Fundamental limits of remote-estimation under communication constraints

Aditya Mahajan
McGill University

Joint work with Jhelum Chakravorty

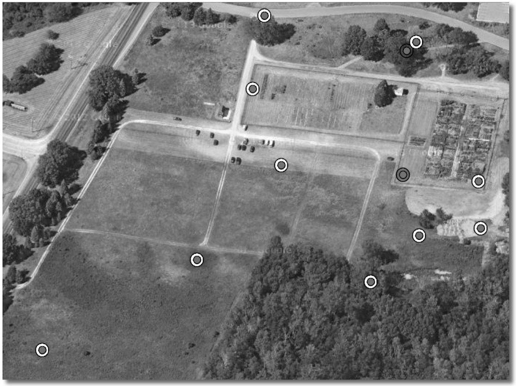
Mathematical Cybernetics: Hybrid, Stochastic and Decentralized Systems
Carleton University, Ottawa, 28–29 May 2015

Motivation

Many applications require:

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

Motivation

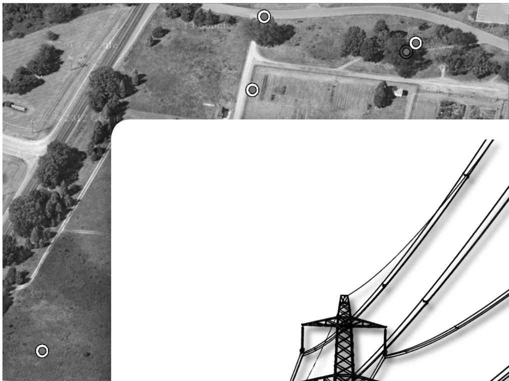


Sensor Networks

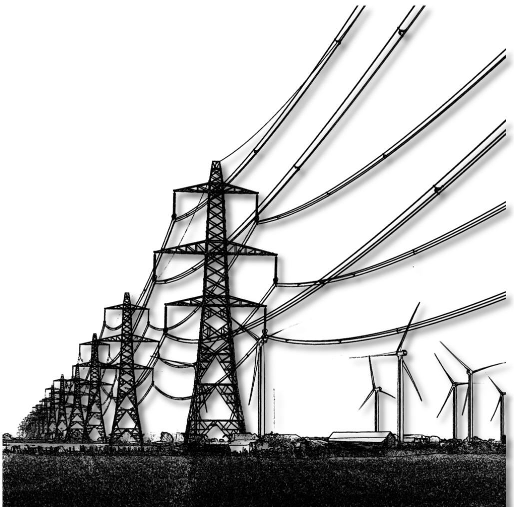
Many applications require:

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

Motivation



S

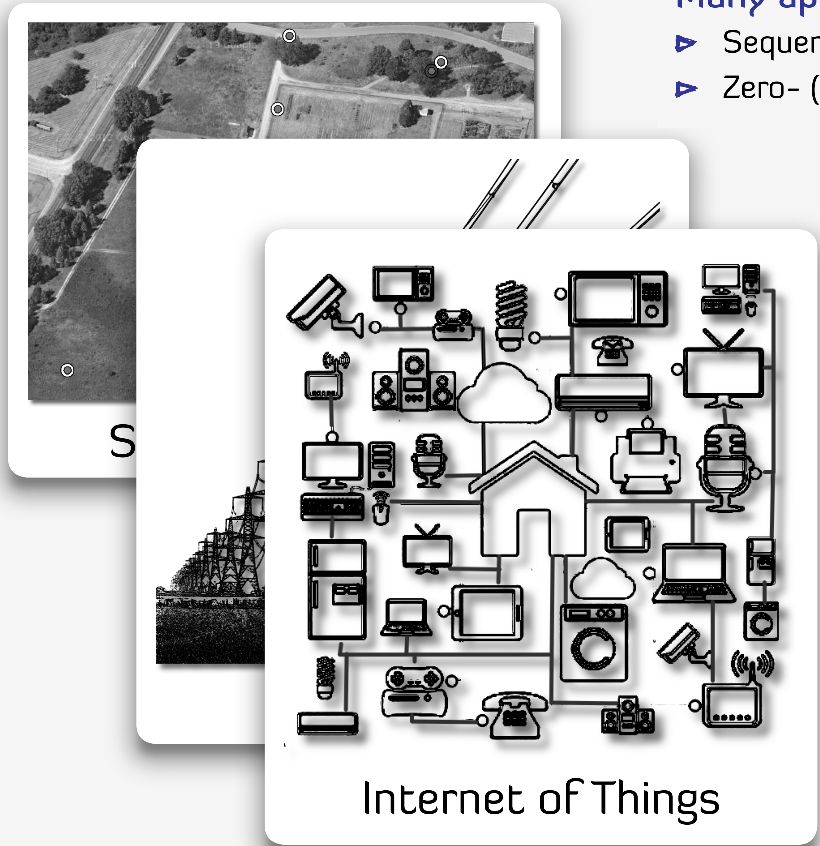


Smart Grids

Many applications require:

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

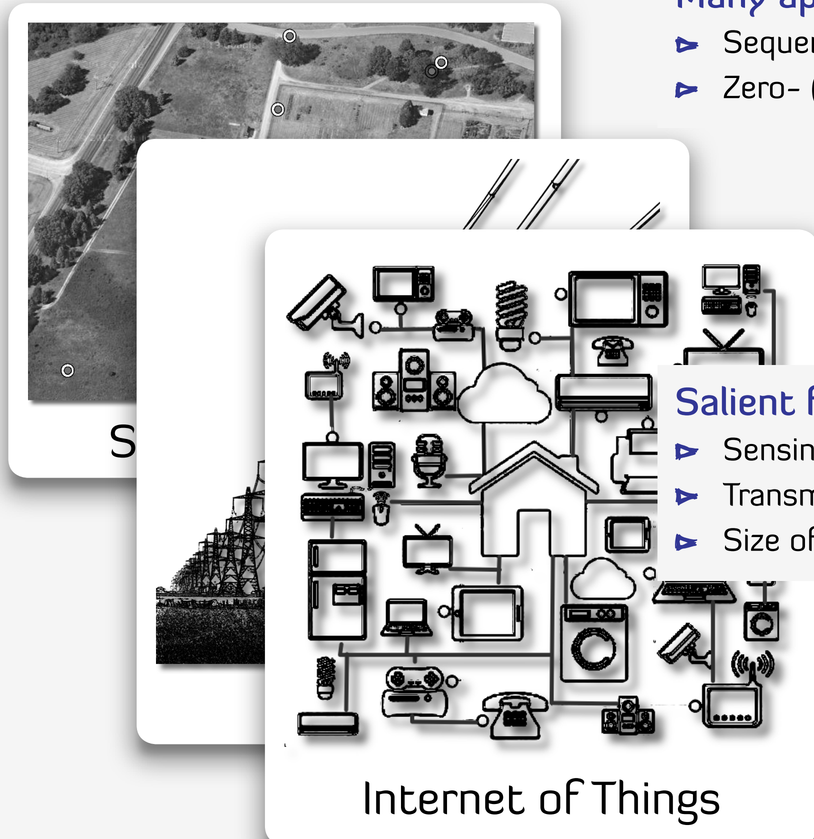
Motivation



Many applications require:

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

Motivation



Many applications require:

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

Salient features

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

Motivation

Many applications require:

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

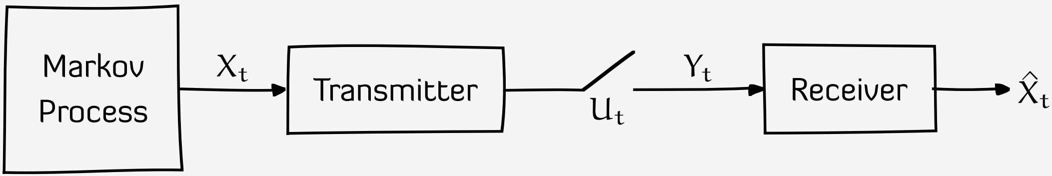
Salient features

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

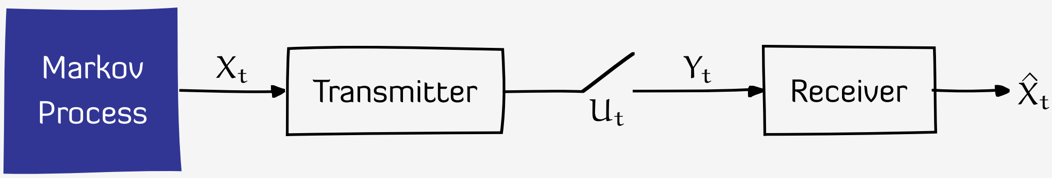
Analyze a stylized model and evaluate fundamental trade-offs

A completely solved example of a “simple” decentralized system with non-classical information structure

The system model

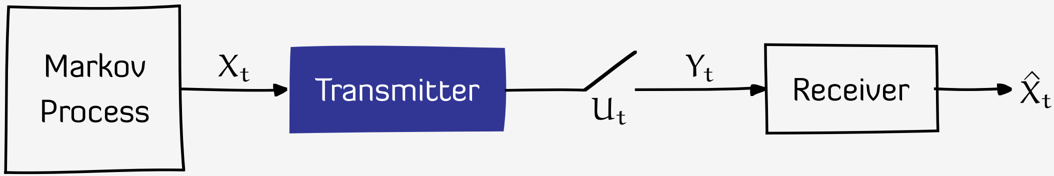


The system model

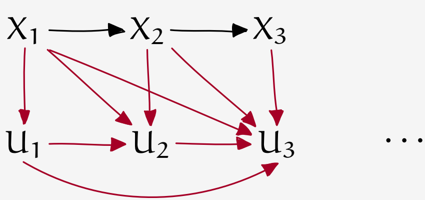


The system model

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



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

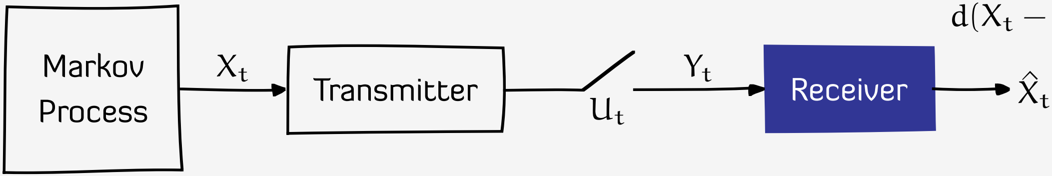


The system model

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

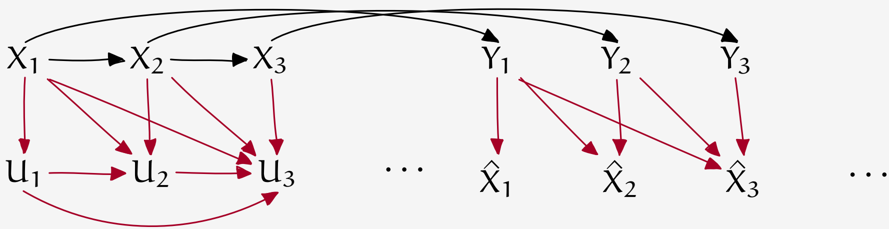
Distortion

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



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

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

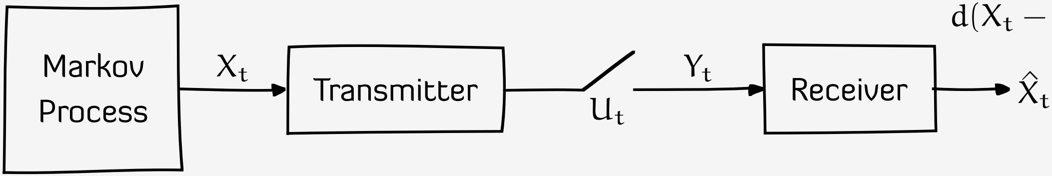


The system model

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

Distortion

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



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

Communication Strategies

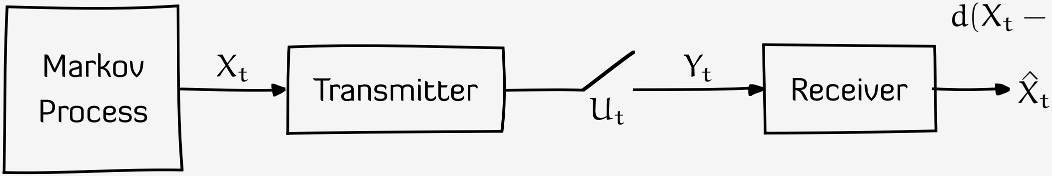
- Transmission strategy $f = \{f_t\}_{t=0}^{\infty}$.
- Estimation strategy $g = \{g_t\}_{t=0}^{\infty}$.

The system model

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

Distortion

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



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

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

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

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

Costly communication

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

Constrained communication

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

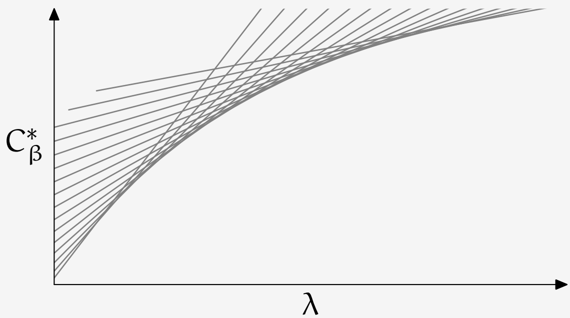
Optimization problems

Costly communication

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

Constrained communication

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



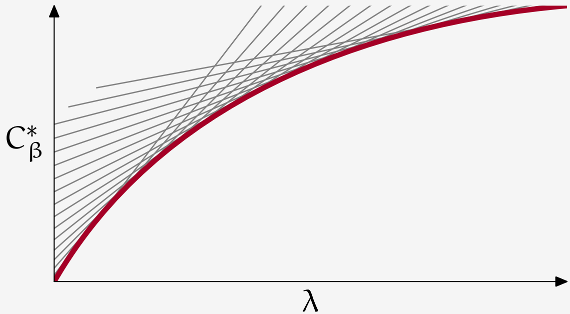
Optimization problems

Costly communication

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

Constrained communication

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



C^*_β is cts, inc, and concave

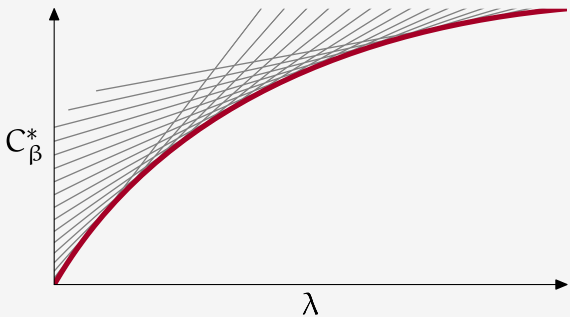
Optimization problems

Costly communication

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

Constrained communication

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



C^*_β is cts, inc, and concave

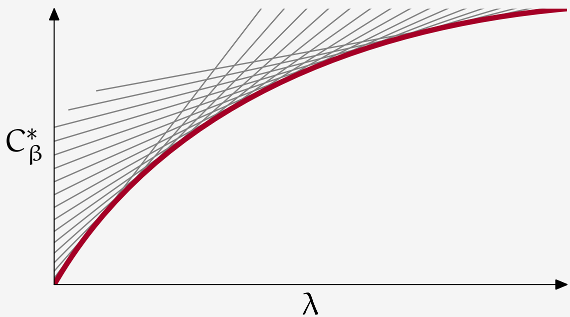
Optimization problems

Costly communication

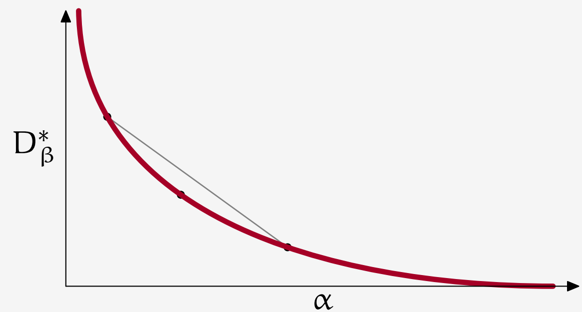
$$\text{For } \lambda \in \mathbb{R}_{>0}, \quad C^*_\beta(\lambda) = C_\beta(f^*, g^*; \lambda) := \inf_{(f,g)} \{D_\beta(f,g) + \lambda N_\beta(f,g)\}$$

Constrained communication

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



C^*_β is cts, inc, and concave



D^*_β is cts, dec, and convex

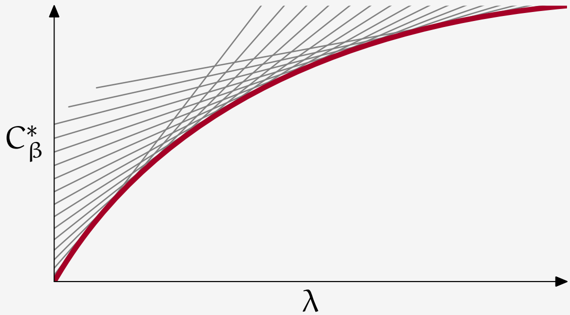
Optimization problems

Costly communication

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

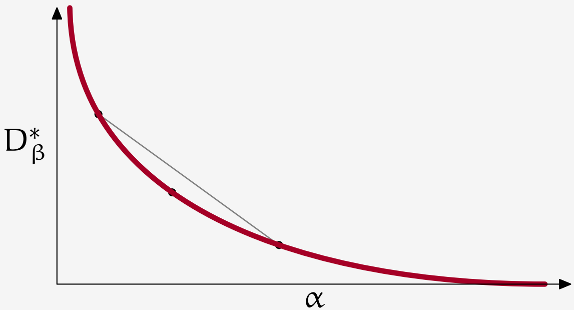
Constrained communication

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



C^*_β is cts, inc, and concave

Distortion-transmission function



D^*_β is cts, dec, and convex

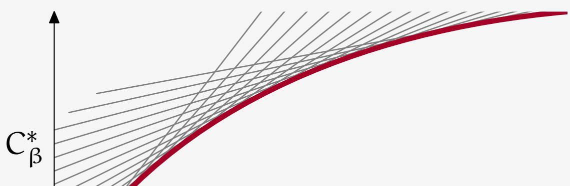
Optimization problems

Costly communication

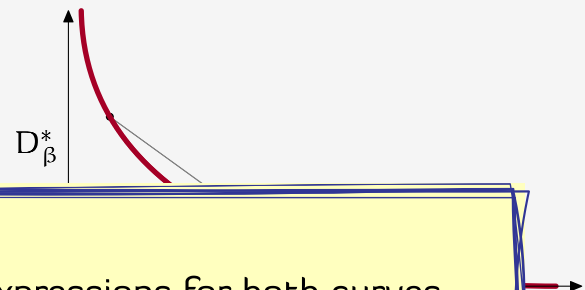
For $\lambda \in \mathbb{R}_{>0}$, $\quad C^*_\beta(\lambda) = C_\beta(f^*, g^*; \lambda) \coloneqq \inf_{(f,g)} \{D_\beta(f,g) + \lambda N_\beta(f,g)\}$

Constrained communication

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



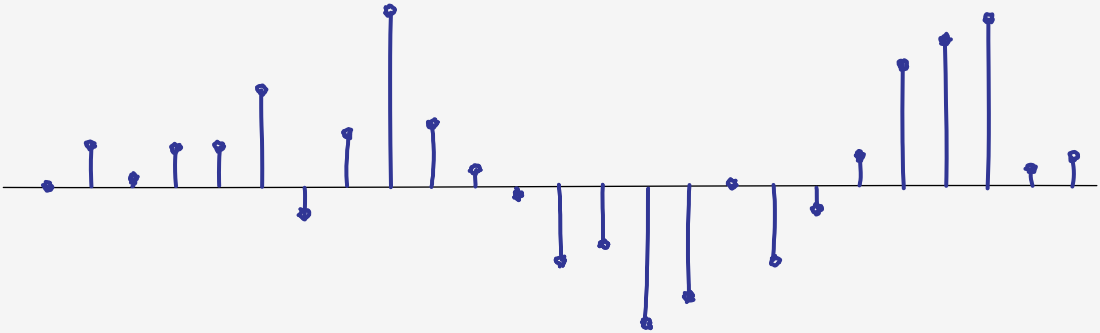
Distortion-transmission function



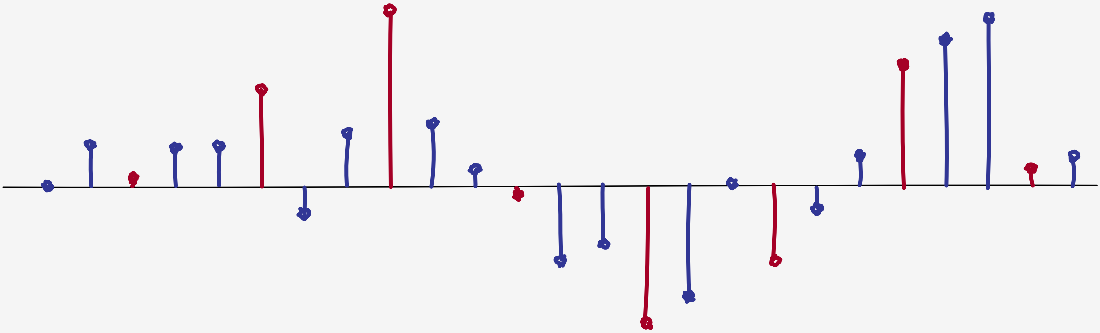
We provide explicit computable expressions for both curves

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

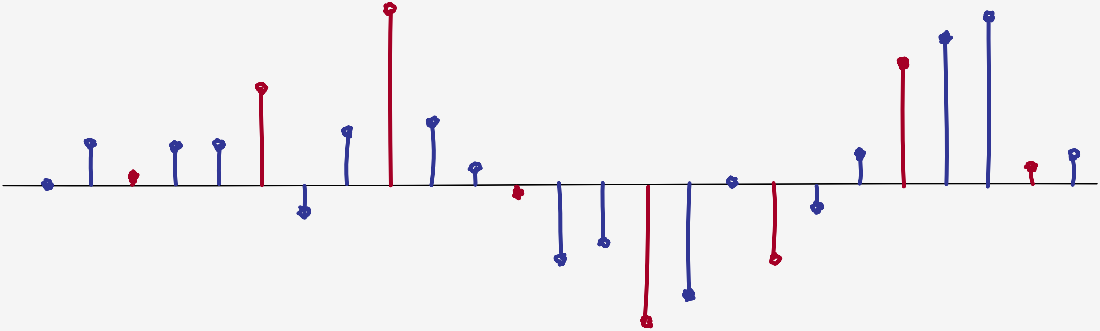
$$X_{t+1} = X_t + W_t,\ W_t \sim \mathcal{N}(0, 1)$$



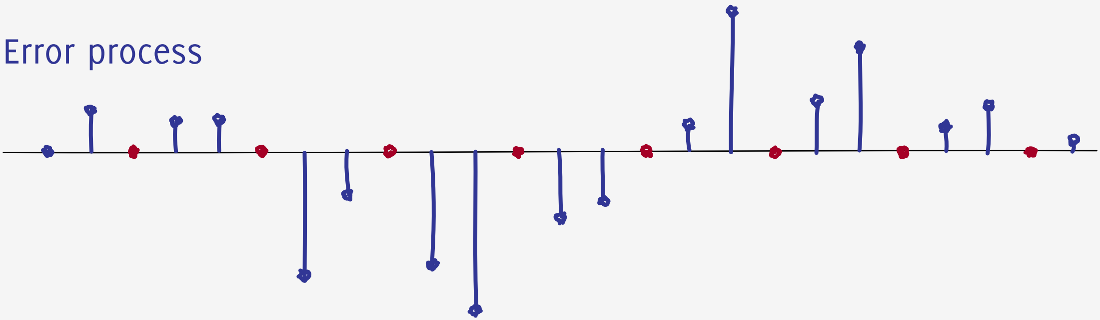
Periodic transmission strategy



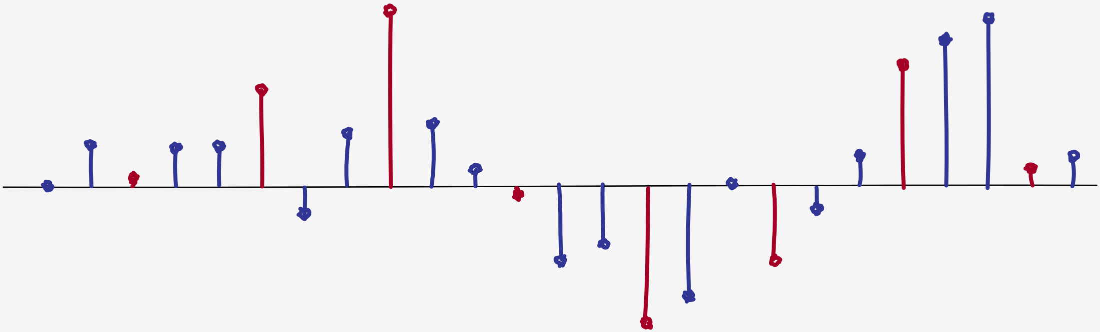
Periodic transmission strategy



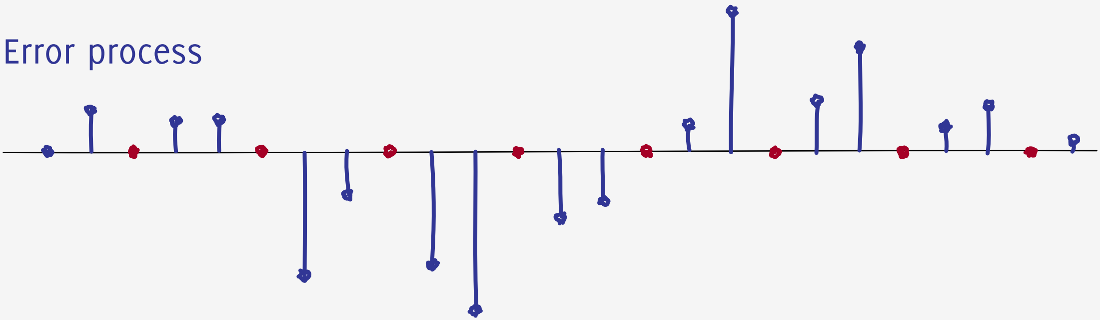
Error process



Periodic transmission strategy

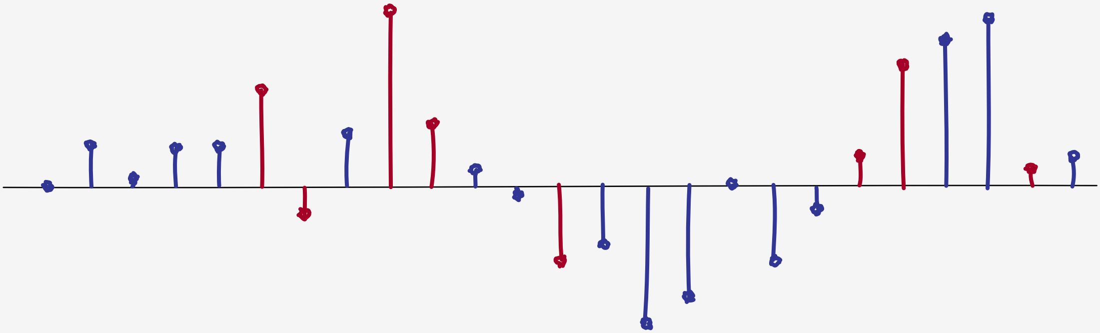


Error process

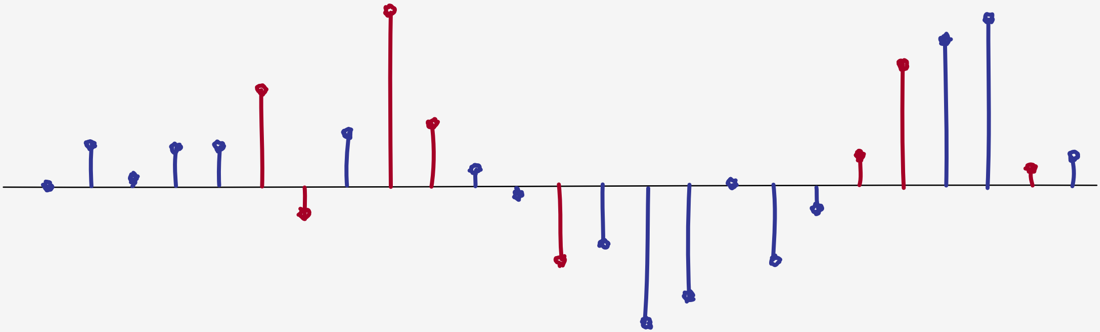


$D = 0.69 \quad N \approx 1/3$

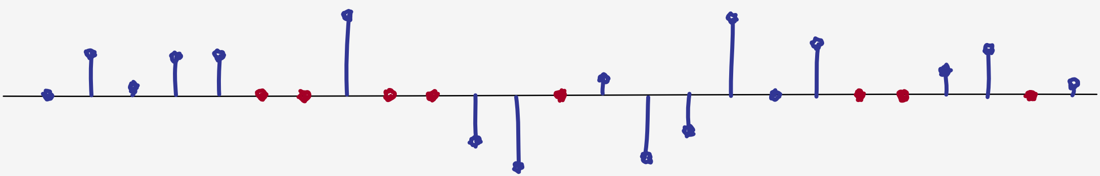
An alternative strategy



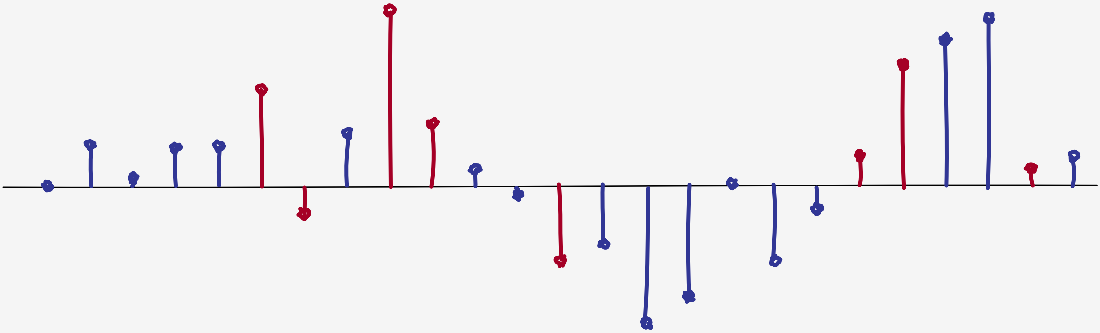
An alternative strategy



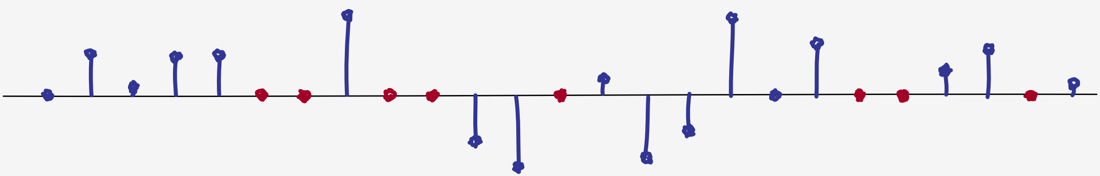
Error process



An alternative strategy

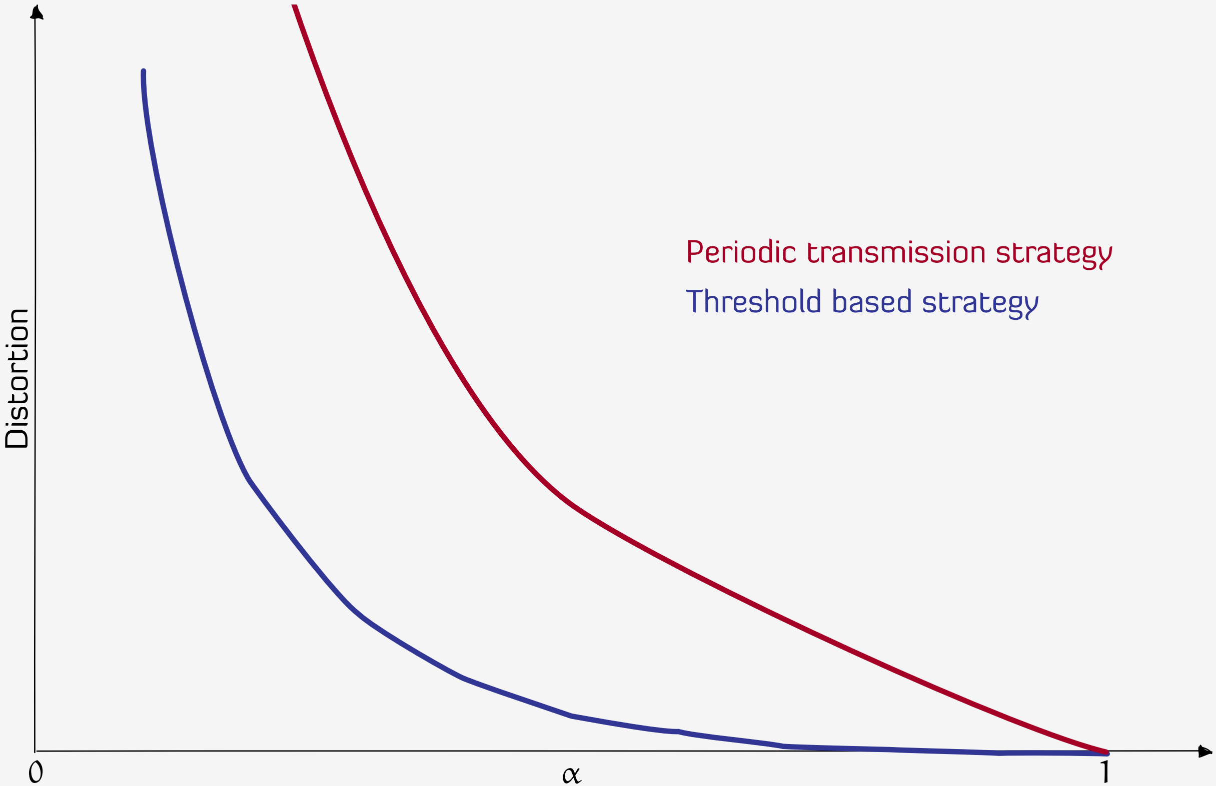


Error process



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

Distortion-transmission function



Identify strategies that achieve the optimal trade-off

Simple and intuitive threshold based strategies are optimal.

Provide computable expressions for distortion-transmission function

Based on simple matrix calculations for discrete Markov processes

Based on solving Fredholm integral equations for Gaussian processes

Identify strategies that achieve the optimal trade-off

Simple and intuitive threshold based strategies are optimal.

Provide computable expressions for distortion-transmission function

Based on simple matrix calculations for discrete Markov processes

Based on solving Fredholm integral equations for Gaussian processes

Beautiful example of stochastics and optimization

Decentralized stochastic control and POMDPs

Stochastic orders and majorization

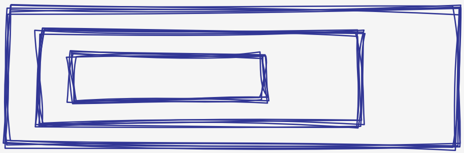
Markov chain analysis, stopping times, and renewal theory

Constrained MDPs and Lagrangian relaxations

So how do we start?

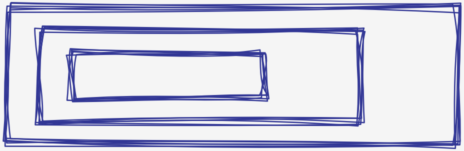
Decentralized stochastic control

Dealing with non-classical information structure



Classical info. struct.

Dealing with non-classical information structure

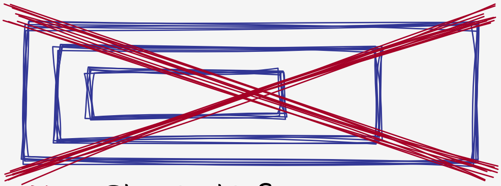


Classical info. struct.

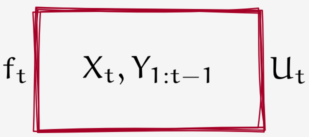
f_t $X_t, Y_{1:t-1}$ U_t

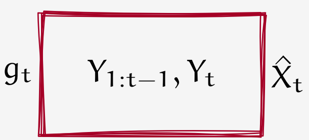
g_t $Y_{1:t-1}, Y_t$ $\hat{X}_t$

Dealing with non-classical information structure

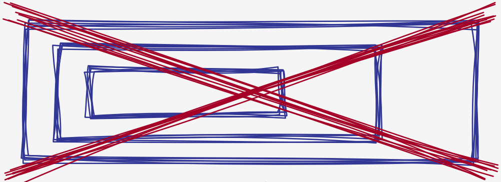


Non-Classical info. struct.





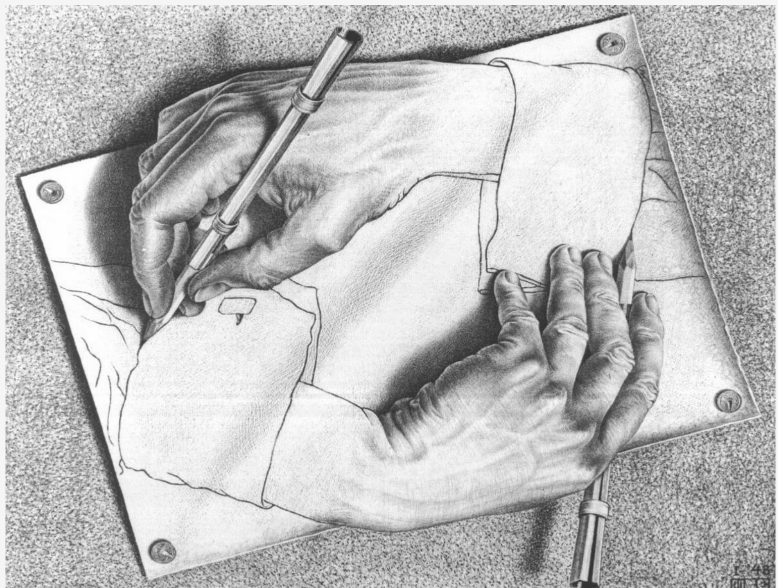
Dealing with non-classical information structure



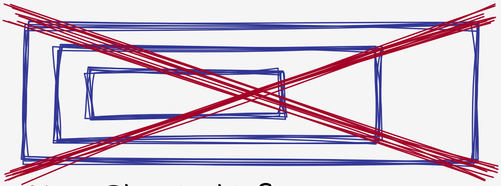
Non-Classical info. struct.

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

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



Dealing with non-classical information structure



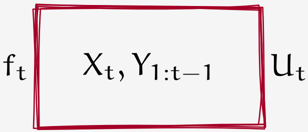
Non-Classical info. struct.

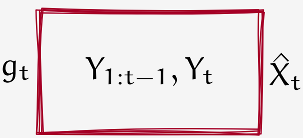
Common info $C_t \coloneqq \bigcap_{s \geq t} \bigcap_{i=1}^{n} I_s^i$

Local info $L_t^i \coloneqq I_t^i \setminus C_t$

$g(C, L) = \psi(C)(L)$

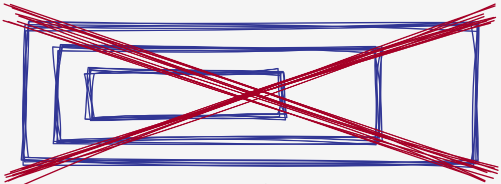
Belongs to the class of **tractable non-classical information structures** (called partial-history sharing) identified in [Mahajan-Nayyar-Teneketzis 2013]





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

Dealing with non-classical information structure



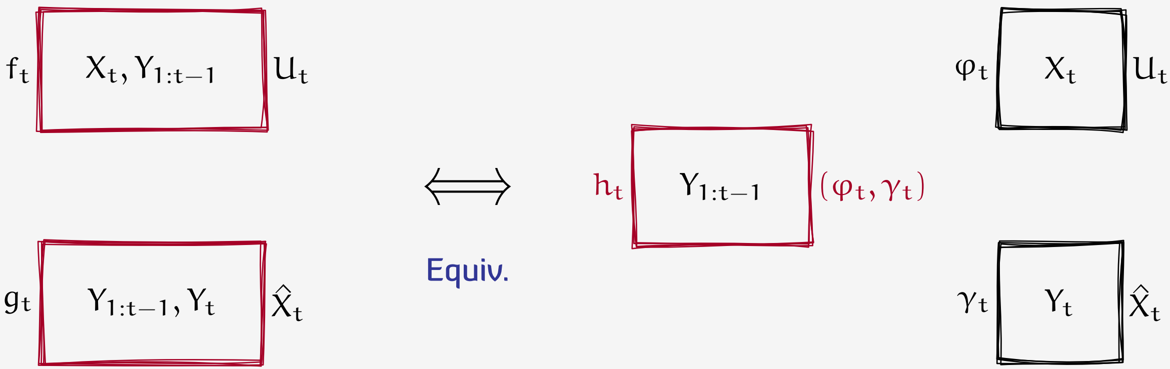
Non-Classical info. struct.

Common info $C_t := \bigcap_{s \geq t} \bigcap_{i=1}^{n} I_s^i$

Local info $L_t^i := I_t^i \setminus C_t$

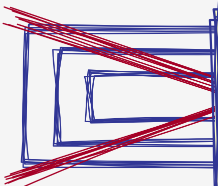
$$g(C, L) = \psi(C)(L)$$

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



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

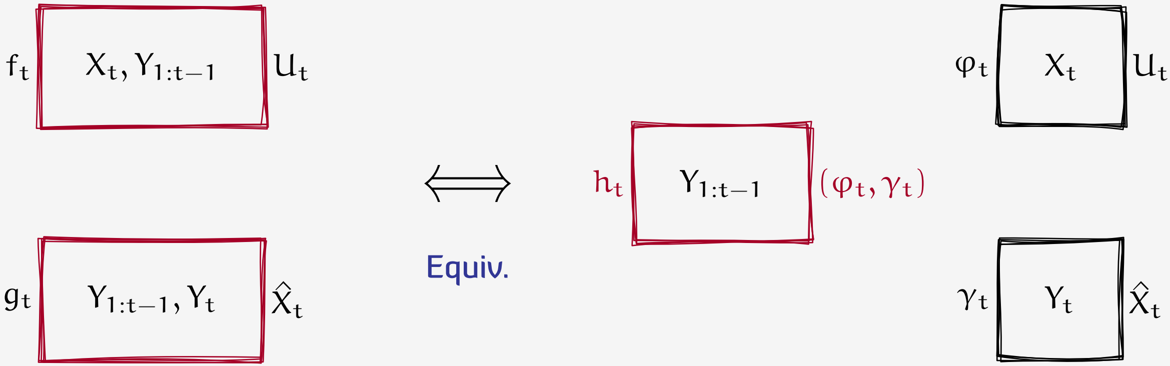
Dealing with non-classical information structure



Non-Classica

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

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



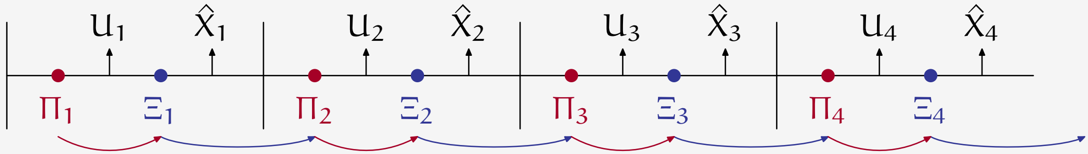
Nayyar, Mahajan and Teneketzis, "Decentralized stochastic control with partial history sharing: A common information approach," IEEE TAC 2013.

Information states and dynamic program

Information states

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

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

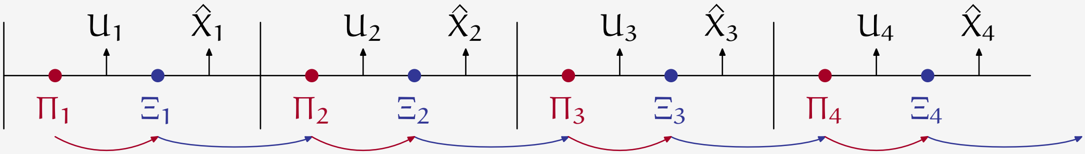


Information states and dynamic program

Information states

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

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



Structural results

There is no loss of optimality in using

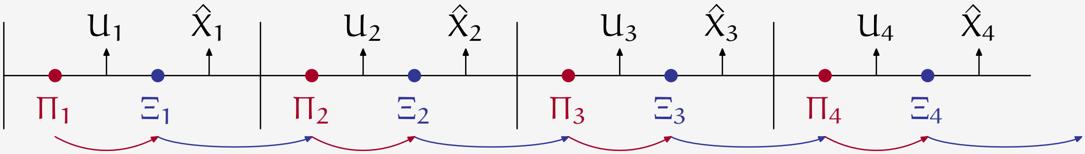
$$U_t = f_t(X_t, \Pi_t) \quad \text{and} \quad \hat{X}_t = g_t(\Xi_t).$$

Information states and dynamic program

Information states

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

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



Structural results

There is no loss of optimality in using

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

Dynamic Program

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

and for $t = T, \dots, 0$

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

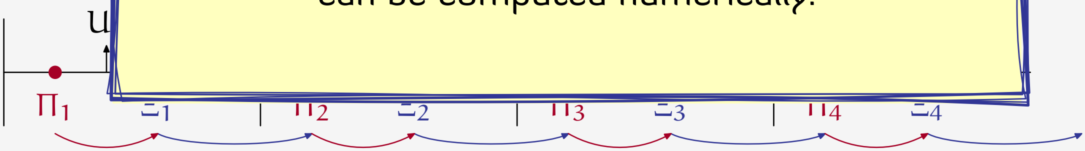
$$W_t(\pi) = \min_{\varphi : \mathcal{X} \to \{0,1\}} \mathbb{E}[\lambda \varphi(X_t) + V_t(\Xi_t) \mid \Pi_t = \pi, \varphi_t = \varphi].$$

Information states and dynamic program

Informa... $x \mid Y_{1:t-1})$.

$x \mid Y_{1:t})$.

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



Structural results

There is no loss of optimality in using

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

Dynamic Program

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

and for $t = T, \dots, 0$

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

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

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

Simplifying modeling assumptions

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

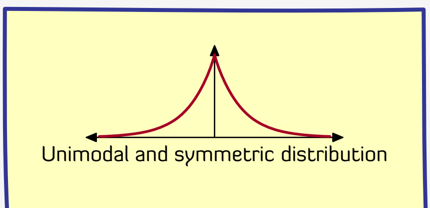
Simplifying modeling assumptions

Markov process	$X_{t+1} = X_t + W_t$	
	Markov chain setup	**Guass-Markov setup**
State spaces	$X_t, W_t \in \mathbb{Z}$	$X_t, W_t \in \mathbb{R}$

Simplifying modeling assumptions

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

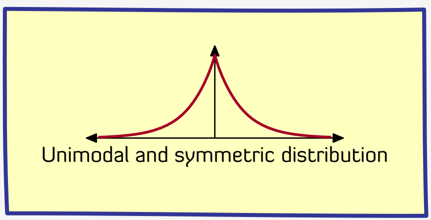
	Markov chain setup	**Guass-Markov setup**
State spaces	$X_t, W_t \in \mathbb{Z}$	$X_t, W_t \in \mathbb{R}$
Noise distribution	Unimodal and symmetric $p_e = p_{-e} \geqslant p_{e+1}$	Zero-mean Gaussian $\varphi_\sigma(\cdot)$

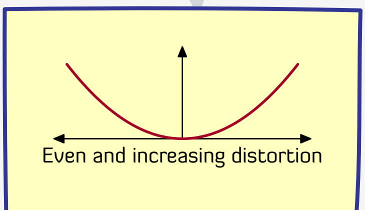


Simplifying modeling assumptions

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

	Markov chain setup	**Guass-Markov setup**
State spaces	$X_t, W_t \in \mathbb{Z}$	$X_t, W_t \in \mathbb{R}$
Noise distribution	Unimodal and symmetric $p_e = p_{-e} \geqslant p_{e+1}$	Zero-mean Gaussian $\varphi_\sigma(\cdot)$
Distortion	Even and increasing $d(e) = d(-e) \leqslant d(e+1)$	Mean-squared $d(e) = \lvert e \rvert^2$





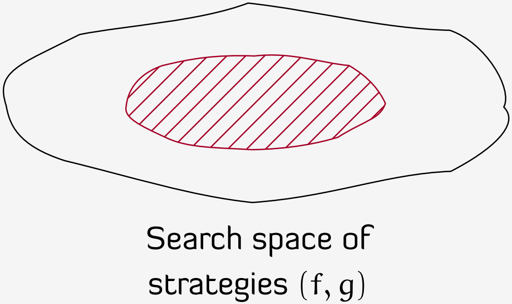
Step 1 Structure of optimal strategies

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

Step 3 Optimal costly comm.

Step 4 Distortion-transmission trade-off

Step 1 Structure of optimal strategies

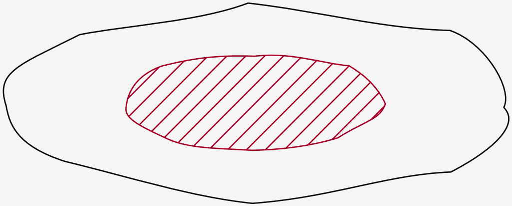


Step 2 Performance of arbitrary threshold strategies $f^{(k)}$

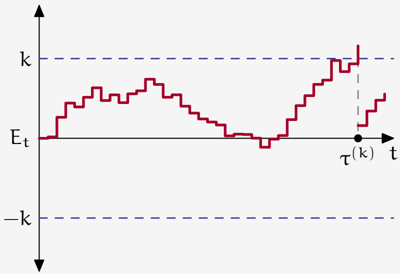
Step 3 Optimal costly comm.

Step 4 Distortion-transmission trade-off

Step 1 Structure of optimal strategies



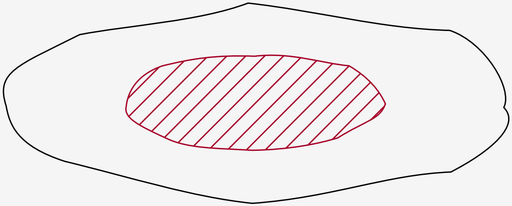
Step 2 Performance of arbitrary threshold strategies $f^{(k)}$



Step 3 Optimal costly comm.

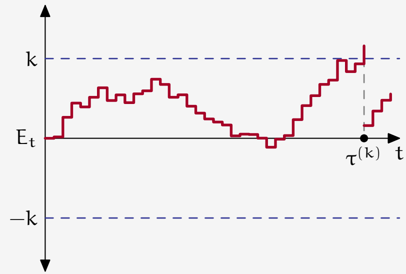
Step 4 Distortion-transmission trade-off

Step 1 Structure of optimal strategies

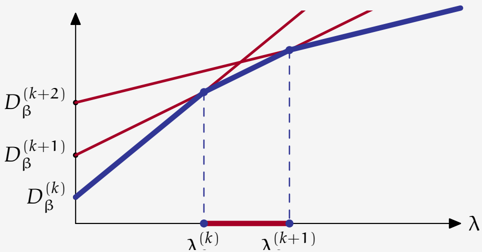


Search space of strategies (f, g)

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

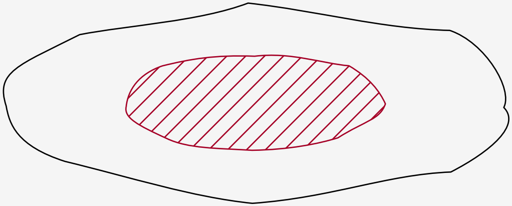


Step 3 Optimal costly comm.



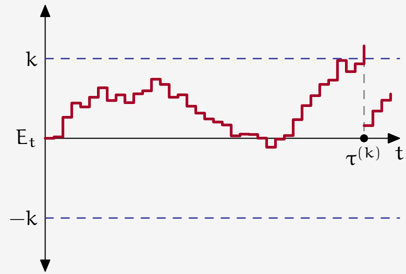
Step 4 Distortion-transmission trade-off

Step 1 Structure of optimal strategies

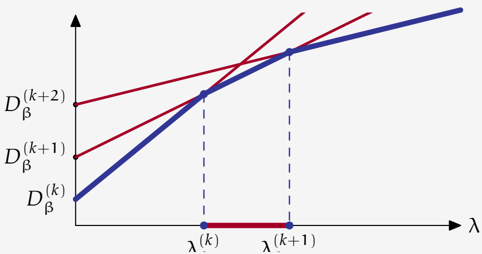


Search space of strategies (f, g)

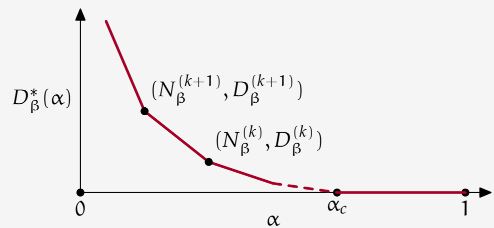
Step 2 Performance of arbitrary threshold strategies $f^{(k)}$



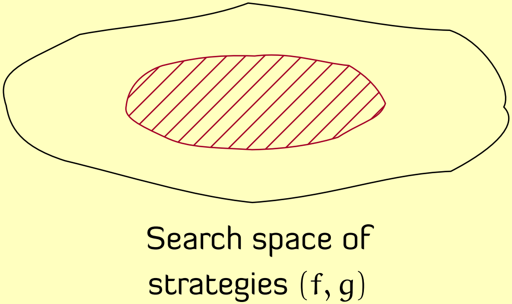
Step 3 Optimal costly comm.



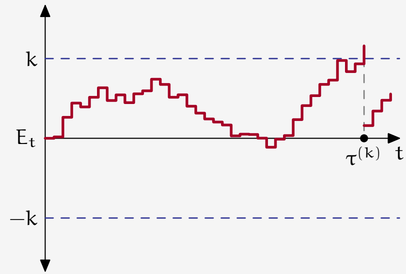
Step 4 Distortion-transmission trade-off



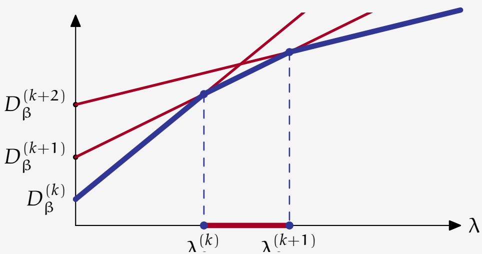
Step 1 Structure of optimal strategies



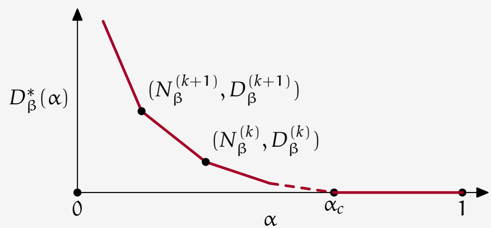
Step 2 Performance of arbitrary threshold strategies $f^{(k)}$



Step 3 Optimal costly comm.



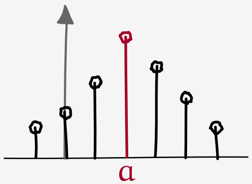
Step 4 Distortion-transmission trade-off



Preliminaries

[LM11, NBTV13]

Almost uniform and unimodal (ASU) distribution about a

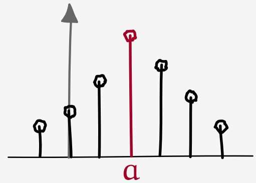


$\pi_a \geqslant \pi_{a+1} \geqslant \pi_{a-1} \geqslant \pi_{a+2} \geqslant \cdots$

Preliminaries

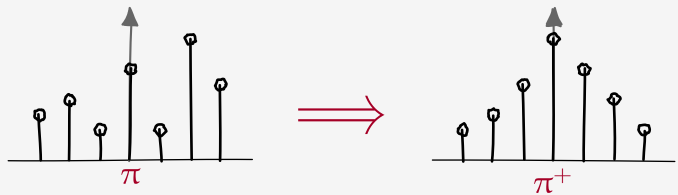
[LM11, NBTV13]

Almost uniform and unimodal (ASU) distribution about a



$$\pi_a \geqslant \pi_{a+1} \geqslant \pi_{a-1} \geqslant \pi_{a+2} \geqslant \cdots$$

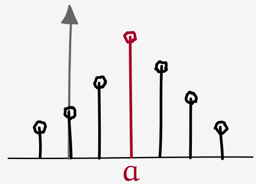
ASU Rearrangement



Preliminaries

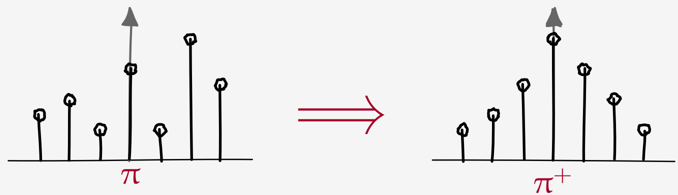
[LM11, NBTV13]

Almost uniform and unimodal (ASU) distribution about a

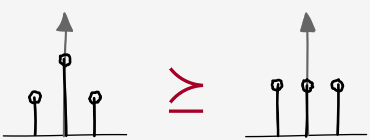


$$\pi_a \geqslant \pi_{a+1} \geqslant \pi_{a-1} \geqslant \pi_{a+2} \geqslant \cdots$$

ASU Rearrangement



Majorization



$\pi \succeq \xi$ iff

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

Invariant to permutations.

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

Definition $\xi \triangleright \tilde{\xi}$ if $\xi \succeq \tilde{\xi}$ and $\tilde{\xi}$ is ASU about some point a

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

Definition $\xi \triangleright \tilde{\xi}$ if $\xi \succeq \tilde{\xi}$ and $\tilde{\xi}$ is ASU about some point a

Lemma
Similar to Schur-concavity

- If $\xi \triangleright \tilde{\xi}$ then $W_t(\xi) \geqslant W_t(\tilde{\xi})$.
- If $\pi \triangleright \tilde{\pi}$ then $V_t(\pi) \geqslant V_t(\tilde{\pi})$.

Properties of the value functions [LM11, NBTV13]

Definition

$\xi \triangleright \tilde{\xi}$ if $\xi \succeq \tilde{\xi}$ and $\tilde{\xi}$ is ASU about some point a

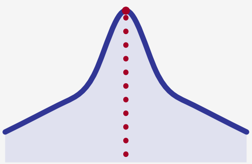
Lemma
Similar to Schur-concavity

- If $\xi \triangleright \tilde{\xi}$ then $W_t(\xi) \geqslant W_t(\tilde{\xi})$.
- If $\pi \triangleright \tilde{\pi}$ then $V_t(\pi) \geqslant V_t(\tilde{\pi})$.

Lemma (Arg min of W)

If ξ is ASU about a then a is the arg min of

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



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

Definition $\xi \rhd \tilde{\xi}$ if $\xi \succeq \tilde{\xi}$ and $\tilde{\xi}$ is ASU about some point a

Lemma
Similar to Schur-concavity

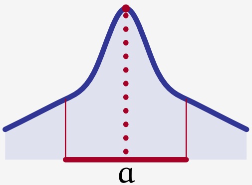
- If $\xi \rhd \tilde{\xi}$ then $W_t(\xi) \geqslant W_t(\tilde{\xi})$.
- If $\pi \rhd \tilde{\pi}$ then $V_t(\pi) \geqslant V_t(\tilde{\pi})$.

Lemma (Arg min of W) If ξ is ASU about a then a is the arg min of

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

Lemma (Arg min of V) If π is ASU about a then the arg min of

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

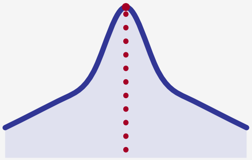


is of the form

$$\varphi(x) = \begin{cases} 1, & \text{if } |x - a| > k(\pi) \\ 0, & \text{if } |x - a| < k(\pi) \\ q_+, & \text{if } x - a = k(\pi) \\ q_-, & \text{if } x - a = -k(\pi) \end{cases}$$

Step 1 Structure of optimal strategies

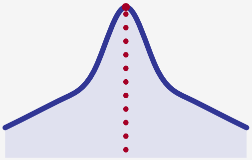
[LM11, NBTV13]



π_1 is ASU about z_0

Step 1 Structure of optimal strategies

[LM11, NBTV13]

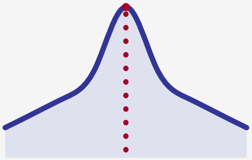


π_1 is ASU about z_0

Is $|x_1 - z_0| > k_1$?

Structure of optimal strategies

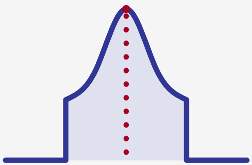
[LM11, NBTV13]



π_1 is ASU about z_0

Is $|x_1 - z_0| > k_1$?

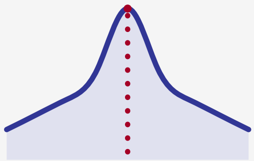
NO. $u_1 = \varepsilon$, $z_1 = z_0$



ξ_1 is ASU about z_1

Step 1 Structure of optimal strategies

[LM11, NBTV13]

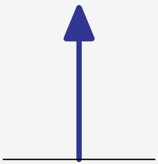


z_0

π_1 is ASU about z_0

Is $|x_1 - z_0| > k_1$?

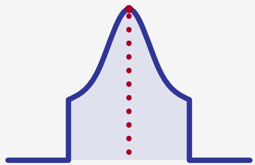
YES. $u_1 = 1$, $z_1 = x_1$



z_1

ξ_1 is ASU about z_1

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

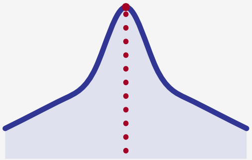


z_1

ξ_1 is ASU about z_1

Step 1 Structure of optimal strategies

[LM11, NBTV13]

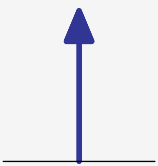


z_0

π_1 is ASU about z_0

Is $|x_1 - z_0| > k_1$?

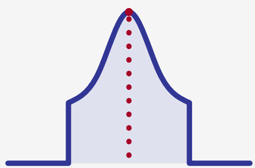
YES. $u_1 = 1$, $z_1 = x_1$



z_1

ξ_1 is ASU about z_1

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



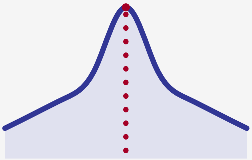
z_1

ξ_1 is ASU about z_1

In both cases: $\hat{x}_1 = z_1$

Step 1 Structure of optimal strategies

[LM11, NBTV13]

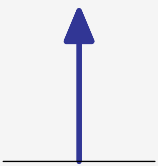


z_0

π_1 is ASU about z_0

Is $|x_1 - z_0| > k_1$?

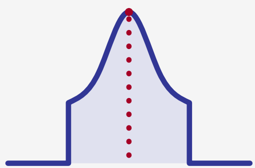
YES. $u_1 = 1$, $z_1 = x_1$



z_1

ξ_1 is ASU about z_1

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



z_1

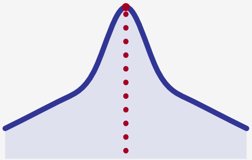
ξ_1 is ASU about z_1

In both cases: $\hat{x}_1 = z_1$

$t = 2$

Step 1 Structure of optimal strategies

[LM11, NBTV13]

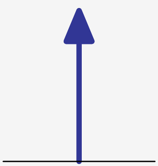


z_0

π_1 is ASU about z_0

Is $|x_1 - z_0| > k_1$?

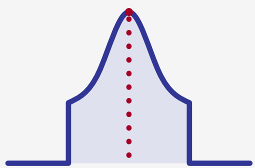
YES. $u_1 = 1$, $z_1 = x_1$



z_1

ξ_1 is ASU about z_1

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



z_1

ξ_1 is ASU about z_1

In both cases: $\hat{x}_1 = z_1$

$t = 2$

$$X_2 = X_1 + W_1 \implies \pi_1 = \xi_1 * p$$

π_1 is ASU about z_1

etc. . . .

Step 1 Structure of optimal estimator

[LM11, NBTV13]

Transmitted Process

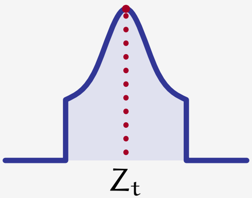
Let Z_t denote the most recently transmitted value of the Markov process.

Step 1 Structure of optimal estimator

[LM11, NBTV13]

Transmitted Process

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



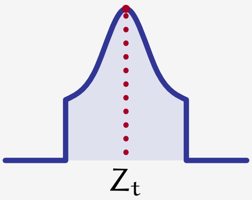
Lemma

Ξ_t is ASU about Z_t

Step 1 Structure of optimal estimator [LM11, NBTV13]

Transmitted Process

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



Lemma

Ξ_t is ASU about Z_t

Theorem

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

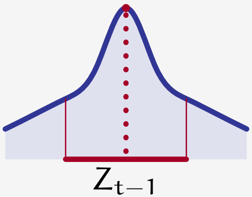
Remark

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

Step 1 Structure of optimal transmitter

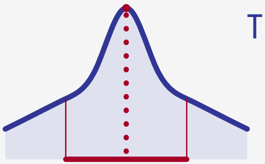
[LM11, NBTV13]

Lemma Π_t is ASU about Z_{t-1}



Step 1 Structure of optimal transmitter [LM11, NBTV13]

Lemma Π_t is ASU about Z_{t-1}

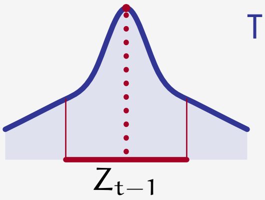


Theorem
$$U_t = f_t(X_t, \Pi_t) = \begin{cases} 1, & \text{if } |X_t - E_t| \geqslant k_t \\ 0, & \text{if } |X_t - E_t| < k_t \end{cases}$$

Step 1 Structure of optimal transmitter

[LM11, NBTV13]

Lemma Π_t is ASU about Z_{t-1}



Theorem

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

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

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

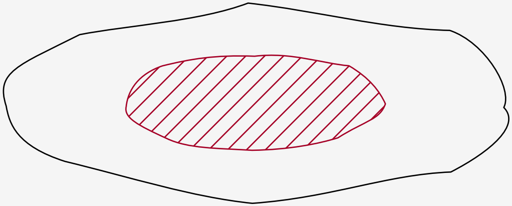
Remark The optimal transmission strategy is a function of the error process.

The results extend to infinite horizon setup under appropriate regularity conditions.

Time-homogeneous **threshold-based strategies are optimal.**

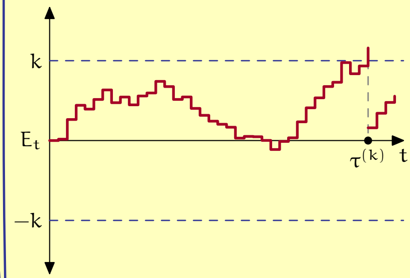
How do we find the optimal threshold-based strategy?

Step 1 Structure of optimal strategies

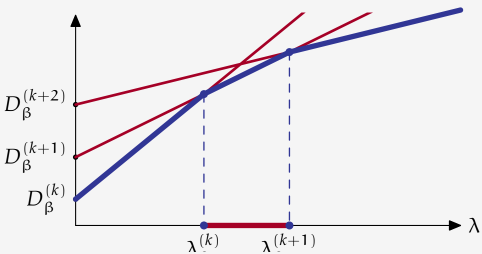


Search space of strategies (f, g)

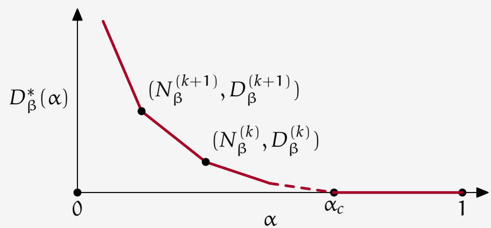
Step 2 Performance of arbitrary threshold strategies $f^{(k)}$



Step 3 Optimal costly comm.



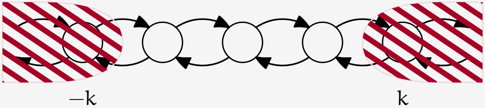
Step 4 Distortion-transmission trade-off



Step 2 Performance of threshold strategies

Consider a threshold-based strategy

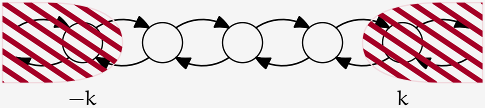
$$f^{(k)}(e) = \begin{cases} 1 & \text{if } |e| \geq k \\ 0 & \text{otherwise} \end{cases}$$



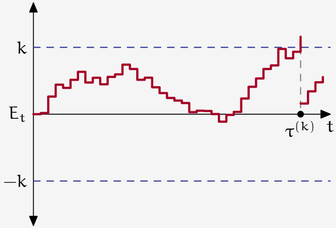
Step 2 Performance of threshold strategies

Consider a threshold-based strategy

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



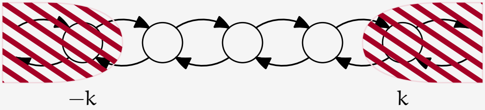
Let $\tau^{(k)}$ denote the stopping time of first transmission (starting at $E_0 = 0$).



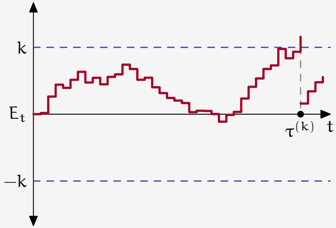
Step 2 Performance of threshold strategies

Consider a threshold-based strategy

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



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



Define

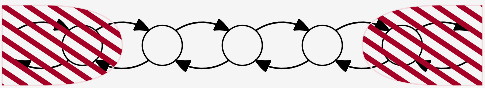
$$L_\beta^{(k)}(e) = (1-\beta)\,\mathbb{E}\Big[\sum_{t=0}^{\tau^{(k)}-1} \beta^t d(E_t) \Big| E_0 = e\Big].$$

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

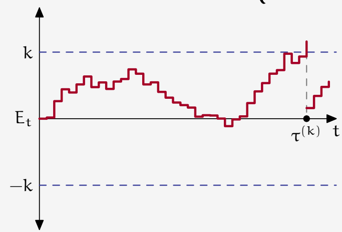
Step 2 Performance of threshold strategies

Consider a threshold-based strategy

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



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



Define

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

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

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

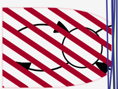
$$D_\beta^{(k)} \coloneqq D_\beta(f^{(k)}, g^*) = \frac{L_\beta^{(k)}(0)}{M_\beta^{(k)}(0)} \quad \text{and} \quad N_\beta^{(k)} \coloneqq N_\beta(f^{(k)}, g^*) = \frac{1}{M_\beta^{(k)}(0)} - (1-\beta).$$

Step 2 Performance of threshold strategies

Consid… of

…0)).

$f^{(k)}$



$-k$

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

Define

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

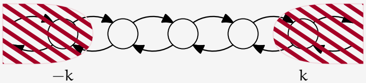
$$M^{(k)}_{\beta}(e) = (1-\beta)\,\mathbb{E}\Big[\sum_{t=0}^{\tau^{(k)}-1} \beta^t \Big| E_0 = e\Big].$$

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

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

Step 2 Computing $L_\beta^{(k)}$ and $M_\beta^{(k)}$

Markov chain setup

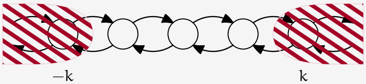


$$L_\beta^{(k)}(e) = d(e) + \beta \sum_{n=-k}^{k} p_{n-e} L_\beta^{(k)}(n)$$

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

Step 2 Computing $L_\beta^{(k)}$ and $M_\beta^{(k)}$

Markov chain setup



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

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

Proposition

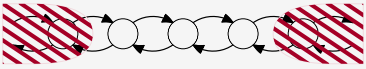
$$L_\beta^{(k)} = \big[[I - \beta P^{(k)}]^{-1} d^{(k)}\big].$$

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

$$M_\beta^{(k)} = \big[[I - \beta P^{(k)}]^{-1} \mathbf{1}^{(k)}\big].$$

Step 2 Computing $L_\beta^{(k)}$ and $M_\beta^{(k)}$

Markov chain setup



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

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

Proposition

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

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

$$M_\beta^{(k)} = \big[[I - \beta P^{(k)}]^{-1} \mathbf{1}^{(k)}\big].$$

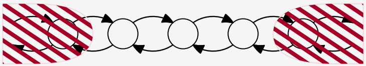
Gauss-Markov setup

$$L_\beta^{(k)}(e) = d(e) + \beta \int_{-k}^{k} \varphi(n-e) L_\beta^{(k)}(n) dn$$

$$M_\beta^{(k)}(e) = 1 + \beta \int_{-k}^{k} \varphi(n-e) M_\beta^{(k)}(n) dn$$

Step 2 Computing $L_\beta^{(k)}$ and $M_\beta^{(k)}$

Markov chain setup



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

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

Proposition

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

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

$$M_\beta^{(k)} = \left[[I - \beta P^{(k)}]^{-1} \mathbf{1}^{(k)}\right].$$

Gauss-Markov setup

$$L_\beta^{(k)}(e) = d(e) + \beta \int_{-k}^{k} \varphi(n-e) L_\beta^{(k)}(n) dn$$

$$M_\beta^{(k)}(e) = 1 + \beta \int_{-k}^{k} \varphi(n-e) M_\beta^{(k)}(n) dn$$

Fredholm Integral Equations of the 2nd kind.

Solutions exist and are unique.

Step 2 Computing $L_\beta^{(k)}$ and $M_\beta^{(k)}$

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

Proposition

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

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

$$M_\beta^{(k)} = \big[[I - \beta P^{(k)}]^{-1} \mathbf{1}^{(k)}\big].$$

Gauss-Markov setup

$$L_\beta^{(k)}(e) = d(e) + \beta \int_{-k}^{k} \varphi(n - e) L_\beta^{(k)}(n)\, dn$$

$$M_\beta^{(k)}(e) = 1 + \beta \int_{-k}^{k} \varphi(n - e) M_\beta^{(k)}(n)\, dn$$

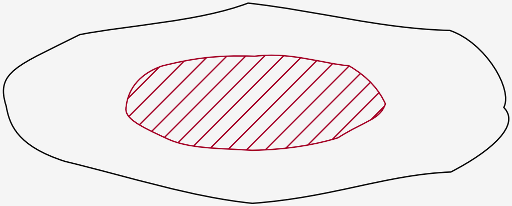
Fredholm Integral Equations of the 2nd kind.

Solutions exist and are unique.

We found the performance of a generic threshold-based strategy

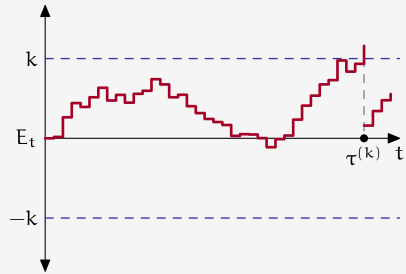
How does this lead to identifying an optimal strategy?

Step 1 Structure of optimal strategies

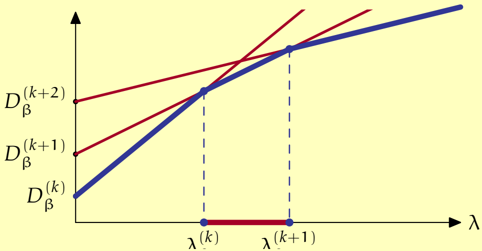


Search space of strategies (f, g)

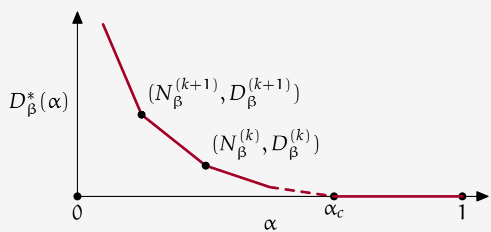
Step 2 Performance of arbitrary threshold strategies $f^{(k)}$



Step 3 Optimal costly comm.



Step 4 Distortion-transmission trade-off



Step 3 Properties of optimal thresholds

Monotonicity $L_\beta^{(k+1)} > L_\beta^{(k)}$ and $M_\beta^{(k+1)} > M_\beta^{(k)}$

Depends on unimodularity of noise

Step 3 Properties of optimal thresholds

Monotonicity $L_\beta^{(k+1)} > L_\beta^{(k)}$ and $M_\beta^{(k+1)} > M_\beta^{(k)}$

Use DP and monotonicity of Bellman operator

Implication:

$D_\beta^{(k+1)} \geqslant D_\beta^{(k)}$ and $N_\beta^{(k+1)} < N_\beta^{(k)}$

Step 3 Properties of optimal thresholds

Monotonicity $L_\beta^{(k+1)} > L_\beta^{(k)}$ and $M_\beta^{(k+1)} > M_\beta^{(k)}$

Implication:

$D_\beta^{(k+1)} \geqslant D_\beta^{(k)}$ and $N_\beta^{(k+1)} < N_\beta^{(k)}$

Submodularity $C_\beta^{(k)}(\lambda) \coloneqq D_\beta^{(k)} + \lambda N_\beta^{(k)}$ is submodular in (k, λ).

Step 3 Properties of optimal thresholds

Monotonicity $L_\beta^{(k+1)} > L_\beta^{(k)}$ and $M_\beta^{(k+1)} > M_\beta^{(k)}$

Implication:

$D_\beta^{(k+1)} \geqslant D_\beta^{(k)}$ and $N_\beta^{(k+1)} < N_\beta^{(k)}$

Submodularity $C_\beta^{(k)}(\lambda) \coloneqq D_\beta^{(k)} + \lambda N_\beta^{(k)}$ is submodular in (k, λ).

Proposition $k_\beta^*(\lambda) \coloneqq \arg\min_{k \in \mathbb{Z}_{\geqslant 0}} C_\beta^{(k)}(\lambda)$ is increasing in λ.

Step 3 Properties of optimal thresholds

Monotonicity $L_\beta^{(k+1)} > L_\beta^{(k)}$ and $M_\beta^{(k+1)} > M_\beta^{(k)}$

Implication:

$$D_\beta^{(k+1)} \geqslant D_\beta^{(k)} \quad \text{and} \quad N_\beta^{(k+1)} < N_\beta^{(k)}$$

Submodularity $C_\beta^{(k)}(\lambda) \coloneqq D_\beta^{(k)} + \lambda N_\beta^{(k)}$ is submodular in (k, λ).

Proposition $k_\beta^*(\lambda) \coloneqq \arg\min_{k \in \mathbb{Z}_{\geqslant 0}} C_\beta^{(k)}(\lambda)$ is increasing in λ.

Thus, optimal threshold increases with increase in λ.

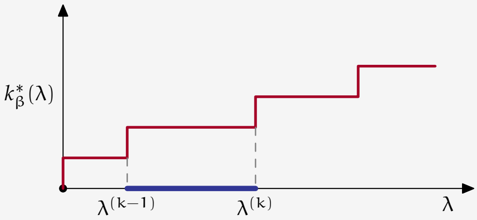
Characterizing the optimal threshold for a given communication cost **is tricky.**

Instead, we will characterize the optimal communication cost for a given threshold.

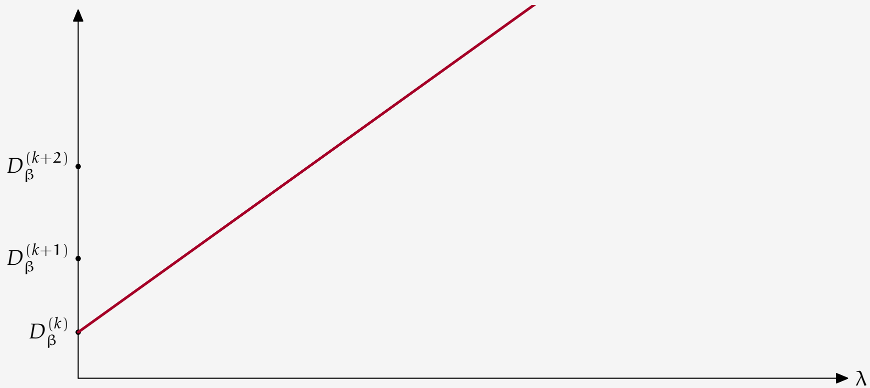
Optimal costly communication: Markov chain

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

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

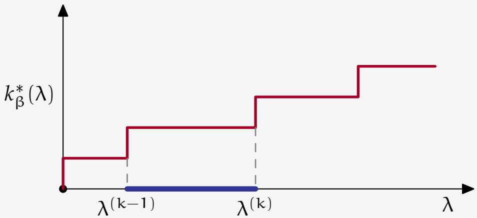


Optimal costly communication: Markov chain

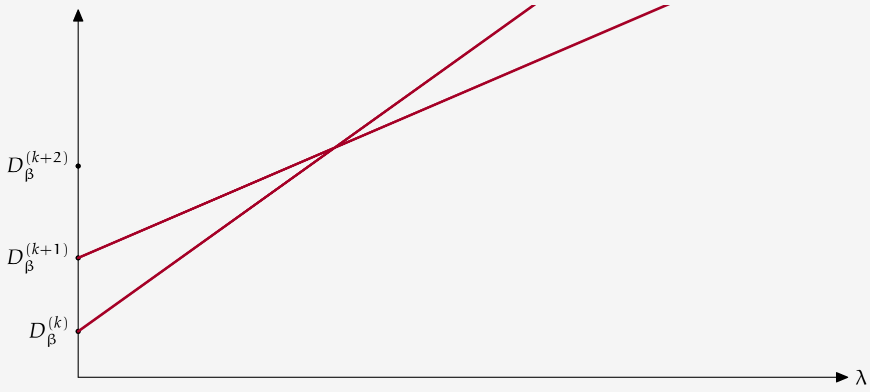


Define $\Lambda_\beta^{(k)} := \{\lambda \in \mathbb{R}_{\geqslant 0} : k_\beta^*(\lambda) = k\}$
$= [\lambda_\beta^{(k-1)}, \lambda_\beta^{(k)}]$.

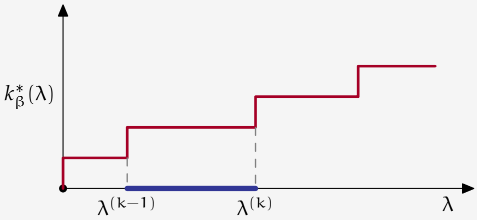
$C_\beta^{(k)}(\lambda_\beta^{(k)}) = C_\beta^{(k+1)}(\lambda_\beta^{(k)})$



Step 3 Optimal costly communication: Markov chain

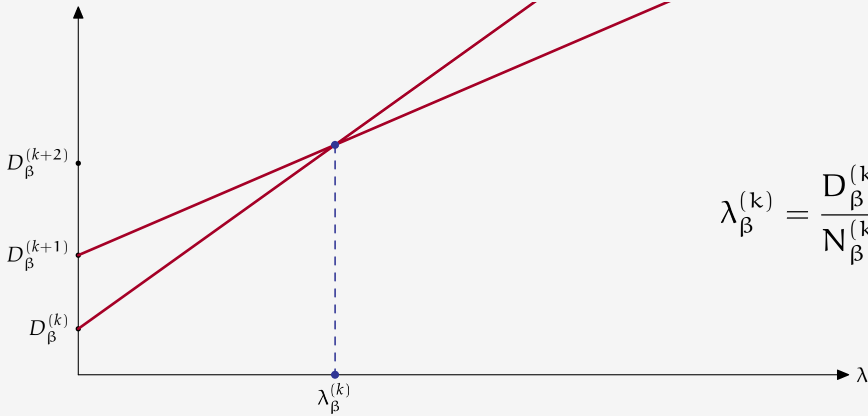


Define $\Lambda_\beta^{(k)} \coloneqq \{\lambda \in \mathbb{R}_{\geqslant 0} : k_\beta^*(\lambda) = k\}$
$= [\lambda_\beta^{(k-1)}, \lambda_\beta^{(k)}]$.



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

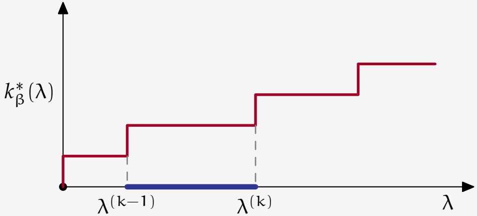
Step 3 Optimal costly communication: Markov chain



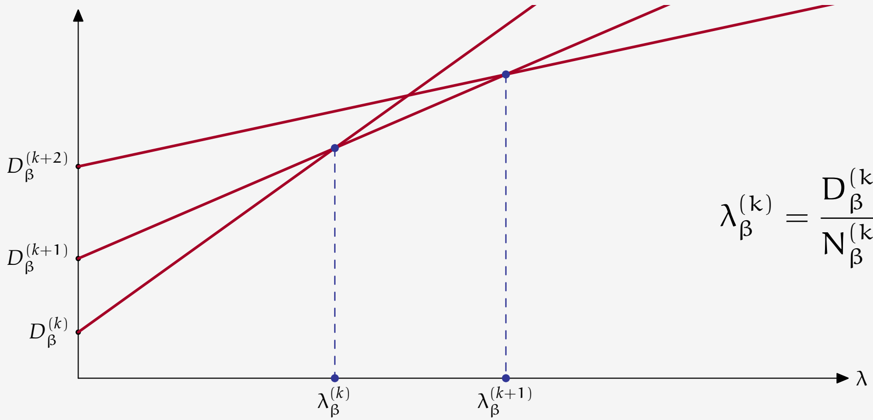
$$\lambda_\beta^{(k)} = \frac{D_\beta^{(k+1)} - D_\beta^{(k)}}{N_\beta^{(k)} - N_\beta^{(k+1)}}$$

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

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



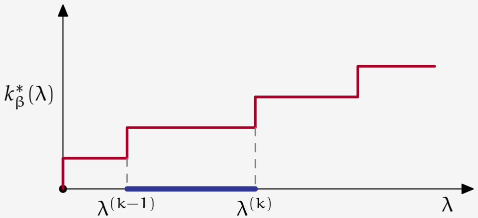
Optimal costly communication: Markov chain



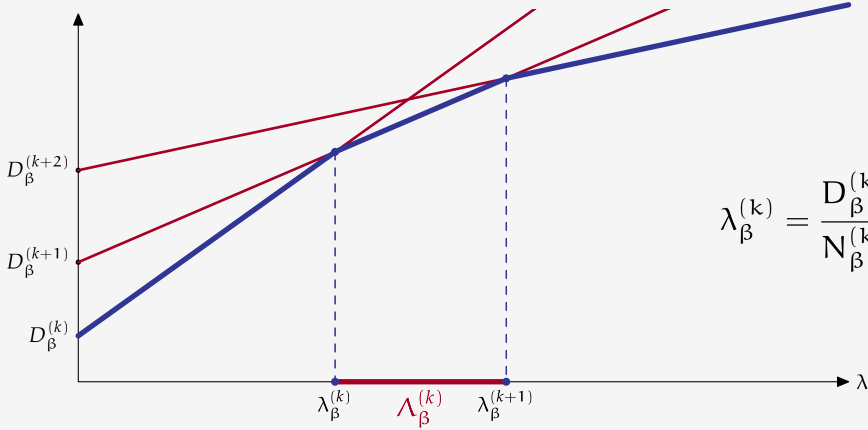
$$\lambda_\beta^{(k)} = \frac{D_\beta^{(k+1)} - D_\beta^{(k)}}{N_\beta^{(k)} - N_\beta^{(k+1)}}$$

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

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



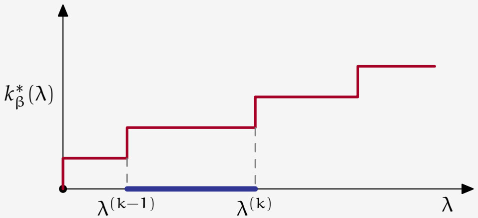
Step 3 Optimal costly communication: Markov chain



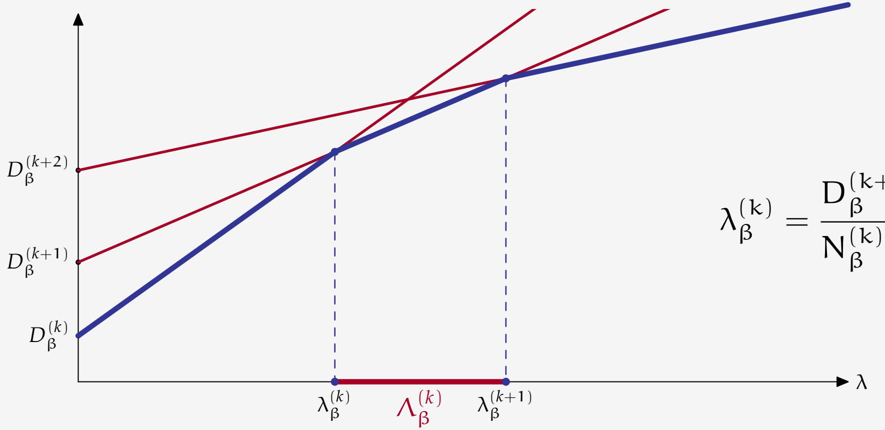
$$\lambda_\beta^{(k)} = \frac{D_\beta^{(k+1)} - D_\beta^{(k)}}{N_\beta^{(k)} - N_\beta^{(k+1)}}$$

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

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



Step 3 Optimal costly communication: Markov chain

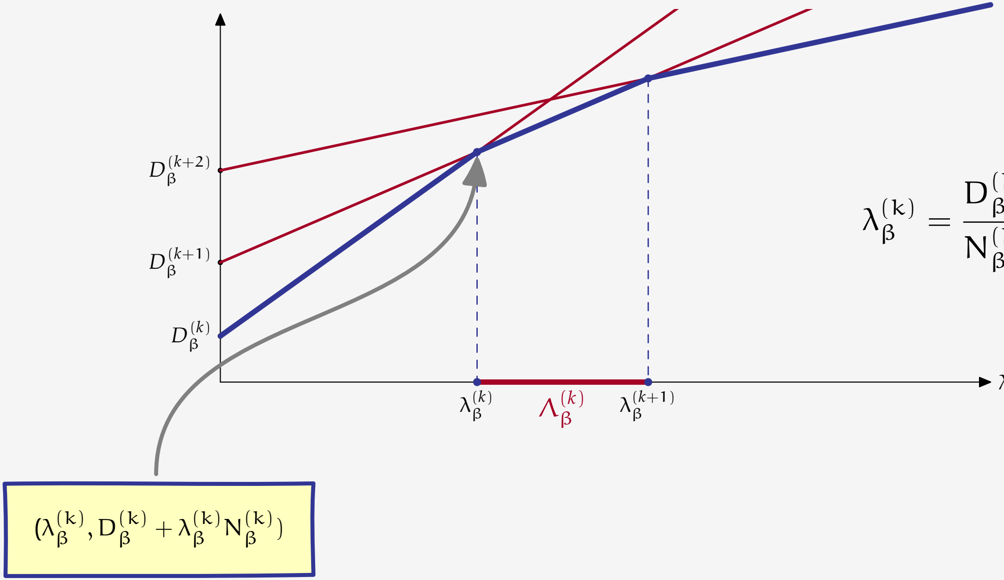


$$\lambda_\beta^{(k)} = \frac{D_\beta^{(k+1)} - D_\beta^{(k)}}{N_\beta^{(k)} - N_\beta^{(k+1)}}$$

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 3 Optimal costly communication: Markov chain

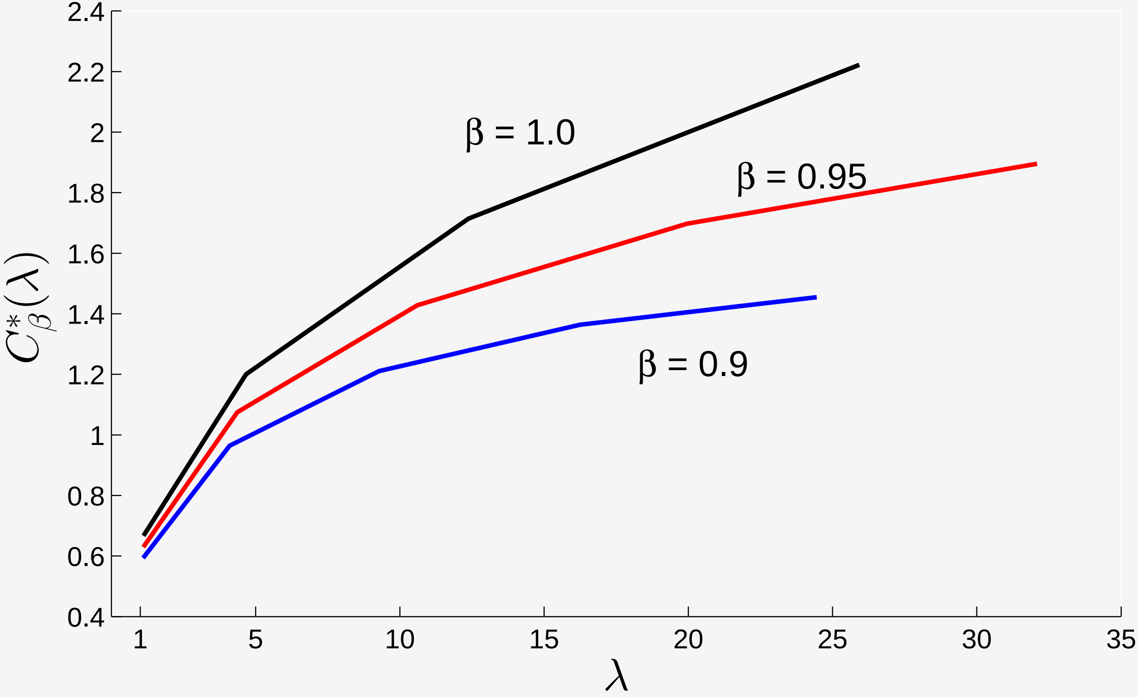


$$\lambda_\beta^{(k)} = \frac{D_\beta^{(k+1)} - D_\beta^{(k)}}{N_\beta^{(k)} - N_\beta^{(k+1)}}$$

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.

Example Symmetric birth-death Markov chain ($p = 0.3$)



Step 3 Optimal costly communication: Gauss-Markov

Lemma

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

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

Step 3 Optimal costly communication: Gauss-Markov

Lemma

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

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

Theorem

Strategy $f^{(k)}$ is optimal for $\lambda = -\dfrac{\partial_k D_\beta^{(k)}}{\partial_k N_\beta^{(k)}}$

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

Step 3 Optimal costly communication: Gauss-Markov

Lemma

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

$D^{(k)}_\beta$ and $N^{(k)}_\beta$ are differentiable in k.

Theorem

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

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

Scaling with variance σ^2

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

Step 3 Optimal costly communication: Gauss-Markov

Lemma

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

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

Theorem

Strategy $f^{(k)}$ is optimal for $\lambda = -\dfrac{\partial_k D_\beta^{(k)}}{\partial_k N_\beta^{(k)}}$

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

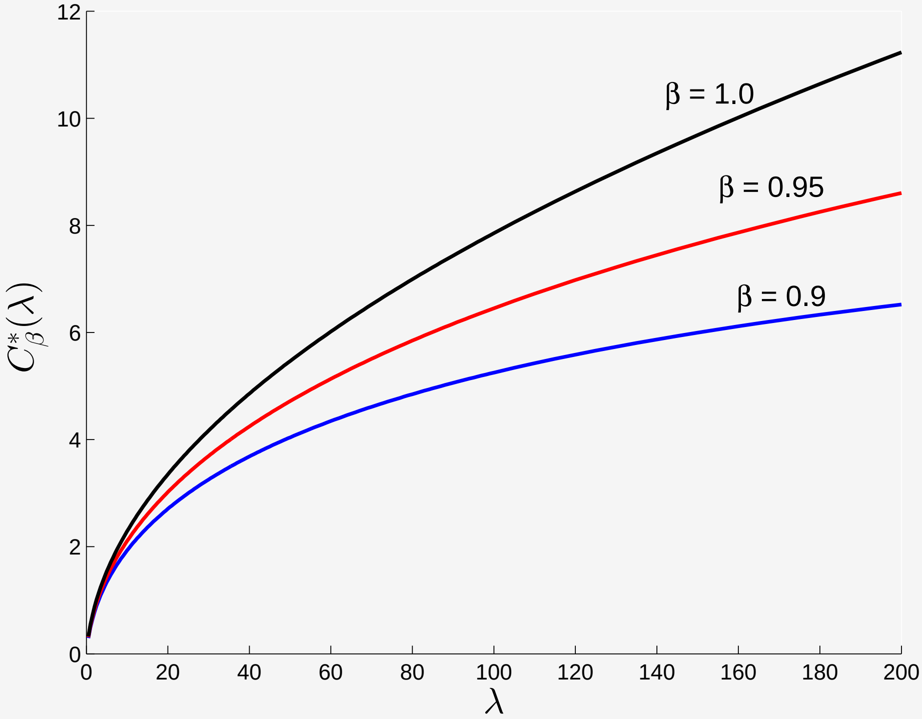
Scaling with variance σ^2

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

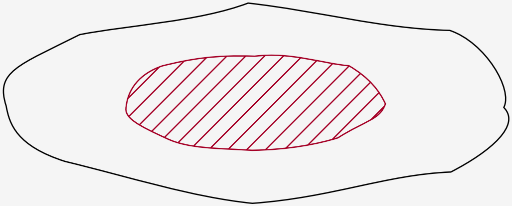
Computation

Use bisection search to find k such that $\lambda = -\dfrac{\partial_k D_\beta^{(k)}}{\partial_k N_\beta^{(k)}}$

Example Gauss-Markov with $\sigma^2 = 1$

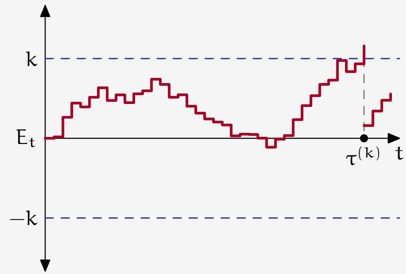


Step 1 Structure of optimal strategies

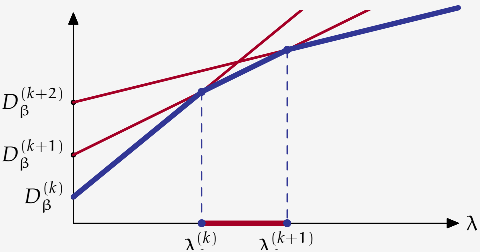


Search space of strategies (f, g)

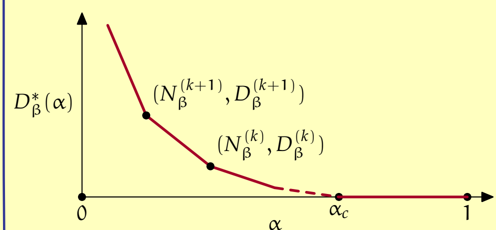
Step 2 Performance of arbitrary threshold strategies $f^{(k)}$



Step 3 Optimal costly comm.



Step 4 Distortion-transmission trade-off



Sufficient conditions for constrained optimality

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

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

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

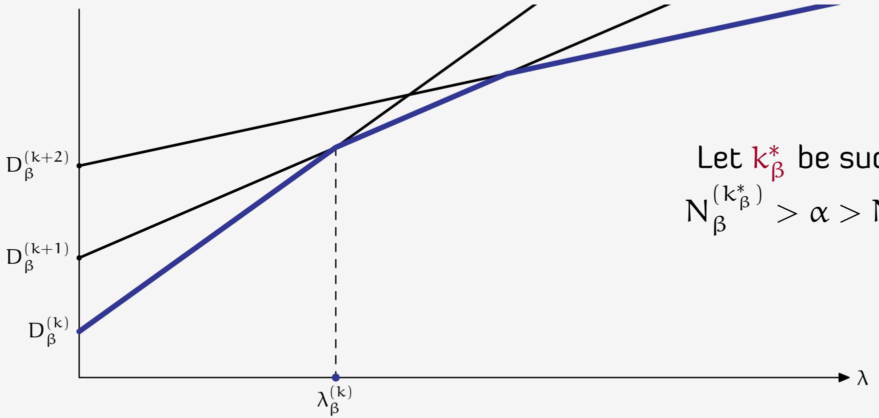
Step 4 Distortion-transmission trade-off: Markov chain

Sufficient conditions for constrained optimality

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

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

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



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

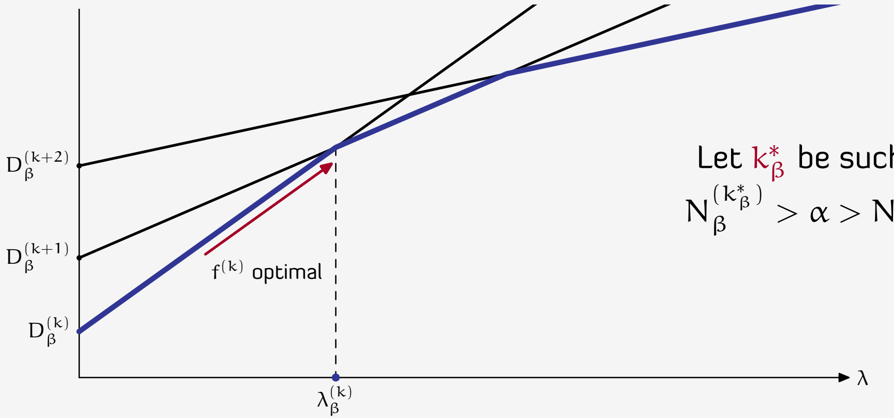
Distortion-transmission trade-off: Markov chain

Sufficient conditions for constrained optimality

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

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

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



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

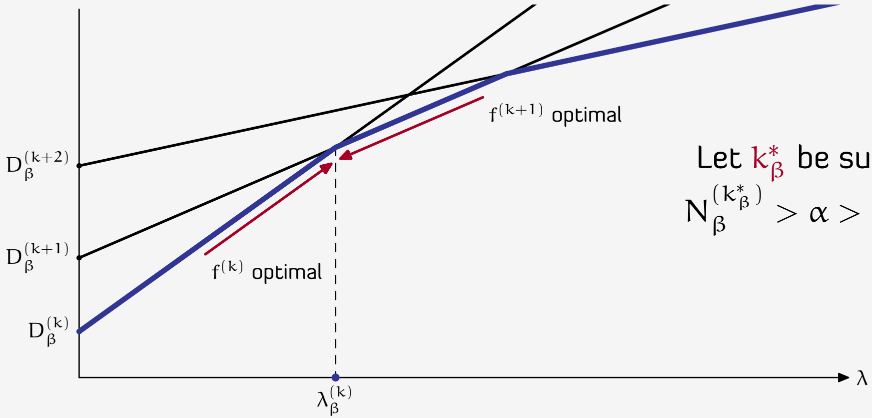
Step 4 Distortion-transmission trade-off: Markov chain

Sufficient conditions for constrained optimality

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

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

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



Let k^*_β be such that

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

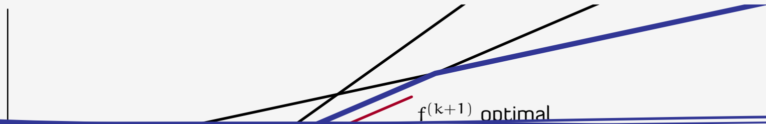
Distortion-transmission trade-off: Markov chain

Sufficient conditions for constrained optimality

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

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

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



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

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

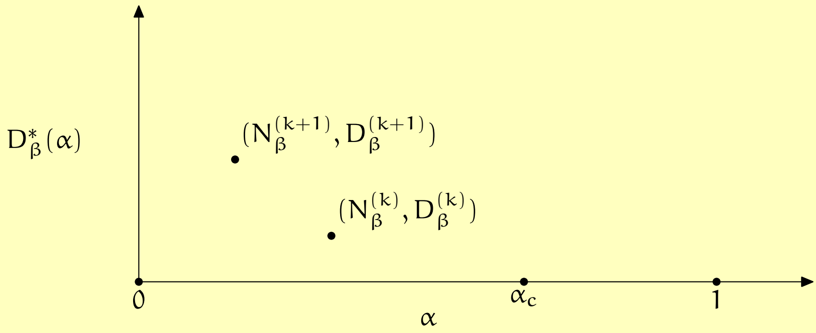
Distortion-transmission trade-off: Markov chain

Sufficient conditions for constrained optimality

A stra … m if

(…

(…).



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

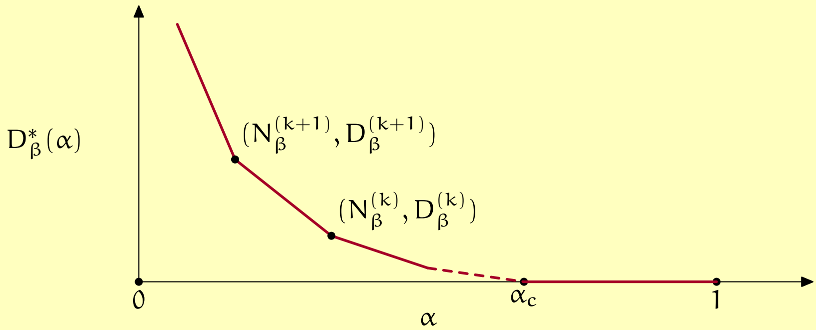
$$\theta^* N^{(k)}_\beta + (1-\theta^*) N^{(k+1)}_\beta = \alpha$$

Sufficient conditions for constrained optimality

A stra... m if

(

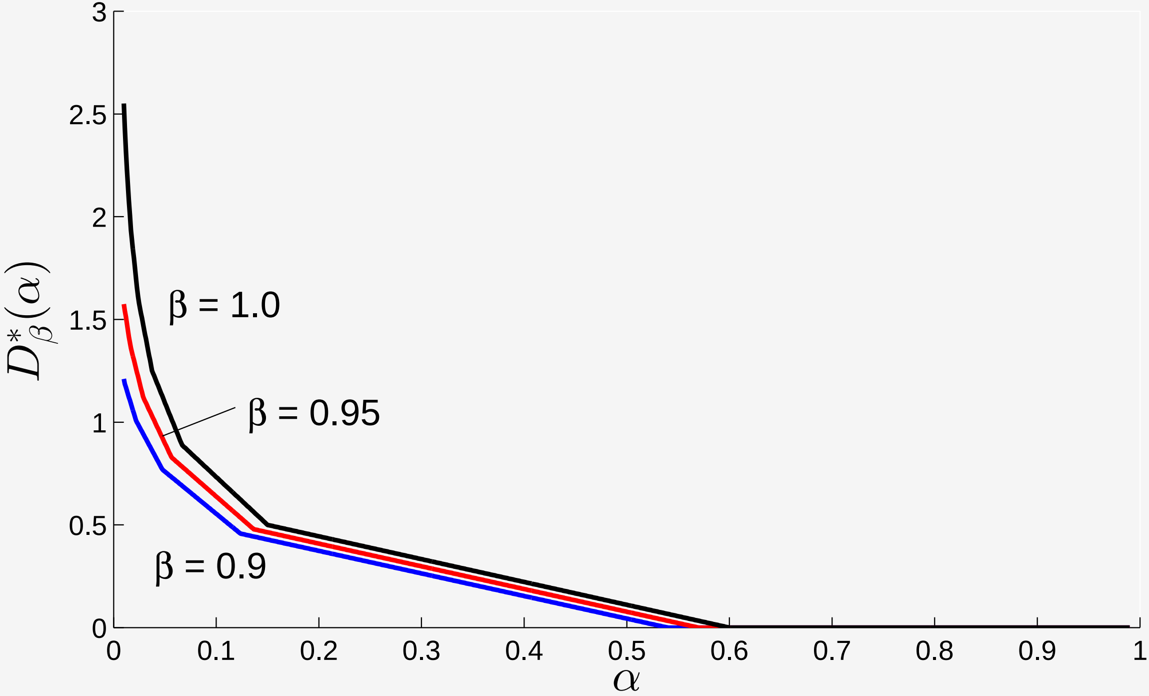
(...).



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

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

Example Symmetric birth-death Markov chain ($p = 0.3$)



Distortion-transmission trade-off: Gauss-Markov

Sufficient conditions for constrained optimality

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

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

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

Step 4 Distortion-transmission trade-off: Gauss-Markov

Sufficient conditions for constrained optimality

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

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

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

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

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

Step 4 Distortion-transmission trade-off: Gauss-Markov

Sufficient conditions for constrained optimality

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

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

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

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

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

Scaling with variance σ^2

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

Step 4 Distortion-transmission trade-off: Gauss-Markov

Sufficient conditions for constrained optimality

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

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

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

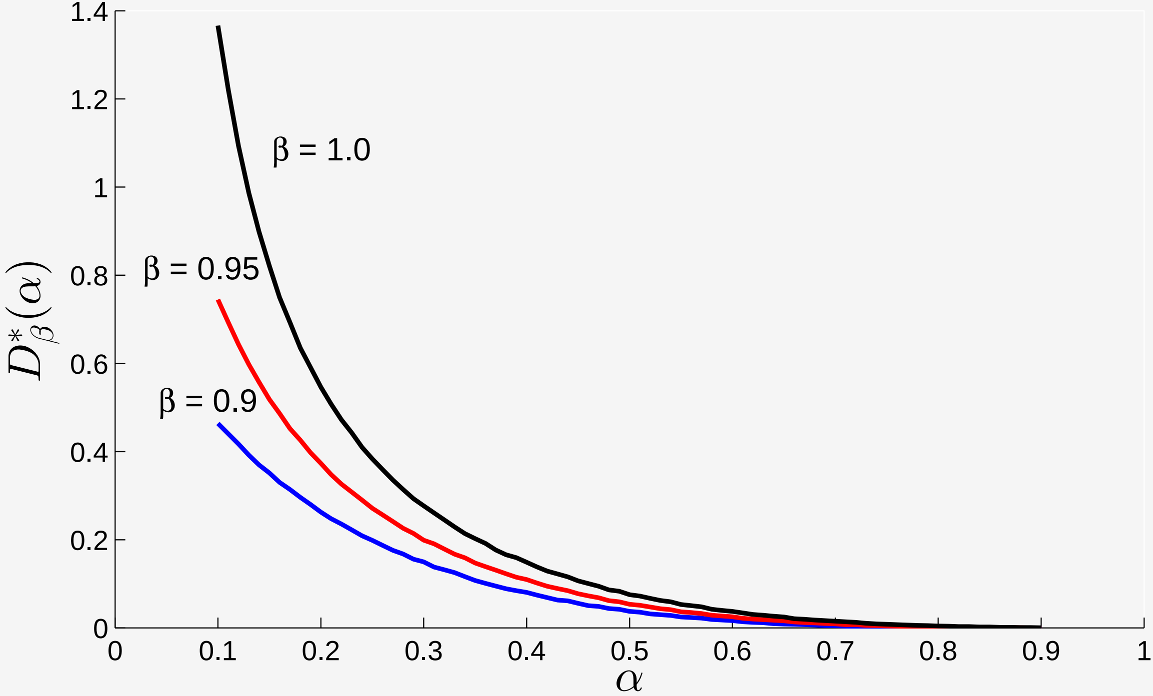
Theorem There exists a $k^*_\beta(\alpha)$ such that $N_\beta^{(k^*_\beta(\alpha))} = \alpha$. Therefore,

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

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

Computation Use bisection search to find k such that $N_\beta^{(k)} = \alpha$.

Example Gauss-Markov with $\sigma^2 = 1$



Conclusion

Analyze fundamental limits of estimation under communication constraints

Conclusion

Analyze fundamental limits of estimation under communication constraints

Possible generalizations to more realistic models

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

Conclusion

Analyze fundamental limits of estimation under communication constraints

Possible generalizations to more realistic models

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

A simple non-trivial "toy-problem" for decentralized control

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

Conclusion

Analyze fundamental limits of estimation under communication constraints

Possible generalizations to more realistic models

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

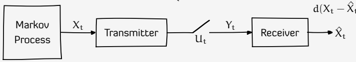
A simple non-trivial "toy-problem" for decentralized control

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

Full version available at arXiv:1505.04829.

The system model

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



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

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

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

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

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

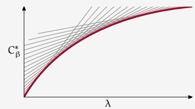
Optimization problems

Costly communication

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

Constrained communication

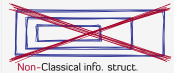
For $\alpha \in (0,1)$, $D^*_\beta(\alpha) := \inf_{(f,g)} \{D_\beta(f,g) : N_\beta(f,g) \leqslant \alpha\}$



C^*_β is cts, inc, and concave

D^*_β is cts, dec, and convex

Dealing with non-classical information structure



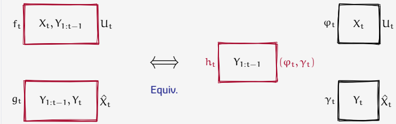
Non-Classical info. struct.

Common info $C_t := \bigcap_{s \geqslant t} \bigcap_{i=1}^{n} I^i_s$

Local info $L^i_t := I^i_t \setminus C_t$

$g(C, L) = \psi(C)(L)$

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



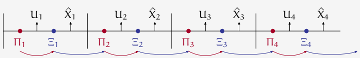
▶ Nayyar, Mahajan and Teneketzis, "Decentralized stochastic control with partial history sharing: A common information approach," IEEE TAC 2013.

Information states and dynamic program

Information states

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

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



Structural results

There is no loss of optimality in using

$U_t = f_t(X_t, \Pi_t)$ and $\hat{X}_t = g_t(\Xi_t)$.

Dynamic Program

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

and for $t = T, \dots, 0$

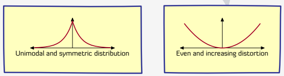
$$V_t(\xi) = \min_{\hat{x} \in \mathcal{X}} \mathbb{E}[d(X_t - \hat{x}) + W_{t+1}(\Pi_{t+1}) \mid \Xi_t = \xi],$$

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

Simplifying modeling assumptions

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

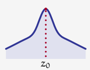
	Markov chain setup	Guass-Markov setup
State spaces	$X_t, W_t \in \mathbb{Z}$	$X_t, W_t \in \mathbb{R}$
Noise distribution	Unimodal and symmetric $p_e = p_{-e} \geqslant p_{e+1}$	Zero-mean Gaussian $\varphi_\sigma(\cdot)$
Distortion	Even and increasing $d(e) = d(-e) \leqslant d(e+1)$	Mean-squared $d(e) = \lvert e\rvert^2$



Step 1 Structure of optimal strategies

[LM11, NBTV13]

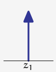
$t = 2$



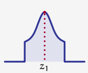
π_1 is ASU about z_0

Is $|x_1 - z_0| > k_1$?

YES. $u_1 = 1, z_1 = x_1$ NO. $u_1 = \varepsilon, z_1 = z_0$



ξ_1 is ASU about z_1



ξ_1 is ASU about z_1

In both cases: $\hat{x}_1 = z_1$

$X_2 = X_1 + W_1 \implies \pi_1 = \xi_1 * p$

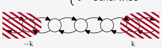
π_1 is ASU about z_1

etc. ...

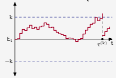
Step 2 Performance of threshold strategies

Consider a threshold-based strategy

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



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



Define

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

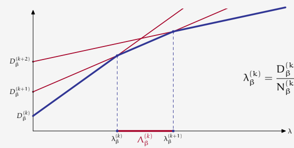
$$M^{(k)}_\beta(e) = (1-\beta)\,\mathbb{E}\Big[\sum_{t=0}^{\tau^{(k)}-1} \beta^t \Big| E_0 = e\Big].$$

Proposition

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

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

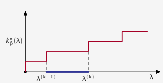
Step 3 Optimal costly communication: Markov chain



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

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

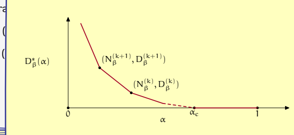
$C^{(k)}_\beta(\lambda^{(k)}_\beta) = C^{(k+1)}_\beta(\lambda^{(k)}_\beta)$



Step 4 Distortion-transmission trade-off: Markov chain

Sufficient conditions for constrained optimality

A stra[...]m if



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

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