

Lecture 3

Introduction to Dynamic Programming (DP)

Outline: Dynamic Programming (DP)

- Notation and Terminology
- Principle of Optimality
- Dynamic Programming Algorithm for Finite Horizon Problems
- Inventory Control Example Revisited

Dynamic Programming

Dynamic Programming: A general mathematical framework for **sequential** decision making under uncertainty. The goal is to minimize a certain (additive) cost subject to the rules of evolution of the system.

“Dynamic programming: a name not even a Congressman could object to.”
—Richard Bellman

Introduction to Dynamic Programming

A mathematical framework of **sequential** decision making under uncertainty

1. State: $s_t \in \mathcal{S}$
2. Dynamics: $s_{t+1} = f_t(s_t, a_t, w_t)$
3. Stage cost: $c_t(s_t, a_t, w_t)$
4. Control constraints: $a_t \in A_t(s_t)$
5. Model of uncertainty: e.g., $\mathbb{P}(\cdot \mid s_t, a_t)$ of w_t

$$\mathbb{E} \left[c_N(s_N) + \sum_{t=0}^{N-1} c_t(s_t, a_t, w_t) \right]$$

■ A feedback policy π , is a sequence of functions: $\pi = \{\mu_0, \dots, \mu_{N-1}\}$
where $\mu_t : s_t \mapsto a_t = \mu_t(s_t) \in A_t(s_t)$

$$V_N^\pi(s_0) = \mathbb{E} \left[c_N(s_N) + \sum_{t=0}^{N-1} c_t(s_t, \mu_t(s_t), w_t) \right]$$

■ $V_N^*(s_0) = \min_{\pi} V_N^\pi(s_0)$

■ $\pi^* \in \arg \min_{\pi} V_N^\pi(s_0) \quad \text{i.e.,} \quad V_N^{\pi^*}(s_0) = V_N^*(s_0)$

DP Variations: Discounted Problems

$$\begin{aligned} \min_{\pi} \quad & \mathbb{E} \left[\gamma^N c_N(s_N) + \sum_{t=0}^{N-1} \gamma^t c_t(s_t, a_t, w_t) \right] \\ \text{s.t.} \quad & s_{t+1} = f_t(s_t, a_t, w_t), \quad a_t = \mu_t(s_t) \in A_t(s_t), \\ & w_t \sim \mathbb{P}(\cdot \mid s_t, a_t) \end{aligned}$$

DP Variations: Infinite Horizon Discounted Cost

$$\begin{aligned} \min_{\pi} \quad & \lim_{N \rightarrow \infty} \mathbb{E} \left[\sum_{t=0}^{N-1} \gamma^t c_t(s_t, a_t, w_t) \right] \\ \text{s.t.} \quad & s_{t+1} = f_t(s_t, a_t, w_t), \quad a_t = \mu_t(s_t) \in A_t(s_t), \\ & w_t \sim \mathbb{P}(\cdot \mid s_t, a_t) \end{aligned}$$

DP Variations: Infinite Horizon Average Cost

$$\min_{\pi} \quad \lim_{N \rightarrow \infty} \frac{1}{N} \quad \mathbb{E} \left[\sum_{t=0}^{N-1} c_t(s_t, a_t, w_t) \right]$$

$$\text{s.t.} \quad s_{t+1} = f_t(s_t, a_t, w_t), \quad a_t = \mu_t(s_t) \in A_t(s_t),$$

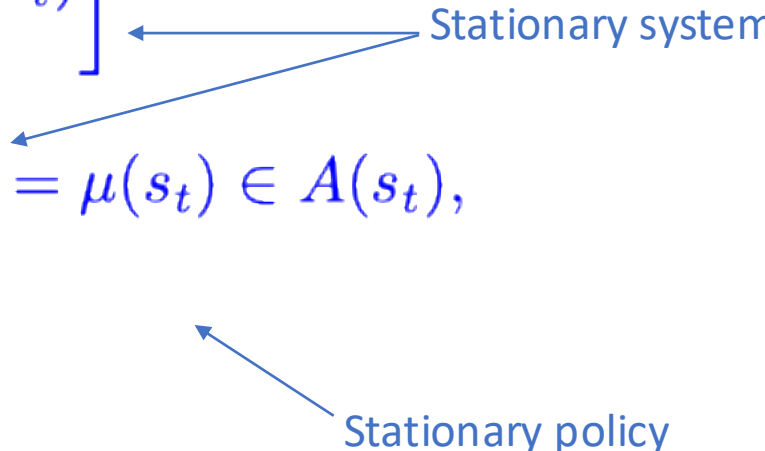
$$w_t \sim \mathbb{P}(\cdot \mid s_t, a_t)$$

DP Simplification: Stationarity

$$\begin{aligned} \min_{\pi} \quad & \mathbb{E} \left[c(s_N) + \sum_{t=0}^{N-1} c(s_t, a_t, w_t) \right] \\ \text{s.t.} \quad & s_{t+1} = f(s_t, a_t, w_t), \quad a_t = \mu(s_t) \in A(s_t), \\ & w_t \sim \mathbb{P}(\cdot \mid s_t, a_t) \end{aligned}$$

Stationary system

Stationary policy



Variations

- Simplification: Assume stationarity!!

The simplest case is finite-horizon, discrete-time, deterministic with perfect information.

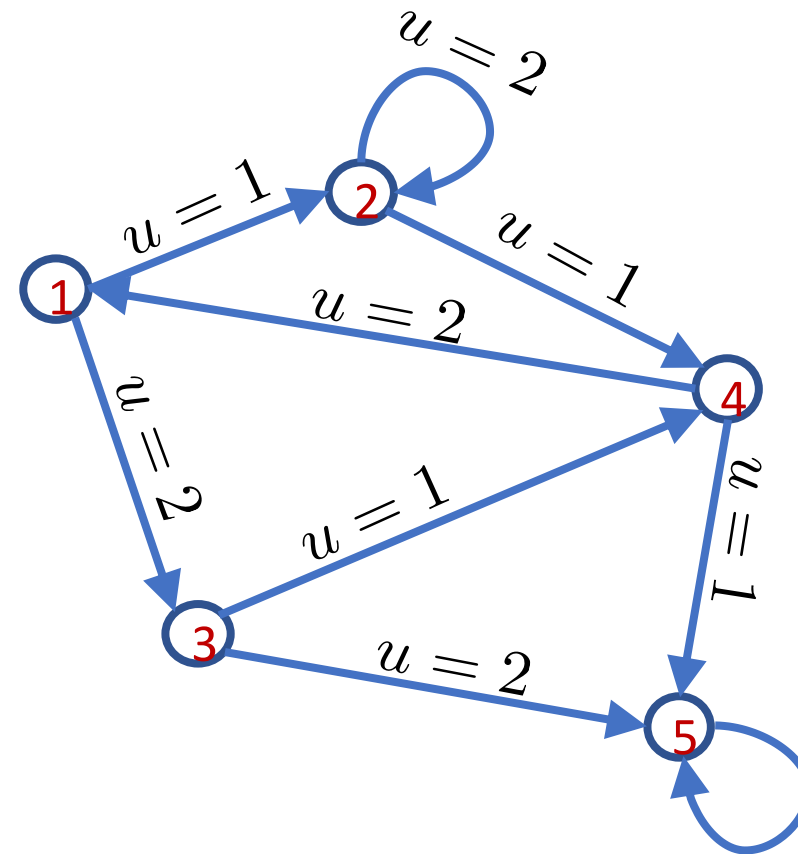
Markov Decision Processes (MDP) : Finite states and actions

$\langle X, U, g, f, P \rangle$

- States
- Control
- Stage Cost
- Dynamics
- Model of Uncertainty

$\langle S, A, g, P \rangle$

- States $\{i_1, \dots, i_n\}$
- Actions $\{a_1, \dots, a_m\}$
- Rewards $g(i, u, j)$
- Transition Probabilities $p_{ij}(a) = p(j; i, a)$

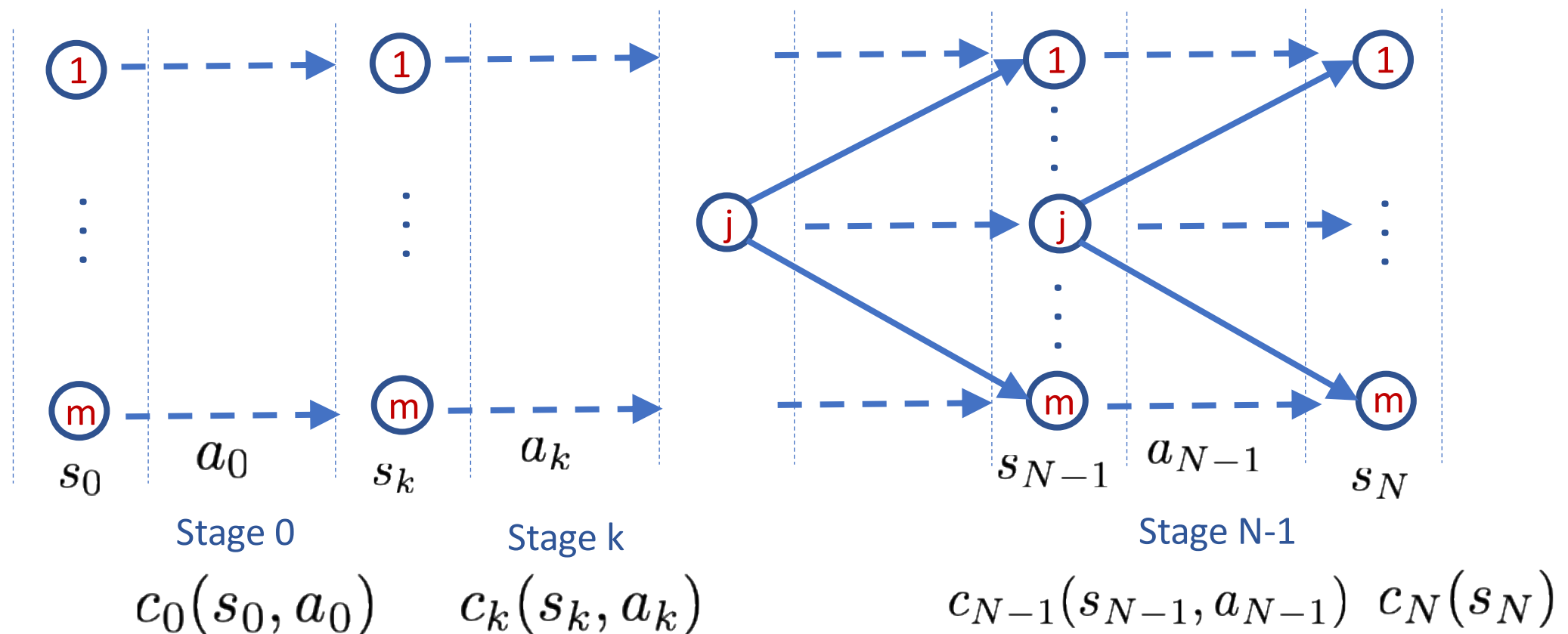


Deterministic Problems

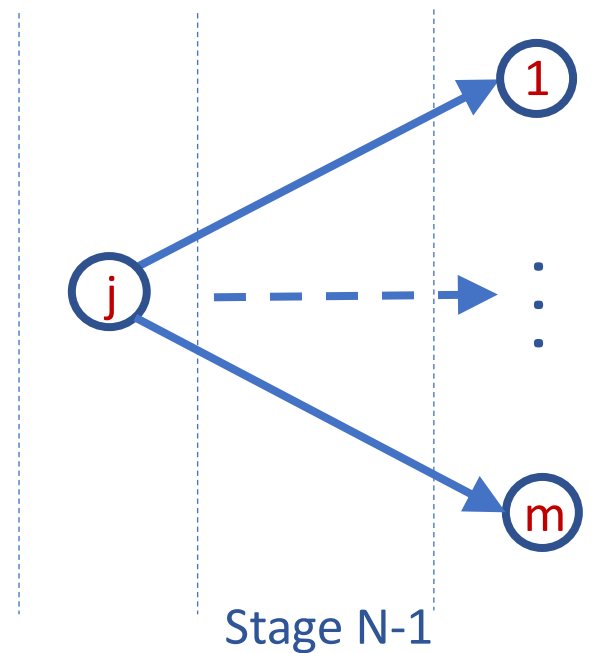
Solving DPs: Deterministic, Finite-Horizon Case

$$\min_{\pi} \left[c_N(s_N) + \sum_{t=0}^{N-1} c_t(s_t, a_t) \right]$$

$$\text{s.t. } s_{t+1} = f_t(s_t, a_t), \quad a_t = \mu_t(s_t) \in A_t(s_t),$$



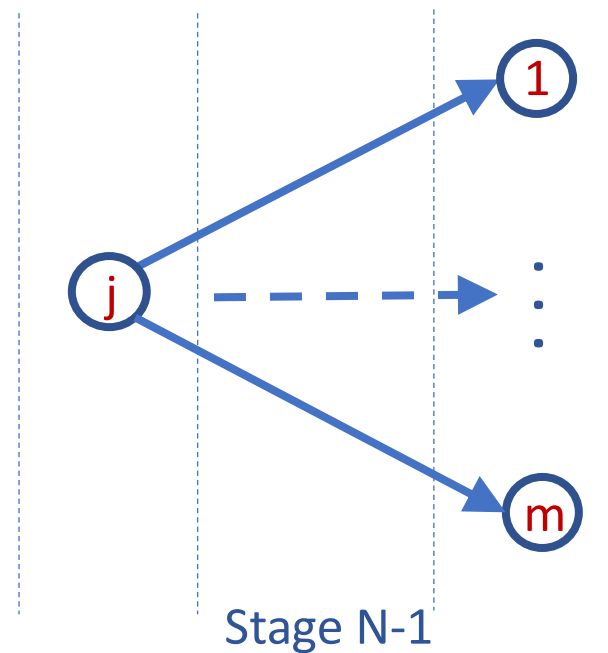
Solving DPs: What would you do?



$$\begin{aligned} \min \quad & c_N(s_N) + c_{N-1}(s_{N-1}, a_{N-1}) \\ \text{s.t.} \quad & s_N = f_t(s_{N-1}, a_{N-1}) \\ & a_{N-1} = \mu_t(s_{N-1}) \in A_t(s_{N-1}) \end{aligned}$$

We have: $V_{N-1}^*(s) = \min_{a \in A} \{c_{N-1}(s, a) + c_N(f(s, a))\}$

Solving DPs: What would you do?



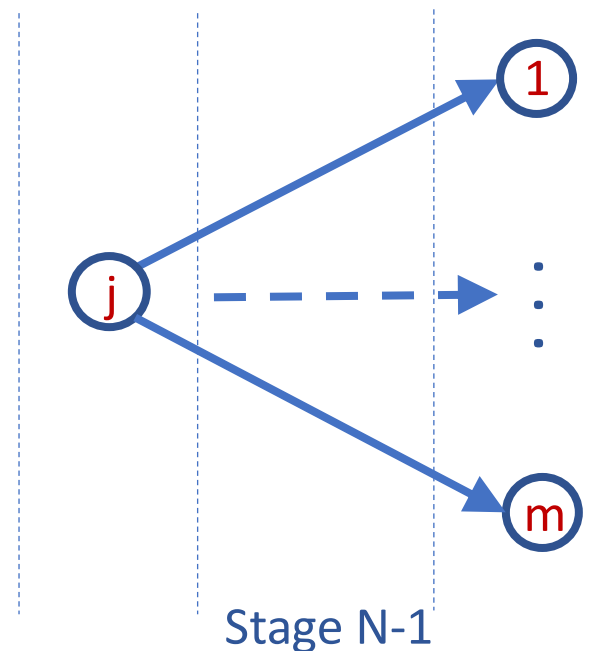
$$\begin{aligned} \min \quad & c_N(s_N) + c_{N-1}(s_{N-1}, a_{N-1}) \\ \text{s.t.} \quad & s_N = f_t(s_{N-1}, a_{N-1}) \\ & a_{N-1} = \mu_t(s_{N-1}) \in A_t(s_{N-1}) \end{aligned}$$

We have: $V_{N-1}^*(s) = \min_{a \in A} \{c_{N-1}(s, a) + c_N(f(s, a))\}$

$$c_{N-1}(j, a_{N-1}^*) + c_N(f(j, a_{N-1}^*)) = \min_a \{c_{N-1}(j, a) + c_N(f(j, a))\}$$

is the optimal cost-to-go associated with state j at time $N - 1$.

Solving DPs: What would you do?



$$\begin{aligned} \min \quad & c_N(s_N) + c_{N-1}(s_{N-1}, a_{N-1}) \\ \text{s.t.} \quad & s_N = f_t(s_{N-1}, a_{N-1}) \\ & a_{N-1} = \mu_t(s_{N-1}) \in A_t(s_{N-1}) \end{aligned}$$

We have: $V_{N-1}^*(s) = \min_{a \in A} \{c_{N-1}(s, a) + c_N(f(s, a))\}$

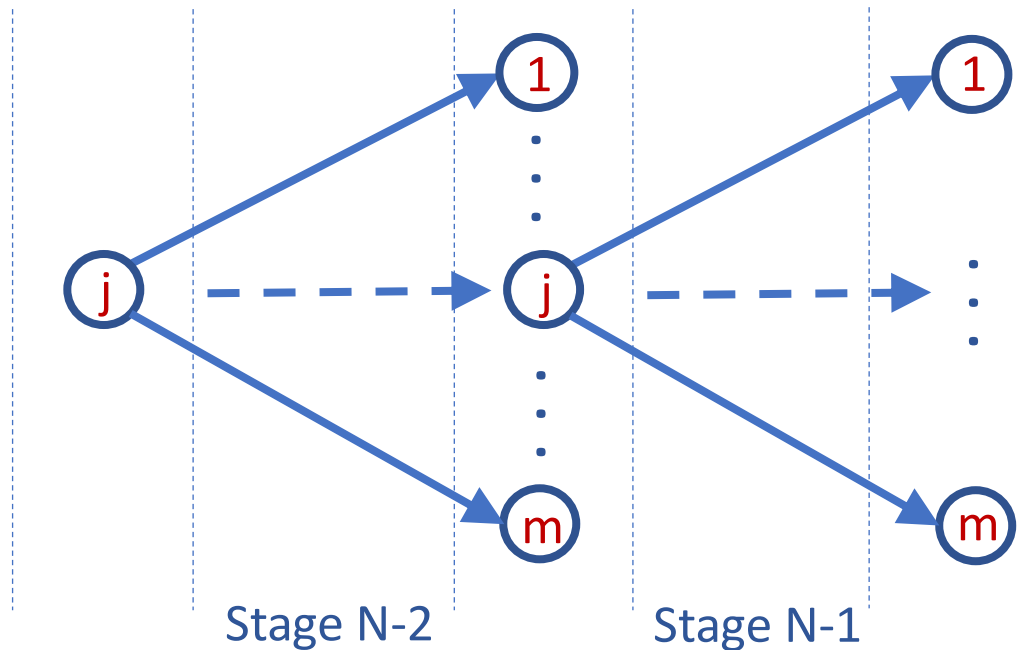
$$c_{N-1}(j, a_{N-1}^*) + c_N(f(j, a_{N-1}^*)) = \min_a \{c_{N-1}(j, a) + c_N(f(j, a))\}$$

is the optimal cost-to-go associated with state j at time $N - 1$.

Let's call this $V_{N-1}^*(j)$.

What about $\mu_{N-1}^*(s)$?

Solving DPs: What would you do?

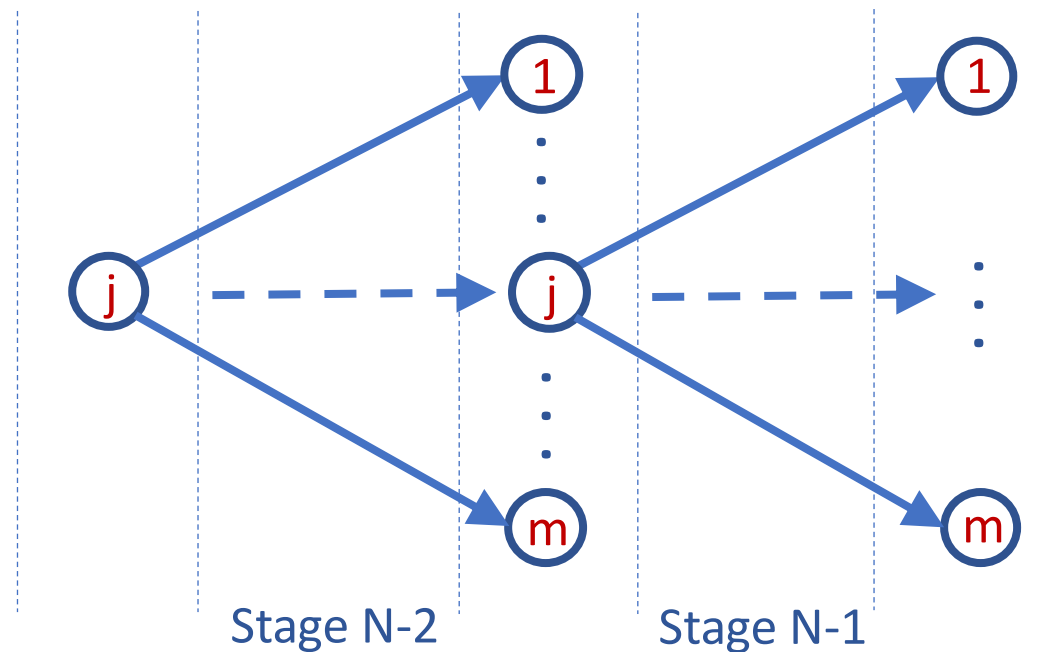


$$\begin{aligned} \min \quad & c_N(s_N) + \sum_{t=0}^{N-1} c_t(s_t, a_t) \\ \text{s.t.} \quad & s_{t+1} = f_t(s_t, a_t) \\ & a_t = \mu_t(s_t) \in A_t(s_t) \end{aligned}$$

$$\begin{aligned} & c_{N-2}(s_{N-2}, a_{N-2}) + c_{N-1}(s_{N-1}, a_{N-1}) + c_N(s_N) \\ & \geq c_{N-2}(s_{N-2}, a_{N-2}) + V_{N-1}^*(s_{N-1}) \\ & = c_{N-2}(s_{N-2}, a_{N-2}) + V_{N-1}^*(f(s_{N-2}, a_{N-2})) \end{aligned}$$

Now minimize both sides!

Solving DPs: What would you do?



$$\begin{aligned} \min \quad & g_N(x_N) + \sum_{t=0}^{N-1} g_t(x_t, u_t) \\ \text{s.t.} \quad & x_{t+1} = f_t(x_t, u_t) \\ & u_t = \mu_t(x_t) \in U_t(x_t) \end{aligned}$$

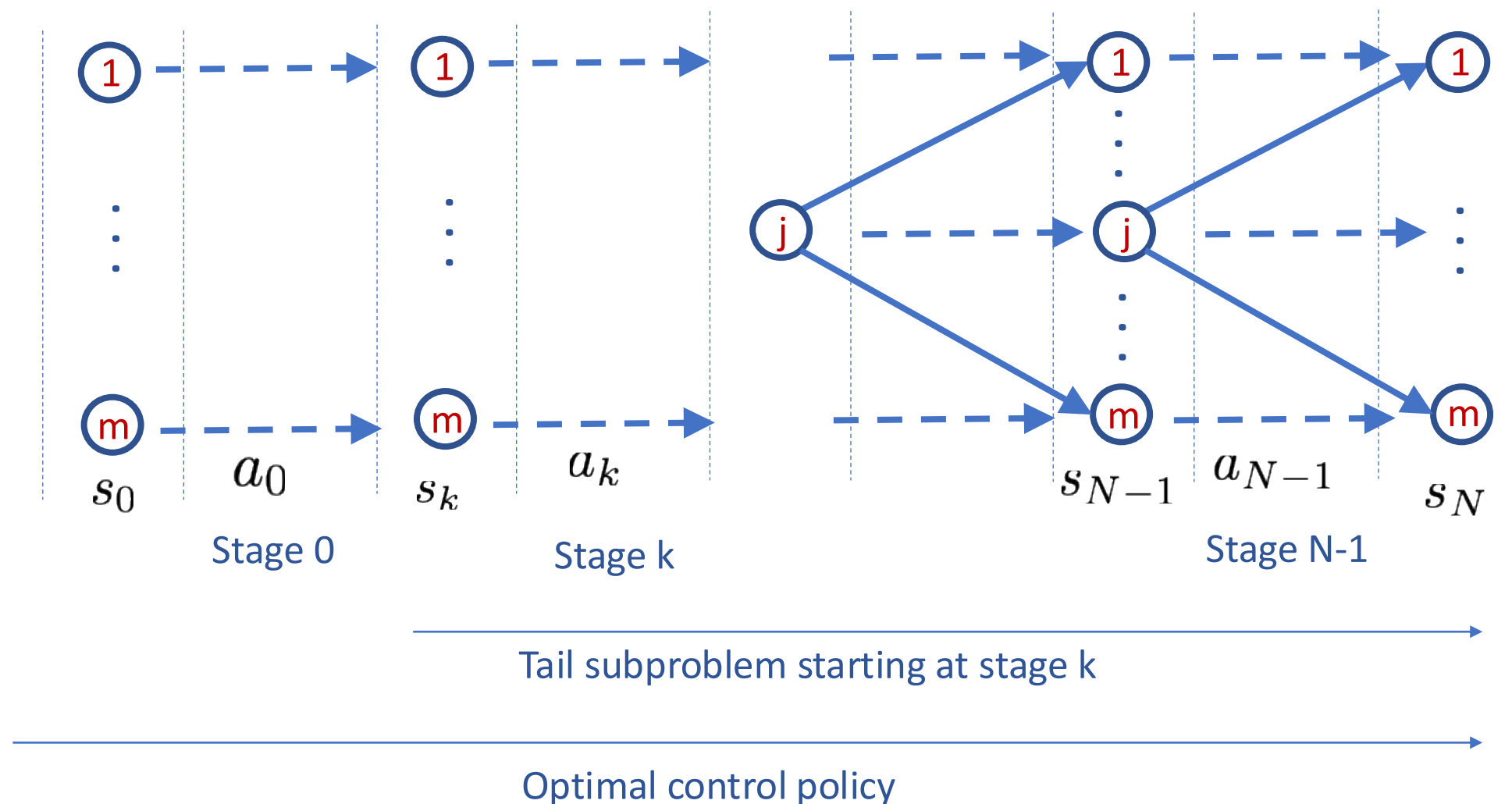
What is the (optimal) cost-to-go of state x at time $N - 2$?

$$V_{N-2}^*(s) = \min_a \{c_{N-2}(s, a) + V_{N-1}^*(f(s, a))\}$$

$$\mu_{N-2}^*(s) = \arg \min_a \{c_{N-2}(s, a) + V_{N-1}^*(f(s, a))\}$$

compare with $V_{N-1}^*(s) = \min_a \{c_{N-1}(s, a) + c_N(f(s, a))\}$

Principle of Optimality



Bellman's Principle of Optimality:

- The truncated (tail) optimal policy is optimal for the truncated sub-problem.
- **Optimization of** the future does not depend on the past.

Dynamic Programming Algorithm

Dynamic Programming: Do what we did already: Solve all tail subproblems of a given length using the solution of shorter tail subproblems!

Initialize backward iteration at $V_N(s) = c_N(s)$.

$$V_t(s) = \min_{a \in A_t(s)} \{c_t(s, a) + V_{t+1}(f_t(s, a))\}$$

$$= c_t(s, \mu_t^*(s)) + V_{t+1}(f_t(s, \mu_t^*(s)))$$

$$\mu_t^*(s) = \arg \min_{a \in A_t(s)} \{c_t(s, a) + V_{t+1}(f_t(s, a))\}$$

Then $V_0(s_0)$ is the optimal cost and the generated policy is optimal.

DP Algorithm: With Stochastic Uncertainty

Good news: The algorithm works in the same way.

$$\begin{aligned} \min_{\pi} \quad & \mathbb{E} \left[c_N(s_N) + \sum_{t=0}^{N-1} c_t(s_t, a_t, w_t) \right] \\ \text{s.t.} \quad & s_{t+1} = f_t(s_t, a_t, w_t), \quad a_t = \mu_t(s_t) \in A_t(s_t), \\ & w_t \sim \mathbb{P}(\cdot \mid s_t, a_t) \end{aligned}$$

- $V_t(s_t)$: optimal cost of tail subproblem starting at s_t .
- Initialize: $V_N(s_N) = c_N(s_N)$
- Backward iteration:

$$V_t(s_t) = \min_{a \in A_t(s_t)} \mathbb{E} [c_t(s_t, a, w_t) + V_{t+1}(f_t(s_t, a, w_t))]$$

- Then $V_0(s_0)$ is the optimal cost and the generated policy is optimal.

DP Algorithm: Computing Cost in Uncertain Case

Define

$$G_t = \sum_{\tau=t}^N c_{\tau} (s_{\tau}, \mu(s_{\tau}))$$

G_t : Total cost from time t to terminal time N

Cost Evaluation

Fix strategy π , then:

$$V_t^\pi(s) = \mathbb{E}[G_t \mid s_t = s]$$

$$V_{N-1}^\pi(s) = c(s, \mu_{N-1}(s)) + \sum_{s'} p(s' \mid s, \mu_{N-1}(s)) \cdot c_N(s')$$

$$V_t^\pi(s) = c(s, \mu(s)) + \mathbb{E}[G_{t+1} \mid s_t = s]$$

$$= c(s, a) + \mathbb{E} [\mathbb{E}[G_{t+1} \mid s_t = s, s_{t+1} = s']]$$

$$= c(s, a) + \sum_{s'} p(s'; s, a) \cdot \mathbb{E}[G_{t+1} \mid s_{t+1} = s'] \quad (\text{Markovian})$$

$$= c(s, a) + \sum_{s'} p(s' \mid s, a) \cdot V_{t+1}^\pi(s')$$

Cost Evaluation: Summary

- Initialize: $V_N(s_N) = c_N(s_N)$
- Backward iteration:

$$V_t(s_t) = \mathbb{E} [c_t(s_t, a, w_t) + V_{t+1}(f_t(s_t, a, w_t))]$$

How about the optimal solution

The Last Step

$$\begin{aligned} V_{N-1}^*(s) &= \min_a \left\{ c(s, a) + \sum_{s'} p(s' \mid s, a) \cdot c_N(s') \right\} \\ &= \min_a \left\{ c(s, a) + \sum_{s'} p(s' \mid s, a) \cdot V_N^*(s') \right\} \end{aligned}$$

Bellman's Equation for Uncertain Problems

$$\pi = \{\mu_{N-1}, \mu_{N-2}\}$$

$$V_{N-2}^*(s) = \min_{\mu_{N-2}, \mu_{N-1}} \left\{ c(s, \mu_{N-2}(s)) + \sum_{s'} p(s' \mid s, \mu_{N-2}(s)) \cdot V_{N-1}(s') \right\}$$
$$V_{N-2}^*(s) = \min_{\mu_{N-2}} \left\{ c(s, \mu_{N-2}(s)) + \min_{\mu_{N-1}} \sum_{s'} p(s' \mid s, \mu_{N-2}(s)) \cdot V_{N-1}(s') \right\}$$

Can we switch averaging with minimization?

Bellman's Equation for Uncertain Problems

Can we switch averaging with minimization?

$$\min_{\mu_{N-1}(\cdot)} \mathbb{E} [F(s, s', \mu_{N-1}(s'))] \stackrel{?}{=} \mathbb{E} \left[\min_{\mu_{N-1}} F(s, s', \mu_{N-1}(s')) \right]$$

Immediate Inequality:

$$\mathbb{E} [F(s, s', \mu_{N-1}(s'))] \geq \mathbb{E} \left[\min_{\mu_{N-1}} F(s, s', \mu_{N-1}(s')) \right]$$

Bellman's Equation for Uncertain Problems

Can we switch averaging with minimization?

$$\min_{\mu_{N-1}(\cdot)} \mathbb{E} [F(s, s', \mu_{N-1}(s'))] \stackrel{?}{=} \mathbb{E} \left[\min_{\mu_{N-1}} F(s, s', \mu_{N-1}(s')) \right]$$

Immediate Inequality:

$$\mathbb{E} [F(s, s', \mu_{N-1}(s'))] \geq \mathbb{E} \left[\min_{\mu_{N-1}} F(s, s', \mu_{N-1}(s')) \right]$$

Suppose the right-hand side is achieved at an optimal policy $\mu_{N-1}^*(\cdot)$

$$\min_{\mu_{N-1}(\cdot)} \mathbb{E} [F(s, s', \mu_{N-1}(s'))] \leq \mathbb{E} [F(s, s', \mu_{N-1}^*(s'))]$$

Equality holds!

Bellman's Equation for Uncertain Problems

$$\pi = \{\mu_{N-1}, \mu_{N-2}\}$$

$$V_{N-2}^*(s) = \min_{\mu_{N-2}, \mu_{N-1}} \left\{ c(s, \mu_{N-2}(s)) + \sum_{s'} p(s' \mid s, \mu_{N-2}(s)) \cdot V_{N-1}(s') \right\}$$
$$V_{N-2}^*(s) = \min_{\mu_{N-2}} \left\{ c(s, \mu_{N-2}(s)) + \min_{\mu_{N-1}} \sum_{s'} p(s' \mid s, \mu_{N-2}(s)) \cdot V_{N-1}(s') \right\}$$

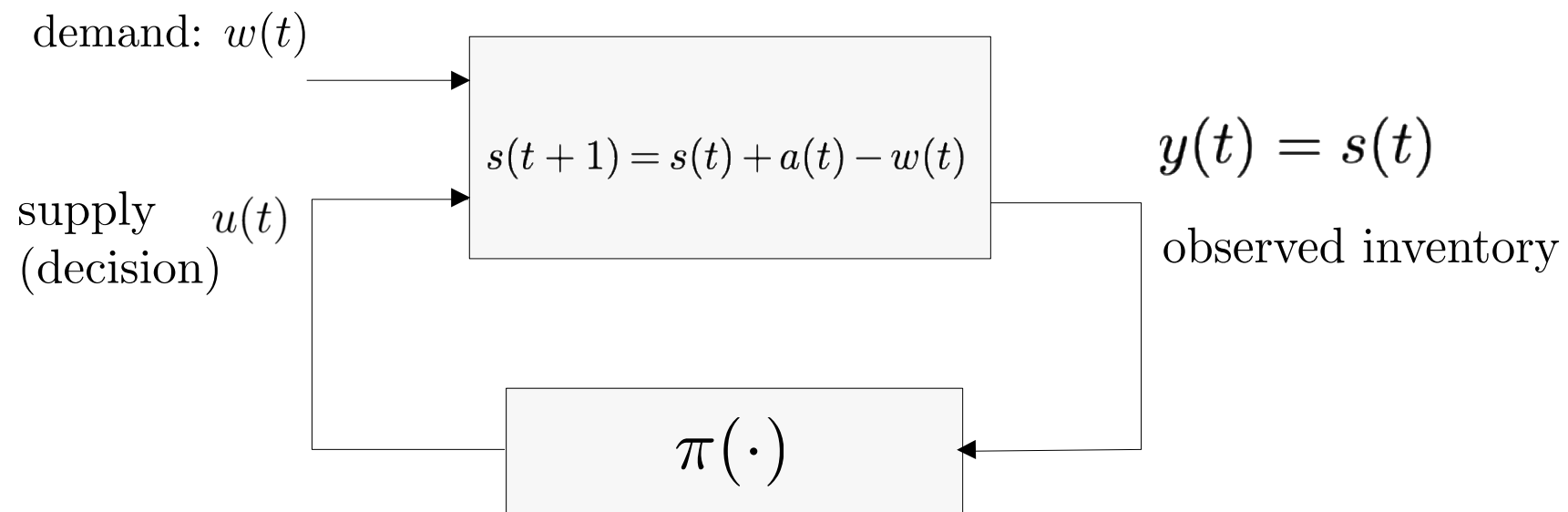
Switching averaging with minimization:

$$V_{N-2}^*(s) = \min_{\mu_{N-1}} \sum_{s'} c(s, \mu_{N-1}(s)) + \min_{\mu_{N-2}} \mathbb{E}_{s'} [V_{N-1}(s')]$$

$$V_{N-2}^*(s) = \min_{\mu_{N-1}} \sum_{s'} c(s, \mu_{N-1}(s)) + \mathbb{E}_{s'} \min_{\mu_{N-2}} [V_{N-1}(s')]$$

$$V_{N-2}^*(s) = \min_{\mu_{N-1}} \sum_{s'} c(s, \mu_{N-1}(s)) + \mathbb{E}_{s'} [V_{N-1}^*(s')]$$

Motivation: Inventory control



Cumulative cost over N periods:

$$\mathbb{E} \left[c(s(N)) + \sum_{t=0}^{N-1} \{c(s_t, a_t)\} \right]$$

How should we make supply order decisions to minimize this cost / maximize profit?

Inventory control example: Convexity of Cost to go

- Induction

$$V_N(s) = 0,$$

$$\begin{aligned} t = N - 1 \quad V_{N-1}(s) &= \min_{a \geq 0} \{ da + g(s) + \mathbb{E}[V_N(s + a - w)] \} \\ &= \min_a \{ da + g(s) \} \end{aligned}$$

- Assume true for t

$$V_{t-1}(s) = \min_{a \geq 0} \{ da + g(s) + \mathbb{E}[V_t(s + a - w)] \}$$

- Expectation preserves convexity.

$$H(s + a) = \mathbb{E}[V_t(s + a - w)]$$

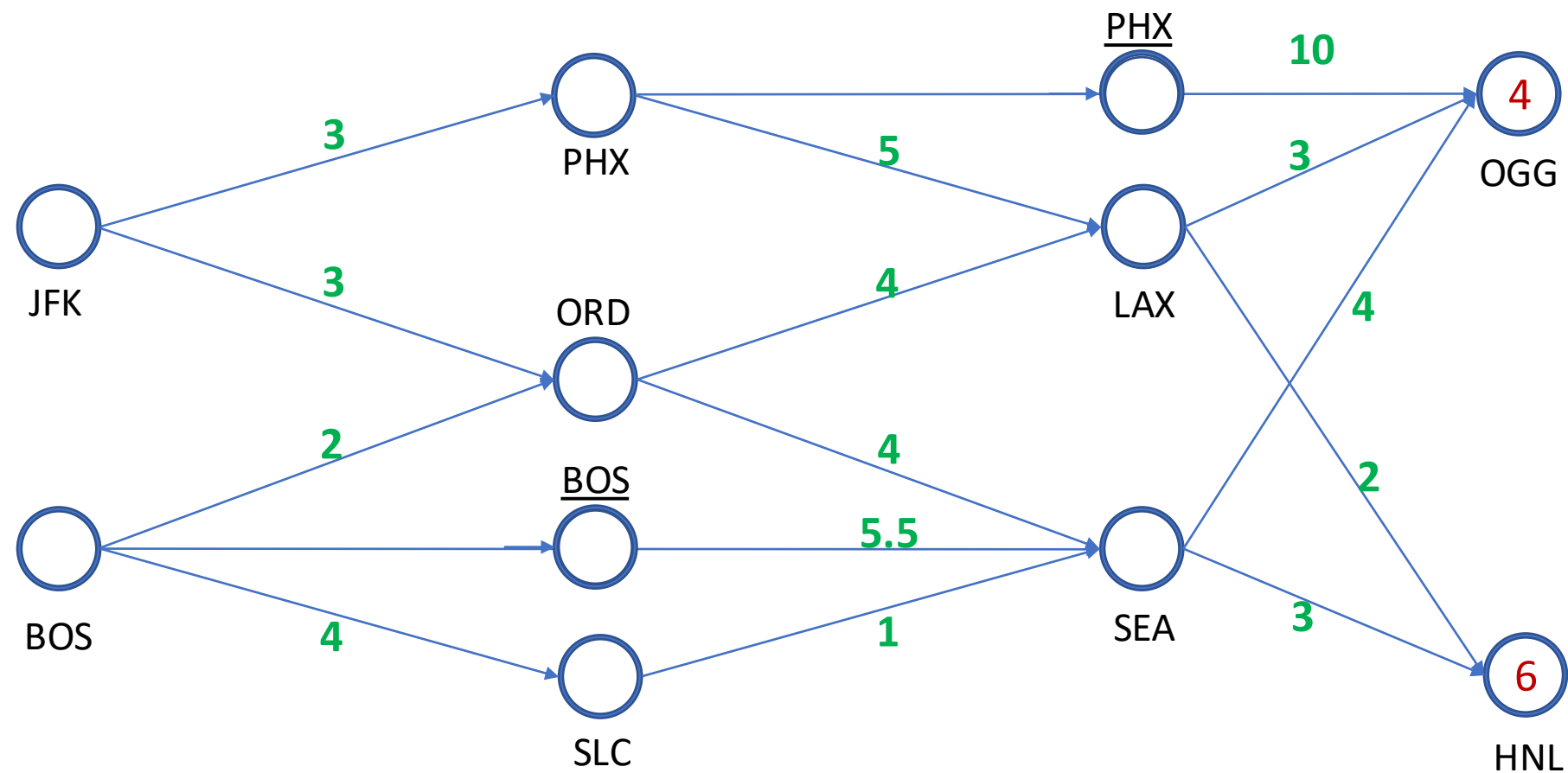
- Need to show the convexity of

$$g(s) + \min_{a \geq 0} \{ da + H(s + a) \} = G(s)$$

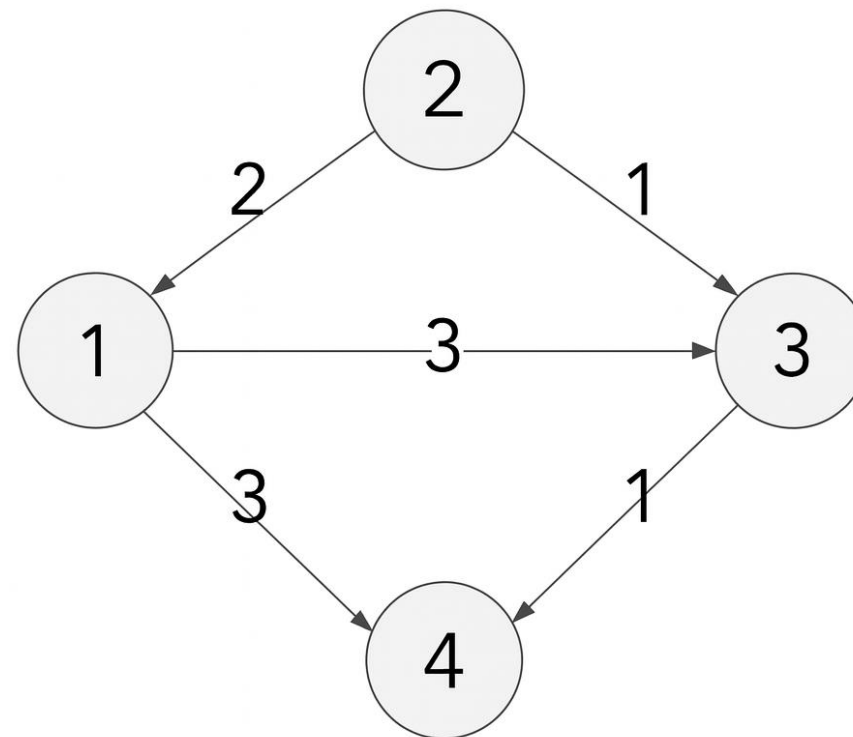
Let $S_t = \arg \min_{y \in \mathbb{R}} G_t(y)$. Then the optimal policy is determined by the sequence of scalars (thresholds) $\{S_0, \dots, S_{N-1}\}$ and is of the form:

$$\mu_t^*(s) = \begin{cases} S_t - s, & \text{if } s < S_t \\ 0, & \text{if } s \geq S_t \end{cases}$$

Shortest Path: Deterministic Case



Shortest Path: Deterministic Case



Shortest path problem

Problem: Find a shortest (i.e., minimum cost) walk of length k that ends at node n

Shortest Path: Deterministic Case

1. Directed graph (S, A) with n nodes. We use i and j to denote generic nodes. Here A is the set of directed arcs (i, j) . We allow self-arcs, of the form (i, i) . Let $A(i)$ be the set of j for the arc (i, j) is present, i.e., $(i, j) \in A$.
2. Node n is a destination node. We have a self-arc (n, n) and node n is the only element of $A(n)$.
3. For every $(i, j) \in A$, we have a cost c_{ij} . If $(i, j) \notin A$, we let $c_{ij} = \infty$. We assume that $c_{nn} = 0$.
4. A walk is a sequence of nodes that can be traversed in order, following the available arcs. Formally, a walk of length t is a finite sequence $i_0, i_1, \dots, i_t$, with $i_{k+1} \in A(i_k)$, for $k = 0, \dots, t - 1$.
5. A path is a walk in which all nodes are distinct.
6. A cycle is a walk in which there are no repeated nodes, except that $i_t = i_0$.
7. The cost of a path or a walk or a cycle is the sum of the arc costs involved.

Problem: Find a shortest (i.e., minimum cost) walk of length k that ends at node n

Shortest Path: State and Time?

- Define

$V_k(i)$ is the length of an optimal cost of a walk of length k .

- Observe

$$V_{k+1}(i) \leq V_k(i)$$

$V_{n-1}(i) < \infty$ If and only if the problem is feasible

- Show

$$V_k(i) = \min_{j \in A(i)} \{c_{ij} + V_{k-1}(j)\} \quad V_0(n) = 0, \quad V_0(i) = \infty, \quad i \neq n$$

Optimal Characterization

- Show

$$V(i) = \min_{j \in A(i)} \{c_{ij} + V(j)\}$$

Questions