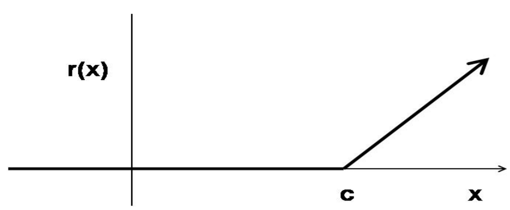
Dynamic Programming:

An Application to Optimal Stopping and American Options

Problem: You have an option on a financial asset (like the price of a stock) that you can exercise any time between now (time 0) and the expiration date of the option (time T). If you exercise the option at time i, you collect $r(X_i)$ dollars, where $r(x) =$"reward function" associated with price x and X_i is the (random) price of the asset at time i. (Assume throughout that $(X_n : n \geq 0)$ is a DTMC with transition matrix $P = (P(x, y) : x, y \in S)$.) A typical form for such a reward function would be

$$r(x) = [x - c]^+,$$

so that r has the following shape:



Note that the decision to exercise at time i can only be based on the prices observed to time i (i.e. no clairvoyance!), namely $X_0, X_1, \ldots, X_i$.

Goal: Compute the optimal exercise strategy for the option, where optimal means "maximize the expected return".

Method: Put

$V(i, x)$ = best possible expected return on the option, given that it was not exercised prior to i, and the price at time i was x.

$V(i, x)$ is called the "value function" associated with being in state x at time i. Clearly, $V(T, x) = r(x)$, since we must exercise the option at time T if it was not exercised earlier. Note that

$$\begin{aligned} V(T-1, x) &= \text{maximum of expected return associated with either exercising at time } T-1 \\ &\quad \text{(and collecting } r(x) \text{ dollars) or continuing to } T \\ &\quad \text{(and getting the best possible return at that time)} \\ &= \max(r(x), \sum_y P(x, y) V(T, y)). \end{aligned}$$

So, $V(T-1,x)$ can be computed from the $V(T,y)$'s. Similarly,

$$V(T-2,x) = \max\,(r(x),\ \sum_y P(x,y)V(T-1,y))$$

so $V(T-2,x)$ can be computed from the $V(T-1,y)$'s. In general,

$$V(i,x) = \max\,(r(x),\ \sum_y P(x,y)V(i+1,y)),$$

so $V(i,x)$ can be computed from the $V(i+1,y)$'s. So, we can compute the $V(i,x)$'s by "backwards recursion" starting at T and recursing back to time 0. Once one has computed the value function $V(i,x)$ $(0 \le i \le T, t \in S)$, one can " read off" the optimal exercise strategy. For example, if we end up at time i at price X_i (having not yet exercised the option), we look at

$$r(X_i)$$

and

$$\sum_y P(X_i,y)V(i+1,y).$$

If $r(X_i)$ is the largest, exercise the option at time i and collect $r(X_i)$ dollars. Otherwise, do not exercise the option. The price then randomly changes to X_{i+1}, and we repeat this process to determine whether to exercise at time $i+1$, etc.

Example: T = 3

$$P = \begin{array}{c} \\ 1 \\ 2 \\ 3 \end{array} \begin{array}{c} \begin{array}{ccc} 1 & 2 & 3 \end{array} \\ \begin{pmatrix} 1/2 & 1/2 & 0 \\ 1/3 & 0 & 2/3 \\ 1/3 & 1/3 & 1/3 \end{pmatrix} \end{array}$$

$$r(1) = 0,\ r(2) = 1,\ r(3) = 2.$$

Start at $T = 3$

$$V(3,1) = 0,\ V(3,2) = 1,\ V(3,3) = 2.$$

Compute $V(2,x)$:

$$\begin{aligned} V(2,1) &= \max\,(0,\ 1/2\cdot 0 + 1/2\cdot 1) \\ &= 1/2 \\ V(2,2) &= \max\,(1,\ 1/3\cdot 0 + 2/3\cdot 2) \\ &= 4/3 \\ V(2,3) &= \max\,(2,\ 1/3\cdot 0 + 1/3\cdot 1 + 1/3\cdot 2) \\ &= 2. \end{aligned}$$

Compute $V(1,x)$:

$$\begin{aligned}
V(1,1) &= \max\left(0,\ 1/2 \cdot 1/2 + 1/2 \cdot 4/3\right) \\
&= 11/12 \\
V(1,2) &= \max\left(1,\ 1/3 \cdot 1/2 + 2/3 \cdot 2\right) \\
&= 3/2 \\
V(1,3) &= \max\left(2,\ 1/3 \cdot 1/2 + 1/3 \cdot 4/3 + 1/3 \cdot 2\right) \\
&= 2.
\end{aligned}$$

Compute $V(0,x)$:

$$\begin{aligned}
V(0,1) &= \max\left(0,\ 1/2 \cdot 11/12 + 1/2 \cdot 3/2\right) \\
&= 29/24 \\
V(0,2) &= \max\left(1,\ 1/3 \cdot 11/12 + 2/3 \cdot 2\right) \\
&= 59/36 \\
V(0,3) &= \max\left(2,\ 1/3 \cdot 11/12 + 1/3 \cdot 3/2 + 1/3 \cdot 2\right) \\
&= 2.
\end{aligned}$$

The following table indicates the optimal exercise strategy (+ means exercise in that state at that time; blank means do not exercise in that state at that time).

	State		
Time	1	2	3
0			+
1			+
2			+
3	+	+	+

If one observes prices 1, 3, 2, 1, we would have exercised at time 1.

If one observes 1, 2, 1, 1, we would have exercised at time 3.