

Dynamic programming

Contents

1	The principle of optimality	2
2	The recurrence relation of dynamic programming	6
2.1	A discrete approximation of the state equation and performance index	6
2.2	The recurrence relation	8
2.3	Computational procedure. Analytical Results	10
2.4	Exercises	13
2.5	Computational procedure for computer implementation	13
2.5.1	Example	14
2.5.2	Interpolation	17
2.5.3	The algorithm	17

The principle of optimality

We defined an optimal control of the form:

$$\mathbf{u}^* = \mathbf{f}(\mathbf{x}(t), t)$$

as being a closed-loop or feedback optimal control. The functional relationship $\mathbf{f}$ is called the *optimal control law*, or the *optimal policy*. Notice that the optimal control law specifies how to generate the control value at time t from the state value at time t . The presence of t as an argument of $\mathbf{f}$ indicates that the optimal control law may be time-varying, (Kirk, 2004).

In the method of dynamic programming, an optimal policy is found by employing the concept called *the principle of optimality*.

The study of dynamic programming dates from Richard Bellman, who wrote the first book on the subject (1957) and gave it its name. A very large number of problems can be treated this way.

The *optimal path* for a multistage decision process is shown in Figure 1.1a.

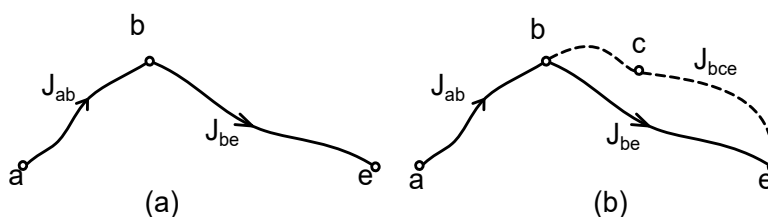


Figure 1.1: (a) Optimal path from a to e. (b) Two possible optimal paths from b to e, (Kirk, 2004)

Suppose that the first decision (made at a) results in segment $a - b$ with cost J_{ab} and that remaining decisions yield segment $b - e$ at a cost J_{be} . The

minimum cost J_{ae}^* from a to e is therefore:

$$J_{ae}^* = J_{ab} + J_{be}$$

If $a - b - e$ is the optimal path from a to e , then $b - e$ is the optimal path from b to e .

Proof by contradiction: Suppose $b - c - e$ in Figure 1.1b is the optimal path from b to e ; then

$$J_{bce} < J_{be}$$

and

$$J_{ab} + J_{bce} < J_{ab} + J_{be} = J_{ae}^*$$

but this last relation can be satisfied only by violating the condition that $a - b - e$ is optimal path from a to e . Thus the assertion is proved.

Bellman has called the above property of an optimal policy the *principle of optimality*:

An optimal policy has the property that whatever the initial state and initial decision are, the remaining decisions must constitute an optimal policy with regard to the state resulting from the first decision.

In other words, an optimal trajectory has the property that at an intermediate point, no matter how it was reached, the rest of trajectory must coincide with an optimal trajectory as computed from this intermediate point as the initial point.

Bellman's principle of optimality serves to "limit the number of potentially optimal control strategies that must be investigated".

Rutherford Aris restates the principle in more colloquial terms:

If you don't do the best with what you have happened to have got, you will never do the best with what you should have had.

The following example illustrates the procedure for making a single optimal decision with the aid of the principle of optimality.

Example 1.1 , (Weber, 2000; Lewis et al., 2012).

Consider the problem traveling salesman problem in which a traveler wishes to minimize the length of a journey from town a to town i (see Figure 1.2). He has to travel first to one of b or c , then to one of d , e or f , then onwards to one of g or h and finally to i . The arcs are marked with distances between towns.

There are 4 "stages" in this problem labeled from $k = 0$ to $k = 4$ (see Figure 1.3). At each stage, $k = 0 \cdots 4$ a decision is required. It can be

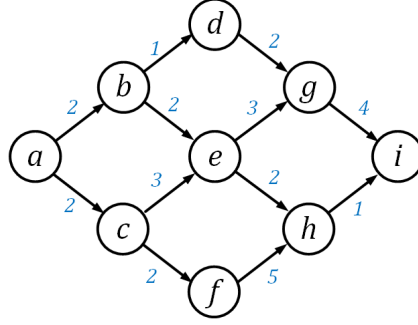


Figure 1.2: The road system

considered as applying a control input $u_k = +1$ or $u_k = -1$ at each stage, where $+1$ means a move up and -1 a move down. The current state is the node (town) where we are making decisions.

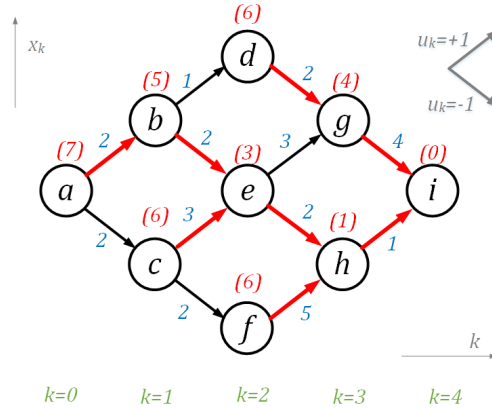


Figure 1.3: The road system and minimum distances to town i

Let $F(x)$ be the minimal distance (or cost) required to reach town i from any town x . We start at stage $k = 4$. Since it is the last stage, no decision is required here, then:

$$F(i) = 0$$

We move then to stage $k = 3$. If the current state is g , the only decision or the only control (optimal control) is $u_3 = -1$ and the cost (optimal cost) is $F(g) = 4$. If the current state is h , the optimal control is $u_3 = +1$ and the optimal cost $F(h) = 1$. These optimal costs (distances) are indicated above each node and the optimal decision (optimal control) is indicated with a red-colored arrow.

$$k = 3, \quad F(g) = \min\{4 + F(i)\} = 4, \text{ and } u_3^* = -1$$

We move now at stage $k = 2$. If the current state is d , the only (optimal) control that can be applied is $u_2^* = -1$ and the only (optimal) cost is $F(d) = 2 + f(g) = 6$. If the current state is e , then we must make a decision: if the control is $u_2 = +1$, the cost is $3 + F(g) = 7$ and if the control is $u_2 = -1$, the cost will be $2 + F(h) = 3$. Hence, at node e , the optimal cost to travel to i is 3, obtained by applying the optimal control $u_2^* = -1$. If the current state is f , the only (optimal) control is $u_2^* = +1$ and the only (optimal) cost is $F(f) = 5 + F(h) = 6$.

$$k = 2, \quad F(d) = \min\{2 + F(g)\} = 2 + 4 = 6, \text{ and } u_2^* = -1$$

$$F(e) = \min\{3 + F(g), 2 + F(h)\} = \min\{3 + 4, 2 + 1\} = 3, \text{ and } u_2^* = -1$$

$$F(f) = \min\{5 + F(h)\} = 5 + 1 = 6, \text{ and } u_2^* = +1$$

With the same strategy, at $k = 1$ we have:

$$k = 1, \quad F(b) = \min\{1 + F(d), 2 + F(e)\} = \min\{1 + 6, 2 + 3\} = 5, \text{ and } u_1^* = -1$$

$$F(c) = \min\{3 + F(e), 2 + F(f)\} = \min\{3 + 3, 2 + 6\} = 6, \text{ and } u_1^* = +1$$

Finally, at the initial stage $k = 0$, the optimal cost to i is computed from:

$$k = 0, \quad F(a) = \min\{2 + F(b), 2 + F(c)\} = \min\{2 + 5, 2 + 6\} = 7, \text{ and } u_0^* = +1$$

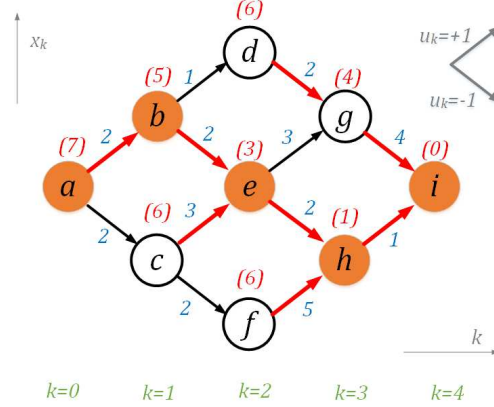


Figure 1.4: The road system

The strategy presented above will determine the optimal strategy to travel from any node x to the final node i , optimal meaning in this case the minimum distance. Starting from any node (or state) the optimal path is marked by red-colored arrows (see Figure 1.4).

The optimal route from node a to node i is then: $a - b - e - h - i$.

The optimal sequence of decisions (controls) to travel from node a to node i is: $u_0^* = +1, u_1^* = -1, u_2^* = -1, u_3^* = +1$.

The recurrence relation of dynamic programming

2.1 A discrete approximation of the state equation and performance index

An n -th order time invariant system is described by the state equation:

$$\dot{\mathbf{x}}(t) = \mathbf{a}(\mathbf{x}(t), \mathbf{u}(t)) \quad (2.1)$$

It is desired to determine the control law which minimizes the performance measure:

$$J = h(\mathbf{x}(t_f)) + \int_0^{t_f} g(\mathbf{x}(t), \mathbf{u}(t)) dt \quad (2.2)$$

where t_f is assumed fixed.

We first approximate the continuous system by a discrete system. This is accomplished by considering N equally spaced time increments in the interval $0 \leq t \leq t_f$ (Figure 2.1).

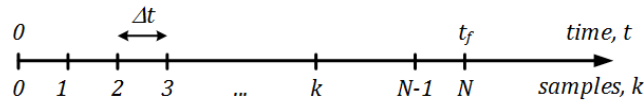


Figure 2.1: N equally spaced time increments

An approximation of (2.1) gives:

$$\frac{\mathbf{x}(t + \Delta t) - \mathbf{x}(t)}{\Delta t} \approx \dot{\mathbf{x}}(t) = \mathbf{a}(\mathbf{x}(t), \mathbf{u}(t))$$

or

$$\mathbf{x}(t + \Delta t) = \mathbf{x}(t) + \Delta t \mathbf{a}(\mathbf{x}(t), \mathbf{u}(t))$$

Using the shorthand notation $\mathbf{x}(k\Delta t) = \mathbf{x}_k$:

$$\mathbf{x}_{k+1} = \mathbf{x}_k + \Delta t \mathbf{a}(\mathbf{x}_k, \mathbf{u}_k)$$

which we will denote by:

$$\mathbf{x}_{k+1} = \mathbf{f}(\mathbf{x}_k, \mathbf{u}_k) \quad (2.3)$$

An approximation of the performance measure (2.2) is obtained as follows:

- The scalar function $h(\mathbf{x}(t_f))$ depends only on the state computed at the final time and will become $h(\mathbf{x}_N)$, where N is the final time sample.
- The integral in (2.2) can be approximated by the rectangle method, as shown in Figure and becomes:

$$\int_0^{t_f} g(\mathbf{x}(t), \mathbf{u}(t)) dt \approx \Delta t \sum_{k=0}^{N-1} g(\mathbf{x}_k, \mathbf{u}_k)$$

where $\mathbf{x}_k$ and $\mathbf{u}_k$ are the values of the states and control inputs, respectively, at each time sample k .

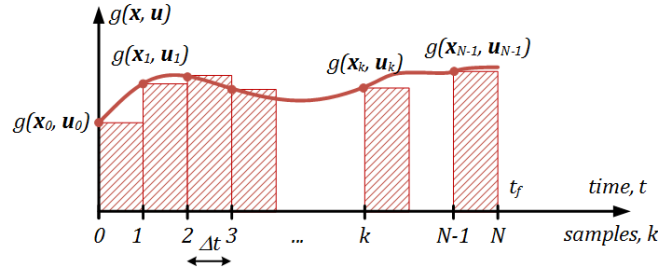


Figure 2.2: Approximation of the integral

The discrete performance measure is then:

$$J = h(\mathbf{x}_N) + \Delta t \sum_{k=0}^{N-1} g(\mathbf{x}_k, \mathbf{u}_k) \quad (2.4)$$

or

$$J = h(\mathbf{x}_N) + \sum_{k=0}^{N-1} L(\mathbf{x}_k, \mathbf{u}_k) \quad (2.5)$$

2.2 The recurrence relation

By making the problem discrete as we have done, it is now required that the optimal control law:

$$\mathbf{u}^*(\mathbf{x}_0, 0), \mathbf{u}^*(\mathbf{x}_1, 1), \dots, \mathbf{u}^*(\mathbf{x}_{N-1}, N-1)$$

be determined for the system given by (2.3) which has the performance measure (2.5).

Begin by defining

$$J_{NN}(\mathbf{x}_N) = h(\mathbf{x}_N)$$

J_{NN} is the cost of reaching the final state value $\mathbf{x}_N$.

Next define:

$$J_{N-1,N}(\mathbf{x}_{N-1}, \mathbf{u}_{N-1}) = L(\mathbf{x}_{N-1}, \mathbf{u}_{N-1}) + h(\mathbf{x}_N) = L(\mathbf{x}_{N-1}, \mathbf{u}_{N-1}) + J_{NN}(\mathbf{x}_N)$$

which is the cost of operation during the interval $(N-1)\Delta t \leq t \leq N\Delta t$. Observe that $J_{N-1,N}$ is also the cost of a *one-stage process* with "initial" state $\mathbf{x}_{N-1}$. The value of $J_{N-1,N}$ is dependent only on $\mathbf{x}_{N-1}$ and $\mathbf{u}_{N-1}$ since $\mathbf{x}_N$ is related to $\mathbf{x}_{N-1}, \mathbf{u}_{N-1}$ through the state equation (2.3).

The optimal cost is then:

$$J_{N-1,N}^*(\mathbf{x}_{N-1}) = \min_{\mathbf{u}_{N-1}} (L(\mathbf{x}_{N-1}, \mathbf{u}_{N-1}) + J_{NN}(\mathbf{f}(\mathbf{x}_{N-1}, \mathbf{u}_{N-1})))$$

We know that the optimal choice of $\mathbf{u}_{N-1}$ will depend on $\mathbf{x}_{N-1}$, so we denote the minimizing control by $\mathbf{u}^*(\mathbf{x}_{N-1})$.

Example 2.1 In Figure 2.3 x_{N-1} is transferred to x_{N1} with the control value u_1 and the cost of operation is J_1 . It is transferred to x_{N2} or x_{N3} with the control value u_2 respectively u_3 and the cost of operation will be J_2 or J_3 . $J_{N-1,N}^*$ is the minimum cost for transferring from x_{N-1} to x_N (which is one of the value x_{N1}, x_{N2}, x_{N3}) given by u_{N-1}^* (one of u_1, u_2, u_3).

The cost of operation over the last two intervals is given by:

$$\begin{aligned} J_{N-2,N}(\mathbf{x}_{N-2}, \mathbf{u}_{N-2}, \mathbf{u}_{N-1}) &= L(\mathbf{x}_{N-2}, \mathbf{u}_{N-2}) + L(\mathbf{x}_{N-1}, \mathbf{u}_{N-1}) + h(\mathbf{x}_N) = \\ &= L(\mathbf{x}_{N-2}, \mathbf{u}_{N-2}) + J_{N-1,N}(\mathbf{x}_{N-1}, \mathbf{u}_{N-1}) \end{aligned}$$

where again we have used the dependence of $\mathbf{x}_N$ on $\mathbf{x}_{N-1}$ and $\mathbf{u}_{N-1}$.

$J_{N-2,N}$ is the cost of a two stage process with the initial state $\mathbf{x}_{N-2}$. The optimal policy during the last two time intervals is found from:

$$J_{N-2,N}^*(\mathbf{x}_{N-2}) = \min_{\mathbf{u}_{N-1}, \mathbf{u}_{N-2}} (L(\mathbf{x}_{N-2}, \mathbf{u}_{N-2}) + J_{N-1,N}(\mathbf{x}_{N-1}, \mathbf{u}_{N-1}))$$

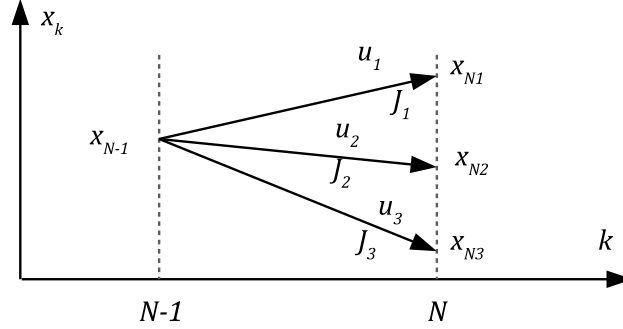


Figure 2.3: Example

The principle of optimality states that for this two stage process, whatever the initial state $\mathbf{x}_{N-2}$ and initial decision $\mathbf{u}_{N-2}$, the remaining decision must be optimal with respect to the value of $\mathbf{x}_{N-1}$ that results from application of $\mathbf{u}_{N-1}$. Therefore:

$$J_{N-2,N}^*(\mathbf{x}_{N-2}) = \min_{\mathbf{u}_{N-2}} \left(L(\mathbf{x}_{N-2}, \mathbf{u}_{N-2}) + J_{N-1,N}^*(\mathbf{x}_{N-1}) \right)$$

For a stage k , the performance measure over the last k stages is defined as:

$$J_{k,N}(\mathbf{x}_k, \mathbf{u}_k) = \sum_{i=k}^{N-1} L(\mathbf{x}_i, \mathbf{u}_i)$$

Suppose the optimal cost-to-go from any $\mathbf{x}_{k+1}$, $J_{k+1,N}^*(\mathbf{x}_{k+1})$ and optimal control sequence $(\mathbf{u}_{k+1}, \dots, \mathbf{u}_{N-1})$ is already computed. Then the principle of optimality states that we only need to minimize the cost $L(\mathbf{x}_k, \mathbf{u}_k) + J_{k+1,N}^*(\mathbf{x}_{k+1})$ i.e. $\mathbf{u}_k$ can be found from :

$$J_{k,N}^*(\mathbf{x}_k) = \min_{\mathbf{u}_k} \left(L(\mathbf{x}_k, \mathbf{u}_k) + J_{k+1,N}^*(\mathbf{x}_{k+1}) \right) \quad (2.6)$$

or

$$J_{k,N}^*(\mathbf{x}_k) = \min_{\mathbf{u}_k} \left(L(\mathbf{x}_k, \mathbf{u}_k) + J_{k+1,N}^*(\mathbf{f}(\mathbf{x}_k, \mathbf{u}_k)) \right) \quad (2.7)$$

To begin the process we simply start with a zero stage process and generate $J_{NN}^* = J_{NN}$ (the * notation is just a notational convenience here; no choice of a control is implied). Next the optimal cost can be found for a one-stage process by using J_{NN}^* .

2.3 Computational procedure. Analytical Results

Problem Given a discrete system described by the state equation

$$\mathbf{x}_{k+1} = \mathbf{f}(\mathbf{x}_k, \mathbf{u}_k), k = 0, 1, \dots, N-1$$

find an optimal policy $\mathbf{u}^*(\mathbf{x}_k, k)$, $k = \overline{0, N-1}$ that minimizes the performance measure:

$$J = h(\mathbf{x}_N) + \sum_{k=0}^{N-1} L(\mathbf{x}_k, \mathbf{u}_k)$$

Suppose the optimal cost-to-go from any $\mathbf{x}_{k+1}, J_{k+1,N}^*(\mathbf{x}_{k+1})$ and optimal control sequence $(\mathbf{u}_{k+1}^* \dots \mathbf{u}_{N-1}^*)$ is already computed.

Then the principle of optimality states that we only need to minimize the cost $L(\mathbf{x}_k, \mathbf{u}_k) + J_{k+1,N}^*(\mathbf{x}_{k+1})$ i.e. $\mathbf{u}_k^*$ can be found from:

$$J_{k,N}^* = \min_{\mathbf{u}_k} \{L(\mathbf{x}_k, \mathbf{u}_k) + J_{k+1,N}^*(\mathbf{x}_{k+1})\} = \min_{\mathbf{u}_k} \{L(\mathbf{x}_k, \mathbf{u}_k) + J_{k+1,N}^*(f(\mathbf{x}_k, \mathbf{u}_k))\}$$

This can be used to solve recursively for the optimal control starting from $J_{N,N}^*(\mathbf{x}_N) = h(\mathbf{x}_N)$

The following algorithm may summarize these results:

Step 0 $J_{N,N}^*(\mathbf{x}_N) = h(\mathbf{x}_N)$

Step 1 $J_{N-1,N}^*(\mathbf{x}_{N-1}) = \min_{\mathbf{u}_{N-1}} \overbrace{\{L(\mathbf{x}_{N-1}, \mathbf{u}_{N-1}) + J_{N,N}^*(f(\mathbf{x}_{N-1}, \mathbf{u}_{N-1}))\}}^{J_{N-1,N}(\mathbf{x}_{N-1}, \mathbf{u}_{N-1})}$

We will call the minimizing control signal $\mathbf{u}_{N-1}^*$. It is a function of $\mathbf{x}_{N-1}$ and can be determined from:

$$\frac{\partial J_{N-1,N}(\mathbf{x}_{N-1}, \mathbf{u}_{N-1})}{\partial \mathbf{u}_{N-1}} = 0$$

Step i $J_{N-i,N}^*(\mathbf{x}_{N-i}) = \min_{\mathbf{u}_{N-i}} \{L(\mathbf{x}_{N-i}, \mathbf{u}_{N-i}) + J_{N-i+1,N}^*(f(\mathbf{x}_{N-i}, \mathbf{u}_{N-i}))\}$

The minimizing control signal will be $\mathbf{u}_{N-i}^*$ and depends on $\mathbf{x}_{N-i}$.

Step N $J_{0,N}^*(\mathbf{x}_0) = \min_{\mathbf{u}_0} \{L(\mathbf{x}_0, \mathbf{u}_0) + J_{1,N}^*(f(\mathbf{x}_0, \mathbf{u}_0))\}$

Example 2.2 Consider a scalar discrete-time dynamical system modeled by:

$$x_{k+1} = 2x_k - 3u_k, \quad x_0 = 4, \quad k = 0, 1, 2 \dots N-1$$

Find the optimal control policy that minimizes the cost function:

$$J = (x_N - 10)^2 + \frac{1}{2} \sum_{k=0}^{N-1} (x_k^2 + u_k^2)$$

Let $N=2$. The cost function is:

$$J = (x_2 - 10)^2 + \frac{1}{2} \sum_{k=0}^1 (x_k^2 + u_k^2)$$

We use the principle of optimality to find an optimal policy u_0^* , u_1^* . There are no constraints on u_k .

Begin with the backward pass. We have:

$$J_{22}^* = (x_2 - 10)^2$$

Then:

$$\begin{aligned} J_{12}^*(x_1) &= \min_{u_1} J_{12}(x_1, u_1) = \min_{u_1} \left\{ \frac{1}{2}(x_1^2 + u_1^2) + J_{22}^*(x_2) \right\} = \\ &= \min_{u_1} \left\{ \frac{1}{2}(x_1^2 + u_1^2) + (x_2 - 10)^2 \right\} \\ &= \min_{u_1} \left\{ \frac{1}{2}(x_1^2 + u_1^2) + (2x_1 - 3u_1 - 10)^2 \right\} \end{aligned}$$

u_1^* can be found from

$$\frac{\partial J_{12}(x_1, u_1)}{\partial u_1} = 0$$

$$\frac{\partial J_{12}(x_1, u_1)}{\partial u_1} = u_1 - 6(2x_1 - 3u_1 - 10) = 0 \Rightarrow$$

$$u_1^* = \frac{12x_1 - 60}{19} \tag{2.8}$$

and

$$\begin{aligned} J_{12}^*(x_1) &= \frac{1}{2} \left[x_1^2 + \left(\frac{12x_1 - 60}{19} \right)^2 \right] + \left[2x_1 - 3 \left(\frac{12x_1 - 60}{19} \right) - 10 \right]^2 \\ &= \frac{27}{38} x_1^2 - \frac{40}{19} x_1 + \frac{100}{19} \end{aligned}$$

$$\begin{aligned}
J_{02}^*(x_0) &= \min_{u_0} J_{02}(x_0, u_0) = \min_{u_0} \left\{ \frac{1}{2}(x_0^2 + u_0^2) + J_{12}^*(x_1) \right\} = \\
&= \min_{u_0} \left\{ \frac{1}{2}(x_0^2 + u_0^2) + J_{12}^*(2x_0 - 3u_0) \right\}
\end{aligned}$$

$$\begin{aligned}
\frac{\partial J_{02}(x_0, u_0)}{\partial u_0} &= \frac{\partial}{\partial u_0} \left\{ \frac{1}{2}(x_0^2 + u_0^2) + \frac{27}{38}(2x_0 - 3u_0)^2 - \frac{40}{19}(2x_0 - 3u_0) + \frac{100}{19} \right\} \\
&= 2u_0 - \frac{6 \cdot 27}{38}u_0(2x_0 - 3u_0) + \frac{120}{19} = 0
\end{aligned}$$

$$u_0^* = \frac{81}{131}x_0 - \frac{60}{131} = 0.618x_0 - 0.458 \quad (2.9)$$

$J_{02}^*(x_0) = J_{02}^*(4) = 13.97$ the minimum value of the performance measure for transferring the system from $x_0 = 4$ to x_2

$$u_0^* = 0.618 \cdot 4 - 0.458 = 2.015$$

$$u_0^* = 2.015 \Rightarrow x_1^* = 2x_0 - 3u_0^* = 1.9541$$

$$u_1^* = \frac{12x_1^* - 60}{19} \Rightarrow u_1^* = -1.923 \Rightarrow x_2^* = 2x_1 - 3u_1^* = 9.679$$

The optimal control policy is then:

$$u_0^* = 2.015 \quad u_1^* = -1.923$$

and the optimal trajectory

$$x_0^* = 4, \quad x_1^* = 1.9541, \quad x_2^* = 9.679$$

$$J^* = J_{02}^* = 13.97$$

Remark. The input sequence may be applied in two ways:

- With a known initial state calculate in advance the entire control sequence

$$u_0^*, \quad u_1^*, \quad \dots \quad u_{N-1}^*$$

and apply it open loop to bring the plant from the initial state to zero.

- Use the variable dependencies of the control value on the states at every discrete time, as calculated for example in (2.8) and (2.9).

This is closed loop control with a time varying control (depending on sample number k)

This closed loop control is preferable because open loop control is vulnerable to disturbances and uncertainties in model parameters.

Note also that this computational procedure is not trivial to be applied for a large number of samples, it may result in very complicated calculations, thus it is applicable for a limited number of situations.

2.4 Exercises

1. Consider a first-order system:

$$x_{k+1} = x_k + u_k; \quad k = 0, 1; \quad x_0 = 1$$

Find the optimal control policy u_0^*, u_1^* that minimizes the performance measure:

$$J = x_2^2 + \sum_{k=0}^1 (x_k^2 + 2u_k^2)$$

Determine the optimal state trajectory $x_0^* = 1, x_1^*, x_2^*$.

2. Consider a first-order system:

$$x_{k+1} = x_k^2 + u_k; \quad k = 0, 1, \dots, N-1; \quad x_0 = 1, \quad N = 2$$

Find the optimal control policy $u_0^*, \dots, u_{N-1}^*$ that minimizes the performance measure:

$$J = 4x_N + \sum_{k=0}^{N-1} u_k^2$$

Determine the optimal state trajectory $x_0^* = 1, x_1^*, \dots, x_N^*$.

2.5 Computational procedure for computer implementation

The computational procedure that is suitable for computer implementation will be introduced first by an example.

2.5.1 Example

Consider a scalar plant with the state equation, (Wen, 2002):

$$x_{k+1} = x_k + u_k, \quad k = 0, 1, \dots, N-1 \quad (2.10)$$

Find the optimal control policy $u_0^*, u_1^*, \dots, u_{N-1}^*$ that minimizes the performance measure:

$$J = x_N^2 + \frac{1}{2} \sum_{k=0}^{N-1} u_k^2 \quad (2.11)$$

Let $N = 2$ and consider that u_k is limited to the discrete values: $(-1, -1/2, 0, 1/2, 1)$ and x_k is limited to $(0, 1/2, 1, 3/2)$. The computational grid (i.e. all the admissible values for all time samples) for u and x is presented in Figure 2.4.

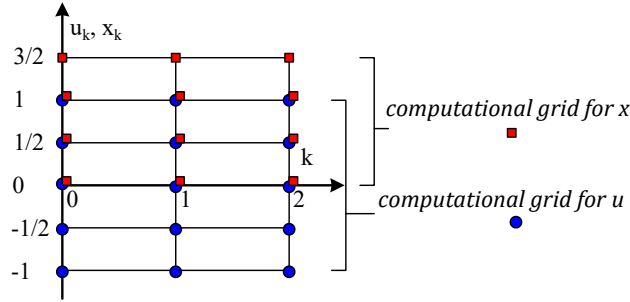


Figure 2.4: Computational grid for u and x

Solution

For $N = 2$ the performance measure becomes:

$$J = x_2^2 + \frac{1}{2} \sum_{k=0}^1 u_k^2 \quad (2.12)$$

$$\underline{k=2}, \quad J_{2,2}^*(x_2) = x_2^2$$

x_2	$J_{2,2}^*(x_2) = x_2^2$
0	0
1/2	1/4
1	1
3/2	9/4

$$\underline{k=1}, \quad J_{1,2}^*(x_1) = \min_{u_1} (J_{2,2}^*(x_2) + \frac{1}{2} u_1^2)$$

x_1	u_1	$x_2 = x_1 + u_1$	$J_{2,2}^*(x_2) + \frac{1}{2}u_1^2$	optimum
0	-1	-1	not admissible	$u_1^* = 0$ $J_{1,2}^*(0) = 0$
	-1/2	-1/2	not admissible	
	0	0	$0 + 1/2 \cdot 0^2 = 0$	
	1/2	1/2	$1/4 + 1/2 \cdot 1/4 = 3/8$	
	1	1	$1 + 1/2 \cdot 1^2 = 3/2$	
1/2	-1	-1/2	not admissible	$u_1^* = -1/2$ $J_{1,2}^*(1/2) = 1/8$
	-1/2	0	$0 + 1/2 \cdot 1/4 = 1/8$	
	0	1/2	$1/4 + 1/2 \cdot 0^2 = 1/4$	
	1/2	1	$1 + 1/2 \cdot 1/4 = 9/8$	
	1	3/2	$9/4 + 1/2 \cdot 1^2 = 11/4$	
1	-1	0	$0 + 1/2 \cdot 1 = 1/2$	$u_1^* = -1/2$ $J_{1,2}^*(1) = 3/8$
	-1/2	1/2	$1/4 + 1/2 \cdot 1/4 = 3/8$	
	0	1	$1 + 1/2 \cdot 0^2 = 1$	
	1/2	3/2	$9/4 + 1/2 \cdot 1/4 = 19/8$	
	1	2	not admissible	
3/2	-1	1/2	$1/4 + 1/2 \cdot 1 = 3/4$	$u_1^* = -1$ $J_{1,2}^*(3/2) = 3/4$
	-1/2	1	$1 + 1/2 \cdot 1/4 = 9/8$	
	0	3/2	$9/4 + 1/2 \cdot 0^2 = 9/4$	
	1/2	2	not admissible	
	1	5/2	not admissible	

$$\underline{k=0}, \quad J_{0,2}^*(x_0) = \min_{u_0} (J_{1,2}^*(x_1) + \frac{1}{2}u_0^2)$$

x_0	u_0	$x_1 = x_0 + u_0$	$J_{1,2}^*(x_1) + \frac{1}{2}u_0^2$	optimum
0	-1	-1	not admissible	$u_0^* = 0$ $J_{0,2}^*(0) = 0$
	-1/2	-1/2	not admissible	
	0	0	$0 + 1/2 \cdot 0^2 = 0$	
	1/2	1/2	$1/8 + 1/2 \cdot 1/4 = 1/4$	
	1	1	$3/8 + 1/2 \cdot 1^2 = 7/8$	
1/2	-1	-1/2	not admissible	$u_0^* = -1/2$ or $u_0^* = 0$ $J_{0,2}^*(1/2) = 1/8$
	-1/2	0	$0 + 1/2 \cdot 1/4 = 1/8$	
	0	1/2	$1/8 + 1/2 \cdot 0^2 = 1/8$	
	1/2	1	$3/8 + 1/2 \cdot 1/4 = 1/2$	
	1	3/2	$3/4 + 1/2 \cdot 1^2 = 5/4$	
1	-1	0	$0 + 1/2 \cdot 1 = 1/2$	$u_0^* = -1/2$ $J_{0,2}^*(1) = 1/4$
	-1/2	1/2	$1/8 + 1/2 \cdot 1/4 = 1/4$	
	0	1	$3/8 + 1/2 \cdot 0^2 = 3/8$	
	1/2	3/2	$3/4 + 1/2 \cdot 1/4 = 7/8$	
	1	2	not admissible	
3/2	-1	1/2	$1/8 + 1/2 \cdot 1 = 5/8$	$u_0^* = -1/2$ $J_{0,2}^*(3/2) = 1/2$
	-1/2	1	$3/8 + 1/2 \cdot 1/4 = 1/2$	
	0	3/2	$3/4 + 1/2 \cdot 0^2 = 3/4$	
	1/2	2	not admissible	
	1	5/2	not admissible	

If the initial state is $x_0 = 1$, the last table indicates the value of the optimal cost from $x_0 = 1$ to the final state x_2 :

$$J_{0,1}^*(1) = \frac{1}{4}$$

and the optimal control value for the last stage ($N = 0$): $u_0^* = -1/2$ transfers the system from state $x_0 = 1$ to the state $x_1^* = x_0 + u_0^* = 1/2$.

The table calculated for $N = 1$ gives the optimal control u_1^* which transfers the system from $x_1 = 1/2$ to $x_2^* = x_1^* + u_1^* = 1 + (-1/2) = 0$.

Thus, the optimal control sequence is:

$$u^* : -\frac{1}{2}, -\frac{1}{2}$$

and the optimal state trajectory starting from 1:

$$x^* : 1, \frac{1}{2}, 0$$

The minimum cost is:

$$J_{0,2}^* = \frac{1}{4}$$

2.5.2 Interpolation

In the preceding example, all of the trial control values drive the state of the system either to a computational "grid" point or to a value outside the allowable range.

If the numerical values were not carefully selected, this happy situation would not have been obtained and interpolation would have been required.

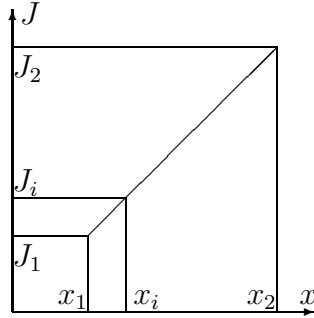


Figure 2.5: Interpolation

Suppose we have two values of the state x_1, x_2 and the values of the performance measure from that state to the final one J_1, J_2 , and suppose x_1 and x_2 are on the computational grid. In case the value of the next state calculated from the plant equation is somewhere between x_1 and x_2 and we do not have a J calculated for this case we shall approximate it by linear approximation. Write the equation of the line between the points (x_1, J_1) and (x_2, J_2) : $J = mx + n$, where

$$m = \frac{J_1 - J_2}{x_1 - x_2}, \quad n = \frac{x_1 J_2 - x_2 J_1}{x_1 - x_2}$$

The calculate J_i for state value x_i (which is not on the computational grid) from

$$J_i = mx_i + n$$

2.5.3 The algorithm

A general algorithm for discrete dynamic programming, for a first order system, may be formulated as in Algorithm 1. The procedure may be extrapolated for higher-order systems with a similar procedure (for details see (Kirk, 2004)).

Algorithm 1 Discrete dynamic programming algorithm

1. Define the discrete system:

$$x_{k+1} = f(x_k, u_k)$$

and the performance measure:

$$J = h(x_N) + \sum_{k=0}^{N-1} L(x_k, u_k)$$

2. Determine a finite number of admissible control values

$$u_{min} \leq u_j \leq u_{max}, \quad j = \overline{1, nr_u}$$

3. Determine a finite number of admissible state values

$$x_{min} \leq x_i \leq x_{max}, \quad i = \overline{1, nr_x}$$

3. Compute $J_N^* = h(x_N)$ for all possible values of the state.

4. Iterate

for every $k = N - 1, N - 2, \dots, 0$ **do**

for every state value $x_{min} \leq x_i \leq x_{max}$ **do**

for every control value $u_{min} \leq u_j \leq u_{max}$ **do**

 a. Calculate the next state $x_{k+1} = f(x_k, u_k)$

 b. Calculate the cost of getting from the next value of the state to the final state $J_{k+1,N}^*$, using the values of J^* obtained at the previous iteration (or, for the beginning, at Step 3)

 c. Calculate the cost of getting from the current state to the final state with the current control value: $J_{k,N} = J_{k+1,N}^*(x_{k+1}) + L(x_k, u_k)$

end for

end for

5. For the current state value calculate the minimum cost of getting to the final state $J_{k,N}^*$ and the control value for which the cost is minimum

end for

6. Given the initial state $x(0) = x_0$, and the optimal control values calculated at each step k , for each value of the states x_j , determine the optimal control sequence: $u_0^*, u_1^*, \dots, u_{N-1}^*$ and reconstruct the state trajectory $X_0^*, x_1^*, \dots, x_N^*$.
-

Bibliography

- Kirk, D. E. (2004). *Optimal Control Theory. An Introduction*. Dover Publications, Inc.
- Lewis, F. L., Vrabie, D., and Syrmos, V. L. (2012). *Optimal Control*. John Wiley & Sons, Inc, 3rd edition.
- Weber, R. (2000). Optimization and control. online at www.statslab.cam.ac.uk.
- Wen, J. T. (2002). Optimal control. online at <http://cats-fs.rpi.edu/~wenj/ECSE644S12/info.htm>.