Discrete-Time Dynamic Programming

Jae Yun JUN KIM*

Reference: Neil Walton's lecture notes

1 Introduction

1.1 Introductory examples

Example 1

In the figure below, there is a tree consisting of a root node labeled R and two leaf nodes colored grey. For each edge, there is a cost. You may turn left or right at the root node. Find the lowest cost path from the root to a leaf.

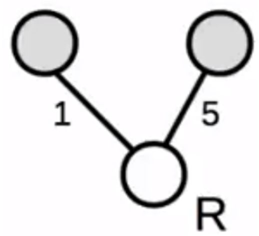

Figure 1: Example 1

Answer: Go to left.

Example 2 (continued)

Again, at each node, you may turn right or left. Find the lowest cost path from a root node (labeled R) to a leaf node (colored gray). [Hint: use your answer from Example 1 - three times]

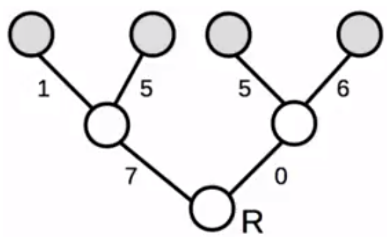

Figure 2: Example 2

Answer: Go to right and then left.

*ECE Paris Graduate School of Engineering, 37 quai de Grenelle 75015 Paris, France; jae-yun.jun-kim@ece.fr

Example 3 (continued)

Again, find the lowest cost path from a root node (labeled R) to a leaf node (colored gray). [Hint: use your answer from [2].]

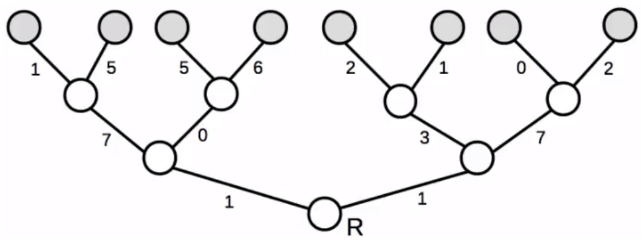

Figure 3: Example 3

Answer: Go to right, then left and then right.

Example 4 (continued)

In the figure below, the tree on the right hand side (rhs) has a lowest cost path of value L_{rhs} and the left hand side tree has lowest cost L_{lhs}. The edges leading to each respective tree have costs l_{rhs} and l_{lhs}. Show that L, the minimal cost path from the root to a leaf node satisfies

$$L = \min_{a \in \{lhs, rhs\}} l_a + L_a. \tag{1}$$

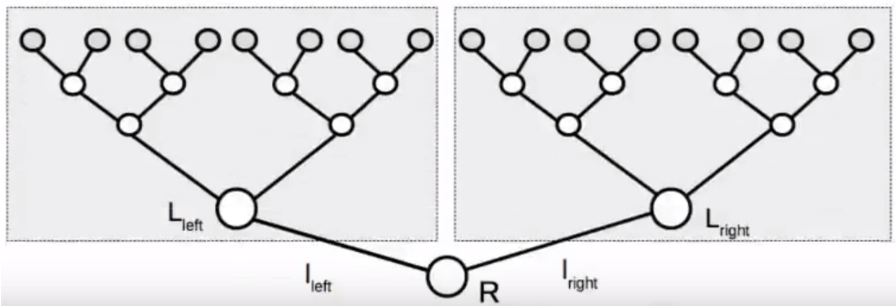

Figure 4: Example 4

Answer:
Left has cost $l_{lhs} + L_{lhs}$ and right has cost $l_{rhs} + L_{rhs}$. So, $L = \min_{a \in \{lhs, rhs\}} l_a + L_a$.

Example 5

Convince yourself that the same argument applies from any node x in the tree network that is

$$L = \min_{a \in \{lhs, rhs\}} l_a + L_x(a), \tag{2}$$

where L_x is the minimum cost from x to a leaf node, and where, for $a \in \{lhs, rhs\}$, $x(a)$ is the node to the left hand side or right hand side of x.

Answer:
Notice that the idea of solving a problem from back to front and the idea of iterating on the above equation to solve an optimization problem lies at the heart of dynamic programming.

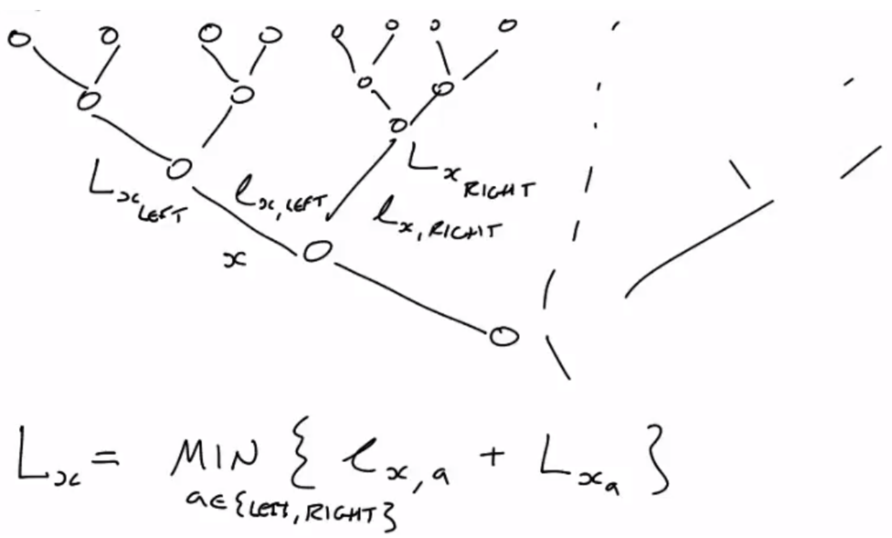

Figure 5: Answer for example 5

2 Dynamic programming

For this section, consider the following *dynamic programming* formulation. Time is discrete $t = 0, 1, \cdots, T$; $x_t \in \mathcal{X}$ is the state at time t; $a_t \in \mathcal{A}_t$ is the action at time t;
Definition (Plant Equation): The state evolves according to functions $f_t : \mathcal{X} \times \mathcal{A}_t \to \mathcal{X}$. Here,

$$x_{t+1} = f(x_t, a_t). \tag{3}$$

This is called the **Plant Equation**.

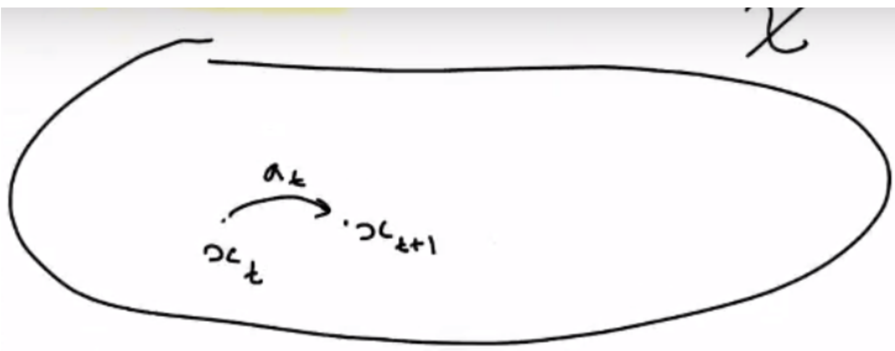

Figure 6: Plant Equation

A policy π chooses an action π_t at each time t. The (instantaneous) reward for taking action a in state x at time t is $r_t(a, x)$, and $r_T(x)$ is the reward for terminating in tate x at time T.

Definition (Dynamic program (DP)):
Given initial state x_0, a dynamic program is the optimization

$$\begin{aligned} W(x_0) :=& \text{Maximize} \quad R(a) := \sum_{t=0}^{T-1} r_t(x_t, a_t) + r_T(x_T) \\ & \text{subject to} \qquad x_{t+1} = f(x_t, a_t) \quad t = 0, \cdots, T-1 \\ & \text{over} \qquad a_t \in \mathcal{A}_t \quad t = 0, \cdots, T-1. \end{aligned} \tag{4}$$

Further, let $R_\tau(a)$ (respectively, $W_\tau(x_\tau)$) be the objective (respectively optimal objective) for DP, when the summation is started from $t = \tau$, rather than $t = 0$.

Notice that

$$R_\tau(a) = \sum_{t=\tau}^{T-1} r_t(x_t, a_t) + r_T(x_T) \tag{5}$$

and

$$W_\tau(x_\tau) = \max_{a_\tau=(a_\tau,\cdots,a_T)} R_\tau(a). \tag{6}$$

Example 6(Bellman's equation)

$W_T(x) = r_T(x)$ and for $t = T-1, \cdots, 0$,

$$W_t(x_t) = \sup_{a_t \in \mathcal{A}_t} r_t(x_t, a_t) + W_{t+1}(x_{t+1}), \tag{7}$$

where $x_t \in \mathcal{X}$ and $x_{t+1} = f_t(x_t, a_t)$.

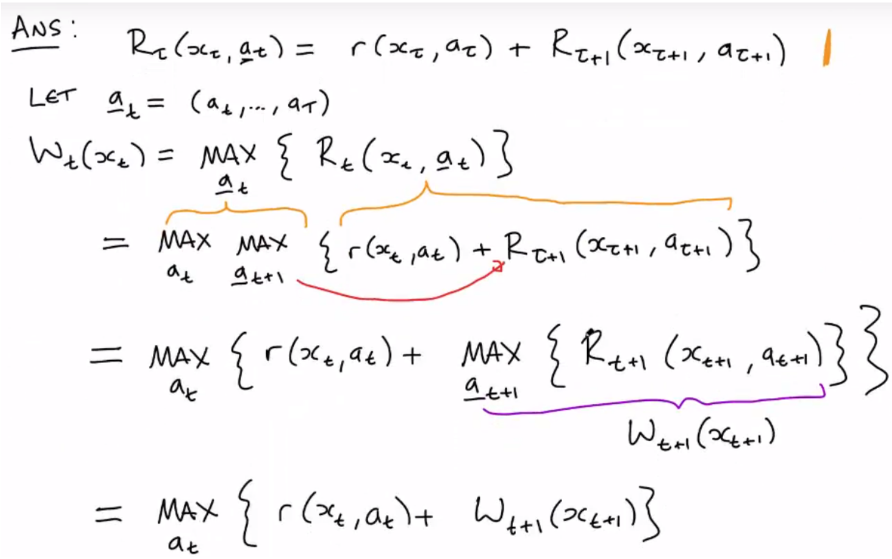

Figure 7: Bellman's equation

Example 7

An investor has a fund: it has x euros at time zero; money can not be withdrawn; it pays $r \times 100\%$ interest per year for T years; the investor consumes proportion a_t of the interest and reinvests the rest. What should the investor do to maximize consumption?

Answer:

Plant equation: $x_{t+1} = x_t + r x_t (1 - a_t)$.
Objective: maximize $\sum_{t=0}^{T-1} r x_t a_t$.
Let us apply the same idea as before and solve backwards in time

- At time $t = T-1$, $r x_{T-1}$, here it must be that $a_{T-1} = 1$ and so $W_{T-1}(x_{T-1}) = r x_{T-1}$.
- At time $t = T-2$: use the Bellman's equation

$$\begin{aligned} W_{T-2}(x_{T-2}) &= \max_{0 \le a_{T-2} \le 1} \{ r x_{T-2} a_{T-2} + W_{T-1}(x_{T-1}) \} \\ &= \max_{0 \le a_{T-2} \le 1} \{ (1+r) + (1-r) a_{T-2} \} \\ &= r x_{T-2} ((1+r) \text{ or } 2) \end{aligned} \tag{8}$$

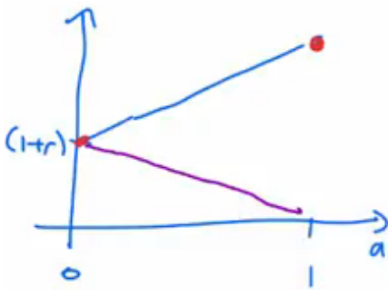

Figure 8: ρ_{T-2}

Let us denote $\rho_{T+2} = (1+r)$ or 2.

- At time $t = T - s$: suppose $W_{T-s+1}(x_{T-s+1}) = rx_{T-s+1}\rho_{T-s+1}$. Following the same steps as above

$$\begin{aligned}
W_{T-s}(x_{T-s}) &= \cdots \\
&= \cdots \\
&= rx_{T-s} \max_{0 \le a_{T-s} \le 1} \{(1+r)\rho_{T-s+1} + (1 - r\rho_{T-s+1}a_{T-s}\} \qquad (9) \\
&= rx_{T-s}\{\rho_{T-s+1}(1+r)\} \text{ or } \{1 + \rho_{T-s+1}\} \\
&= rx_{T-s}\rho_{T-s}
\end{aligned}$$

That is, we define $\rho_{T-s} = \{\rho_{T-s+1}(1+r)\}$ or $\{1 + \rho_{T-s+1}\}$, where the first term means to save and consume nothing, while the second term means to consume everything.

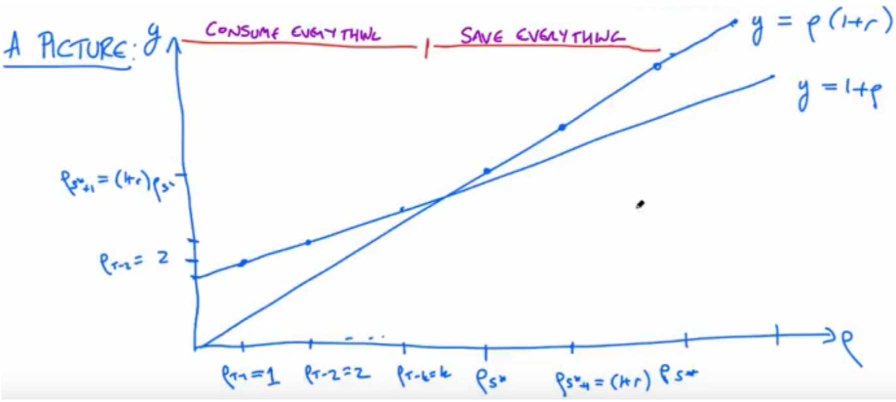

Figure 9: ρ_{T-s}