Lec36

Monday, November 27, 2023 4:42 PM

Please complete the online course evaluation.

Example: Age of Information (AoI)

Friday, April 14, 2023 12:01 PM

Age of Information (AoI)

- M sensors send status update to the BS
- Due to interference, only 1 sensor can transmit at a time

$$u_k^m = \begin{cases} 1 & \text{sensor } m \text{ transmits at slot } k \\ 0 & \text{o/w} \end{cases}$$

- Further, transmission of sensor m will succeed with prob. p_m
- Want information to be as fresh as possible
- Define AoI d_k^m as the elapsed time from the last successful update
  - $$d_{k+1}^m = \begin{cases} 1 & \text{if sensor } m \text{ transmits \& succeeds} \\ d_k^m + 1 & \text{o/w} \end{cases}$$
- Goal: minimize time-average of total AoI

- Formulate the constrained MDP

$$\min \lim_{t \to +\infty} \frac{1}{t} E\left[\sum_{m=1}^{M} \sum_{k=0}^{t-1} d_k^m \right]$$

Sub to $\sum_{m=1}^{M} u_k^m \leq 1$ for all k

- This is a global MDP that suffers from the curse-of-dimensionality

  - Further, we can not apply duality directly because there are too many constraints!

- Instead, let us replace the "hard" constraint by a "soft" constraint first

$$\min \lim_{t\to+\infty} \frac{1}{t} E\left[\sum_{m=1}^{M}\sum_{t=0}^{k-1} d_k^m\right]$$

$$\text{Sub to } \lim_{t\to+\infty} \frac{1}{t} E\left[\sum_{m=1}^{M}\sum_{k=0}^{t-1} u_k^m\right] \leq 1 \quad (\lambda)$$

  - Why this is meaningful will be clear shortly when we discuss Whittle's Index policies.

- We can now apply duality on the single constraint

  - Each sensor should optimize

$$\min_{\pi_m} \lim_{t\to+\infty} \frac{1}{t} E\left[\sum_{k=1}^{t-1}\left(d_k^m + \lambda u_k^m\right)\right]$$

  - This is an average-value problem without constraints

  - The per-stage cost is now

$$d_k^m + \lambda u_k^m$$

$d_k^m + \lambda u_k^m$

- Note: this is not the same as λ^* that we defined earlier for the SSP!

- λ can be thought of as the price to be scheduled.

- Finally, the value of λ can be updated by dual gradient algorithm

$$\lambda(l+1) = \left[\lambda(l) + \alpha\left(\lim_{t\to+\infty} \frac{1}{t} E\left[\sum_{k=0}^{t-1} (d_k^m + \lambda u_k^m)\right] - 1\right)\right]^+$$

- To solve the per-sensor problem, the Bellman equation is

$$h(d) + \nu^* = \min\left\{ (d+\lambda) + p_m h(1) + (1-p_m) h(d+1), \quad d + h(d+1) \right\}, \quad d = 1, 2, \cdots,$$

use ν^* for λ^*

drop the superscript m for simplicity.

- The decision is transmit if

$$d+\lambda + p_m h(1) + (1-p_m) h(d+1) \le d + h(d+1) \qquad (*)$$

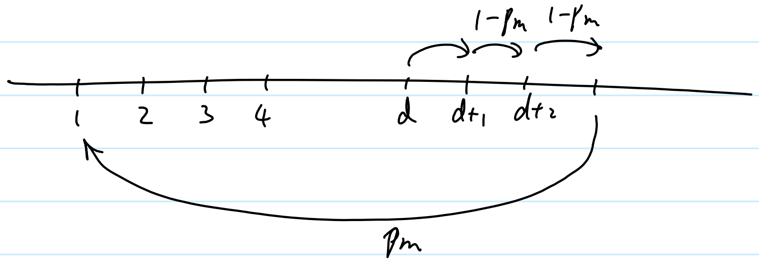
- We can choose $h(d) = 0$ for one of the states.
  - Let $h(1) = 0$

$$(*) \iff p_m h(d+1) \ge \lambda$$

$$h(d+1) \ge \frac{\lambda}{p_m}$$

- We can show that $h(d)$ is non-decreasing in d.

    - can be seen by looking at the SSP to state $d=1$.



where the cost at each stage (either $d+\lambda-\nu^*$ or $d-\nu^*$) is increasing in d.

- This then implies a threshold policy. There exists d_H such that the decision is to transmit if & only if $d \geq d_H$

    - The threshold d_H depends on both p_m & λ!

Go back to hard constraints

Friday, April 14, 2023 12:15 PM

- The relaxation from

$$\min \lim_{t\to+\infty} \frac{1}{t} E\left[\sum_{m=1}^{M}\sum_{k=0}^{t-1} d_k^m\right]$$

$$\text{sub to} \quad \sum_{m=1}^{M} u_k^m \le 1 \quad \text{for all } k$$

to

$$\min \lim_{t\to+\infty} \frac{1}{t} E\left[\sum_{m=1}^{M}\sum_{t=0}^{k-1} d_k^m\right]$$

$$\text{sub to} \quad \lim_{t\to+\infty} \frac{1}{t} E\left[\sum_{m=1}^{M}\sum_{k=0}^{t-1} u_k^m\right] \le 1 \quad (\lambda)$$

are however unsatisfactory:

- The per-sensor decisions, when put together, may be infeasible, i.e., they violate

$$\sum_{m=1}^{M} u_k^m \le 1$$

- For some classes of problems, the per-agent problem has a desirable structure, in which case a so-called index policies can be derived to always meet the original "hard" constraint.

- Roughly speaking, this structural property states

- Roughly speaking, this structural property states that, for each state d, there is a threshold value of dual price λ_{th}, such that,

  - when $\lambda \geq \lambda_{th}$, the optimal decision to the per-agent problem will be to use $u=0$ at state d.

  - when $\lambda < \lambda_{th}$, the optimal decision to the per-agent problem will be to use $u=1$ at state d.

  - Do not confuse this with the threshold policy in d. Here, as λ changes, the per-sensor problem itself changes!

- If this property holds, then we can use the threshold λ_{th} to denote the "urgency" of state d.

  - The higher the threshold, the more urgent this agent at state d will transmit

- Note that this threshold λ_{th} only depends on d, not time.

- Intuitively, we can, at each time, schedule the sensor with the highest λ_{th}

  - The "hard" constraint is then always respected!

- The hard constraint is then always respected!

- This threshold λ_{th} is called the Whittle's Index for state d

  - The corresponding policy is called Whittle's Index policy.

Reference:

1. J. Sun, Z. Jiang, B. Krishnamachari, S. Zhou and Z. Niu, "Closed-Form Whittle's Index-Enabled Random Access for Timely Status Update," https://arxiv.org/abs/1910.13871
2. P. Whittle, "Restless Bandits: Activity Allocation in a Changing World," Journal of Applied Probability, 1988

Indexability

Friday, April 14, 2023 12:16 PM

- Note that the Whittle's Index can only be properly defined when the indexability condition holds

- For a given λ, define the set of states for which the optimal per-agent decision is to use $u=0$ (i.e. do not transmit) as the passive set $P(\lambda)$

- The indexability condition requires that, as λ increases, $P(\lambda)$ must only increase in size, all the way to when $P(\lambda)$ includes all states

- When this indexability condition holds, then for each state d, there is a λ_{th} such that the state d first enters the passive set $P(\lambda_{th})$.

  $\Rightarrow$ This λ_{th} is then the Whittle's index for state d

- For $\lambda \geq \lambda_{th}$, d will always in $P(\lambda)$.

- Let us verify this for the AoI problem

- Recall the Bellman Equation:

$$h(d) + \nu^* = \min\Big\{ (d+\lambda) + p_m h(1) + (1-p_m) h(d+1), \quad d + h(d+1) \Big\}, \quad d = 1, 2, \cdots,$$

use ν^* for λ^*

drop the superscript m for simplicity.

- The decision is not to transmit

$$d+\lambda + p_m h(1) + (1-p_m) h(d+1) \geq d + h(d+1) \qquad (*)$$

- We can choose $h(d)=0$ for one of the states.
  - Let $h(1) = 0$

$$(*) \iff p_m h(d+1) \leq \lambda$$

$$h(d+1) \leq \frac{\lambda}{p_m}$$

- It may seem obvious that, if the decision at state d is $u=0$ for some λ, then as λ increases, the decision for state d should still be $u=0$.

  - Not quite as easy! The problem is that, as λ increases, the relative value function $h(d)$ also changes

  - Denote this dependence by $h^\lambda(d)$. We need to show that $h^\lambda(d+1)$ increases slower than $\frac{\lambda}{p_m}$, i.e., increases slower than $\frac{\lambda}{p_m}$, i.e., for $\Delta > 0$

$$h^{\lambda+\Delta}(d+1) - h^{\lambda}(d+1) \le \frac{\Delta}{p_m}$$

Why is that true?

- Consider two versions of the per-sensor MDP problems

- The two versions have the same state transitions.

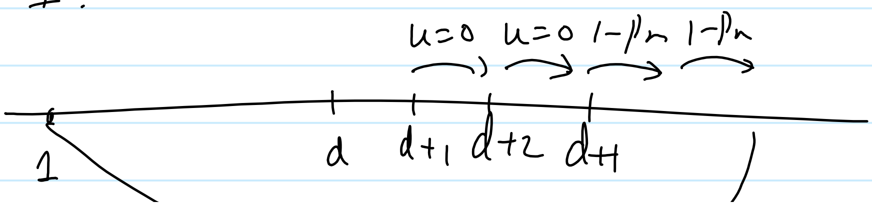
- The only difference is the the dual price λ & $\lambda+\Delta$.

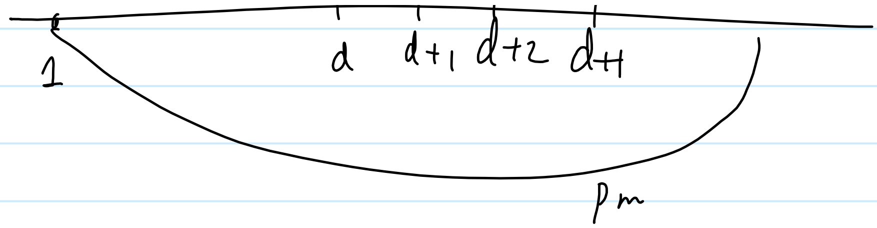
- We can then look at their corresponding SSP (to the termination state 1)

  - The two SSP subtract different values of $v^*(\lambda)$ & $v^*(\lambda+\Delta)$ at each time

  - Further, when $u=1$, they add different dual prices (differing by Δ).

- Suppose that, for λ, the optimal decision is to use $u=0$ until d increases to d_H, and then use $u=1$ until a success occurs and d reduces to 1.





- If we apply exactly the same decision to the per-agent problem with $\lambda+\Delta$, the increase in cost is

$$= \Delta \cdot (\#\text{ of times } u=1 \text{ until success}) - \left[v^*(\lambda+\Delta) - v^*(\lambda)\right] \cdot \left(\begin{array}{c}\#\text{ of times} \\ \text{until success}\end{array}\right)$$

- If I use the optimal decision for $\lambda+\Delta$, the cost should be even smaller.

$$\Rightarrow \quad h^{\lambda+\Delta}(d+1) - h^{\lambda}(d+1)$$

$$\leq \Delta \cdot \frac{1}{p_m} - \left(v^*(\lambda+\Delta) - v^*(\lambda)\right)\left(\begin{array}{c}\#\text{ of times} \\ \text{until success}\end{array}\right)$$

- But $v^*(\lambda+\Delta) \geq v^*(\lambda)$. Hence

$$h^{\lambda+\Delta}(d+1) - h^{\lambda}(d+1) \leq \frac{\Delta}{p_m}.$$

$\Rightarrow$ Whittle's indexability holds! We can then use Whittle's index policy for our problem

---

In general, establishing indexability is not an easy task.

Optimality of Whittle's Index Policy

Friday, April 14, 2023 12:16 PM

- Why is Whittle's Index policy good?
- It is not an optimal policy. But, it is asymptotically optimal for a "large" system.
- Such a large system will scale ~~both~~ the # of agents/sensors & the # of transmission opportunities by $\bar{C}$, and let $\bar{C} \to +\infty$.
  - Each agent becomes $\bar{C}$ identical agents.
- The original MDP becomes

$$v_0^*(c) = \min \lim_{t\to+\infty} \frac{1}{t} E\left[\sum_{m=1}^{M} \sum_{c=1}^{\bar{C}} \sum_{k=0}^{t-1} d_{kc}^{m}\right]$$

$$\text{sub to} \quad \sum_{m-1}^{M} \sum_{c=1}^{\bar{C}} u_{kc}^{m} \le \bar{c} \text{ for all } k$$

- The per-agent problem, however, does not change with $\bar{C}$ (recall that we relax the hard constraints first)

$$\min_{\lambda_m} \lim_{t\to+\infty} \frac{1}{t} E\left[\sum_{k=1}^{t-1} \left(d_k^m + \lambda u_k^m\right)\right]$$

- Therefore, the Lagrange-relaxed problem will have the same decisions. The total cost will be $v^*(\bar{C}) = \bar{C}\, v^*(1)$, when $v^*(1)$ is the solution to the relaxed version of the base problem

$$\min \lim_{t\to+\infty} \frac{1}{t} E\left[\sum_{m=1}^{M} \sum_{t=0}^{k-1} d_k^m\right]$$

$$\min \lim_{t \to +\infty} \frac{1}{t} E\left[\sum_{m=1}^{M} \sum_{t=0}^{K-1} d_k^m \right]$$

$$\text{Sub to } \lim_{t \to +\infty} \frac{1}{t} E\left[\sum_{m=1}^{M} \sum_{k=0}^{t-1} u_k^m \right] \leq 1$$

- By relaxation, we must have

$$\bar{c}\, v^*(1) \leq \underset{\substack{\uparrow \\ \text{True optimal for the MDP.}}}{v_0^*(\bar{c})}$$

- However, as $\bar{c}$ increases, when we follow the decisions of the per-agent problems, the chance of violating the hard constraint

$$\sum_{m=1}^{M} \sum_{c=1}^{\bar{c}} u_{kc}^m \leq \bar{c}$$

by a margin $(1+\varepsilon)$ would be smaller & smaller

- Think the sum of $\bar{c}$ i.i.d random variables, each of mean 1.

- Thus, as $\bar{c} \Uparrow$, we will have

$$v_0^*(\bar{c})\Big|_{1 \to 1+\varepsilon} \leq \bar{c}\, v^*(1)$$

↑ replace $\bar{c}$ by $\bar{c}(1+\varepsilon)$.

- This makes us believe that, as $\bar{c} \to +\infty$

$$\frac{v_0^*(\bar{c})}{\bar{c}} \longrightarrow \underset{\substack{\uparrow \\ \text{solution to the relaxed problem.}}}{v^*(1)}$$

- In other words, the decisions of the per-sensor MDP is not too bad, if we have the optimal λ.

- It remains to see that the Whittle's index policy is very similar to the decision of $\nu^*(1)$.

  - Let λ be the optimal dual variable.

  - The decision of $\nu^*(1)$ basically solves the per-sensor MDP with λ

- This means that

  ① For some states d, the decision is $u=1$.

  → But by the definition of Whittle's index, we must have

$$\lambda_{th}(d) \geq \lambda,$$

Whittle's index of state d

  ② For some states d, the decision is $u=0$

$$\rightarrow \quad \lambda_{th}(d) < \lambda$$

  ③ There may be some remaining states, such that the Cost-to-go $(u=1)$ = Cost-to-go $(u=0)$.

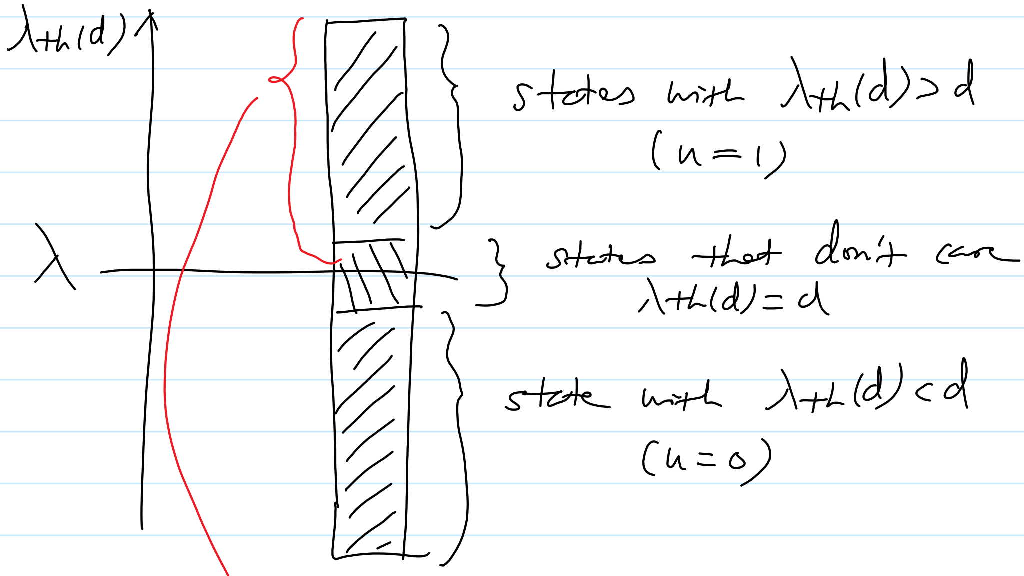
  - "Don't care"

  - To meet the KKT condition, these states should us some kind of .

- To meet the KKT condition, these states should us some kind of probablistic decision, so that the total resource constraint is met with equality on average

- These states will have $\lambda_{th}(d) = \lambda$

- Further, when $\bar{C}$ is large, the total resource constraint is approximately met at each time.



Total # of sensors with $u=1$
- on average $= \bar{C}$
- with high prob. $\le \bar{C}(1+\varepsilon)$

- In contrast, Whittle's index policy will sort the sensors by their current indices, and chose $\bar{C}$ of them with the highest indices.

- That means that most of the decisions will be the same as that of $\nu^*(1)$.

- The above is a sketch of the intuitions why

- The above is a sketch of the intuitions why the Whittle's Index policy is asymptotically optimal.

- For further details, see the paper

R. R. Weber and G. Weiss, "On an index policy for restless bandits," Journal of Applied Probability, vol. 27, no. 3, p.637–648, 1990