak{E}_0, \mathfrak{E}_1\} \\ Y_t, & \text{if } Y_t \in \mathcal{X} \end{cases}$$

Performance of threshold based strategies

- $K^{(k)}_\beta$: Expected discounted number of transmissions until first successful reception.
- $L^{(k)}_\beta$: Expected discounted distortion until first successful reception.
- $M^{(k)}_\beta$: Expected discounted time until first successful reception.

Then, $D^{(k)}_\beta = \dfrac{L^{(k)}_\beta}{M^{(k)}_\beta}$ and $N^{(k)}_\beta = \dfrac{K^{(k)}_\beta}{M^{(k)}_\beta}$. (Renewal Relationships)

Dynamic program

$$V^1_{T+1}(s, \pi^1) = 0$$

and for $t \in \{T, \dots, 0\}$

$$V^1_t(s,\pi^1) = \min_{\varphi: \mathcal{X} \to \{0,1\}} \{\lambda \pi^1(B_1(\varphi)) + \pi^1(B_0(\varphi)) W^0_t(\pi^1, \varphi) + \sum_{x \in B_1(\varphi)} \pi^1(x) W^1_t(\pi^1, \varphi, x)\}$$

$$V^2_t(s,\pi^2) = \min_{\hat{x} \in \mathcal{X}} \sum_{x \in \mathcal{X}} \pi^2(x) d(x, \hat{x}) + V^1_{t+1}(s, \pi^2 P)$$

where $W^0_t(\pi^1, \varphi) = Q_{s0} V^2_t(0, \pi^1) + Q_{s1} V^2_t(1, \pi^1|_\varphi)$

$$W^1_t(\pi^2, \varphi, x) = Q_{s0} V^2_t(0, \pi^1) + Q_{s1} V^2_t(1, \delta_x)$$

Step 1 A change of variables

Define $Z_0 = 0$ and $Z_t = \begin{cases} aZ_{t-1}, & \text{if } Y_t \in \{\mathfrak{E}_0, \mathfrak{E}_1\} \\ Y_t, & \text{if } Y_t \in \mathcal{X} \end{cases}$ (Observable at both Tx and Rx)

$$E_t = X_t - aZ_{t-1}, \quad E^+_t = X_t - Z_t, \quad \hat{E}_t = \hat{X}_t - Z_t$$

Thus, these are related as

$$E^+_t = \begin{cases} E_t, & \text{if } Y_t \in \{\mathfrak{E}_0, \mathfrak{E}_1\} \\ 0, & \text{if } Y_t \in \mathcal{X} \end{cases} \quad \text{and} \quad E_{t+1} = aE^+_t + W_t$$

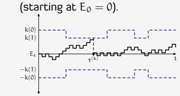
Note $X_t - \hat{X}_t = E^+_t - \hat{E}_t$ and hence $d(X_t - \hat{X}_t) = d(E^+_t - \hat{E}_t)$.

Step 2 Performance of threshold-based strategies

Consider a threshold-based strategy

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

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

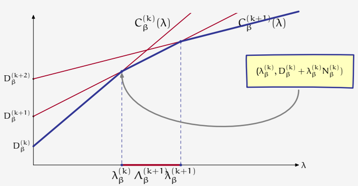


Define $L^{(k)}_\beta(e) = \mathbb{E}\Big[\sum_{t=0}^{\tau^{(k)}-1} \beta^t d(E_t) \Big| E_0 = e\Big]$. (Distortion until first reception)

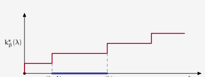
$M^{(k)}_\beta(e) = \mathbb{E}\Big[\sum_{t=0}^{\tau^{(k)}-1} \beta^t \Big| E_0 = e\Big]$. (Time until the first reception)

$K^{(k)}_\beta(e) = \mathbb{E}\Big[\sum_{t=0}^{\tau^{(k)}} \beta^t U_t \Big| E_0 = e\Big]$. (Transmissions until the first reception)

Step 3 Solution to costly comm. for discrete sources



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



$C^{(k)}_\beta(\lambda^{(k)}_\beta) = C^{(k+1)}_\beta(\lambda^{(k)}_\beta)$
$\implies \lambda^{(k)}_\beta = (D^{(k+1)}_\beta - D^{(k)}_\beta)/(N^{(k)}_\beta - N^{(k+1)}_\beta)$.

Step 4 Solution to constrained comm. for discrete sources

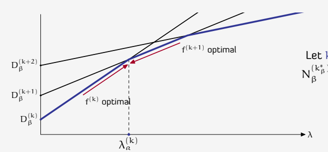
Sufficient conditi...

A strategy (f°, g°) i...

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

(C2) There exi... ...meter λ°.

Randomized strategy $(\theta^*, f^{(k)}, f^{k+1})$ is optimal where
$\theta^* N^{(k)}_\beta + (1 - \theta^*) N^{(k+1)}_\beta = \alpha$



Let k^*_β be such that
$N^{(k^*_\beta)}_\beta > \alpha > N^{(k^*_\beta+1)}_\beta$

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.