# Uncertainty of Information: An Exploration of a Broader Notion of Timing in Networked Systems

Gongpu Chen

Department of Electrical and Electronic Engineering
Imperial College London

November, 2023

# Outline

1

## UoI of Binary Markov Sources

A Whittle Index Policy [1]

2

## UoI of Finite-state Markov Sources

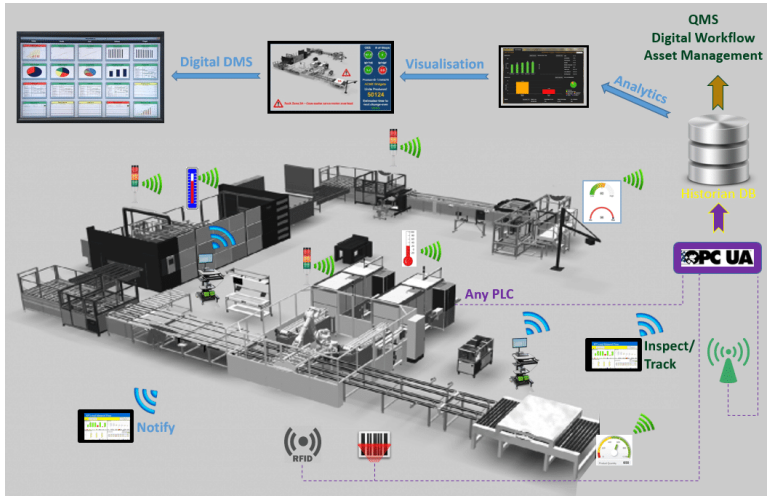
A Gain index policy [2]


[1] **G. Chen**, S. C. Liew and Y. Shao, "Uncertainty-of-Information Scheduling: A Restless Multiarmed Bandit Framework," in *IEEE Transactions on Information Theory*, vol. 68, no. 9, pp. 6151-6173, Sept. 2022, doi: 10.1109/TIT.2022.3177891.

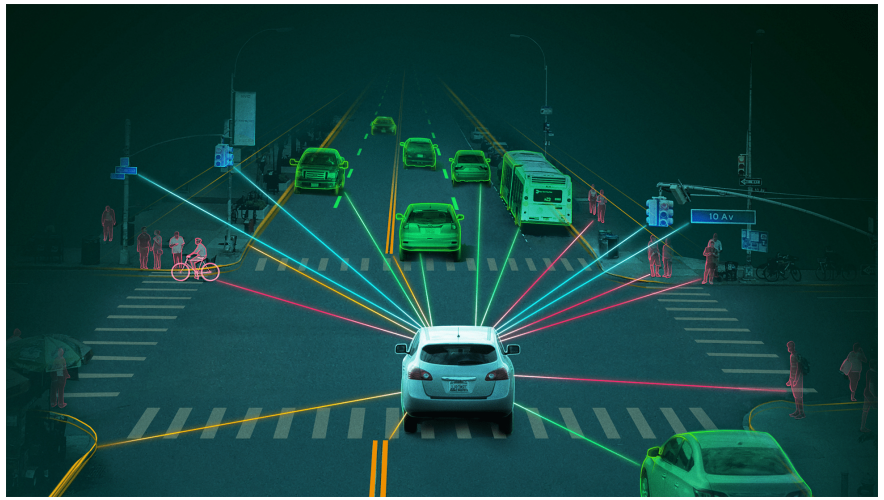
[2] **G. Chen** and S. C. Liew, "An Index Policy for Minimizing the Uncertainty-of-Information of Markov Sources," in *IEEE Transactions on Information Theory*, doi: 10.1109/TIT.2023.3315459.

# Background

Real-time information delivery is important in modern networked systems




Smart Factory



Autonomous Driving

Multi-agent System

The classic latency metric is inadequate to reflect the timeliness requirements

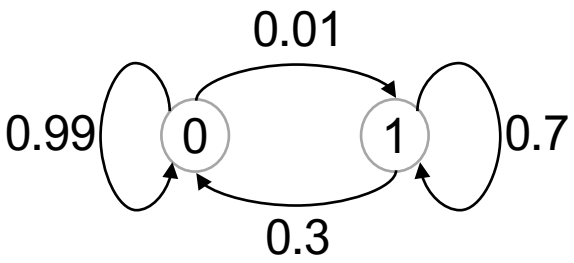
Two underlying assumptions of latency-centric designs:

- A1: The quality of information decreases linearly over time
- A2: The quality of information decreases with time in a way that is independent of the content of the information

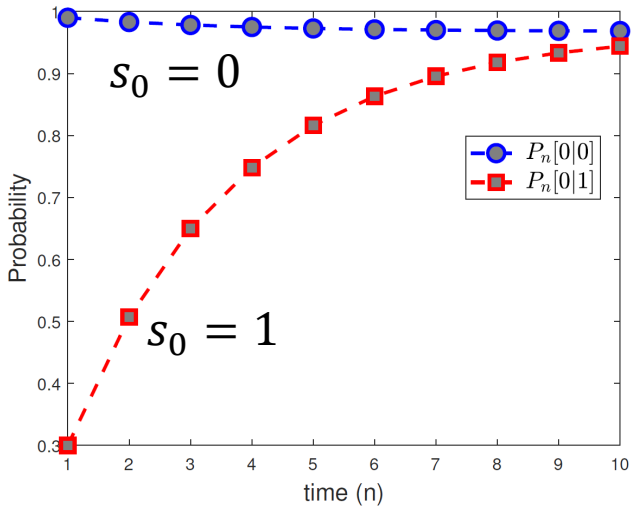
**Usually NOT valid!**

# Motivation

*Example*: A binary Markov chain being observed by a remote monitor.




A binary Markov Chain




$n$-step transition probabilities

**$s_0 = 0$ is still useful in the next few time steps**

If $s_0 = 0$, then the state remains in 0 with high probability in the next few time steps.

**$s_0 = 1$ quickly becomes outdated**

If $s_0 = 1$, we can hardly infer the states of the next few time steps without new updates.

From the perspective of the remote monitor:

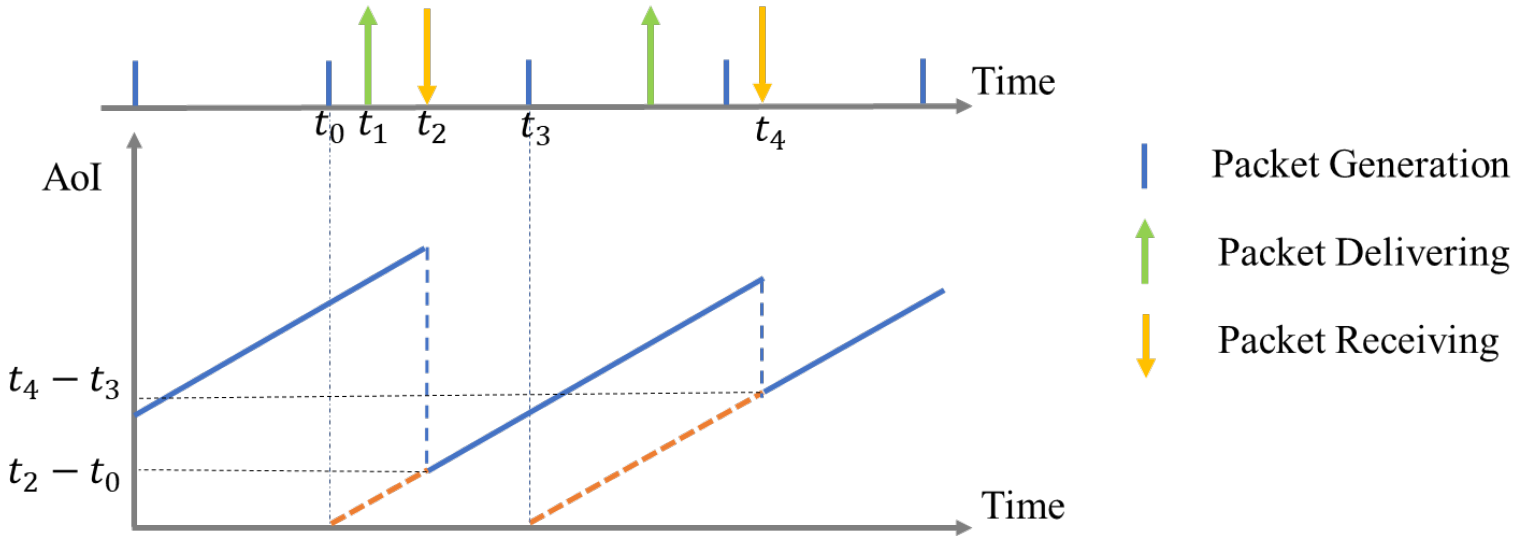
- Even from the same source, some information becomes outdated quickly while others slowly.
- The information quality evolves over time in a way that depends on the value of the last update $s_0$.

*Neither the classic latency metric nor the more contemporary age-of-information (AoI) metric capture this attribute.*

# Age of Information

An emerging metric for information freshness: **Age of Information (AoI)**

*The time elapsed since the generation of the last packet delivered to the receiver*

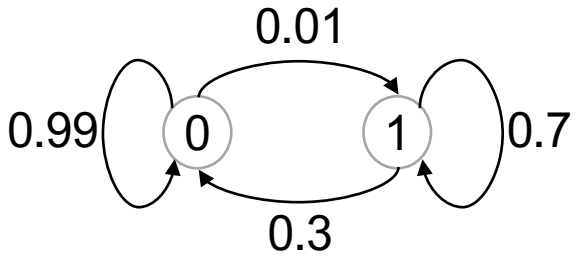



| AoI-related metrics |
| --- |
| Various increasing (possibly nonlinear) functions of AoI are used to capture the system dynamics. |
| Information content is still not considered in these metrics. |

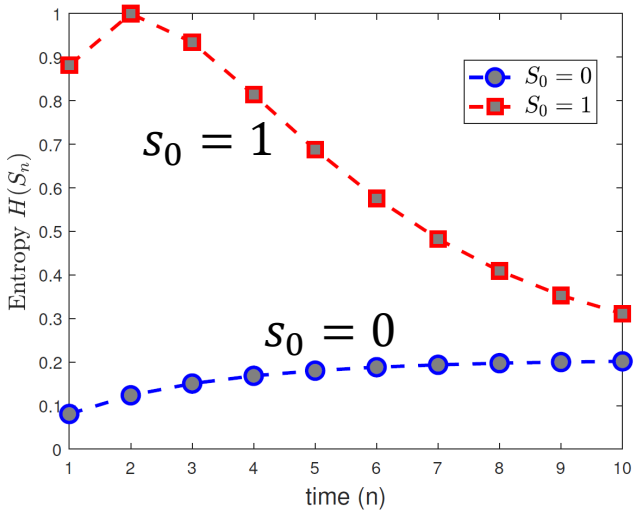
- AoI metric still assumes A1 and A2
- AoI-related metrics partially relax A1 (nonlinear increasing)

# Uncertainty of Information: An Information-theoretical metric

Shannon's entropy measures how much we do not know about the current state




A binary Markov Chain




Uncertainty (entropy) over time

Given the latest observation $s_0$, the remote monitor's uncertainty about the states of the next few time steps are

- low, if $s_0 = 0$
- high, if $s_0 = 1$

$s_0 = 1$ quickly becomes outdated because it contains less information about the future states

**Uncertainty of Information (UoI):**
The entropy of the current state conditioned on the latest observation

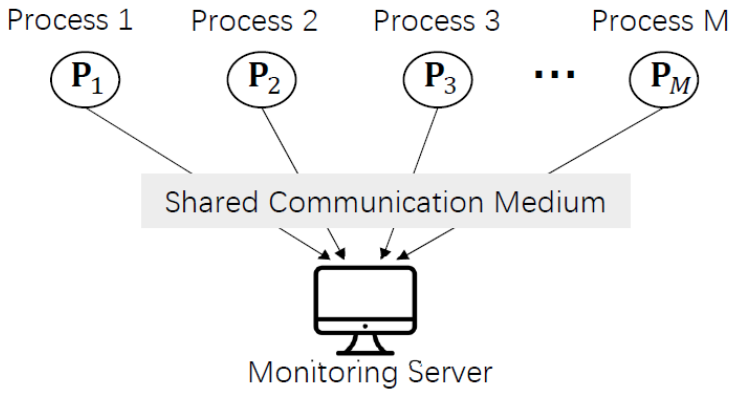
$$U[t] = -\sum_{x\in\{0,1\}} P[S_t = x \mid S_{t'} = s] \log_2 P[S_t = x \mid S_{t'} = s]$$

(Assume the latest observation at time $t$ is $S_{t'} = s, t' < t$)

- UoI may NOT monotonically increase with time!
- UoI is dependent of the value of the last observation.

# UoI Scheduling for Binary Markov Sources

*System Model*: $M$ binary Markov processes being observed by a remote monitor




*Markov sources*:

- One-step transition matrix of process $i$ (known to the monitor):

$$\mathbf{P}_i \triangleq \begin{bmatrix} P[0|0] & P[0|1] \\ P[1|0] & P[1|1] \end{bmatrix} = \begin{bmatrix} 1-p_i & q_i \\ p_i & 1-q_i \end{bmatrix}$$

- All Markov processes are ergodic and stationary

*Channel model*:

- Time-slotted communications scheduled by the monitor
- Reliable communication medium used by $m$ users each time (m $< M$)
- Each transmission takes one time slot

*Objective*: minimize the long-term average sum-UoI

$$\min \lim_{T\to\infty} \frac{1}{T} E\Big[\sum_{t=1}^{T}\sum_{i=1}^{M} U_i[t]\Big].$$

# Multi-armed Bandit (MAB) Problem: History

**Thompson 1933: Clinical trial**

- Two experimental treatment/medicines are available for a certain disease
- Each treatment is modeled as a Bernoulli random variable with an unknown parameter
- Each time a patient comes, select a treatment to use for the patient
- Goal: maximize the (expected) cure rate

→ **Frequentist Framework** (Online learning under the MAB model is a hot topic recent years)

- $M$ bandit processes with unknown parameters, no prior distributions are assumed
- Maximize the expected reward (alternatively, minimize the regret) via online learning

→ **Bellman 1956: Belief MDP**

- Assume a prior distribution (belief) for each treatment
- Update the beliefs with arriving samples (each belief process is Markovian)
- Solved by dynamic programming

→ **Bayesian Framework**

# Multi-armed Bandit (MAB) Problem: Definition

- $M$ discrete-time Markov bandit process with countable states
- Controls applied at time $t$:
  - $u(t) = 0$ freezes the process (no state transition and no reward)
  - $u(t) = 1$ activates the process, leading to a state transition and an instantaneous reward
- At each time step, only one bandit can be activated
- Goal: maximize the cumulative reward

⚠ Challenge: curse of dimensionality (the state space increases exponentially with $M$)

**Gittins 1976: Gittins Index Policy**

- The optimal policy for MAB is an index policy
- Index policy: there exists a function $G_i(s_i)$, computed separately for each bandit $i$, such that for every state $S = [s_1, s_2, \cdots, s_M]$, the optimal policy activates the bandit:

$$i = \underset{i \in [M]}{\text{argmax}}\{G_i(s_i)\}$$

# Restless Multi-armed Bandit (RMAB)

Whittle (1988) generalized MAB to RMAB in two ways:

- $u(t) = 0$ does not freezes the process. The unactivated processes may also have rewards and state transitions (hence, the term "restless" in RMAB)
- At each time step, $k$ bandits can be activated ($1 \leq k < M$)

Some facts:

- RMAB is P-SPACE hard (Papadimitriou 1999)
- Gittins index policy is not optimal for RMAB.
- Whittle proposed a new index policy, which is now referred to as Whittle's index policy
- Whittle's index policy is asymptotically optimal (Weber and Weiss 1990)

# RMAB Formulation

Formulate the UoI scheduling problem as a Restless Multi-armed Bandit (RMAB):

- $M$ Markov bandit processes
  - Define a belief MDP for each Markov process (i.e., a bandit)
  - Belief state of the $i$-th belief MDP: $x_i(t) \triangleq P\{S_i(t) = 1 | \text{value of } S_i(t')\}$
- At each time step, only m $< M$ bandits can be selected
  - Action of the $i$-th belief MDP: $u_i(t) \in \{0,1\}$, where $u_i(t) = 1$ (active action) means that this process is selected to transmit at time $t$.
  - The unselected bandits are said to be applied with passive action ($u_i(t) = 0$)
- Belief state transition under different actions:

$$x_i(t+1) = \begin{cases} p_i, & \text{if } u_i(t) = 1 \text{ and } S_i(t) = 0 \\ 1 - q_i, & \text{if } u_i(t) = 1 \text{ and } S_i(t) = 1 \\ \tau(x_i(t)), & \text{if } u_i(t) = 0 \end{cases}$$

$\tau(\cdot)$ is the one-step belief state evolution under the passive action

$$\tau(x_i(t)) \triangleq p_i + x_i(t)(1 - p_i - q_i).$$

- Cost function of the $i$-th belief MDP in belief state $x_i(t)$ : $U_i[t] = H(x_i(t))$ with $H(\cdot)$ being the entropy function

# RMAB Formulation

UoI scheduling problem in RMAB form:

$$\text{P1}: \quad \min_{\{u_i(t)\}} \lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} \sum_{i=1}^{M} H\left(x_i(t)\right)$$

$$s.t. \quad \sum_{i=1}^{M} u_i(t) = m, \ \forall t$$

$$u_i(t) \in \{0,1\}, \ \forall i,t.$$

# RMAB Decomposition: Whittle's Method

UoI scheduling problem in RMAB form:

$$\text{P1}: \min_{\{u_i(t)\}} \lim_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T}\sum_{i=1}^{M} H\left(x_i(t)\right)$$

$$s.t. \quad \boxed{\sum_{i=1}^{M} u_i(t) = m, \ \forall t} \xrightarrow{\text{Relaxation}} \lim_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T}\sum_{i=1}^{M} u_i(t) = m.$$

$$u_i(t) \in \{0,1\}, \ \forall i, t.$$

Lagrange multiplier method:

$$\text{P2}: \min_{\{u_i(t)\}} \lim_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T}\left[\sum_{i=1}^{M} H\left(x_i(t)\right) + \lambda \sum_{i=1}^{M} u_i(t)\right] - m\lambda$$

For any fixed $\lambda$, P2 can be decoupled into $M$ subproblems:

$$J_i := \min_{\{u_i(t)\}} \lim_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T}\left[H\left(x_i(t)\right) + \lambda u_i(t)\right], i \in [M]$$

Single bandit problem

# The Single Bandit Problem

Each single bandit problem with fixed $\lambda$ is an MDP

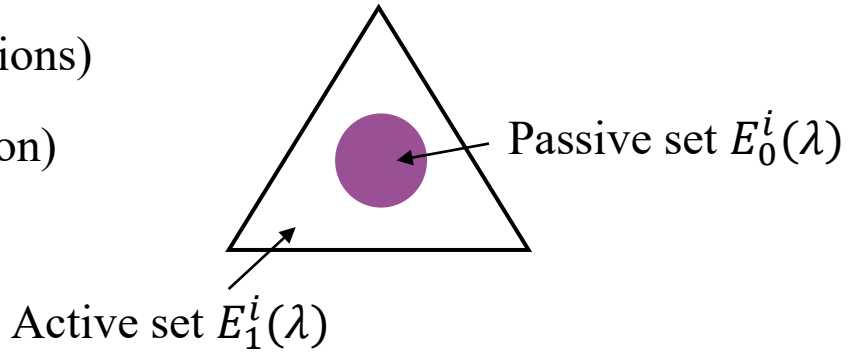
$$J_i := \min_{\{u_i(t)\}} \lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} [H(x_i(t)) + \lambda u_i(t)], i \in [M]$$

- Cost for state-action pair $(x, u)$: $H(x) + \lambda u$ ($\lambda$ is called the service charge)
- Optimality equation (Bellman equation):

$$V(x) + g = H(x) + \min\{\underbrace{\lambda + xV(1-q) + (1-x)V(p)}_{a(x)}, \underbrace{V(\tau(x))}_{r(x)}\}$$

- The optimal policy partitions the belief state space into two sets:
  - Passive set $E_0^i(\lambda) := \{x: a(x) \geq r(x)\}$ (i.e., $u = 0$ is the optimal actions)
  - Active set $E_1^i(\lambda) := \{x: a(x) < r(x)\}$ (i.e., $u = 1$ is the optimal action)

Note: $E_0^i(\lambda)$ and $E_1^i(\lambda)$ vary with $\lambda$



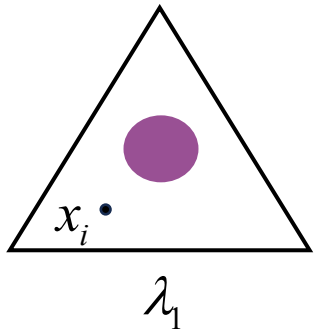
# Whittle Index Policy is Near-optimal

Whittle index policy only applies to indexable problems

Two steps for developing the Whittle index policy:

Step 1: Establish the indexability

**Indexability**: A bandit $i$ is indexable if the passive set $E_0^i(\lambda)$ monotonically expands from the empty set to the whole state space as $\lambda$ increases from 0 to $+\infty$. An RMAB is indexable if all bandits are indexable.

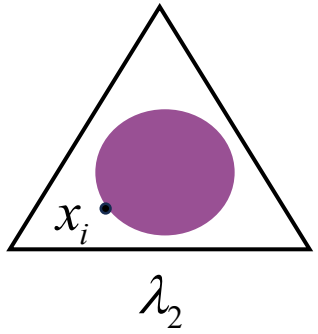



Step 2: Compute Whittle indices

- For every bandit, assign an index for each belief state $x_i$: $W(x_i) = \inf_{\lambda} \{\lambda : x_i \in E_0^i(\lambda)\}$

  $W(x_i)$ is the service charge making the two actions equally rewarding in state $x_i$

- At each time, select the $m$ bandits with the greatest $m$ indices.




We aim to develop the Whittle index policy for the UoI scheduling problem!

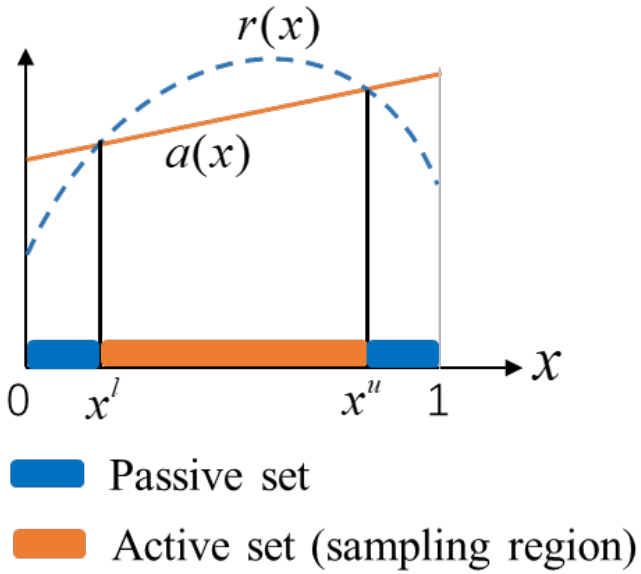
# Threshold Structure of the Optimal Policy

Bellman equation: $V(x)+g=H(x)+\min\{\underbrace{\lambda+xV(1-q)+(1-x)V(p)}_{a(x)},\underbrace{V(\tau(x))}_{r(x)}\}$

The active set is $E_1=\{x: a(x)<r(x)\}$; the passive set is $E_0=\{x: a(x)\geq r(x)\}$

**Proposition 2.10**

For any $\lambda\in(0,+\infty)$, there exists two thresholds $0\leq x^l\leq x^u\leq 1$ such that $a(x^l)=r(x^l), a(x^u)=r(x^l)$, and that $E_1(\lambda)=(x^l,x^u)$. If $(x^l,x^u)$ is empty, then $E_0(\lambda)=(0,1)$.



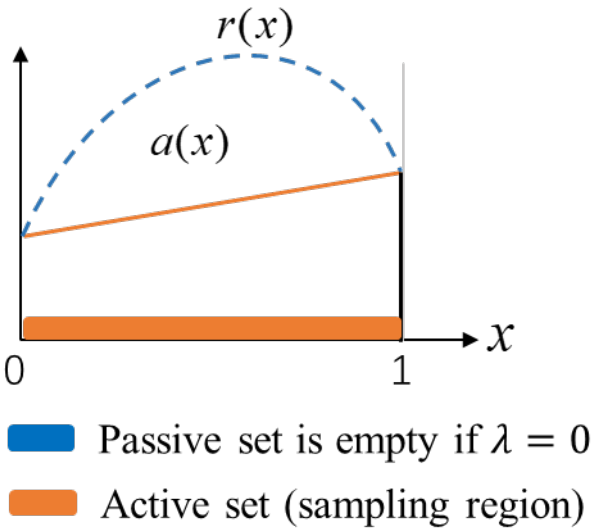

The active set $E_1(\lambda)$ is always a single continuous interval (if not empty)

# Transmission is Always Desired

Bellman equation: $V(x)+g=H(x)+\min\left\{\underbrace{\lambda+xV(1-q)+(1-x)V(p)}_{a(x)},\underbrace{V\left(\tau(x)\right)}_{r(x)}\right\}$

The active set is $E_1=\{x: a(x)<r(x)\}$; the passive set is $E_0=\{x: a(x)\geq r(x)\}$

**Corollary 2.11.** For any single-bandit problem with zero service charge (i.e., $\lambda=0$), the sampling region of the optimal policy is $(0,1)$.




**Implication of Corollary 2.11**

Although UoI is not a monotonically increasing function of age, a specific remote process may still want to update its observation at every available opportunity to minimize its average UoI.

# A Sufficient Condition for Indexability

Bellman equation: $V(x)+g=H(x)+\min\{\underbrace{\lambda+xV(1-q)+(1-x)V(p)}_{a(x)},\underbrace{V(\tau(x))}_{r(x)}\}$

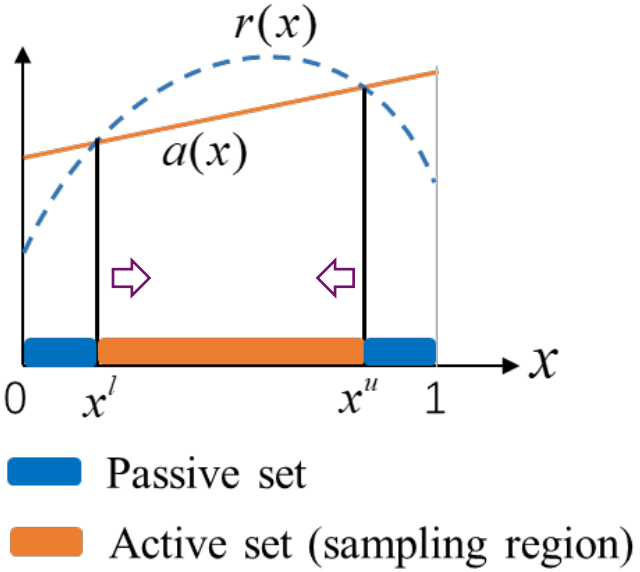
The active set is $E_1=\{x: a(x)<r(x)\}$; the passive set is $E_0=\{x: a(x)\geq r(x)\}$

**Lemma 2.13 (Monotonicity Condition)**

Suppose that for any $\lambda\geq 0$ such that $(x^l,x^u)$ is non-empty, we have

$$\frac{\partial a(x^l,\lambda)}{\partial\lambda}>\frac{\partial r(x^l,\lambda)}{\partial\lambda} \text{ and } \frac{\partial a(x^u,\lambda)}{\partial\lambda}>\frac{\partial r(x^u,\lambda)}{\partial\lambda}$$

Then $x^l$ monotonically increases with $\lambda$, while $x^u$ monotonically decreases with $\lambda$.



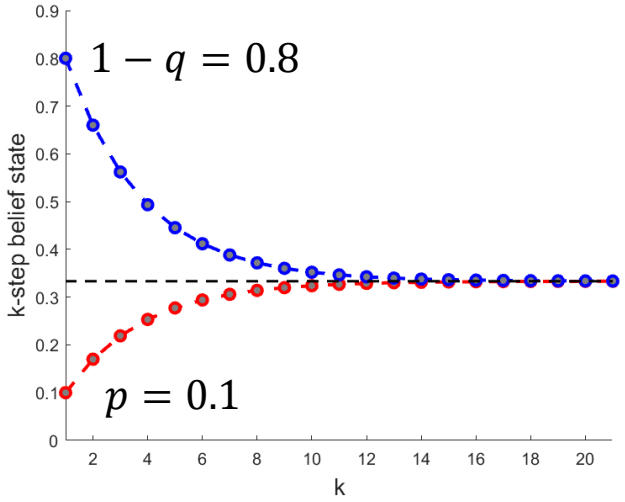

If the condition holds, then the active set shrinks monotonically. Consequently, the passive set expands monotonically, implying the indexability.

# The UoI Scheduling Problem is Indexable

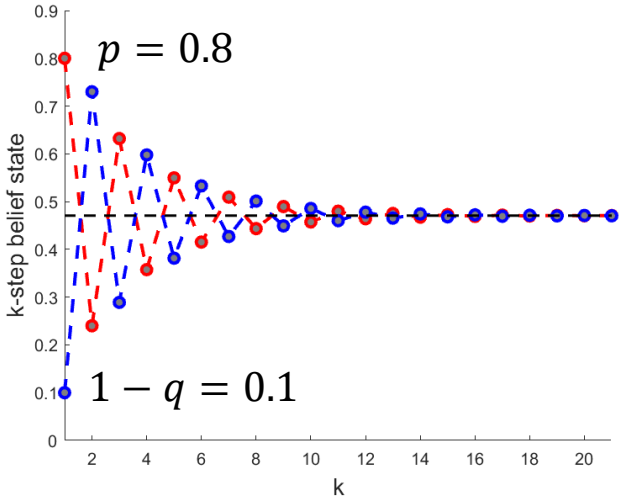
**Proposition 2.18**: A monotonic bandit ($p + q < 1$) is indexable.

**Proposition 2.19**: An oscillating bandit ($p + q > 1$) is indexable.

Belief state evolution under the passive action




Monotonic bandit ($p + q < 1$)




Oscillating bandit ($p + q > 1$)

**Theorem 2.17**: The RMAB is indexable.

# Algorithm to Compute Whittle Index

The RMAB is indexable, but the closed-form expression for Whittle index is not available.

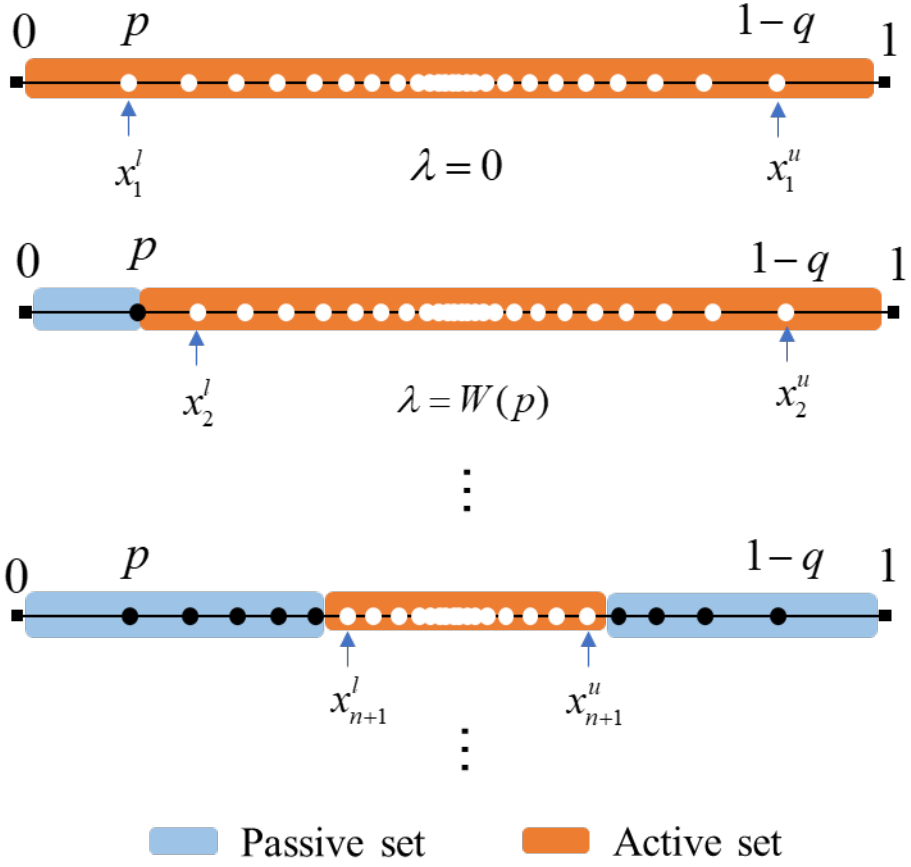
$E_0(0) = \varnothing$

$E_0$ expends as $\lambda$ increases

**Whittle index definition**

$$W(x) = \inf_{\lambda} \{\lambda : x \in E_0(\lambda)\}$$

- Denote by $z_k$ the $k$-th belief state that enters passive set.




$z_1$ has two candidates: $p$ and $1-q$

If $z_1$ is determined, then $z_2$ has two candidates: $x_2^l$ and $x_2^u$

If $\{z_1, \cdots, z_n\}$ are determined, then $z_{n+1}$ has two candidates: $x_{n+1}^l$ and $x_{n+1}^u$

How to identify $z_k$ from its two candidates?

Compute $W(x_k^l)$ and $W(x_k^u)$, then $z_k = \text{argmin}\{W(x_k^l), W(x_k^u)\}$

# Algorithm to Compute Whittle Index

**Proposition 2.20.** For a monotonic bandit $(p+q<1)$, the Whittle index of $\mathtt{x}_{n+1}$ can be computed as follows:

1. For $0 \leq n \leq e-1$,

$$W(\mathtt{x}_{n+1}) = \frac{B_n H(\tau_{n+1}) - G_n - (\mathtt{x}_{n+1} - \tau_{n+1})\, C_n}{(\mathtt{x}_{n+1} - \tau_{n+1})\,(L_n - K_n) + q^{(K_n)} + p^{(L_n)}},$$

2. For $e \leq n < |E|-1$, let $(x^l, x^u)$ denote the sampling region of the optimal policy with $\lambda = W(\mathtt{x}_{n+1})$. If $\mathtt{x}_{n+1} = x^l$, then $W(\mathtt{x}_{n+1}) =$

$$\frac{q^{(K_n)} \sum_{k=1}^{F} \left[H\left(\tau^k(\mathtt{x}_{n+1})\right) - H\left(p^{(k)}\right)\right] - \mathtt{x}_{n+1} \sum_{k=1}^{K_n} \left[H\left(1-q^{(k)}\right) - H(\omega)\right]}{q^{(K_n)} + \mathtt{x}_{n+1}}.$$

If $\mathtt{x}_{n+1} = x^u$, then $W(\mathtt{x}_{n+1}) =$

$$\frac{q^{(K_n)} \left[H\left(\tau_{n+1}\right) - H(\omega)\right] - (\mathtt{x}_{n+1} - \tau_{n+1}) \sum_{k=1}^{K_n} \left[H\left(1-q^{(k)}\right) - H(\omega)\right]}{\mathtt{x}_{n+1} - \tau_{n+1}}.$$

# Algorithm to Compute Whittle Index

**Proposition 2.21.** For an oscillating bandit ($p+q>1$), the Whittle index of $\mathtt{x}_{n+1}$ can be computed as follows:

1. For $0 \le n \le e$. Let $\tau_{n+1} \triangleq \tau(\mathtt{x}_{n+1}), \tau^{(2)}_{n+1} \triangleq \tau^2(\mathtt{x}_{n+1})$. If $\tau_{n+1} \in \{\mathtt{x}_1, \cdots, \mathtt{x}_n\}$, $W(\mathtt{x}_{n+1})$ is given by

$$W(\mathtt{x}_{n+1}) = \frac{B_n\left[H(\tau_{n+1}) + H(\tau^{(2)}_{n+1})\right] - 2G_n - \left(\mathtt{x}_{n+1} - \tau^{(2)}_{n+1}\right) C_n}{\left(\mathtt{x}_{n+1} - \tau^{(2)}_{n+1}\right)(L_n - K_n) + 2\left(q^{(K_n)} + p^{(L_n)}\right)}.$$

   If $\tau_{n+1} \notin \{\mathtt{x}_1, \cdots, \mathtt{x}_n\}$, then

$$W(\mathtt{x}_{n+1}) = \frac{B_n H(\tau_{n+1}) - G_n - (\mathtt{x}_{n+1} - \tau_{n+1}) C_n}{(\mathtt{x}_{n+1} - \tau_{n+1})(L_n - K_n) + q^{(K_n)} + p^{(L_n)}},$$

   where $B_n, C_n$ and $G_n$ are defined in Proposition 2.20.

2. For $e \le n < |E|$, $W(\mathtt{x}_n) = W(\mathtt{x}_e) = W(\omega)$.

3. Finally, $W(\mathtt{x}_{|E|})$ can be computed by (2.51).

# Simulations

Whittle index policy achieves near-optimal performance

In UoI scheduling problems

| | $(p_i, q_i)$ | Optimal UoI | WI UoI | Myopic UoI | WI regret | Myopic regret |
|---|---|---|---|---|---|---|
| A1 | (0.05,0.2) (0.2,0.4) | 1.2866 | 1.2867 | 1.527 | 0.01% | 18.7% |
| A2 | (0.2,0.2) (0.4,0.4) | 1.7219 | 1.7219 | 1.873 | 0 | 8.8% |
| A3 | (0.95,0.95) (0.7,0.7) | 1.2864 | 1.2864 | 1.5668 | 0 | 21.8% |
| A4 | (0.05,0.1) (0.2,0.9) | 1.0309 | 1.0318 | 1.2424 | 0.09% | 20.5% |

2-process systems

| | $(p_i, q_i)$ | Optimal UoI | WI UoI | Myopic UoI | WI regret | Myopic regret |
|---|---|---|---|---|---|---|
| B1 | (0.1,0.1) (0.6,0.6) (0.3,0.3) | 2.469 | 2.469 | 2.792 | 0 | 13.8% |
| B2 | (0.1,0.3) (0.6,0.6) (0.1,0.2) | 2.2963 | 2.2968 | 2.7005 | 0.02% | 17.6% |
| B3 | (0.1,0.3) (0.5,0.6) (0.9,0.9) | 2.2158 | 2.2179 | 2.6506 | 0.1% | 19.6% |

3-process systems

- Optimal UoI: the optimal policy is determined by the value iteration algorithm
- WI UoI: results of the Whittle index policy
- Myopic UoI: results of the myopic (one-step greedy) policy
- WI regret = (WI UoI – Optimal UoI)/Optimal UoI

# Simulations

Whittle index policy achieves near-optimal performance in other RMABs

Our method is applicable to any C-type RMAB:

**C-type RMAB**

An RMAB is called C-type if it is of the form of problem P1 and the penalty function $H(x)$ is a concave function of the belief state $x$.

| | $(p_i, q_i)$ | Optimal penalty | WI penalty | Myopic penalty | WI regret | Myopic regret |
|---|---|---|---|---|---|---|
| C1 | (0.05,0.2) (0.4,0.5) | 1.057 | 1.064 | 1.275 | 0.7% | 20.6% |
| C2 | (0.05,0.1) (0.5,0.6) | 1.480 | 1.482 | 1.814 | 0.1% | 22.6% |
| D1 | (0.05,0.2) (0.1,0.3) (0.4,0.7) | 1.1467 | 1.1485 | 1.4079 | 0.2% | 22.8% |
| D2 | (0.1,0.2) (0.1,0.8) (0.4,0.5) | 1.3843 | 1.3845 | 1.587 | 0.01% | 14.6% |

| | $(p_i, q_i)$ | Optimal penalty | WI penalty | Myopic penalty | WI regret | Myopic regret |
|---|---|---|---|---|---|---|
| E1 | (0.05,0.2) (0.4,0.5) | 1.2677 | 1.268 | 1.618 | 0.02% | 27.6% |
| E2 | (0.05,0.2) (0.4,0.5) (0.1,0.2) | 1.904 | 1.906 | 2.507 | 0.1% | 31.7% |
| F1 | (0.05,0.2) (0.4,0.5) | 21.466 | 21.622 | 32.722 | 0.7% | 52.4% |
| F2 | (0.05,0.2) (0.4,0.5) (0.1,0.2) | 37.875 | 38.225 | 49.722 | 0.9% | 32.3% |

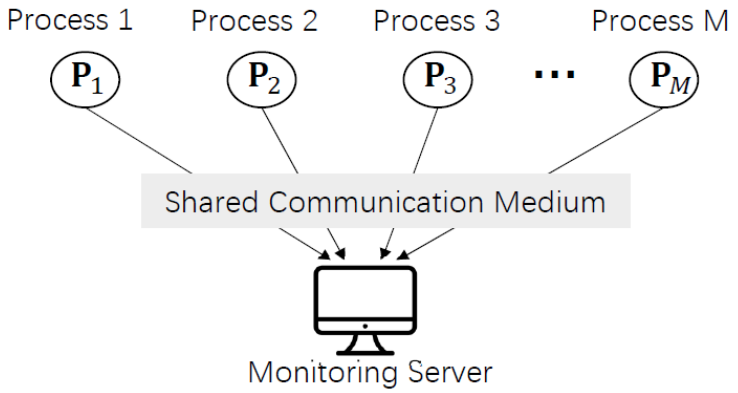
$$H_1(x) = \alpha_1 x + \alpha_0(1-x) + \beta\sqrt{\alpha_1^2 x + \alpha_0^2(1-x) - (\alpha_1 x + \alpha_0(1-x))^2}$$

E1 and E2: $H_2(x) = 1-(2x-1)^2$

F1 and F2: $H_3(x) = 20 - 1/x$

# UoI Scheduling for Finite-state Markov Sources

*System Model*: $M$ Markov processes being observed by a remote monitor




Generalizations:

- Each process is an $N$-state Markov chain with transition matrix $T^{(i)}$
- Unreliable channel: process $i$ succeeds with probability $\rho_i \in (0,1]$
- Objectives: min total discounted UoI and long-term average UoI

UoI: $U_i(t) = -\sum_{S_i(t)} P[S_i(t) \mid S_i(t')] \log_2 P[S_i(t) \mid S_i(t')]$

*Objective 1*: minimize the expected total discounted sum-UoI

$$\min \quad E\left[\sum_{t=1}^{\infty}\sum_{i=1}^{M}\beta^{t-1}U_i(t)|\chi\right],$$

*Objective 2*: minimize the long-term average sum-UoI

$$\min \quad E\left[\lim_{n\to\infty}\frac{1}{n}\sum_{t=1}^{n}\sum_{i=1}^{M}U_i(t)|\chi\right]$$

# Whittle Index Policy: Pros and Cons

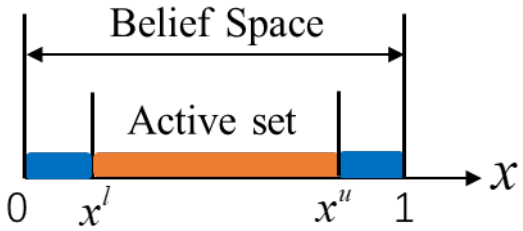
Whittle index policy is a famous policy for RMABs with many advantages:

- Offline computation, simple for execution
- Computing the Whittle indices of a bandit is independent of other bandits (relatively low complexity)
- Near-optimal for a large class of problems

Limitations:

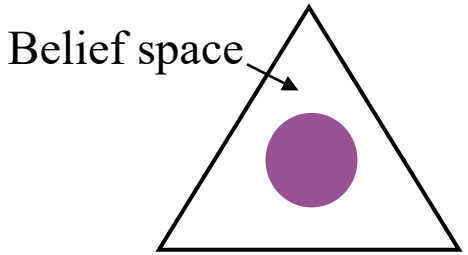
- It only applies to indexable RMABs. NOT all RMABs are indexable
- Establishing indexability is challenging
- Computing the Whittle indices is usually difficult due to the lack of closed-form expression

Establishing indexability for the generalized UoI scheduling problem is difficult!




- Threshold structure
- Active set is determined by two points

Binary Markov chain




- Active set is a convex set
- Infinite boundary points

3-state Markov chain

# Total Discounted UoI Scheduling

Develop a new index policy that does not require indexability

- Relaxation:

$$\text{P1: } \min \quad \mathbb{E}\left[\sum_{t=1}^{\infty}\sum_{i=1}^{M}\beta^{t-1}H\left(X_i(t)\right)|\chi\right]$$

$$s.t. \quad \sum_{i=1}^{M}u_i(t)=m, \quad \forall t \xrightarrow{\text{Relaxation}} \mathbb{E}\left[\sum_{t=1}^{\infty}\beta^{t-1}\sum_{i=1}^{M}u_i(t)\right]=\sum_{t=1}^{\infty}\beta^{t-1}m=\frac{m}{1-\beta}.$$

$$u_i(t)\in\{0,1\}, \quad \forall i,t.$$

- Lagrange multiplier method:

$$\text{P1(a): } \min_{\{u_i(t)\}}\sup_{\lambda\geq 0}\mathbb{E}\left[\sum_{t=1}^{\infty}\sum_{i=1}^{M}\beta^{t-1}\left[H\left(X_i(t)\right)+\lambda u_i(t)\right]|\chi\right]-\frac{m\lambda}{1-\beta}.$$

- Lagrange dual problem:

$$\text{P1(b): } \sup_{\lambda\geq 0}\min_{\{u_i(t)\}}\mathbb{E}\left[\sum_{t=1}^{\infty}\sum_{i=1}^{M}\beta^{t-1}\left[H\left(X_i(t)\right)+\lambda u_i(t)\right]|\chi\right]-\frac{m\lambda}{1-\beta}.$$

According to LP theory or constrained MDP theory, P1(a)=P1(b).

# Relaxation and Decomposition

- The Lagrange dual problem (equivalent to the relaxed problem):

$$\text{P1(b)}: \sup_{\lambda \geq 0} \min_{\{u_i(t)\}} \mathbb{E}\left[\sum_{t=1}^{\infty}\sum_{i=1}^{M} \beta^{t-1}\left[H\left(X_i(t)\right) + \lambda u_i(t)\right] | \chi\right] - \frac{m\lambda}{1-\beta}.$$

- Decomposition of the inner min problem (fix $\lambda$):

$$J_i(\lambda) := \min_{\{u_i(t)\}} \mathbb{E}\left[\sum_{t=1}^{\infty} \beta^{t-1}\left[H\left(X_i(t)\right) + \lambda u_i(t)\right] | \chi^{(i)}\right], \quad i \in [M].$$

Single bandit problem

Basic idea of developing the novel index policy

1. Solve the relaxed problem and obtain its optimal policy
   - Compute the optimal $\lambda^*$
   - Determine the optimal policy for each single bandit problem $J_i(\lambda^*)$, combining them yields the optimal policy for the relaxed problem
2. Derive a policy for the original RMAB by rounding up the optimal relaxed policy

*Core issue: does $J_i(\lambda)$ possess good properties that allow the Lagrange dual problem to be solved efficiently?*

# The Single Bandit Problem

Bellman equation of the $i$-th single bandit problem:

$$V_i(X,\lambda) = H(X) + \min\left\{\lambda + \beta\rho\sum_{k=1}^{N} x_k V_i(\mathbf{T}_k^{(i)},\lambda) + \beta(1-\rho)V_i(\mathbf{T}^{(i)}X,\lambda), \beta V_i(\mathbf{T}^{(i)}X,\lambda)\right\}, \forall X \in \Omega_i$$

**Theorem 3.3**. For any $i$ and $X \in \Omega_i$, $V_i(X,\lambda)$ is an increasing, concave, and piecewise linear function of $\lambda \in [0,\infty)$. In addition,

$$0 \le \frac{\partial V_i(X,\lambda)}{\partial \lambda} \le \frac{1}{1-\beta}$$

The Lagrange dual problem reduces to maximizing a concave function

$$\text{P1(b)}: \sup_{\lambda\ge 0}\min_{\{u_i(t)\}} \mathbb{E}\left[\sum_{t=1}^{\infty}\sum_{i=1}^{M}\beta^{t-1}\left[H\left(X_i(t)\right) + \lambda u_i(t)\right]|\chi\right] - \frac{m\lambda}{1-\beta}.$$

$$\Downarrow$$

$$\text{P1(c)}: \sup_{\lambda\ge 0}\left\{\sum_{i=1}^{M} V_i(\chi^{(i)},\lambda) - \frac{m\lambda}{1-\beta}\right\} = \sup_{\lambda\ge 0} f(\lambda).$$

$f(\lambda)$ is concave and piecewise linear

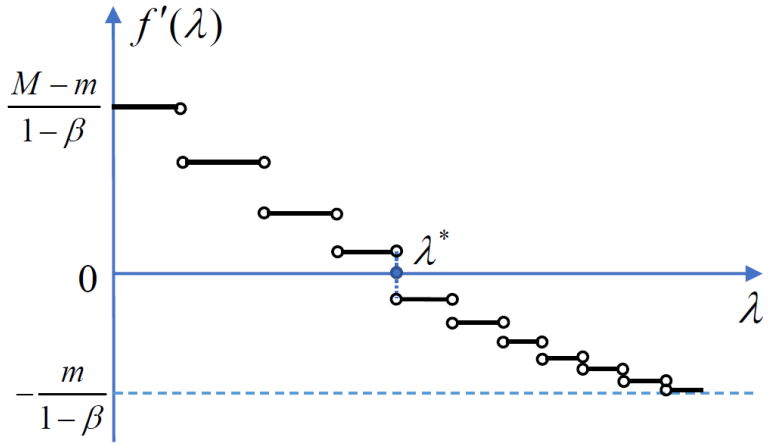
# The Lagrange Dual Problem

**A gradient method to compute $\lambda^*$**

- Step1: Let $\lambda_0 = 0$ and $k = 0$. Specify $\epsilon > 0$.
- Step2: Select a stepsize $a_k$ and compute $\lambda_{k+1} = \lambda_k + a_k f'(\lambda_k)$
- Step3: Stop if

$$f'(\lambda_k) f'(\lambda_{k+1}) \le 0 \text{ and } |\lambda_{k+1} - \lambda_k| < \epsilon.$$

Otherwise, increment $k$ by 1 and return to step 2.




- The stopping criterion:

**Lemma 3.7.** Denote by $\lambda^*$ an optimal solution to P1(c), then $\lambda^* > 0$ and

$$f'(\lambda_-^*) \ge 0, \quad f'(\lambda_+^*) \le 0.$$

- $f'(\lambda)$ may not exist for some $\lambda$ because $f(\lambda)$ is piecewise linear
  - Using $f'(\lambda_-)$ or $f'(\lambda_+)$ instead, which are always well-defined.

# Convergence of the Gradient Method

The algorithm stops within a finite number of iterations

**Theorem 3.8.** Choosing $a_k = c/k$ with $c$ being a positive constant, and let $\{\lambda_k\}$ be the sequence generated by (3.31). Then given any $\epsilon > 0$, there exists a finite integer $D_\epsilon$ such that

$$f'(\lambda_{D_\epsilon}) f'(\lambda_{D_\epsilon+1}) \leq 0 \text{ and } |\lambda_{D_\epsilon+1} - \lambda_{D_\epsilon}| < \epsilon.$$

The interval between $\lambda_{D_\epsilon+1}$ and $\lambda_{D_\epsilon}$ contains at least one optimal solution of P1(c).

Note: (3.31) refers to $\lambda_{k+1} = \lambda_k + a_k f'(\lambda_k)$.

A remaining issue: how to compute $f'(\lambda)$ for a given $\lambda$?

- For each bandit $i$, determine the optimal policy for $J_i(\lambda)$, denoted by $\pi_{i,\lambda}$
- $\partial V_i / \partial \lambda$ can be derived from the Markov chain generated by policy $\pi_{i,\lambda}$

$$\frac{df(\lambda)}{d\lambda} = \sum_{i=1}^{M} \frac{\partial V_i(\chi^{(i)}, \lambda)}{\partial \lambda} - \frac{m}{1-\beta}$$

# Finite-state Approximation

The single bandit problem has countably infinite belief states

**$L$-truncated belief MDP of the single bandit problem**

For a given positive integer $L$, the state space of the $L$-truncated belief MDP only consists of the equilibrium distribution and the $n$-step transition distributions for $1 \leq n \leq L$.

**Theorem 3.6.** Let $L$ be a positive integer such that

$$\max_{i\in[N]} ||\mathbf{T}_i^L - \omega||_\infty \leq \eta_L \text{ and } \max_{i\in[N],k\geq 0} \left|H(\mathbf{T}_i^{L+k}) - H(\omega)\right| \leq \sigma_L,$$

where $\eta_L, \sigma_L > 0$, $||\cdot||_\infty$ denotes the max norm. In addition, suppose that $H(X) \leq B_H$ for $X \in \mathcal{S}$. Then for any $\lambda \in [0, \infty)$ and $X \in \mathcal{S}^L$,

$$|V(X,\lambda) - \phi^L(X,\lambda)| \leq \frac{\beta\sigma_L}{1-\beta} + \beta\rho\eta_L N \frac{B_H + \lambda}{(1-\beta)^2}.$$

- $\eta_L \to 0$ as $L \to \infty$
- $\sigma_L \to 0$ as $L \to \infty$
- The approximation error can be arbitrarily small if $L$ is large enough

# Gain Index Policy

Bellman equation for the $i$-th single bandit problem:

$$V_i(X,\lambda) = H(X) + \min\left\{ \underbrace{\lambda + \beta\rho\sum_{k=1}^{N} x_k V_i(\mathbf{T}_k^{(i)},\lambda) + \beta(1-\rho)V_i(\mathbf{T}^{(i)}X,\lambda)}_{a_i(X,\lambda)}, \underbrace{\beta V_i(\mathbf{T}^{(i)}X,\lambda)}_{r_i(X,\lambda)} \right\}$$

**Optimal policy for the relaxed problem (OR policy)**

Denote by $\lambda^*$ the optimal solution to Lagrange dual problem. At each time, the $i$-th Markov source is selected to transmit if and only if $a_i(X_i(t),\lambda^*) \le r_i(X_i(t),\lambda^*)$.

**The Gain Index Policy (A rounded-up version of the OR policy)**

For each Markov source $i \in [M]$, define for each belief state of $J_i(\lambda^*)$, say $X = [x_1, x_2, \cdots, x_N]$, a gain index as follows:

$$W_i(X) = \rho_i\left[ V_i(\mathbf{T}^{(i)}X,\lambda^*) - \sum_{j=1}^{N} x_j V_i(\mathbf{T}_j^{(i)},\lambda^*) \right]$$

$$d_i(t) \triangleq r_i\left(X_i(t),\lambda^*\right) - a_i\left(X_i(t),\lambda^*\right) = \beta W_i\left(X_i(t)\right) - \lambda^*$$

$d_i(t)$ can be interpreted as the gain of applying active instead of passive action to the $i$-th Markov source

# The Gain Index Policy is Asymptotically Optimal

Assumptions:

- $\frac{m}{M} = \alpha$ is fixed.
- The bandits of the RMAB can be divided into $Q$ classes, where $Q < \infty$.
- The bandits of the same class are stochastically identical. The proportion of the $k$-th class is $q_k \in (0,1]$.

**Proposition 3.18.** Suppose Assumption 1 holds. Then for any $M$ such that $\{Mq_k : k \in [Q]\}$ are positive integers,

$$\Pr\left\{\left|\frac{y_t^{(M)}}{M} - \frac{m}{M}\right| \geq M^{-\frac{1}{4}}\right\} \leq 2\exp\left(-2M^{\frac{1}{2}}\right).$$

The OR policy performs almost the same as the gain index (GI) policy as $M \to \infty$

$y_t^{(M)} = \sum_{i=1}^{M} u_i(t)$ under the OR policy

$m = \sum_{i=1}^{M} u_i(t)$ under the GI policy

**Theorem 3.19.** Suppose Assumption 1 holds. Then

$$\lim_{M\to\infty} \frac{1}{M} J_M^{ind} = \lim_{M\to\infty} \frac{1}{M} J_M^{opt} = \lim_{M\to\infty} \frac{1}{M} J_M^{rel}.$$

GI policy value | Optimal value | Relaxed optimal value

For any $M$, $J_M^{ind} \geq J_M^{opt} \geq J_M^{rel}$

# UoI Scheduling with Average Cost Criterion

All the previous results can be extended to the UoI scheduling with average cost criterion

Bellman equation for the $i$-th single bandit problem:

$$Z_i(X,\lambda)+g(\lambda)=H(X)+\min\left\{\lambda+\rho\sum_{k=1}^{N}x_k Z_i(\mathbf{T}_k^{(i)},\lambda)+(1-\rho)Z_i(\mathbf{T}^{(i)}X,\lambda),Z_i(\mathbf{T}^{(i)}X,\lambda)\right\}$$

**The Gain Index Policy (average cost criterion)**

For each Markov source $i\in[M]$, define for each belief state of $J_i(\lambda^*)$, say $X=[x_1,x_2,\cdots,x_N]$, a gain index as follows:

$$W_i(X)=\rho_i\left[Z_i(\mathbf{T}^{(i)}X,\lambda^*)-\sum_{j=1}^{N}x_j Z_i(\mathbf{T}_j^{(i)},\lambda^*)\right]$$

**Theorem 3.20.** The gain index policy is asymptotically optimal for the RMAB with average cost criterion:

$$\lim_{M\to\infty}\frac{1}{M}G_M^{ind}=\lim_{M\to\infty}\frac{1}{M}G_M^{opt}=\lim_{M\to\infty}\frac{1}{M}G_M^{rel}.$$

# Simulations

The gain index policy shows near-optimal performance

| Group | The Discounted Cost Criterion | | The Average Cost Criterion | |
|---|---|---|---|---|
| | Optimal Policy | Gain Index Policy | Optimal Policy | Gain Index Policy |
| A1 | 27.31 | 27.35 | 2.71 | 2.71 |
| A2 | 25.17 | 25.23 | 2.47 | 2.47 |
| A3 | 24.28 | 24.29 | 2.358 | 2.359 |

UoI scheduling of 2 remote Markov sources with 3 states

| Group | The Discounted Cost Criterion | | The Average Cost Criterion | |
|---|---|---|---|---|
| | Optimal Policy | Gain Index Policy | Optimal Policy | Gain Index Policy |
| B1 | 15.48 | 15.59 | 2.94 | 2.94 |
| B2 | 15.63 | 15.74 | 3.02 | 3.04 |
| B3 | 16.32 | 16.36 | 3.14 | 3.16 |

UoI scheduling 2 remote Markov sources with 4 states

# Simulations

The gain index policy performs as well as the Whittle index policy for Whittle-indexable problems

| Group | $(M, m)$ | Gain Index | Whittle Index | Myopic | Round-robin |
|---|---|---|---|---|---|
| C1 ($\rho = 1$) | (5,2) | 3.73 | 3.73 | 3.87 | 3.96 |
| C2 ($\rho = 1$) | (10,3) | 7.36 | 7.37 | 7.87 | 7.73 |
| C3 ($\rho < 1$) | (5,2) | 3.85 | 3.85 | 4.31 | 4.25 |
| C4 ($\rho < 1$) | (8,4) | 5.73 | 5.73 | 6.35 | 6.52 |

Average UoI of different policies under different (indexable) settings

The UoI scheduling problem for binary Markov sources are Whittle-indexable

$M$: The number of Markov sources

$m$: $m$ Markov sources can be selected at each time

Myopic policy: compute the current UoI at each time and select the $m$ Markov sources with the largest $m$ UoIs

# Simulations

The gain index policy applies to any RMAB with bounded cost functions

| Group | $(M, m)$ | Optimal Policy | Gain Index | Whittle Index |
|---|---|---|---|---|
| F1 | (3,1) | 53.2 | 53.56 | 53.72 |
| F2 | (4,1) | 88.7 | 91.18 | 92.06 |
| G1 | (2,1) | 5.03 | 5.07 | 5.08 |
| G2 | (3,1) | 12.65 | 13.12 | 12.98 |

Performance of the gain index policy for general RMABs

Groups F1 and F2: the RMABs is a scheduling problem that minimizes the discounted total error covariance of multiple remote Kalman filters [1].

Groups G1 and G2: the RMABs is a minimum-AoI scheduling problem [2].

[1] J. Wang, X. Ren, Y. Mo, and L. Shi, "Whittle index policy for dynamic multichannel allocation in remote state estimation," IEEE Transactions on Automatic Control, vol. 65, no. 2, pp. 591–603, 2020.
[2] V. Tripathi and E. Modiano, "A whittle index approach to minimizing functions of age of information," in 57th Annual Allerton Conference on Communication, Control, and Computing (Allerton). IEEE Press, 2019, pp. 1160–1167.

# Conclusion

This work explored a broader notion of timing in networked systems:

- Proposed using the uncertainty of information as a metric for information update systems
- Studied the UoI scheduling for binary Markov sources and developed a Whittle index policy
  - Established the indexability
  - Proposed an algorithm to compute the Whittle index
- Studied the UoI scheduling for finite-state Markov sources with unreliable channels
  - Developed a gain index policy that does not require indexability (good universality)
  - Proposed an algorithm to compute the gain index
  - Proved the asymptotic optimality of the gain index policy

# Future Work

A limitation: The effectiveness of the UoI metric has not been validated in practical systems

Experimental study: Finding a suitable application and implementing a real system to assess the performance of a UoI-based design against latency-based and AoI-based designs

A valuable point: The information quality evolves over time in a way that depends both on the value and age of information

Generalization: explore a broader notion of timing for networked systems. Define a general metric

$$q(x,t), \text{ where } x \text{ is value, } t \text{ is age}$$

- What assumptions are needed for $q(x,t)$?
- Answer these questions: When should the transmitter send a new update? How urgent is it for the receiver to obtain a new update?

*Discovering fundamental results, e.g., establishing connections between $q(x,t)$ and the performance of upper layer applications (control, inference, forecasting…)*

Other potential extensions:

- Distributed transmission policy (random access)
- UoI-scheduling with fairness in mind