Dynamic Programming and Optimal Control

Study Note

Author: Jinyi Liu

Institute: Haslam Business School, UTK

Date: October 29, 2022

Bio: A first-year Ph.D. student in Business Analytics.

Victory won't come to us unless we go to it.

Contents

Chapter 1 The Dynamic Programming Algorithm

1.1 Introduction

1.1.1 General Structure of Finite Horizon Optimal Control Problems

Our finite horizon model has two principal features: (1) a *discrete-time dynamic system,* and (2) a *cost function that is additive over time.* The system has the form

$$x_{k+1} = f_k(x_k, u_k, w_k), \quad k = 0, 1, \ldots, N-1,$$

where

x_k	state variable
u_k	control variable
w_k	random parameter,

and f_k is a function the describes the system.

The cost function is additive. The total cost is

$$g_N(x_N) + \sum_{i=0}^{N-1} g_k(x_k, u_k, w_k).$$

Since w_k is random, we formulate the problem as an optimization of the *expected cost*

$$E\left\{ g_N(x_N) + \sum_{i=0}^{N-1} g_k(x_k, u_k, w_k) \right\}.$$

1.2 The Basic Problem

Basic Problem

We are given a discrete-time dynamic system

$$x_{k+1} = f_k(x_k, u_k, w_k), \quad k = 0, 1, \ldots, N-1,$$

where the state $x_k \in S_k$, the control $u_k \in C_k$ and the random "disturbance" w_k is an element of a space D_k.

The control u_k is constrained to be $u_k \in U_k(x_k) \subset C_k$ for all $x_k \in S_k$ and k.

w_k is characterized by a probability distribution $P_k(\cdot|x_k, u_k)$ that may explicitly on x_k and u_k but not on values of prior disturbances $w_{k-1}, \ldots, w_0$.

We consider the class of polices

$$\pi = \{\mu_0, \ldots, \mu_{N-1}\}$$

, where μ_k maps x_k into controls $u_k = \mu_k(x_k)$ and is such that $\mu_k(x_k) \in U_k(x_k)$ for all $x_k \in S_k$. Such polices will be called *admissible.*

Given x_0 and admissible π, we have

$$x_{k+1} = f_k(x_k, \mu_k(x_k), w_k), \quad k = 0, 1, \ldots, N-1 \tag{1.1}$$

Thus, for given function g_k, we have the expected cost of π starting at x_0:

$$J_\pi(x_0) = E\left\{ g_N(x_N) + \sum_{i=0}^{N-1} g_k(x_k, \mu_k(x_k), w_k) \right\}$$

where the expectation is taken over x_k and w_k. An optimal policy π^* is one such that

$$J_{\pi^*}(x_0) = \min_{\pi \in \Pi} J_\pi(x_0).$$

The Role and Value of Information

Encoding Risk in the Cost Function

1.3 The Dynamic Programming Algorithm

The DP algorithm rests on the *principle of optimality*.

The DP Algorithm

Proposition 1.3.1

For every initial state x_0, the optimal cost $J^*(x_0)$ of the basic problem is equal to $J_0(x_0)$, given by the last step of the following algorithm, which proceeds backward in time from period $N-1$ to period 0 :

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

$$J_k(x_k) = \min_{u_k \in U_k(x_k)} \underset{w_k}{E} \left\{ g_k(x_k, u_k, w_k) + J_{k+1}(f_k(x_k, u_k, w_k)) \right\}$$

$$k = 0, 1, \ldots, N-1,$$

where the expectation is taken with respect to the probability distribution of w_k, which depends on x_k and u_k. Furthermore, if $u_k^* = \mu_k^*(x_k)$ minimizes the right side of Eq. (1.6) for each x_k and k, the policy $\pi^* = \{\mu_0^*, \ldots, \mu_{N-1}^*\}$ is optimal.

1.4 State Augmentation and Other Reformulations

The general guideline in *state augmentation* is to *include in the enlarged state at time k all the information that is known to the controller at time k and can be used with advantage in selecting u_k.*

Time Delays

1.5 Some Mathematical Issues

Well-defined random variables.

1.6 Dynamic Programming and Minimax Control

Consider a triplet (Π, W, J), where π is the set of policies under consideration, W is the set in which the uncertain quantities are known to belong, and $J : \Pi \times W \to [-\infty, +\infty]$ is a given cost function. The objective is to

$$\min_{w \in W} \max J(\pi, w)$$

over all $\pi \in \Pi$.

Lemma 1.6.1

Let $f : W \to X$ be a function, and M be the set of all functions $\mu : X \to U$, where W, X, and U are some sets. Then for any functions $G_0 : W \to (-\infty, \infty]$ and $G_1 : X \times U \to (-\infty, \infty]$ such that

$$\min_{u \in U} G_1(f(w), u) > -\infty, \quad \textit{for all } w \in W,$$

we have

$$\min_{\mu \in M} \max_{w \in W} [G_0(w) + G_1(f(w), \mu(f(w)))] = \max_{w \in W} \left[G_0(w) + \min_{u \in U} G_1(f(w), u) \right].$$

♡

Chapter 2 Deterministic Systems and the Shortest Path Problem

In this chapter we focus on deterministic problems, i.e., w_k can take only one value. In contrast with stochastic problems, *using feedback results in no advantage in terms of cost reduction.*

2.1 Finite-State Systems and Shortest Paths

The DP algorithm takes the form

$$J_N(i) = a_{it}^N, \quad i \in S_N, \tag{2.1}$$

$$J_k(i) = \min_{j \in S_{k+1}} \left[a_{ij}^k + J_{k+1}(j) \right], \quad i \in S_k, \quad k = 0, 1, \ldots, N-1. \tag{2.2}$$

A Forward DP Algorithm for Shortest Path Problems

An optimal path from s to t is also an optimal path from t to s in a "reverse" shortest path problem. It is given by

$$\tilde{J}_N(j) = a_{sj}^0, \quad j \in S_1, \tag{2.3}$$

$$\tilde{J}_k(j) = \min_{i \in S_{N-k}} \left[a_{ij}^{N-k} + _{k+1}(i) \right], \quad j \in S_{N-k+1}, \quad k = 1, 2, \ldots, N-1. \tag{2.4}$$

The optimal cost is

$$\tilde{J}_0(t) = \min_{i \in S_N} \left[a_{ij}^N + \tilde{J}_1(i) \right].$$

The above equations yield the same result

$$J_0(s) = \tilde{J}_0(t).$$

Note that there is no analog of forward DP algorithm for stochastic problems.

Converting a Shortest Path Problem to a Deterministic Finite-Stage Problem

2.2 Some Shortest Path Applications

2.2.1 Critical Path Analysis

2.2.2 Hidden Markov Models and the Vaterbi Algorithm

We are given the probability $r(z; i, j)$ of an observation taking value z when the state transition is from i to j. We assume independent observations; i.e., an observation depends only on its corresponding transition and not on other transitions. Time independent. π initial state's probability.

Given the observation sequence $Z_N = \{z_1, z_2, \ldots, z_N\}$, we adopt $\hat{X}_N = \{\hat{x}_0, \hat{x}_1, \ldots, \hat{x}_N\}$ that maximizes over all $X_N = \{x_1, x_2, \ldots, x_N\}$ the conditional probability $p(X_N | Z_N)$. This is called the

maximum a posteriori probability approach.

Using independence, we have

$$p(X_N, Z_N) = \pi_{x_0} \prod_{k=1}^{N} p_{x_{k-1}x_k} r(z_k; x_{k-1}, x_k) \tag{2.5}$$

Trellis diagram and Viterbi algorithm.

The problem of maximizing $p(X_N, Z_N)$ is equivalent to the problem

$$\text{minimize } -\ln(\pi_{x_0}) - \sum_{k=1}^{N} \ln\left(p_{x_{k-1}x_k} r(z_k; x_{k-1}, x_k)\right)$$

over all possible sequences $\{x_0, x_1, \ldots, x_N\}$.

2.3 Shortest Path Algorithms

To be continued.

Chapter 3 Problesm with Perfect State Information

In this chapter we consider a number of applications of discrete-tiem stochastic optimal control with perfect state infomation.

3.1 Linear Systems and Quadratic Cost

In this section we consider the special case of a linear Systems

$$x_{k+1} = A_k x_k + B_k u_k + w_k, \quad k = 0, 1, \ldots, N-1,$$

and the quadratic cost

$$\underset{w_k, k=0,1,\ldots,N-1}{E} \left\{ x_N' Q_N x_N + \sum_{k=0}^{N-1} (x_k' Q_k x_k + u_k' R_k u_k) \right\}.$$

We assume that Q_k are postive semidefinite symmetric and R_k are positive definite symmetric. w_k has 0 mean and finite second moment.

Applying the DP algorithm, we have

$$\begin{aligned} J_N(x_N) &= x_N' Q_N x_N, \\ J_k(x_k) &= \min_{u_k} E\{x_k' Q_k x_k + u_k' R_k u_k + J_{k+1}(A_k x_k + B_k u_k + w_k)\}. \end{aligned} \tag{3.1}$$

We have the optimal control law for every k:

$$\mu_k^*(x_k) = L_k x_k, \tag{3.2}$$

where

$$L_k = -(B_k' K_{k+1} B_k + R_k)^{-1} B_k' K_{k+1} A_k,$$

and where the symmetric positive semidefinite matrices K_k are given by

$$K_N = Q_N, \tag{3.3}$$

$$K_k = A_k'(K_{k+1} - K_{k+1} B_k (B_k' K_{k+1} B_k + R_k)^{-1} B_k' K_{k+1}) A_k + Q_k \tag{3.4}$$

The optimal cost is then given by

$$J_0(x_0) = x_0' K_0 x_0 + \sum_{k=0}^{N-1} E\{w_k' K_{k+1} w_k\}.$$

The Riccati Equation and Its Asymptotic Behavior

Eq. (3.4) is called the *discrete-time Riccati equation*. If the matrices are constant, then as $k \to \infty$, K satisfies the *algebraic Riccati equation*

$$K = A'(K - KB(B'KB + R)^{-1} B'K)A + Q. \tag{3.5}$$

This property indicates that for a large N, one can approximate the control law Eq. (3.2) by $\{\mu^*, \ldots, \mu^*\}$, where

$$\mu^*(x) = Lx, \tag{3.6}$$

$$L = -(B'KB + R)^{-1}B'KA,$$

and K solves Eq. (3.5). This control law is *stationary*.

Definition 3.1.1

A pair (A, B), where A is an $n \times n$ matrix and B is an $n \times m$ matrix, is said to be controllable if the $n \times nm$ matrix

$$\left[B, AB, A^2B, \ldots, A^{n-1}B\right]$$

has full rank (i.e., has linearly independent rows). A pair (A, C), where A is an $n \times n$ matrix and C an $m \times n$ matrix, is said to be observable if the pair (A', C') is controllable, where A' and C' denote the transposes of A and C, respectively. ♣

We have

$$\begin{aligned} & x_{k+1} = Ax_k + Bu_k \\ \Rightarrow & x_n = A^n x_0 + Bu_{n-1} + ABu_{n-2} + \cdots + A^{n-1}Bu_0 \end{aligned}$$

or equivalently

$$x_n - A^n x_0 = (B, AB, \ldots, A^{n-1}B) \begin{pmatrix} u_{n-1} \\ u_{n-2} \\ \vdots \\ u_0 \end{pmatrix}. \tag{3.7}$$

Since (A, B) is controllable, the right-hand side of Eq. (3.7) can be made equal to any vector in $\mathbb{R}^n$. This explains the name of "controllable pair".

Observability: given measurements $z_0, z_1, \ldots, z_{n-1}$ of the form $z_k = Cx_k$, we can infer the initial state x_0 of the system $x_{k+1} = Ax_k$, since

$$\begin{pmatrix} z_{n-1} \\ \vdots \\ z_1 \\ z_0 \end{pmatrix} = \begin{pmatrix} CA^{n-1} \\ \vdots \\ CA \\ C \end{pmatrix} x_0.$$

It's also equivalent to that in the absence of control, if $Cx_k \to 0$ then $x_k \to 0$.

To simplify notation, we denote K_{N-k} in Eq. (3.4) by P_k.

Proposition 3.1.1

Let A be an $n \times n$ matrix, B be an $n \times m$ matrix, Q be an $n \times n$ positive semidefinite symmetric matrix, and R be an $m \times m$ positive definite symmetric matrix. Consider

$$P_{k+1} = A'(P_k - P_kB(B'P_kB + R)^{-1}B'P_k)A + Q, \quad k = 0, 1, \ldots, \tag{3.8}$$

where P_0 is an arbitrary positive semidefinite symmetric matrix. Assume that (A, B) is controllable. Assume also that $Q = C'C$, where (A, C) is observable. Then

(a) $\exists!$ positive definite symmetric matrix P such that for every positive semidefinite symmetric

matrix P_0 we have

$$\lim_{k\to\infty} P_k = P.$$

Furthermore, P is the unique solution of

$$P = A'(P - PB(B'PB + R)^{-1}B'P)A + Q \tag{3.9}$$

within the class of positive semidefinite symmetric matrices.

(b) The corresponding closed-loop system is stable; i.e., the eigenvalues of the matrix

$$D = A + BL, \tag{3.10}$$

where

$$L = -(B'PB + R)^{-1}B'PA, \tag{3.11}$$

are strictly within the unit circle.

Proof *Initial Matrix $P_0 = 0$.* Consider the optimal control problem of finding $u_0, u_1, \ldots, u_{k-1}$ that minimize

$$\sum_{i=0}^{k-1}(x_i'Qx_i + u_i'Ru_i)$$

subject to

$$x_{i+1} = Ax_i + Bu_i, \quad i = 0, 1, \ldots k-1,$$

where x_0 is given. The optimal value for this problem, by Eq. (3.4), $x_0'P_k(0)x_0$, is given by Eq. (3.8) with $P_0 = 0$ (since $K_N = Q_N$). For any control sequence $(u_0, u_1, \ldots, u_k)$ we have

$$\sum_{i=0}^{k-1}(x_i'Qx_i + u_i'Ru_i) \le \sum_{i=0}^{k}(x_i'Qx_i + u_i'Ru_i)$$

and hence

$$\begin{aligned} x_0'P_k(0)x_0 &= \min_{u_i, i=0,\ldots,k-1} \sum_{i=0}^{k-1}(x_i'Qx_i + u_i'Ru_i) \\ &\le \min_{u_i, i=0,\ldots,k} \sum_{i=0}^{k}(x_i'Qx_i + u_i'Ru_i) \\ &= x_0'P_{k+1}(0)x_0, \end{aligned}$$

where both minimizations are subject to the system equation constraint $x_{i+1} = Ax_i + Bu_i$. Furthermore, for a fixed x_0 and for every k, $x_0'P_k(0)x_0$ is bounded from above by the cost corresponding to a control sequence that forces x_0 to the origin in n steps and applies zero control after that. Such a sequence exists by the controllability assumption. Thus the sequence $\{x_0'P_k(0)x_0\}$ is nondecreasing with respect to k and bounded from above, and therefore it converges for every $x_0 \in \mathbb{R}^n$. Then $P_k(0)$ converges to a P by choosing $x_0 = (1, 0, \ldots, 0), (0, 1, 0, \ldots, 0), \ldots, (1, 1, 0, \ldots, 0)$. So we have

$$\lim_{k\to\infty} P_k(0) = P,$$

where P is generated by Eq. (3.8) with $P_0 = 0$. Then we have Eq. (3.9). By direct calculation we have

$$P = D'PD + Q + L'RL, \tag{3.12}$$

where D and L are given by Eq. (3.10) and Eq. (3.11).

Stability of the Closed-Loop System. Consider the system

$$x_{k+1} = (A + BL)x_k = Dx_k \tag{3.13}$$

for an arbitrary initial state x_0. We will show $x_k \to 0$. By Eq. (3.12), we have

$$x'_{k+1}Px_{k+1} - x'_kPx_k = x'_k(D'PD - P)x_k = -x'_k(Q + L'RL)x_k.$$

hence

$$x'_{k+1}Px_{k+1} = x'_0Px_0 - \sum_{i=0}^{k} x'_i(Q + L'RL)x_i. \tag{3.14}$$

The left-hand side is bounded below by zero, so it follows that

$$\lim_{k\to\infty} x'_k(Q + L'RL)x_k = 0.$$

Since R is positive definite and $Q = C'C$, we have

$$\lim_{k\to\infty} Cx_k = 0, \quad \lim_{k\to\infty} Lx_k = \lim_{k\to\infty} \mu^*(x_k) = 0. \tag{3.15}$$

The preceding relations imply that as the control asymptotically becomes negligible, we have $\lim_{k\to\infty} Cx_k = 0$, and in view of the observability assumption, this implies that $x_k \to 0$. To express this argument more precisely, using Eq. (3.13), we have

$$\begin{pmatrix} C\left(x_{k+n-1} - \sum_{i=1}^{n-1} A^{i-1}BLx_{k+n-i-1}\right) \\ C\left(x_{k+n-2} - \sum_{i=1}^{n-2} A^{i-1}BLx_{k+n-i-2}\right) \\ \vdots \\ C\left(x_{k+1} - BLx_k\right) \\ Cx_k \end{pmatrix} = \begin{pmatrix} CA^{n-1} \\ CA^{n-2} \\ \vdots \\ CA \\ C \end{pmatrix} x_k. \tag{3.16}$$

Since $Lx_k \to 0$ by Eq. (3.15), the left-hand side tends to zero and hence the right-hand side tends to zero also. By the observability assumption, the matrix on the right right of Eq. (3.16) has full rank to that $x_k \to 0$.

Positive Definiteness of P. Assume the contrary, i.e., there exists some $x_0 \neq 0$ such that $x'_0Px_0 = 0$. Since P is positive semidefinite, from Eq. (3.14) we obtain

$$x'_k(Q + L'RL)x_k = 0, \quad k = 0, 1, \ldots$$

Since $x_k \to 0$, we obtain $x'_kQx_k = x'_kC'Cx_k = 0$ and $x'_kL'RLx_k = 0$, or

$$Cx_k = 0, \quad Lx_k = 0, \quad k = 0, 1, \ldots$$

Consider Eq. (3.16) for $k = 0$. By the precreding equalities, we then have

$$0 = \begin{pmatrix} CA^{n-1} \\ \vdots \\ CA \\ C \end{pmatrix} x_0.$$

Then we have $x_0 = 0$ since the matrix has full rank, which contradicts to the hypothesis $x_0 \neq 0$. Thus, P is positive definite.

Arbitrary Initial Matrix P_0. The optimal cost of the problem of minimizing

$$x_k' P_0 x_k + \sum_{i=0}^{k-1} (x_i' Q x_i + u_i' R u_i) \tag{3.17}$$

subject to $x_{i+1}' = A x_i' + B u_i$ is equal to $x_0' P_k(P_0) x_0$. Hence we have for every $x_0 \in \mathbb{R}^n$

$$x_0' P_k(0) x_0 \leq x_0' P_k(P_0) x_0.$$

Consider now the cost Eq. (3.17) corresponding to $\mu(x_k) = u_k = L x_k$ defined by Eq. (3.11). We have

$$x_0' \left(D^{k\prime} P_0 D^k + \sum_{i=0}^{k-1} D^{i\prime} (Q + L' R L) D^i \right) x_0 \geq x_0' P_k(P_0) x_0$$

since $x_0' P_k(P_0) x_0$ is the optimal value of Eq. (3.17). Hence we have

$$x' P_k(0) x \leq x_0' P_k(P_0) x_0 \leq x_0' \left(D^{k\prime} P_0 D^k + \sum_{i=0}^{k-1} D^{i\prime} (Q + L' R L) D^i \right) x_0$$

We have proved that

$$\lim_{k \to \infty} P_k(0) = P,$$

and we also have, using the fact $\lim_{k\to\infty} D^{k\prime} P_0 D^k = 0$, and Eq. (3.12),

$$\begin{aligned} &\lim_{k \to \infty} \left\{ D^{k\prime} P_0 D^k + \sum_{i=0}^{k-1} D^{i\prime} (Q + L' R L) D^i \right\} \\ &= \lim_{k \to \infty} \left\{ \sum_{i=0}^{k-1} D^{i\prime} (Q + L' R L) D^i \right\} \\ &= \lim_{k \to \infty} \left\{ \sum_{i=0}^{k-1} D^{i\prime} (P - D' P D) D^i \right\} \\ &= P. \end{aligned} \tag{3.18}$$

Combining the preceding three equations, we obtain

$$\lim_{k \to \infty} P_k(P_0) = P,$$

for any arbitrary positive semidefinite symmetric initial matrix P_0.

Uniqueness of Solution. If $\tilde{P}$ is another positive semidefinite symmetric solution of Eq. (3.9), we have $P_k(\tilde{P}) = \tilde{P}$ for all $k = 0, 1, \ldots$ From the convergence result just proved, we then have

$$\lim_{k \to \infty} P_k(\tilde{P}) = P,$$

implying that $\tilde{P} = P$.

Random System Matrices

On Certainty Equivalence

3.2 Inventory control

We assume that excess demand at each period is backlogged and is filled when additional inventory becomes available. This is represented by negative inventory in the system equation

$$x_{k+1} = x_k + u_k - w_k, \quad k = 0, 1, \ldots, N-1.$$

We assume that w_k are bounded and independent. Consider a holding/shortage cost of the form

$$r(x) = p \max(0, -x) + h \max(0, x),$$

where p and h are nonnegative scalars. Thus we have

$$\min E \left\{ \sum_{i=0}^{N-1} (cu_k + r(x_k + u_k - w)k) \right\}.$$

We assume that the perchase cost per unit stock c has the property $p > c > 0$.

By applying the DP algorithm, we have

$$J_N(x_N) = 0$$

$$J_k(x_k) = \min_{u_k \geq 0} \left[cu_k + H(x_k + u_k) + E\{J_{k+1}(x_k + u_k - w_k)\} \right], \tag{3.19}$$

where H is defined by

$$H(y) = E\left\{r(y - w_k)\right\} = pE\{\max(0, w_k - y)\} + hE\{\max(0, y - w_k)\}.$$