Aalto University
School of Science

Dynamic Programming Algorithm

Hilkka Hännikäinen
Presentation *8*
9.10.2020

MS-E2191 Graduate Seminar on Operations Research
Fall 2020

Content

i. Formulation of the basic problem
ii. Policy, principle of optimality
iii. Dynamic programming (DP) algorithm
iv. State augmentation
v. Other types of problems

Motivation

How the system should be controlled for optimal performance?

- How much should be ordered for restocking?
- What is the optimal route?
- How to choose a strategy to win a sequence of games?

Basic problem

Discrete-time dynamic system

$$
\begin{aligned}
x_{t+1} &= f_t(x_t, u_t, w_t),\ \ t = 0, 1, \ldots, N \\
x_t &\in X_t \\
u_t &\in U(x_t) \subset C_t \\
w_t &\sim P(\cdot | x_t, u_t), w_t \in D_t
\end{aligned}
$$

$$
\begin{aligned}
t &= \text{index of time} \\
N &= \text{the time horizon} \\
x_t &= \text{the state of the system} \\
u_t &= \text{the control} \\
w_t &= \text{a random parameter} \\
f_t &= \text{a function describing the system dynamics}
\end{aligned}
$$

Additive cost function

$$J_0(x_0) = \mathbb{E}\left[g_N(x_N) + \sum_{t=0}^{N-1} g_t(x_t, u_t, w_t)\right]$$

$$J_t = \text{cost-to-go starting from state } t$$
$$g_t = \text{cost at state } t$$

Inventory control example

- What are the system variables? Do we have some constraints?
- What is the state equation?
- What could be our cost function?

$$x_t = \text{stock at time } t,\ \ u_t = \text{inventory at time } t,\ \ w_t = \text{demand}$$

$$x_{t+1} = x_t + u_t - w_t$$

$$J_0(x_0) = \mathbb{E}\Big[\sum_{t=0}^{N-1} (cu_t + r(x_t + u_t - w_t)^2)\Big]$$

Policy

Solving a set of functions called policy $\pi = \{\mu_0, \dots, \mu_{N-1}\}$ where $\mu_t \colon x_t \to u_t$. If the policy is such that $\mu_t(x_t) \in U(x_t)\ \forall\ x_t \in X_t$, it is called admissible

Optimization problem

$$
\begin{aligned}
J^*(x_0) = \min_{\pi \in \Pi} \ & \mathbb{E}\left[g_N(x_N) + \textstyle\sum_{t=0}^{N-1} g_t(x_t, \mu_t(x_t), w_t)\right] \\
\text{s.t.} \quad & x_{t+1} = f_t(x_t, \mu_t(x_t), w_t) \\
& x_t \in X_t \ \forall t = 0, \ldots, N \\
& \mu_t(x_t) \in U(x_t) \ \forall t = 0, \ldots, N \\
& w_t \in D_t \ \forall t = 0, \ldots, N \\
& w_t \sim P(w_t | x_t, u_t)
\end{aligned}
$$

Principle of optimality

Let $\pi^* = \{\mu_0^*, \mu_i^*, \ldots, \mu_N^*\}$ be an optimal policy. Considering a subproblem at state x_i at time i in which the cost-to-go

$$J_i(x_i) = \mathbb{E}\left[g_N(x_N) + \sum_{j=i}^{N-1} g_j\left(x_j, u_j, w_j\right)\right]$$

is wished to be minimized. Then the truncated policy $\{\mu_i^*, \mu_{i+1}^*, \ldots, \mu_{N-1}^*\}$ is optimal for the subproblem.

DP algorithm

For the initial state x_0, the optimal cost $J^*(x_0)$ for the basic problem can be solved by traversing through the steps backward in time starting from period N-1 to 0:

$$J_N(x_N) = g_N(x_N)$$

$$J_i(x_i) = \min_{u_i \in U(x_i)} \mathbb{E}_{w_i}\left[g_i(x_i, u_i, w_i) + J_{i+1}\big(f_i(x_i, u_i, w_i)\big)\right], i = 0, \dots, N-1$$

If $u_i^* = \mu_i^*(x_i)$ minimizes the right side of equation above, the policy $\pi = \{\mu_0^*, \dots, \mu_{N-1}^*\}$ is optimal

Solving inventory control problem

First as a deterministic problem… And then stochastic!

State augmentation

What if some of the assumptions of the basic problem are violated?

- Time lags in system dynamics or cost function
- The random variables are correlated
- Decision maker can access a forecast of the future randomness

Other types of problems

- Continuous-time
- Different constraints: initial state unknown, initial time unknown, final time unknown, or some combination of these
- Infinite time horizon
- Imperfect state information

Summary

- Formulated a basic problem for time-evolving, controllable system
- Defined the concepts of policy and cost-to-go for the optimization problem
- Got familiar with DP algorithm
- Discussed about the assumptions of the basic problem and other types of dynamic programming problems

References

Dimitri P. Bertsekas. *Dynamic Programming and Optimal Control Volume I*, chapter 1, pages 2-43. Athena Scientific, Belmont, USA, 2000.

Homework

You have taken a step in your career and you're in charge for the inventory. You want to succeed better than the last responsible, so you decide to develop the model a bit further.

The system dynamics are as follows. All demand at period t can be satisfied with the available stock, but any excess demand will be lost. In this case, both the stock and inventory variables are constrained to be positive. Furthermore, assume that the maximum storage capacity is two units. The cost function is the same as in the previous example without terminal cost with ordering cost c equal to 1 and additional storage cost r equal to 2. The demand is random, and follows a distribution

$$P(w = 0) = 0.1,\ P(w = 1) = 0.6,\ P(w = 2) = 0.3$$

for all time instances.

Now your task is to solve the optimal policy for three time periods $N = 3$

A. Formulate the optimization problem (2 pts)

B. Solve the optimal policy, namely the optimal controls for the problem (7 pts)

C. What is the optimal cost for the problem if $x_0 = 1$? (1 pts)

DL 16.9.2020 at 9.15

Submissions: hilkka.hannikainen@aalto.fi

Questions etc.: email or at Telegram @Hilimaaa (during office hours ☺)