Chapter 4. Total Expected (Discounted) Cost Problem

Total expected (discounted) cost criteria come up quite often in the literature. The properties of optimization problems of this type are relatively well understood under the Markov setting. However, no convenient theory is available for general non-Markovian setup, with the exception of the duality approach in some specific problem settings when the cost functions are concave or convex.

1 Markov control models

There are several different ways to define a *Markov control process* (MCP). A commonly used framework is as follows. The state process $\{X_n : n = 0, 1, \ldots\}$ evolves by

$$X_{n+1} = h(X_n, u_n; \xi_{n+1}), \qquad n = 0, 1, \ldots \tag{1}$$

where $\{\xi_n\}$ is a sequence of iid random *disturbances* with common distribution, say μ. The sequence $\{u_n\}$ is the control.

Unless specified, we will always assume that the state process $\{X_n\}$ takes value in a subset S of the Euclidean space $\mathbb{R}^d$. For each state $x \in S$, the set of *feasible controls* when the system is at this state is denoted by $U(x)$. A (pure) control policy $\{u_n\}$ is said to be *admissible* if for each $n = 0, 1, \ldots$, the control u_n can be expressed as a function of $(X_0, u_0, X_1, u_1, \ldots, X_n)$ taking value in the set $U(X_n)$. An admissible policy $\{u_n\}$ is said to be a *Markov policy* if for each n, the control u_n can be expressed as a function of X_n alone.

Remark **1** The above "definition" of the MCP is at best a formal description. A mathematically sound definition will involve all sorts of measurability conditions on the components of MCP. Also, the controls we consider are sometimes termed as *deterministic* (or *pure*) policies, as opposed to the more general case where u_n is a realization of a probability measure (the *randomized control*) on $U(X_n)$.

2 Finite-horizon

Given a Markov control process (1), the total expected cost associated with a control policy $\{u_n : n = 0, 1, \ldots, N-1\}$ is

$$J(x; \{u_n\}) \doteq E_x \left[g(X_N) + \sum_{j=0}^{N-1} c(X_j, u_j) \right];$$

where E_x just means the expectation is taken under the initial condition $X_0 = x \in S$. The function $c : S \to \mathbb{R}$ is said to be the *running cost* and $g : S \to \mathbb{R}$ the *terminal cost*. The objective of the control problem is to judiciously choose an admissible control policy so as to minimize the total expected cost. The value function is thus

$$v_N(x) \doteq \inf_{\{u_n\}} J(x; \{u_n\}).$$

The DPE associated with this problem is as follows.

$$V_N(x) \doteq g(x) \tag{2}$$

$$V_n(x) \doteq \inf_{u \in U(x)} \left[c(x; u) + \int V_{n+1}(h(x, u; \xi)) \mu(d\xi) \right] \tag{3}$$

for all $n = 0, 1, \ldots, N-1$. We have the following result

Proposition 1 *The value function $v_N(x) = V_0(x)$. If there is a sequence $\{u_n^* = u_n^*(x) \in U(x)\}$ achieving the infimum on the RHS of the equation (3), then $\{u_n^*\}$ defines an optimal Markov control policy with the corresponding dynamics*

$$\begin{aligned} X_0^* &= X_0 \\ X_{n+1}^* &= h(X_n^*, u_n^*(X_n^*); \xi_{n+1}), \qquad n = 0, 1, \ldots, N-1. \end{aligned}$$

Proof. Consider an arbitrary control sequence $\{u_n\}$. Let $\{X_n\}$ be the corresponding state process, and define

$$Z_n \doteq V_n(X_n) + \sum_{j=0}^{n-1} c(X_j; u_j), \qquad n = 0, 1, \ldots, N.$$

It is not difficult to see that $\{Z_n\}$ is a submartingale. Indeed,

$$\begin{aligned} E[Z_{n+1} \,|X_n, u_n, \ldots, X_0, u_0] &= \sum_{j=0}^{n} c(X_j; u_j) + E[V_{n+1}(X_{n+1}) \,|X_n, u_n] \\ &= \sum_{j=0}^{n} c(X_j; u_j) + \int V_{n+1}(h(X_n, u_n; \xi) \mu(d\xi) \\ &\geq \sum_{j=0}^{n-1} c(X_j; u_j) + V_n(X_n) \\ &= Z_n. \end{aligned}$$

In particular, we have $V_0(X_0) = E[Z_0] \leq E[Z_N]$. Thanks to the terminal condition (2), we have

$$V_0(X_0) \leq E\left[g(X_N) + \sum_{j=0}^{n-1} c(X_j; u_j)\right] = J(X_0; \{u_n\}).$$

Since the control $\{u_n\}$ is arbitrary, we have $V_0(X_0) \leq v(X_0)$.

Now assume that there exists a minimizer $\{u_n^* \doteq u_n^*(x) \in U(x)\}$ to the RHS of the DPE (3) with the corresponding state process $\{X_n^*\}$. Define the process $\{Z_n^*\}$ in a similar way, and it is not difficult to see that $\{Z_n^*\}$ is indeed a martingale. In particular,

$$V_0(X_0) = E[Z_0^*] = E[Z_N^*] = E\left[g(X_N^*) + \sum_{j=0}^{n-1} c(X_j^*; u_j^*)\right] = J(X_0; \{u_n^*\}).$$

Thus $V_0(X_0) \geq v(X_0)$ by definition, from which it follows readily that $V_0(X_0) = v(X_0)$ and $\{u_n^*\}$ is optimal.

It remains to show that $v(X_0) \leq V_0(X_0)$ when $\{u_n^*\}$ may not exist. Let $\varepsilon > 0$ be arbitrary, and let $u_n^\varepsilon \doteq u_n^\varepsilon(x)$ be such that

$$c(x; u_n^\varepsilon(x)) + \int V_{n+1}(h(x, u_n^\varepsilon(x); \xi))\mu(d\xi) \leq V_n(x) + \varepsilon$$

Abusing the notation a bit, let X_n be the state process corresponding to the control $\{u_n^\varepsilon\}$, and define Z_n as before. We have

$$\begin{aligned}
E[Z_{n+1} \,|X_n, u_n^\varepsilon, \ldots, X_0, u_0^\varepsilon] &= \sum_{j=0}^{n} c(X_j; u_j^\varepsilon) + E[V_{n+1}(X_{n+1}) \,|X_n, u_n^\varepsilon] \\
&= \sum_{j=0}^{n} c(X_j; u_j^\varepsilon) + \int V_{n+1}(h(X_n, u_n^\varepsilon; \xi)\mu(d\xi) \\
&\leq \sum_{j=0}^{n-1} c(X_j; u_j^\varepsilon) + V_n(X_n) + \varepsilon \\
&= Z_n + \varepsilon.
\end{aligned}$$

In particular, $E[Z_n] \geq E[Z_{n+1}] - \varepsilon$. It follows that

$$V_0(X_0) = E[Z_0] \geq E[Z_N] - N\varepsilon \geq v(X_0) - N\varepsilon.$$

Letting $\varepsilon \to 0$, we complete the proof. ■

Remark **2** The value function $\{v_n\}$ satisfies the DPE of *forward* form

$$\begin{aligned} v_0(x) &= g(x) \\ v_{n+1}(x) &= \inf_{u\in U(x)} \left[c(x;u) + \int v_n(h(x,u;\xi))\mu(d\xi) \right] \end{aligned}$$

for all $n \geq 0$.

Remark **3** Suppose we want to minimize the total expected discount cost

$$E_x \left[\beta^N g(X_N) + \sum_{j=0}^{N-1} \beta^j c(X_j, u_j) \right];$$

for some constant $\beta \in (0, 1]$. Then all the DPE becomes

$$\begin{aligned} V_N(x) &= g(x) \\ V_N(x) &= \inf_{u\in U(x)} \left[c(x;u) + \beta \int V_{n+1}(h(x,u;\xi))\mu(d\xi) \right] \end{aligned}$$

for all $n \geq 0$. The value function $v(x)$ again, equals $V_0(x)$.

Remark **4** Sometimes the cost involved in the control may depend on the disturbances. For example, suppose the running cost takes form $c(X_n, u_n; \xi_{n+1})$. In this case, the DPE can be written as

$$\begin{aligned} V_N(x) &\doteq g(x) \\ V_n(x) &\doteq \inf_{u\in U(x)} \left[\bar{c}(x;u) + \int V_{n+1}(h(x,u;\xi))\mu(d\xi) \right]. \end{aligned}$$

where

$$\bar{c}(x;u) \doteq \int c(x,u;\xi)\mu(d\xi).$$

Thus the problem is reduced to a problem with running cost of form $\bar{c}(x;u)$.

2.1 Words of caution on measurability

The proof we have given for Proposition 1 is not completely rigorous. Remember that the probability theory is established on some probability space, with all operations taken on functions or sets that are measurable. We intentionally put the measurability out of the picture, because otherwise many annoying problems will obscure the idea of DP. For example, the definition of $V_n(x)$ by equation (3) does not necessarily imply that the function V_n is a

suitable measurable function. Without such measurability, its integral will not be well defined. Also, the minimizer u_n^*, if exists, may not be measurable, which leads to the difficulty that the resulting state process $\{X_n\}$ may not be legitimate random variables, and the expectation of the total cost will not be well defined. Moreover, in the proof, we construct the *ε-optimal* policy $\{u_n^\varepsilon\}$. But there is no guarantee that the policy $\{u_n^\varepsilon\}$ is measurable, and thus all the following lines are not entirely justifiable.

In most applications, one can directly check that the functions V_n is measurable from the DPE, and that there exists a measurable (in some proper sense) minimizer $\{u_n^*\}$. However, to resolve the measurability concerns for a general control problem is not an easy task. One approach is to extend the notion of Borel measurability so as to include more admissible controls, such that $\{V_n\}$ is measurable in this generalized sense and that the existence of ε-optimal policies (and the validity of DPE) can be proved; see [1]. Another approach is to put some constraints on the cost functions c and g, the control space $U(x)$, and the dynamics of the system, so that the equation (3) will admit a measurable minimizer, and the functions $\{V_n\}$ defined through DPE will be measurable; see [4]. For example, one type of such conditions is

1. For every $x \in S$, the control space $U(x)$ is compact.
2. For every $x \in S$, the function $c(x;\cdot)$ is non-negative and lower semicontinuous on the control space $U(x)$.
3. For a function $V : S \to \mathbb{R}$, define
$$\bar{V}(x;u) \doteq \int V(h(x,u;\xi))\mu(d\xi).$$
Then either one of the two following conditions hold.
   (a) $\bar{V}(x;\cdot)$ is lower semicontinuous on $U(x)$ for every $x \in S$ and every continuous bounded function V on S, and function g is non-negative and lower seimcontinuous.
   (b) $\bar{V}(x;\cdot)$ is lower semicontinuous on $U(x)$ for every $x \in S$ and every bounded measurable function V on S, and function g is non-negative and measurable.

2.2 Examples

Example 1 *(Inventory control)* Consider an inventory production system in which the state variable X_n is the stock level at the beginning of period

n for $n = 0, 1, \ldots, N$. The control $u_n \geq 0$ is the quantity ordered (or produced) and immediately supplied at the beginning of period n, and the disturbance ξ_{n+1} is the demand during this period. We assume $\{\xi_n\}$ is a sequence of iid non-negative, integrable random variables. For the simplicity of computations, we will further assume that $\{\xi_n\}$ have a common density f and denote by F the cumulative distribution function.

The dynamics of the system follow

$$X_{n+1} = X_n + u_n - \xi_{n+1}, \qquad n = 0, 1, \ldots, N-1.$$

Here we allow a "negative" inventory level by assuming that excess demand is backlogged and filled when additional inventory becomes available. We also have assumed that the capacity of the system is infinity. One wishes to minimize the total expected operation cost

$$E\left[\sum_{n=0}^{N-1} \beta^n c(X_n, u_n; \xi_{n+1})\right]$$

with a discount factor $\beta \in (0, 1]$ and a cost function

$$c(X_n, u_n; \xi_{n+1}) \doteq b \cdot u_n + h \cdot (X_n + u_n - \xi_{n+1})^+ + p \cdot (X_n + u_n - \xi_{n+1})^-,$$

where b is the unit production cost, h the unit holding cost for excessive inventory, and p the unit shortage cost for unfilled demands. We assume $p > b$ (otherwise there is no incentive for production).

The value function will be denoted by $v_N(x)$ given $X_0 = x$. The corresponding DPE is (see Remark 4)

$$\begin{aligned} V_N(x) &\doteq 0 \\ V_n(x) &\doteq \inf_{u \geq 0} \left[\bar{c}(x; u) + \beta \int_{\mathbb{R}} V_{n+1}(x + u - \xi) f(\xi) d\xi\right]. \end{aligned}$$

where

$$\begin{aligned} \bar{c}(x; u) &\doteq bu + h \int_{\mathbb{R}} (x + u - \xi)^+ f(\xi)\, d\xi + p \int_{\mathbb{R}} (x + u - \xi)^- f(\xi)\, d\xi \\ &\doteq bu + L(x + u). \end{aligned}$$

Note that the function L is a convex function, and that

$$L'(x + u) = (h + p) \int_{-\infty}^{x+u} f(\xi)\, d\xi - p = (h + p) F(x + u) - p.$$

Let us first take a look of the case $n = N - 1$, in which case

$$V_{N-1}(x) = \inf_{u \geq 0} \bar{c}(x; u).$$

It follows trivially from the convexity of L that

$$V_{N-1}(x) = \begin{cases} L(x) & ; \quad \text{if } x \geq x^*_{N-1} \\ L(x^*_{N-1}) + b(x^*_{N-1} - x) & ; \quad \text{if } x < x^*_{N-1} \end{cases}$$

with $x^*_{N-1} \doteq F^{-1}\left[(p-b)/(p+h)\right]$ and the minimizer

$$u^*_{N-1}(x) = \begin{cases} 0 & ; \quad \text{if } x \geq x^*_{N-1} \\ x^*_{N-1} - x & ; \quad \text{if } x < x^*_{N-1} \end{cases}$$

Clearly, V_{N-1} is convex (even though it is the minimum of convex functions).

Now consider $n = N - 2$. The DPE implies

$$V_{N-2}(x) \doteq \inf_{u \geq 0} \left[bu + L(x+u) + \beta \int_{\mathbb{R}} V_{N-1}(x + u - \xi) f(\xi) d\xi \right].$$

The argument will be exactly the same except that function $L(x+u)$ is replaced by

$$\bar{L}(x+u) \doteq L(x+u) + \beta \int_{\mathbb{R}} V_{N-1}(x + u - \xi) f(\xi),$$

observing that $\bar{L}$ is again a convex function. In particular, there exists a x^*_{N-2} such that

$$V_{N-2}(x) = \begin{cases} \bar{L}(x) & ; \quad \text{if } x \geq x^*_{N-2} \\ \bar{L}(x^*_{N-2}) + b(x^*_{N-2} - x) & ; \quad \text{if } x < x^*_{N-2} \end{cases}$$

and

$$u^*_{N-2}(x) = \begin{cases} 0 & ; \quad \text{if } x \geq x^*_{N-2} \\ x^*_{N-2} - x & ; \quad \text{if } x < x^*_{N-2} \end{cases}$$

Again V_{N-2} is convex.

It is not difficult to see that for an arbitrary $n = 0, 1, \ldots$, the function V_n is convex. In particular $v_N = V_0$ is convex. The optimal policy is determined by

$$u^*_n(x) = \begin{cases} 0 & ; \quad \text{if } x \geq x^*_n \\ x^*_n - x & ; \quad \text{if } x < x^*_n \end{cases}$$

for a sequence of thresholds $\{x^*_n\}$. Sometimes it is called a *threshold-type* policy. Many control problems have optimal policies of this type. ■

3 Infinite-horizon: an overview

Given the Markov control process (1), the total expected discounted cost associated with control $\{u_n : n = 0, 1, \ldots\}$ on the infinite horizon is defined by

$$J(x; \{u_n\}) \doteq E_x \left[\sum_{j=0}^{\infty} \beta^j c(X_j, u_j) \right]$$

for some discount factor $\beta \in (0, 1)$. The objective is to minimize J over all control sequences $\{u_n\}$ such that $u_n \in U(X_n)$ for every n. We will denote the value function by U, or

$$U(x) = \inf_{\{u_n\}} J(x; \{u_n\}).$$

The DPE associated with this problem is

$$W(x) = \inf_{u \in U(x)} \left[c(x; u) + \beta \int W(h(x, u; \xi)) \mu(d\xi) \right] \tag{4}$$

There are three main questions that we are interested in.

1. Characterize the value function u through the DPE. More precisely, does the value function U satisfy the DPE?

2. Characterize the optimal control policy through the DPE. More precisely, does the minimizer to the RHS of the DPE induce an optimal control policy?

3. Convergence of the DP algorithm. More precisely, does the finite horizon value function v_n (with terminal cost $g = 0$ and discount factor β) converge to U as n tends to infinity? This is related to the computational methods of solving for the value function U.

More often than not, the answers to the above three questions are affirmative. However, there are pathological counter examples that the answer to these questions are negative. We list below one such example. For more counter examples, see [2, 5].

Example 2 Consider a deterministic control problem with dynamics

$$X_{n+1} = \frac{2}{\beta} X_n + u_n, \qquad n = 0, 1, \ldots$$

with control $u_n \in (0, \infty)$ and initial condition $X_0 = 0$. We want to minimize the cost defined by

$$J(x; \{u_n\}) \doteq \sum_{j=0}^{\infty} \beta^j X_j.$$

It can be easily argued that the corresponding finite horizon problem has a value function $v_N(0) \equiv 0$ for all $N \in \mathbb{N}$. However, we claim that $U(0) = \infty$, and thus $v_N(0)$ does not converge to $U(0)$. Indeed, for any control sequence $\{u_n\}$, we have

$$X_{n+1} \geq \frac{2}{\beta} X_n \geq \cdots \geq \left(\frac{2}{\beta}\right)^n X_1 = \left(\frac{2}{\beta}\right)^n u_1$$

Thus

$$J(0; \{u_n\}) \geq \sum_{j=0}^{\infty} \beta^j X_j \geq u_1 \sum_{j=1}^{\infty} \beta^j \left(\frac{2}{\beta}\right)^{j-1} = \infty.$$

This complete the proof. ■

4 Infinite-horizon: the contraction mapping approach

We should assume throughout this section that the boundedness condition.

Condition **1** *The cost function $c(x; u)$ is uniformly bounded.*

Under this assumption, the value function U is bounded.

Let f be an arbitrary bounded function from the state space S, and define an operator $\mathbb{L}$ by

$$\mathbb{L}f(x) \doteq \inf_{u \in U(x)} \left[c(x; u) + \beta \int f(h(x, u; \xi)) \mu(d\xi) \right]. \tag{5}$$

Then $\mathbb{L}$ is again a bounded function. The DPE can be rewritten as

$$W = \mathbb{L}W.$$

In other words, a function W is a solution to the DPE if and only if the function W is a *fixed point* of the operator $\mathbb{L}$. We want to establish that the operator $\mathbb{L}$ is a *contraction* mapping.

4.1 The characterization of value function and optimal policy

Let $\mathbb{L}$ be the operator defined by the equation (5). Let $M_b(S)$ be the space of all bounded (measurable) functions on S, endowed with the sup norm. The operator $\mathbb{L}$ defines a mapping from space $M_b(S)$ to $M_b(S)$; see Remark 5 for more discussion on this issue.

Furthermore, let v_n be the value function for the corresponding discount cost control problem with finite horizon n and terminal cost 0, then the forward DP algorithm for the $\{v_n\}$ (see Remark 2 and take into consideration of the discount factor) is

$$\begin{aligned} v_0 &= 0 \\ v_{n+1} &= \mathbb{L}v_n, \qquad n = 0, 1, \ldots \end{aligned}$$

Lemma 1 *Under Condition 1, the operator $\mathbb{L} : M_b(S) \to M_b(S)$ is a contraction, where $M_b(S)$ in endowed with the sup norm.*

The proof is an easy corollary of the Blackwell's sufficient condition for contraction (Theorem 7). We omit the proof.

Theorem 1 *Under Condition 1, we have*

1. *The value function U is the unique fixed point of the operator $\mathbb{L}$; i.e., U is the unique bounded solution to the DPE (4).*
2. *The DP algorithm converges, that is, $v_n(x) \to U(x)$ as $n \to \infty$. Indeed, the convergence is uniform over $x \in S$.*
3. *If $u^* \doteq u^*(x)$ is a minimizer of the RHS of the DPE (4), then u^* defines an optimal Markov control policy with the corresponding dynamics*

$$\begin{aligned} X_0^* &= X_0 \\ X_{n+1}^* &= h(X_n^*, u^*(X_n^*); \xi_{n+1}), \qquad n = 0, 1, \ldots. \end{aligned}$$

Before we give a proof of the theorem, it worth pointing out that even though the DPE (4) has unique bounded solution, it may admit other solutions that are unbounded.

Example 3 Consider a Markov control process with state space $S = \{1, 2, \ldots\}$, control space $U(x) \equiv \{0\}$, and running cost $c(x; u) \equiv 0$. The transition probability for this chain is

$$P\left(X_{n+1} = 1 \,|X_n, \{u_n = 1\}\right) = \frac{2X_n}{3(2X_n - 1)}$$

and

$$P\left(X_{n+1} = 2X_n \,|X_n, \{u_n = 1\}\right) = \frac{4X_n - 3}{3(2X_n - 1)}.$$

Let the discount factor $\beta = 3/4$. The DPE associated with this problem is

$$W(x) = \beta\left[\frac{2x}{3(2x-1)}W(1) + \frac{4x-3}{3(2x-1)}W(2x)\right], \quad x = 0, 1, \ldots$$

Clearly the value function $U(x) \equiv 0$ is a solution to the DPE, but $W(x) \equiv x$ is also an (unbounded) solution to the DPE. ■

Proof of Theorem 1. Under Condition 1, the operator $\mathbb{L} : M_b(S) \to M_b(S)$ is a contraction when $M_b(S)$ is endowed with the sup norm. Thus there exists a unique fixed point, say $U^* \in M_b(S)$, such that $U^* = \mathbb{L}U^*$.

We prove now $U^* = U$. Assume $|c(x;u)| \le K$ for every x and u. By definition, for any control sequence $\{u_n\}$, we have

$$E_x\left[\sum_{j=0}^{\infty}\beta^j c(X_j, u_j)\right] = E_x\left[\sum_{j=0}^{N-1}\beta^j c(X_j, u_j)\right] + E_x\left[\sum_{j=N}^{\infty}\beta^j c(X_j, u_j)\right].$$

But the absolute value of the last term is bounded by

$$\sum_{j=N}^{\infty}\beta^j K = K\beta^N/(1-\beta).$$

Taking infimum over $\{u_n\}$, we have $|U(x) - v_N(x)| \le K\beta^N/(1-\beta)$. Letting $n \to \infty$, we arrive at $U(x) = \lim_n v_n(x) = U^*(x)$.

It remains to show that $\{u_n^*\}$, if exists, defines an optimal control sequence. Consider the process

$$Z_n^* \doteq \sum_{j=0}^{n-1}\beta^j c(X_j^*, u_j^*) + \beta^n U(X_n^*); \quad n = 0, 1, \ldots.$$

It is not difficult to verify that $\{Z_n^*\}$ defines a martingale. In particular,

$$U(x) = E_x[Z_0^*] = E_x[Z_n^*] = E_x\left[\sum_{j=0}^{n-1}\beta^j c(X_j^*, u_j^*) + \beta^n U(X_n^*)\right]$$

for every n. Letting $n \to \infty$, thanks to the boundedness of c and the DCT,

$$U(x) = E_x\left[\sum_{j=0}^{\infty}\beta^j c(X_j^*, u_j^*)\right].$$

This completes the proof. ■

Remark **5** As usual, we are intentionally avoid any measurability concerns. Given a function $f \in M_b(S)$, the image $\mathbb{L}f$ is a bounded function for sure, but may *not* be a measurable. However, there are conditions like those in Section 2.1 to guarantee the measurability of $\mathbb{L}f$. In most applications, however, one can directly check the measurability of $\mathbb{L}f$. Since $\{v_n\}$ satisfy $v_n = \mathbb{L}v_{n-1} = \mathbb{L}^n v_0$, Theorem 6 implies that v_n converges to U^* uniformly.

4.2 Example: An optimal stopping problem

Let the dynamics be $X_{n+1} = h(X_n; \xi_{n+1})$ where $\{\xi_n\}$ is a sequence of iid radon variable with common distribution μ. Let $c : S \to \mathbb{R}$ be a non-negative, bounded function, and $k \geq 0$ a constant. The problem is to find a stopping time τ taking values in $\{0, 1, \ldots, \infty\}$ so as to maximize the expectation

$$E_x \left[\sum_{j=0}^{\tau-1} \beta^j c(X_j) + \beta^\tau k \right].$$

Let the value function be $\Phi(x)$.

One can put this problem into the framework of infinite horizon total expected discounted problems. Introduce a generic state $\bar{x}$ and expend the state space to $\bar{S} = S \cup \bar{x}$. The control will be $u_n \in \{0, 1\}$ with "0" for continuation, and "1" for stop. The state process will be

$$X_{n+1} = \begin{cases} h(X_n; \xi_{n+1}) & ; \quad \text{if } u_n = 0 \text{ and } X_n \neq \bar{x} \\ \bar{x} & ; \quad \text{if } u_n = 1 \text{ or } X_n = \bar{x} \end{cases}.$$

The payoff associated with the problem is

$$c(x; u) = \begin{cases} c(x) & ; \quad \text{if } u_n = 0 \text{ and } X_n \neq \bar{x} \\ (1-\beta)k & ; \quad \text{if } u_n = 1 \text{ or } X_n = \bar{x} \end{cases}.$$

Then it is not difficult to check that

$$\Phi(x) = \sup_{\{u_n\}} E_x \left[\sum_{j=0}^{\infty} \beta^j c(X_j; u_j) \right],$$

and Φ the unique bounded solution (fixed point) to the DPE

$$\Phi(x) = \max \left\{ k, \quad c(x) + \beta \int \Phi(h(x; \xi)) \mu(d\xi) \right\}$$

for $x \in S$. This is exactly the DPE we should be expecting from the outset using the theory of optimal stopping!

We will state the following result, which is needed for the next example. For convenience, we denote the value function $\Phi(x)$ by $\Phi(x;k)$ for the obvious reason.

Lemma 2 *The mapping $k \mapsto \Phi(x;k)$ is non-negative, convex, non-decreasing, and satisfies $\Phi(x;k) \geq k$. Denote its right-derivative by $D_k^+\Phi(x;k)$. Then $0 \leq D_k^+\Phi(x;k) \leq 1$ for every x and k. Furthermore, for $k \geq \|c\|_\infty/(1-\beta)$, we have $\Phi(x;k) \equiv k$ and $D_k^+\Phi(x;k) = 1$.*

Proof. Clearly the mapping $k \mapsto \Phi(x;k)$ is non-negative, non-decreasing, and satisfies $\Phi(x;k) \geq k$. The convexity is also trivial from the definition since the value function can be regarded as the supremum of a collection of linear functions of k.

For the remaining claims, all we need to show is that $\Phi(x;k) = k$ for $k \geq \|c\|_\infty/(1-\beta)$. Indeed, it follows that

$$\Phi(x;k) \leq \sup_\tau E_x \left[\sum_{j=0}^{\tau-1} \beta^j k(1-\beta) + \beta^\tau k \right] = k.$$

We complete the proof. ■

4.3 Example: Multiarmed bandit problem

The example we are going to discuss is the classical "multiarmed bandit problem". Consider the following scenario. There are d projects of which one can be worked on at any time period. For simplicity, we assume $d = 2$. The same argument works for general d. We will denote by $X_{1,n}$ and $X_{2,n}$ the state of the two projects at time n. If project i $(i = 1, 2)$ is worked on, it will produce a reward $\beta^n R_i(X_{i,n})$ with $\beta \in (0,1)$. The state of project i at the next time step will become

$$X_{i,n+1} = h_1(X_{i,n}\ ;\ \xi_{i,n+1}),$$

while the state of the other project will stay the same; i.e.

$$X_{j,n+1} = X_{j,n}, \qquad \text{for } j \neq i.$$

The sequences $\{\xi_{1,n}\}$ and $\{\xi_{2,n}\}$ are assumed to be independent, iid sequences with distribution μ_1 and μ_2 respectively. We will further assume that $0 \leq R_i(x) \leq B$; i.e. the reward function is non-negative and bounded. The question is to find a policy so as to maximize the total expected discounted reward.

Let $X_n \doteq (X_{1,n}, X_{2,n})$ be the state at n, and the initial state $X_0 = x = (x_1, x_2)$. The control u_n can only take two value, say $u_n = 1$ if project 1 is worked on, and $u_n = 2$ if project 2 is worked on. The problem falls into the framework of maximizing

$$V(x) \doteq V(x_1, x_2) = \sup_{\{u_n\}} E_x \left[\sum_{n=0}^{\infty} \beta^n c(X_n; u_n) \right]$$

with the payoff $c(X_n, u_n) = R_1(X_{1,n})$ if $u_n = 1$ and $c(X_n, u_n) = R_2(X_{2,n})$ if $u_n = 2$. Theorem 1 immediately yields

Lemma 3 *The value function V is the unique bounded function satisfying the DPE*

$$\begin{aligned} V(x_1, x_2) \quad = \quad & \max \Big\{ R_1(x_1) + \beta \int V(h_1(x_1; \xi_1), x_2)\, \mu_1(d\xi_1), \\ & R_2(x_2) + \beta \int V(x_1, h_2(x_2; \xi_2))\, \mu_2(d\xi_2) \Big\}. \end{aligned} \tag{6}$$

Let $\{u_n^* \doteq u_n^*(x_1, x_2)\}$ be index of the maximizing term on the RHS of the DPE. Then $\{u_n^*\}$ defines an optimal policy.

This result is not entirely satisfactory because it does not provide much information on the value function and the optimal policy.

Instead, we consider the following auxiliary optimal stopping problems. For $i = 1, 2$, and constant $k \geq 0$, define

$$\Phi_i(x_i; k) = E_{x_i} \left[\sum_{n=0}^{\tau - 1} \beta^n R_i(X_{i,n}) + \beta^\tau k \right],$$

where the supremum is taken over all stopping times taking value in $\{0, 1, \ldots, \infty\}$ and the dynamics of the system is $X_{i,n+1} = h_i(X_{i,n}\ ;\ \xi_{i,n+1})$. The properties of Φ_i has been presented in Lemma 2.

Define the *Gittins index*

$$m_i(x_i) \doteq \inf \{k \geq 0 :\ \Phi(x_i; k) = k\}, \quad i = 1, 2.$$

Clearly, for every $k \geq m_i(x_i)$, $\Phi(x_i; k) = k$. We have the following result.

Theorem 2 *Let C be an arbitrary constant such that $C \geq (\|R_1\|_\infty \vee \|R_2\|_\infty)/(1-\beta)$. The value function V admits the following representation.*

$$V(x_1, x_2) = C - \int_0^C \prod_{i=1}^{2} D_k^+ \Phi_i(x_i; k)\, dk$$

for every (x_1, x_2). *An optimal policy is determined by* $\{u^* \doteq u^*(x_1, x_2)\}$ *which is the maximizing index for* $m_1(x_1) \vee m_2(x_2)$. *In other words, it is optimal to work on the project with the maximal current Gittins index.*

Proof. Let

$$U(x_1, x_2) \doteq C - \int_0^C \prod_{i=1}^{2} D_k^+ \Phi_i(x_i; k)\, dk.$$

We wish to prove that U is a bounded solution to the DPE (6) and u^* is the maximizer. To this end, we first rewrite U by integration by parts:

$$U(x_1, x_2) = C - D_k^+ \Phi_1(x_1; k)\Phi_2(x_2; k)\Big|_0^C + \int_0^C \Phi_2(x_2; k) d_k D_k^+ \Phi_1(x_1; k).$$

Thanks to Lemma 2, $\Phi_2(x_2; C) = C$ and $D_k^+ \Phi_1(x_1; C) = 1$. Thus

$$U(x_1, x_2) = D_k^+ \Phi_1(x_1; 0)\Phi_2(x_2; 0) + \int_0^C \Phi_2(x_2; k) d_k D_k^+ \Phi_1(x_1; k).$$

It follows that

$$\begin{aligned}\int U(x_1, h_2(x_2; \xi_2))\, \mu_2(d\xi_2) &= D_k^+ \Phi_1(x_1; 0) \int \Phi_2(h_2(x_2; \xi_2); 0)\, \mu_2(d\xi_2) \\ &\quad + \int \int_0^C \Phi_2(h(x_2, \xi_2); k) d_k D_k^+ \Phi_1(x_1; k)\, \mu_2(d\xi_2).\end{aligned}$$

By Tonelli theorem, one can switch the order of integration in the last term, and we arrive at

$$\begin{aligned}&U(x_1, x_2) - R_2(x_2) - \beta \int U(x_1, h_2(x_2; \xi_2))\, \mu_2(d\xi_2) \\ &= D_k^+ \Phi_1(x_1; 0) \left[\Phi_2(x_2; 0) - R_2(x_2) - \beta \int \Phi_2(h_2(x_2; \xi_2); 0)\, \mu_2(d\xi_2)\right] \\ &\quad + \int_0^C \left[\Phi_2(x_2; k) - R_2(x_2) - \beta \int \Phi_2(h(x_2, \xi_2); k)\, \mu_2(d\xi_2)\right] d_k D_k^+ \Phi_1(x_1; k),\end{aligned}$$

thanks again to $D_k^+ \Phi_1(x_1; C) = 1$. To ease notation, we will use "LHS" for the left hand side of this equation in this proof.

Write, for $i = 1, 2$,

$$\phi_i(x_i; k) \doteq \Phi_i(x_i; k) - R_i(x_i) - \beta \int \Phi_i(h(x_i, \xi_i); k)\, \mu_i(d\xi_i).$$

The DPE for Φ_i implies that $\phi_i(x_i; k) \geq 0$, $\phi_i(x_i, 0) = 0$, and the definition of Gittins index implies that $\phi_i(x_i; k) = 0$ if $k < m_i(x_i)$. Since $D_k^+ \Phi_1(x_1; k)$ is non-decreasing, we have

$$\text{LHS} = \int_0^C \phi_2(x_2; k) d_k D_k^+ \Phi_1(x_1; k) \geq 0.$$

However, on set $\{m_2(x_2) \geq m_1(x_1)\}$, the above inequality is indeed equality, since $\phi_2(x_2;k) = 0$ for $k < m_1(x_1)$, and $D_k^+\phi_1(x_1;k) \equiv 1$ for $k \geq m_1(x_1)$. In other words,

$$U(x_1,x_2) - R_2(x_2) - \beta \int U(x_1, h_2(x_2;\xi_2))\, \mu_2(d\xi_2)$$

is always non-negative, and 0 on set $\{m_2(x_2) \geq m_1(x_1)\}$. Similarly,

$$U(x_1,x_2) - R_1(x_1) - \beta \int U(h_1(x_1;\xi_1), x_2)\, \mu_1(d\xi_1)$$

is always non-negative, and 0 on set $\{m_2(x_2) \leq m_1(x_1)\}$. This completes the proof. ■

Example 4 Assume that we have two gold mines, and a gold mining machine. Let x_1, x_2 be the current amount of gold in Gold Mine 1 and Gold Mine 2, respectively. When the machine is used in Gold Mine i, there is a probability p that $r_i x_i$ of gold will be mined without damaging the machine, and a probability $1-p$ that the machine will be damaged beyond repair and no gold will be mined. Assume $r_1, r_2 \in (0,1)$. Find the mine selecting policy that maximize the total expected discounted (with discount factor β) amounted of gold mined before the machine breaks down.

Solution: Let the value function be $V(x_1,x_2)$. Then the value function V satisfies the DPE (see Remark 6)

$$V(x_1,x_2) = \max\left\{pr_1x_1 + p\beta V((1-r_1)x_1;x_2),\ pr_2x_2 + p\beta V(x_1;(1-r_2)x_2)\right\}.$$

Compare this equation with the DPE (6), the value function corresponding to a multiarmed bandit problem with

$$R_i(x_i) = pr_ix_i, \quad i = 1,2,$$

and the discount factor $p\beta$, and the dynamics for project i (if project i is worked on)

$$X_{i,n+1} = (1-r_i)X_{i,n}, \qquad i = 1,2.$$

Thus, the corresponding auxiliary optimal stopping problems are

$$\Phi_i(x_i;k) = \sup_\tau E_{x_i}\left[\sum_{n=0}^{\tau-1}(p\beta)^n R_i(X_{i,n}) + (p\beta)^\tau k\right]$$

with $X_{i,n+1} = (1-r_i)X_{i,n}$. It is not difficult to check that the value function is (check!)

$$\Phi_i(x_i;k) = \begin{cases} k & ; \quad \text{if } x_i \le x_i^* \\ \varepsilon_i x_i & ; \quad \text{if } x_i \ge x_i^* \end{cases}$$

with

$$\varepsilon_i = \frac{pr_i}{1-p\beta(1-r_i)}, \qquad x_i^* = \frac{k}{\varepsilon_i} = \frac{k[1-p\beta(1-r_i)]}{pr_i}.$$

The Gittins index is thus

$$m_i(x_i) \doteq \inf\{k : \ \Phi_i(x_i;k) = k\} = \varepsilon_i x_i.$$

Thus the optimal control is to use the machine in the mine with the maximal $\varepsilon_i x_i$.

In order to compute the value function V, note that

$$D_k^+ \Phi_i(x_i;k) = 1_{\{k \ge \varepsilon_i x_i\}}.$$

Take the constant C, say as $x_1 + x_2$ (why we can do so?). We have

$$V(x_1,x_2) = C - \int_0^C 1_{\{k \ge \varepsilon_1 x_1 \vee \varepsilon_2 x_2\}}\, dk = \varepsilon_1 x_1 \vee \varepsilon_2 x_2.$$

This complete the solution. ■

Remark **6** In the above example, one may protest that the payoff $R_i(x_i) = pr_i x_i$ is not a bounded function, and thus we cannot apply the existing result. However, since in this example the dynamics is such that $\{X_{i,n}\}$ is non-negative and non-increasing, one can literally put the problem into the bounded-cost framework.

4.4 Application: Asset price in an exchange economy

Consider an economy with a single consumer, interpreted as a representative for a collection of identical consumers. There are a finite number, say d, of (after normalization) *unit* distinctive productive assets. There is also a single type of consumption good. Each asset produces a random quantity of this consumption good each period; we call these *dividends*. Denote by

$$Y_n \doteq (Y_n^1, Y_n^2, \ldots, Y_n^d)$$

the vector of dividends produced by unit productive assets at period n. We assume that $\{Y_0, Y_1, \ldots\}$ form a non-negative, time homogeneous, Markov chain with transition probability function $P(dy|x)$.

Suppose at the beginning of period n, the consumer owns

$$X_n \doteq (X_n^1, X_n^2, \ldots, X_n^d)$$

units of productive assets, which will produce for this period a total dividend of

$$\sum_{j=1}^{d} X_n^j Y_n^j \doteq X_n \cdot Y_n.$$

These dividend can be used for two purposes: (1) consumption; (2) reallocation of the ownership of the productive assets through a competitive stock market. At period n, let the price vector for unit productive assets be

$$p_n \doteq (p_n^1, p_n^2, \ldots, p_n^d).$$

Suppose that the consumer wants to consume c_n dividends for period n and owns X_{n+1} productive assets at the beginning of period $n+1$, he or she can do so as long as the constraints

$$c_n + p_n \cdot X_{n+1} \leq X_n \cdot Y_n + p_n \cdot X_n,$$

as well as

$$c_n \geq 0, \qquad 0 \leq X_{n+1} \leq (1+\delta, 1+\delta, \ldots, 1+\delta) \doteq 1+\delta.$$

is satisfied. The value of the positive constant δ is not essential, since it will not affect the equilibrium state as we will see below. The consumer wishes to maximize the quantity

$$E\left[\sum_{n=0}^{\infty} \beta^n U(c_n)\right],$$

where $\beta \in (0,1)$ is a discounting factor, and $U : \mathbb{R}^+ \to \mathbb{R}^+$ is a bounded *utility* function.

We will study the determination of *equilibrium* asset prices in the economy, in other words, the behavior of $\{p_n\}$ in the equilibrium. The equilibrium state in this economy is fairly easy: all productive asset will be held in each time period, that is,

$$\bar{X}_n \equiv 1,$$

and all dividends have to be consumed, that is,

$$\bar{c}_n \equiv \bar{X}_n \cdot Y_n = 1 \cdot Y_n.$$

But this says that in equilibrium all the information of the system at period n is contained in the random variable Y_n. Since the problem is of infinite horizon, the asset price p_n in equilibrium should have form

$$p_n = \bar{p}(Y_n) = [\bar{p}^1(Y_n), \bar{p}^2(Y_n), \ldots, \bar{p}^d(Y_n)]$$

for some *price* function $\bar{p}$. The question is what characterizes the equilibrium price function $\bar{p}$.

Now given this equilibrium, the consumer will be able to determine the optimal consumption rule $\{c_n^*\}$ and the optimal state process $\{X_n^*\}$. Thus the price function $\bar{p}$ defines the behavior of the consumer. On the other hand, given the optimal consumption rule $\{c_n^*\}$ and the optimal state process $\{X_n^*\}$, the market clearing conditions determines a price function p. In this sense, the consumer's behavior determines the price function. We close the system by an assumption of *rational expectations:* the market clearing price function p implied by consumer behavior is the same as the price function $\bar{p}$ on which consumer's decisions are based.

Given a price function p, denote by $v^p(x,y)$ the value function with initial condition $X_0 = x$ and $Y_0 = y$. Then v^p satisfies the DPE

$$v^p(x,y) = \sup_{(c,z)} \left[U(c) + \beta \int v^p(z, y_1) P(dy_1|y) \right] \doteq (\mathbb{L}v^p)(x,y)$$

where the supremum is over all $c \in \mathbb{R}$ and $z \in \mathbb{R}^d$ such that

$$c + p(y) \cdot z \leq y \cdot x + p(y) \cdot x$$

and

$$c \geq 0, \quad 0 \leq z \leq 1 + \delta.$$

Thus an equilibrium price function $\bar{p}$ is such that, for $v^{\bar{p}}(1, y)$, the supremum on the RHS of the DPE is attained at

$$c^* = 1 \cdot y, \qquad z^* = 1.$$

Our discussion will be taken upon the space of bounded continuous functions. For that, we will impose the following regularity conditions.

Condition **2** *1. The Markov chain $\{Y_n\}$ is non-negative, takes value in a compact subset $S \subset \mathbb{R}^d$, and $P(1 \cdot y > 0|x) = 1$ for every $x \in S$.*

2. The Markov chain $\{Y_n\}$ satisfies the Feller property*; i.e., for any bounded continuous function $h : S \to \mathbb{R}$, the function*

$$(\mathbb{T}h)(x) \doteq \int_S h(y)P(dy|x)$$

is a bounded continuous function.

3. The utility function $U : \mathbb{R}^+ \to \mathbb{R}^+$ is bounded, strictly increasing, strictly concave, continuously differentiable, with $U(0) = 0$.

We are only going to consider price function that are continuous (and non-negative).

Definition 1 *An equilibrium price function $\bar{p} : S \to \mathbb{R}^d$ is a continuous, non-negative function such that, for the consumer with initial state $X_0 = 1$, the optimal policy is*

$$c^* = 1 \cdot y, \qquad z^* \doteq 1.$$

We will start with the following lemma that characterizes the value function v^p for any price function p.

Lemma 4 *For every continuous, non-negative, price function $p : S \to \mathbb{R}^d$, the value function v^p is the unique bounded, non-negative, continuous function satisfying equation*

$$v = \mathbb{L}v.$$

Furthermore, for each $y \in S$, the function $v^p(x, y)$ is an increasing, concave function of x.

Proof. For the first part, it suffices to show that $\mathbb{L}$ maps any bounded continuous function to a bounded continuous function, and it is a contraction. It is immediate from Lemma 6 and Theorem 8 (Appendix B) that $\mathbb{L}v$ is

bounded and continuous (and the supremum is attained). As for the contraction property, observe that the operator $\mathbb{L}$ is monotone in that for any $v_1 \leq v_2$, we have $\mathbb{L}v_1 \leq \mathbb{L}v_2$, and for any constant $c \geq 0$, $\mathbb{L}(v+c) = \mathbb{L}v + \beta c$. Thus $\mathbb{L}$ is a contraction, thanks to Theorem 7. It follows that $\mathbb{L}v = v$ has a unique bounded continuous solution, which is clearly v^p.

Note that $\mathbb{L}v$ is increasing with respect to x for every function v, thus $v^p = \mathbb{L}v^p$ is increasing with respect to x.

It remains to show that v^p is concave with respect to x. We first show that for any function $v(x,y)$ that is concave in x, the function $\mathbb{L}v$ is also concave in x. Fix arbitrarily x_1, x_2 and $0 \leq \theta \leq 1$, and let $x = \theta x_1 + (1 - \theta_2)x_2$. Let (c_i, z_i) be arbitrary but

$$c_i + p(y) \cdot z_i \leq y \cdot x_i + p(y) \cdot x_i,$$

and let

$$c \doteq \theta c_1 + (1-\theta)c_2, \qquad z \doteq \theta z_1 + (1-\theta)z_2.$$

Then

$$c + p(y) \cdot z \leq y \cdot x + p(y) \cdot x.$$

It follows that

$$\begin{aligned} \mathbb{L}v(x,y) &\geq U(c) + \beta \int v(z, y_1) P(dy_1|y) \\ &\geq \theta U(c_1) + \theta \int v(z_1, y_1) P(dy_1|y) \\ &\qquad + (1-\theta)U(c_2) + (1-\theta) \int v(z_2, y_1) P(dy_1|y). \end{aligned}$$

Taking supremum at the RHS over (c_i, z_i), we have

$$\mathbb{L}v(x,y) \geq \theta\, \mathbb{L}\, v(x_1, y) + (1-\theta)\, \mathbb{L}\, v(x_2, y).$$

This completes the proof. ■.

The next lemma is concerned with the derivatives of v^p in x.

Lemma 5 *For every continuous, non-negative, price function $p : S \to \mathbb{R}^d$, the value function v^p is differentiable with respect to x, and its derivative is*

$$\frac{\partial v^p}{\partial x^i}(x,y) = U'(c^*) \left[z^{*,i} + p^i(y) \right],$$

where c^* and $z^* = (z^{*,1}, z^{*,2}, \ldots, z^{*,d})$ is the maximizer for the DPE at (x,y), provided c^* is strictly positive.

Proof. Fix y. Define for every $\alpha \geq 0$,

$$f(\alpha) \doteq \sup_{(c,z)} \left[U(c) + \beta \int v^p(z, y_1) P(dy_1|y) \right]$$

where the supremum is taken over

$$c + p(y) \cdot z \leq \alpha, \quad c \geq 0, \quad 0 \leq z \leq 1.$$

Then $v^p(x, y)$ equals the value of f at $\alpha = y \cdot x + p(y) \cdot x$, and the function f is concave in α. The supremum will be attained since (c, z) is in a compact set, say at c^* and z^*. Because U is strictly concave in c and $v^p(z, y)$ is concave in z, it is not difficult to see that c^* is uniquely determined.

Assume now that $c^* > 0$ (thus $\alpha > 0$). Let ε be an arbitrarily small positive number. Then $(c^* + \varepsilon, z^*)$ is feasible at $\alpha + \varepsilon$, which implies that

$$\begin{aligned} f(\alpha + \varepsilon) &\geq U(c^* + \varepsilon) + \beta \int v^p(z^*, y_1) P(dy_1|y) \\ &= f(\alpha) - U(c^*) + U(c^* + \varepsilon). \end{aligned}$$

It follows that

$$D^+ f(\alpha) \doteq \lim_{\varepsilon \downarrow 0} \frac{f(\alpha + \varepsilon) - f(\alpha)}{\varepsilon} \geq \lim_{\varepsilon \downarrow 0} \frac{U(c^* + \varepsilon) - U(c^*)}{\varepsilon} = U'(c^*).$$

Similarly, $(c^* - \varepsilon, z^*)$ is feasible at $\alpha - \varepsilon$, which implies that

$$\begin{aligned} f(\alpha - \varepsilon) &\geq U(c^* - \varepsilon) + \beta \int v^p(z^*, y_1) P(dy_1|y) \\ &= f(\alpha) - U(c^*) + U(c^* - \varepsilon). \end{aligned}$$

Thus

$$D^- f(\alpha) \doteq \lim_{\varepsilon \downarrow 0} \frac{f(\alpha) - f(\alpha - \varepsilon)}{\varepsilon} \leq \lim_{\varepsilon \downarrow 0} \frac{U(c^*) - U(c^* - \varepsilon)}{\varepsilon} = U'(c^*).$$

But concavity implies that $D^+ f(\alpha) \leq D^- f(\alpha)$. Hence f is differentiable at α with derivative $U'(c^*)$. The rest is just chain rule. ■

Now we come back and work on the equilibrium state. Suppose the equilibrium price function is $\bar{p}(y)$. For $v^{\bar{p}}(x, y)$, denote the maximizer for the DPE by $(c^*, z^*) = (1 \cdot y, 1)$. It is easy to see that

$$U'(c^*) \bar{p}^i(y) = \beta \int \frac{\partial v^{\bar{p}}}{\partial z^i}(z^*, y_1) P(dy_1|y).$$

However, for $v^{\bar{p}}(z^*, y_1)$ the maximum of DPE is attained $c = 1 \cdot y_1$ which is strictly positive with probability 1, thanks to Condition 2. Thus Lemma 5 implies that

$$\frac{\partial v^{\bar{p}}}{\partial z^i}(z^*, y_1) = U'(1 \cdot y_1) \left[y_1^i + \bar{p}^i(y_1) \right].$$

It follows readily that

$$U'(1 \cdot y) p^i(y) = \beta \int U'(1 \cdot y_1) \left[y_1^i + \bar{p}^i(y_1) \right] P(dy_1|y).$$

Let

$$g = (g^1, g^2, \ldots, g^d), \quad g^i(y) \doteq \beta \int U'(1 \cdot y_1) y_1^i P(dy_1|y),$$

and

$$f \doteq (f^1, f^2, \ldots, f^d), \quad f^i(y) \doteq U'(1 \cdot y) \bar{p}^i(y)$$

Then the equation becomes

$$f(y) = g(y) + \beta \int f(y_1) P(dy_1|y). \tag{7}$$

Note g only depends on U and the transition probability of $\{Y_n\}$.

Theorem 3 *There exists a unique non-negative, bounded continuous function f satisfies equation (7), and the equilibrium price function is given by*

$$\bar{p}(y) = \frac{f(y)}{U'(1 \cdot y)}.$$

Proof. It can be easily shown that

$$f(y) \mapsto g(y) + \beta \int f(y_1) P(dy_1|y)$$

defines a contraction from the space of non-negative bounded continuous function to itself. We complete the proof. ■

We will give a simple example with one asset ($d = 1$), assuming that $\{Y_n\}$ is an iid sequence with common distribution μ. Then function g is a constant

$$g(y) \equiv \bar{g} = \int U'(z) \mu(dz),$$

and easily the solution to (7) is

$$f(y) \equiv \bar{f} = \frac{\bar{g}}{1 - \beta}.$$

The equilibrium price function is

$$\bar{p}(y) = \frac{\bar{g}}{1 - \beta} \frac{1}{U'(y)}.$$

4.5 Computational issues

The dynamics programming idea already gives an iterative method for computing the infinite horizon value function; see Theorem 1. This successive approximation is sometimes called *value iteration*, and its convergence rate is β^n since $\mathbb{L}$ is a contraction with modulus β.

In this section, we will discuss another approximation method, sometimes called *policy iteration*. At each step, the algorithm generates a new (stationary) control policy whose associated cost improves over the preceding one.

Definition 2 *A control policy $\{u_n\}$ is said to be (pure) stationary if there exists a function ϕ such that*

$$u_n = \phi(X_n).$$

When no confusion is incurred, we denote by $J(x;\phi)$ the cost associated with a stationary cost.

The algorithm roughly go as follows. Given a stationary policy ϕ^i, then an improved policy is

$$\phi^{i+1}(x) \doteq \operatorname{argmin}\left[c(x;u) + \beta \int J(h(x,u;\xi);\phi^i)\mu(d\xi)\right].$$

In other words, given ϕ^i, the algorithm compute the value $J(x;\phi^i)$, and ϕ^{i+1} is the optimizer for the RHS of the DPE save that the value function is replaced by its approximation $J(x;\phi^i)$.

Proposition 2 *Given any initial policy ϕ^0, the policy iteration yields a sequence of controls $\{\phi^i\}$ such that*

$$J(x;\phi^{i+1}) \le J(x;\phi^i), \qquad \forall\ x \in S.$$

The equality holds for all x if and only if ϕ^i is optimal. Furthermore, $J(x;\phi^i) \to v(x)$ as $i \to \infty$.

Proof. It is not hard to see that, for any stationary control policy ϕ, the corresponding cost $J(x;\phi)$ is the unique bounded solution to the equation

$$U = \mathbb{L}_\phi U, \qquad \mathbb{L}_\phi\, U(x) \doteq c(x;\phi(x)) + \beta \int U(h(x,\phi(x);\xi))\mu(d\xi).$$

Furthermore, given any bounded function U, we have

$$J(x;\phi) = \lim_n \mathbb{L}_\phi^n U(x).$$

To ease notation, let $J_i \doteq J(x;\phi^i)$. By construction, we have

$$J_i = \mathbb{L}_{\phi^i} J_i \geq \mathbb{L}_{\phi^{i+1}} J_i.$$

This, combined with the monotonicity of $\mathbb{L}_{\phi^{i+1}}$, yields

$$J_i \geq \mathbb{L}_{\phi^{i+1}} J_i \geq \cdots \geq \mathbb{L}_{\phi^{i+1}}^n J_i.$$

Letting $n \to \infty$, we have $J_i \geq J_{i+1}$.

Now assume that $J_i = J_{i+1}$, then we have

$$J_i = \mathbb{L}_{\phi^i} J_i \geq \mathbb{L}_{\phi^{i+1}} J_i \geq \mathbb{L}_{\phi^{i+1}} J_{i+1} = J_{i+1} = J_i.$$

Thus

$$J_i = \mathbb{L}_{\phi^{i+1}} J_i = \mathbb{L} J_i.$$

or $J_i = v$ and ϕ^i is optimal. On the other hand, if ϕ^i is optimal, $J_i = v \geq J_{i+1} \geq v$, thus $J_{i+1} = v = J_i$.

It remains to show that $J_i \downarrow v$. Clearly $J_i \geq v$. Define

$$A_i \doteq \sup_{x \in S} \left[J_i(x) - v(x) \right].$$

Since $\mathbb{L}$ is a contraction with modulus β, we have, for every $x \in S$,

$$\sup_{x \in S} \left[\mathbb{L} J_i(x) - \mathbb{L} v(x) \right] = \sup_{x \in S} \left[\mathbb{L} J_i(x) - v(x) \right] \leq \beta A_i.$$

But

$$\mathbb{L} J_i = \mathbb{L}_{\phi^{i+1}} J_i \geq \mathbb{L}_{\phi^{i+1}} J_{i+1} = J_{i+1}.$$

It follows that

$$\sup_{x \in S} \left[J_{i+1}(x) - v(x) \right] \leq \beta A_i.$$

This implies that $J_i \downarrow v$. We complete the proof. ■

Corollary 1 *Suppose the state space S is a finite set and for each $x \in S$, the control set $U(x)$ is finite. Given any initial condition, the policy iteration algorithm yields a stationary optimal policy after a finite number of iterations.*

5 Infinite-horizon: The monotonicity assumption

In many control problems, the boundedness assumption (Condition 1) is violated; e.g., the linear-quadratic case. The analysis for the unbounded cost is more sophisticated than that for the bounded cost.

We will consider the case where the cost is either non-negative or non-positive. When the cost is non-positive, the control problem is equivalent to a maximizing problem with non-negative running cost.

5.1 Infinite-horizon: A maximizing problem

Given the Markov control process (1), the control problem under consideration is

$$U(x) \doteq \sup_{\{u_n\}} J(x; \{u_n\}) \doteq \sup_{\{u_n\}} E_x \left[\sum_{j=0}^{\infty} \beta^j c(X_j, u_j) \right]$$

for some discount factor $\beta \in (0, 1)$. We have the following assumption.

Condition **3** *The running cost $c(x, u)$ is non-negative.*

The value function U is clearly non-negative. It is possible, however, that U takes value $+\infty$. The associated DPE is

$$W \equiv \mathbb{L}W, \tag{8}$$

where

$$\mathbb{L}W(x) = \max_{u \in U(x)} \left[c(x; u) + \int W(h(x, u; \xi)) \mu(d\xi) \right].$$

Again, the corresponding finite horizon value functions are denoted by $\{v_n\}$, which satisfy the forward DPE

$$\begin{aligned} v_0(x) &= 0 \\ v_{n+1}(x) &= \mathbb{L}v_n(x), \qquad n = 0, 1, \ldots \end{aligned}$$

A similar result to Theorem 1 holds here.

Theorem 4 *Under Condition 3, we have*

1. *The value function U is the (pointwise) smallest, nonnegative solution to the DPE (8).*
2. *The DP algorithms converges. That is $v_n(x) \to U(x)$ for every $x \in S$.*

3. *A stationary policy* $\{u^* \doteq u^*(x)\}$ *is optimal if and only if* $J(x;\{u^*\})$ *is a solution to the DPE (8). Or, if* u^* *is a maximizer of the RHS of the DPE (8) with* U *in place of* W*, and*

$$\lim_n \beta^n E_x\left[U(X_n^*)\right] = 0$$

then u^* *is optimal. Here* $\{X_n^*\}$ *is the MCP corresponding to* $\{u^*\}$*:*

$$\begin{aligned} X_0^* &= X_0 \\ X_{n+1}^* &= h(X_n^*, u^*(X_n^*); \xi_{n+1}), \qquad n = 0, 1, \ldots. \end{aligned}$$

Before the proof, we will first give the following example.

Example 5 Let $S = [0, \infty)$, the discount factor $\beta \in (0, 1)$, and the dynamics

$$X_{n+1} = X_n/\sqrt{\beta},$$

the control space $U(x) \equiv \{0\}$, with cost

$$c(x; u) \doteq (1 - \sqrt{\beta})x.$$

The DPE is

$$W(x) = (1 - \sqrt{\beta})x + \beta W(x/\sqrt{\beta}).$$

The value function $U(x) \equiv x$ is a solution to the DPE. There are other solutions to the DPE,

$$W_1(x) = x + x^2, \qquad W_2(x) = x - x^2.$$

But $W_1(x) \geq U(x)$, and $W_2(x)$ is not non-negative for all x.

Example 6 This example shows that a control policy u^* maximizing the DPE with U in place of W alone is not sufficient for optimality. Let $S = [0, \infty)$ and $U(x) = [0, \infty)$ for every $x \in S$. The dynamics

$$X_{n+1} = u_n$$

with value function

$$U(x) = \sup_{\{u_n\}} \left[\sum_{j=0}^{\infty} \beta^j u_j\right].$$

The associated DPE is

$$W(x) = \sup_{u \geq 0} \left[u + \beta W(u)\right].$$

Clearly, the value function $U(x) = \infty$ for all x, and $\{u^* = u^*(x) \equiv 0\}$ is a maximizer, but with the corresponding cost 0. ■

Proof of Theorem 4: By definition $v_{n+1}(x) \geq v_n(x)$. Thus $v_n \uparrow U^*$ for some non-negative function v. By MCT, we have, for every $x \in S$, and $u \in U(x)$,

$$c(x;u) + \int v_n(h(x,u;\xi))\mu(d\xi) \;\uparrow\; c(x;u) + \int U^*(h(x,u;\xi))\mu(d\xi).$$

This implies that (why?)

$$U^* = \lim_n v_{n+1} = \lim_n \mathbb{L}v_n = \mathbb{L}U^*.$$

Thus U^* is a solution to the DPE (8).

Next we show that $U = \lim_n v_n = U^*$. Assume $U(x)$ is finite. The case for $U(x) = \infty$ is completely similar and thus omitted. Indeed, for any control process $\{u_n\}$ and $\varepsilon > 0$, there exists N such that

$$J(x;\{u_n\}) = E_x\left[\sum_{j=0}^{\infty} \beta^j c(X_j;u_j)\right] \leq E_x\left[\sum_{j=0}^{N-1} \beta^j c(X_j;u_j)\right] + \varepsilon,$$

thanks to MCT. Thus,

$$J(x;\{u_n\}) \leq v_N(x) + \varepsilon \leq U^*(x) + \varepsilon.$$

Taking supremum over $\{u_n\}$ and letting $\varepsilon \to 0$, one has $U(x) \leq U^*(x)$. The reverse inequality is trivial since $U(x) \geq v_n(x)$ by definition. It follows that $U(x) = U^*(x)$.

We now show that $U(x)$ is the smallest, non-negative solution to the DPE. Let W be an arbitrary non-negative solution. Then for any control $\{u_n\}$, consider the process

$$Z_n \doteq \sum_{j=0}^{n-1} \beta^j c(X_j;u_j) + \beta^n W(X_n).$$

It is not difficult to see that $\{Z_n\}$ forms a supermartingale, whence

$$W(x) = E[Z_0] \geq E[Z_n] \geq E\left[\sum_{j=0}^{n-1} \beta^j c(X_j;u_j)\right].$$

Letting $n \to \infty$, the MCT implies $W(x) \geq J(x;\{u_n\})$. It follows readily that $U(x) \leq W(x)$.

It remains to show that u^* defines an optimal control policy. The if and only if part is simple – if $\{u^*\}$ is optimal then $J(x;\{u^*\}) = U(x)$ is

a solution to the DPE, and if $J(x;\{u^*\})$ is a solution to the DPE then $J(x;\{u^*\}) \geq U(x)$, whence $J(x;\{u^*\}) = U(x)$ or $\{u^*\}$ is optimal.

Now assume u^* is the maximizer in the DPE with U in place of W. Define

$$Z_n^* \doteq \sum_{j=0}^{n-1} \beta^j c(X_j^*; u^*(X_j^*)) + \beta^n U(X_n^*).$$

Then $\{Z_n^*\}$ defines a martingale, and

$$U(x) = E[Z_0^*] = E[Z_n^*] = E_x \left[\sum_{j=0}^{n-1} \beta^j c(X_j^*; u^*(X_j^*)) + \beta^n U(X_n^*) \right].$$

Letting $n \to \infty$, by assumption and MCT, we have

$$U(x) = E_x \left[\sum_{j=0}^{\infty} \beta^j c(X_j^*; u^*(X_j^*)) \right] = J(x;\{u^*\}).$$

This completes the proof. ■

Corollary 2 *Let $\bar{U}$ be an arbitrary non-negative solution to the DPE (8), and $\{\bar{u} \doteq \bar{u}(x)\}$ the maximizer for this DPE with $\bar{U}$ in place for W. If*

$$\lim_n \beta^n E_x \left[\bar{U}(\bar{X}_n) \right] = 0,$$

then $\bar{U} = U$ and $\bar{u}$ is optimal. Here $\{\bar{X}_n\}$ is the MCP corresponding to $\{\bar{u}\}$.

In particular, if $\bar{U}$ is a non-negative, bounded solution to the DPE, then $\bar{U} = U$ and $\bar{u}$ is optimal.

Proof. The proof is just a simple verification argument analogous to the preceding theorem, and is omitted. ■

Example 7 *(pure credit economy).* Consider a household in an economy that wants to the following problem: At each time period $n = 0, 1, \ldots$, the economy has an endowment of non-storable consumption goods (which is indeed for both consumption and trading). The household has X_n shares of certain "claim" at time n. One unit of this claim will allow the household to obtain one unit of consumption good for each time period $\{n, n+1, \ldots\}$. This claim can also be traded with price Q in unit of consumption good. The dynamics for $\{X_n\}$ is then

$$X_{n+1} \doteq \frac{QX_n + X_n - c_n}{Q}; \qquad n = 0, 1, \ldots$$

where $\{c_n\}$ is the consumption for each period. The constraints are that $c_n, X_n \geq 0$ for each n. The optimization problem for this household is to find a consumption process $\{c_n\}$ so as to maximize the total expected utility

$$E\left[\sum_{j=0}^{\infty} \beta^j U(c_j)\theta_j\right],$$

where $\{\theta_n\}$ is an iid, non-negative, bounded, "taste shock" sequence with common distribution μ, and U is the power utility function

$$U(c) \doteq c^\alpha/\alpha, \qquad \text{for some } \alpha \in (0,1).$$

The taste shock θ_n is known to the household at the beginning of period n. The value function will be denoted by $v(x;\theta)$ given $X_0 = x$ and $\theta_0 = \theta$, and it satisfies the DPE

$$v(x,\theta) = \max_c \left[U(c)\theta + \beta \int v\left(x + Q^{-1}(x-c); z\right)\mu(dz)\right]$$

where the maximization is over c such that $0 \leq c \leq Qx + x$.

The DPE admit an explicit solution for every α (when β is appropriated bounded from 1). For illustration, we only consider $\alpha = 1/2$, and assume that $\beta < \sqrt{Q/(Q+1)}$. An explicit solution is of form

$$\bar{v}(x,\theta) = B(\theta)\sqrt{x},$$

with

$$B(\theta) \doteq \sqrt{Q+1} \cdot \sqrt{\beta^2 \bar{B}^2 + \theta^2}$$

where $\bar{B}$ is unique positive solution to the equation

$$\int \sqrt{\frac{Q+1}{Q}} \sqrt{\beta^2 + \frac{\theta^2}{\bar{B}^2}}\mu(d\theta) = 1.$$

The maximizer is

$$c^* = \frac{\theta^2}{\beta^2 \bar{B}^2 + \theta^2} \cdot (Q+1)x,$$

and the corresponding controlled process is

$$X^*_{n+1} = \frac{\beta^2 \bar{B}^2}{\beta^2 \bar{B}^2 + \theta_n^2} \cdot \frac{Q+1}{Q} X^*_n.$$

In order to show that $\bar{v} = v$ and that c^* is optimal, it suffices to show that

$$\lim_n \beta^n E\left[\bar{v}(X^*_n, \theta_n)\right] = 0$$

This is trivial following from $X^*_{n+1} \le (Q+1)/QX^*_n$.

An *equilibrium price* Q is such that the average of $\{X^*_n\}$ stays the same. In other words,

$$E\left[\frac{\beta^2\bar{B}^2}{\beta^2\bar{B}^2+\theta^2}\cdot\frac{Q+1}{Q}\right]=1$$

More precisely, does there exist a (unique) Q such that $\beta < \sqrt{Q/(Q+1)}$ (or $Q > \beta^2/[1-\beta^2]$) and the preceding equality holds. Note that $\bar{B}$ is a function depending on Q. This is true. As $Q \downarrow \beta^2/[1-\beta^2]$ one has $\bar{B} \uparrow \infty$, and by MCT, LHS $\uparrow \beta^{-2}$. However, as $Q \uparrow \infty$, clearly the LHS decreases to some constant less than 1.

Suppose now that the price Q is the equilibrium price, and define

$$h(\theta)=\frac{\beta^2\bar{B}^2}{\beta^2\bar{B}^2+\theta^2}\cdot\frac{Q+1}{Q}.$$

Thus $Eh(\theta)=1$. The optimal process $\{X^*_n\}$ can then be written as

$$X^*_{n+1}=h(\theta_n)X^*_n \qquad \Leftrightarrow \qquad \log X^*_{n+1}=\log X^*_n-\log h(\theta_n).$$

If the distribution of $\{\theta_n\}$ is not degenerate, then Jensen inequality implies that

$$E[\log h(\theta_n)]<0, \qquad \forall\ n.$$

In other words, the process $\{\log X^*_n\}$ is a simple random walk with negative drift, and it follows that

$$X^*_n \to 0$$

with probability 1.

Now consider a large collection of such households, each starts with the same share of claims, say one unit, and each solves its own optimization problem as above. Then for each household, the number of shares will converge to zero. However, the average number of shares across households will stay at one unit. This says that "wealth concentrates in an ever shrinking number of ever wealthier households".

5.2 Infinite-horizon: A minimizing problem

Given the Markov control process (1), the control problem under consideration is

$$v(x) \doteq \inf_{\{u_n\}} J(x;\{u_n\}) \doteq \inf_{\{u_n\}} E_x\left[\sum_{j=0}^{\infty}\beta^j c(X_j,u_j)\right]$$

for some discount factor $\beta \in (0,1)$. The DPE for this control problem is

$$W = \mathbb{L}W \tag{9}$$

where

$$\mathbb{L}W(x) \doteq \inf_{u \in U(x)} \left[c(x;u) + \beta \int W(h(x,u;\xi))\mu(d\xi) \right].$$

Condition **4** *The running cost $c(x,u)$ is non-negative, and $v(x)$ is finite for every $x \in S$.*

The following result is immediate.

Proposition 3 *Suppose Condition 4 holds. If $\bar{v}$ is a non-negative, bounded solution to the DPE (9), and $\{u^* \doteq u^*(x)\}$ the minimizer for this DPE with $\bar{v}$ in place for W. then $\bar{v} = v$ and u^* is optimal.*

Proof. The proof is just a straightforward verification argument, and is thus omitted. ■

The general theory for this type of optimization problem is quite involved. We will need more assumptions.

Condition **5**

1. *The set $U(x)$ is compact for every $x \in S$, and the correspondence $x \mapsto U(x)$ is upper-semicontinuous.*
2. *The running cost $c(\cdot,\cdot)$ is lower-semicontinuous.*
3. *The MCP satisfies the Feller property; i.e., for every bounded continuous function W on S, the function*

$$\bar{W}(x,u) \doteq \int W(h(x,u;\xi))\mu(d\xi)$$

is bounded continuous.

This assumption is not necessary to establish the desired result, but is convenient to work with.

Again, we will denote by $\{v_n\}$ the corresponding finite horizon value function. They satisfies the forward form DPE:

$$\begin{aligned} v_0(x) &= 0 \\ v_{n+1}(x) &= \mathbb{L}v_n(x), \quad n = 0,1,\ldots \end{aligned}$$

We have the following result.

Theorem 5 *Assume Condition 4 and Condition 5 holds.*

1. *The value function v is the (pointwise) smallest, nonnegative, lower-semicontinuous solution to the DPE (9).*
2. *The DP algorithms converges. That is $v_n(x) \to v(x)$ for every $x \in S$. Furthermore $\{v_n\}$ are all lower-semicontinuous.*
3. *There exists a stationary policy $\{u^* \doteq u^*(x)\}$ minimizing the RHS of the DPE (9) with v in place of W. This stationary policy u^* is optimal.*

Proof. Clearly by definition, we have $0 \le v_n \le v_{n+1} \le v$. Assume $v_n \uparrow v^*$, whence $0 \le v^* \le v$.

We now show that $\{v_n\}$ are all lower-semicontinuous by induction, which in turn implies that v^* is lower-semicontinuous. The claim is true for $n = 0$. Suppose now v_n is lower-semicontinuous, then

$$v_{n+1}(x) = \inf_{u \in U(x)} \left[c(x,u) + \beta \int v_n(h(x,u;\xi))\mu(d\xi) \right].$$

The function v_n can be approximated from below by a non-decreasing sequence of non-negative bounded continuous functions; see Lemma 8. It is easy to see then the function

$$\beta \int v_n(h(x,u;\xi))\mu(d\xi),$$

as the limit of a non-decreasing sequence of continuous functions (thanks to Feller property), is lower-semicontinuous. By Corollary 3, the infimum is achieved at some $u^* \doteq u^*(x)$ and the infimum v_{n+1} is lower-semicontiuous. Taking limit as $n \to \infty$, by MCT,

$$c(x,u) + \beta \int v_n(h(x,u;\xi))\mu(d\xi) \ \uparrow \ c(x,u) + \beta \int v^*(h(x,u;\xi))\mu(d\xi)$$

for any $u \in U(x)$. By Lemma 7, the infimum of the LHS also converges to the infimum of the RHS, or

$$v^* = \lim_n v_{n+1} = \mathbb{L}v^*.$$

Thus v^* solves the DPE.

To argue that $v^* = v$ it suffices to show that $v^* \ge v$. Abusing the notation a bit, let $\{u^* = u^*(x)\}$ be the minimizer for the DPE with v^* in place for W. Note u^* always exists. Consider the process

$$Z_n^* \doteq \sum_{j=0}^{n-1} \beta^j c(X_j^*, u_j^*) + \beta^n v^*(X_n^*).$$

It is not difficult to show that $\{Z_n^*\}$ is indeed a martingale, and thus

$$v^*(x) = E[Z_0^*] = E[Z_n^*] \geq E\left[\sum_{j=0}^{n-1} \beta^j c(X_j^*, u_j^*)\right].$$

Letting $n \to \infty$, we have

$$v^*(x) \geq J(x; \{u^*\}) \geq v(x).$$

Thus $v^*(x) = v(x)$ and u^* is optimal. The same argument can be used to show that v is the smallest non-negative, lower-semicontinuous solution to the DPE. We complete the proof. ■

Example 8 *(Production and inventory accumulation)* In markets for many agricultural commodities, inventories play an important role in smoothing the stochastic shocks to supply that result from weather fluctuations. One model is as follows.

Suppose the demand is constant over time, and the market-clearing price when the supply is s is denoted by $D(s)$. The function $D : \mathbb{R}_+ \to \mathbb{R}_+$ is assumed to be continuous, strictly decreasing and $\lim_{s\to\infty} D(s) = 0$. The *consumers' surplus* is defined as

$$U(x) \doteq \int_0^x D(s)\, ds.$$

Clearly U is a non-negative, strictly increasing, strictly concave function. Assume $\lim_{x\to\infty} U(x) < \infty$; i.e., U is bounded.

Let X_n denote the stock of goods at the beginning of period n, which is the stock carried from the last period plus the current harvest). The planner must decide the consumption $0 \leq c_n \leq X_n$ for this period and the input $z_n \geq 0$ for production. The size of z_n will yield a harvest of size $z_n\xi_{n+1}$. Thus the dynamics is

$$X_{n+1} = X_n - c_n + z_n\xi_{n+1}.$$

The holding cost is given by function ϕ and the cost for production is given by function f. Both functions are assumed to be non-negative, strictly convex, strictly increasing, continuously differentiable, and $\phi(0) = \phi'(0) = f(0) = f'(0) = 0$.

Assuming that $\{\xi_n\}$ is a sequence of iid, non-negative, bounded random variables with common distribution μ, the goal for the planner is to solve the optimization problem

$$v(x) = \sup_{\{c_n, z_n\}} E_x\left\{\sum_{n=0}^{\infty} \beta^n \left[U(c_n) - \phi(X_n - c_n) - f(z_n)\right]\right\}.$$

Since each summand is bounded from above, it can be regarded as an minimization problem with cost bounded from below. However, the control set is not compact because z_n can take any non-negative value.

For $M \geq 0$, let $v_M(x)$ be the value function for the same optimization problem save that an additional constraint $0 \leq z_n \leq M$ is imposed. Note that v and v_M are all non-negative (taking $c_n = z_n = 0$) and bounded from above by constant $\lim_{x\to\infty} U(x)/(1-\beta)$.

Theorem 5 implies that v_M is a solution to the DPE

$$v_M(x) = \sup_{\{0\leq c\leq x, 0\leq z\leq M\}} \left[U(c) - \phi(x-c) - f(z) + \beta \int v_M(x-c+z\xi)\mu(d\xi)\right].$$

Moreover, it is not difficult to see that v_M is strictly increasing and strictly concave (whence continuous). For M large enough, it follows that v_M satisfies the DPE of the original control problem; i.e.,

$$v_M(x) = \sup_{\{0\leq c\leq x, 0\leq z\}} \left[U(c) - \phi(x-c) - f(z) + \beta \int v_M(x-c+z\xi)\mu(d\xi)\right].$$

But from this equation, the boundedness of v_M and Proposition 3, we have $v_M \equiv v$. Thus v is the (unique) bounded continuous solution satisfying the DPE

$$v(x) = \sup_{\{0\leq c\leq x, 0\leq z\}} \left[U(c) - \phi(x-c) - f(z) + \beta \int v(x-c+z\xi)\mu(d\xi)\right],$$

and v is strictly increasing and strictly concave.

Many properties of the optimality can be derived from this DPE. One may easily check that for each x, there exists a unique $c^*(x)$ and a unique $z^*(x)$ that attains the maximum over the RHS, that c^* and z^* are both continuous functions, and that v is continuously differentiable with derivative $v'(x) = U'(c^*(x)) = D(c^*(x))$. This yields that $c^*(x)$ is strictly increasing.

With some effort, it can also be shown that $x - c^*(x)$ is non-decreasing with respect to x and $z^*(x)$ is strictly positive and strictly decreasing. Indeed, it is not difficult to see that

$$U'(c^*(x)) + \phi'[x - c^*(x)] \geq \beta \int v'[x - c^*(x) + z^*(x)\xi]\mu(d\xi)$$

with equality if $c^*(x) < x$, and

$$f'(z^*(x)) \geq \beta \int \xi v'[x - c^*(x) + z^*(x)\xi]\mu(d\xi)$$

with equality if $z^*(x) > 0$. However, this inequality automatically yields that $z^*(x)$ is strictly positive since $f'(0) = 0$. Thus

$$f'(z^*(x)) = \beta \int \xi v'[x - c^*(x) + z^*(x)\xi]\mu(d\xi).$$

Now let $0 < x < y$. We want to show $x - c^*(x) \leq y - c^*(y)$. Assume otherwise. Then $x - c^*(x) > 0$ and it follows that

$$\int v'[x - c^*(x) + z^*(x)\xi]\mu(d\xi) > \int v'[y - c^*(y) + z^*(y)\xi]\mu(d\xi).$$

But this implies that $z^*(x) < z^*(y)$. But this implies that $z^*(y) > 0$ and

$$\int \xi v'[y - c^*(y) + z^*(y)\xi]\mu(d\xi) > \int \xi v'[x - c^*(x) + z^*(x)\xi]\mu(d\xi).$$

Let

$$Z \doteq v'[x - c^*(x) + z^*(x)\xi] - v'[y - c^*(y) + z^*(y)\xi].$$

Then the above two inequalities are equivalent to

$$E[Z] > 0, \qquad E[Z\xi] < 0.$$

Define

$$A \doteq \frac{[x - c^*(x)] - [y - c^*(y)]}{z^*(y) - z^*(x)} > 0.$$

Then $Z > (=, <)0$ if and only if $\xi > (=, <)A$. But this implies that

$$E[Z^+\xi] \geq AE[Z^+] > AE[Z^-] \geq E[Z^-\xi],$$

or $E[Z\xi] > 0$, a contradiction. Thus $x - c^*(x) \leq y - c^*(y)$ as $x < y$. The fact that $z^*(x) > z^*(y)$ is then trivial by contradiction. ■

A [Appendix] A brief review of contraction mapping

We give a brief review of some of the basics of functional analysis in this section. The contraction mapping yields the most elementary fixed point theorem [3].

Definition 3 A *metric space* is a pair (X, ρ) where X is a set and the metric ρ is a mapping from $X \times X$ to $\mathbb{R}_+$. The metric ρ satisfies

1. $\rho(x, y) \geq 0$ for all $x, y \in X$, and the equality holds if and only if $x = y$.
2. $\rho(x, y) = \rho(y, x)$.
3. $\rho(x, y) \leq \rho(x, z) + \rho(z, y)$ for all $x, y, z \in X$.

Definition 4 A sequence $\{x_n\} \in X$ is said to be a *Cauchy sequence* if for any $\varepsilon > 0$ there exists an $N \in \mathbb{N}$ such that $\rho(x_n, x_m) \leq \varepsilon$ for every $n, m \geq N$. The metric space (X, ρ) is said to be *complete* if each Cauchy sequence converges to a point in X.

Example 9 The following metric spaces are complete.

1. Let S be a Borel set in $\mathbb{R}^d$. Let $X = C_b(S)$ be the set of all bounded continuous functions on S. Then $C_b(S)$, equipped with the metric induced by sup norm, is a complete metric space. The metric induced from sup norm is, for every $f, g \in C_b(S)$,
$$\rho(f, g) \doteq \sup_{x \in S} \|f(x) - g(x)\|.$$
2. Let S be a Borel set in $\mathbb{R}^d$. Let $X = M_b(S)$ be the set of all bounded (measurable) functions on S. Then $M_b(S)$, equipped with the metric induced by sup norm, is a complete metric space. ■

Definition 5 Let (X, ρ) be a metric space. A mapping $\mathbb{L} : X \to X$ is said to be a *contraction* (with *modulus* α) if there exists a real number $\alpha \in (0, 1)$ such that
$$\rho(\mathbb{L}x, \mathbb{L}y) \leq \alpha \cdot \rho(x, y),$$
for all $x, y \in X$.

The following result is elementary yet very powerful.

Theorem 6 *Let (X, ρ) be a complete metric space, and $\mathbb{L} : X \to X$ a contraction. Then there exists a unique fixed point $x^* \in X$ satisfying $\mathbb{L}x^* = x^*$. Furthermore, for any $x \in X$, $\mathbb{L}^n x \to x^*$ as n tends to infinity. Here $\mathbb{L}^n x \doteq \mathbb{L}(\mathbb{L}(\cdots \mathbb{L}(x)))$ for any $x \in X$.*

Proof. Consider an arbitrary $x \in X$, and define $x_n \doteq \mathbb{L}^n x$ for $n = 1, 2, \ldots$. It is not difficult to see that

$$\rho(x_n, x_{n+1}) \leq \alpha \cdot \rho(x_{n-1}, x_n) \leq \cdots \leq \alpha^n \cdot \rho(x, x_1).$$

This, together with the triangle inequality, yield that $\{x_n\}$ is a Cauchy sequence. Since X is complete, there exists an $x^* \in X$ such that $x_n \to x^*$. But then we must have $x_{n+1} = \mathbb{L}x_n \to \mathbb{L}x^*$ since $\mathbb{L}$ is necessarily continuous. This implies $x^* = \mathbb{L}x^*$. Furthermore,

$$\rho(\mathbb{L}^n x, x^*) = \rho(\mathbb{L}^n x, \mathbb{L}^n x^*) \leq \alpha^n \cdot \rho(x, x^*),$$

for every $n \in \mathbb{N}$. Thus $\mathbb{L}^n x \to x^*$. As for the uniqueness, suppose $\bar{x}$ is another fixed point. Then

$$\rho(\bar{x}, x^*) = \rho(\mathbb{L}\bar{x}, \mathbb{L}x^*) \leq \alpha \cdot \rho(\bar{x}, x^*),$$

or $\rho(\bar{x}, x^*) = 0$. This completes the proof. ■

Theorem 7 *(Blackwell's sufficient condition for a contraction). Let $S \in \mathbb{R}^d$ be a Borel set, and $X = C_b(S)$ or $M_b(S)$ with the metric induced by sup norm. Let $\mathbb{L} : X \to X$ be a mapping satisfying*

1. *$\mathbb{L}f \leq \mathbb{L}g$ for all $f, g \in X$ such that $f \leq g$.*
2. *there exists a constant $\alpha \in (0, 1)$ such that*

$$\mathbb{L}(f + c) \leq \mathbb{L}f + \alpha c$$

for all $f \in X$, and constants $c \geq 0$.

Then $\mathbb{L}$ is a contraction with modulus α.

Proof. For any $f, g \in X$, we have $f \leq g + \rho(f, g)$. Then we have $\mathbb{L}f \leq \mathbb{L}g + \alpha \cdot \rho(f, g)$. Reversing the roles of f and g, we have $\mathbb{L}g \leq \mathbb{L}f + \alpha \cdot \rho(f, g)$. Thus

$$\rho(\mathbb{L}f, \mathbb{L}g) \leq \alpha \cdot \rho(f, g).$$

We complete the proof. ■

B [Appendix] Miscellaneous results

Lemma 6 *Assume Condition 2 holds. Then for any bounded continuous function $f(x,y)$, the function*

$$(\mathbb{T}f)(x,y) \doteq \int f(x,z)P(dz|x), \quad \forall\ x,y$$

is also bounded and continuous.

Proof. The boundedness of $\mathbb{T}f$ is trivial. Let (x_n, y_n) be a sequence such that $(x_n, y_n) \to (x, y)$. Then

$$\begin{aligned} &|(\mathbb{T}f)(x,y) - (\mathbb{T}f)(x_n,y_n)| \\ &\quad \leq \ |(\mathbb{T}f)(x,y) - (\mathbb{T}f)(x,y_n)| + |(\mathbb{T}f)(x,y_n) - (\mathbb{T}f)(x_n,y_n)|\,. \end{aligned}$$

The Feller property implies that the first term converges to zero as $n \to \infty$. As for the second term, since $x_n \to x$, then there exists a compact set, say Q, such that $x \in Q$ and $x_n \in Q$ for sufficiently large n. The product $Q \times S$ is also a compact set, and thus f is uniformly continuous on this set. But

$$|(\mathbb{T}f)(x,y_n) - (\mathbb{T}f)(x_n,y_n)| \leq \int |f(x,y_n) - f(x_n,y_n)|\, P(dz|y_n).$$

Thanks to uniform continuity, this term converges to zero. ■

The next result is concerned with the continuity of the maximum (or minimum) of a collection of continuous functions. Abusing the notation a bit, let X and Y be two Borel sets in some metric spaces, and $f : X \times Y \to \mathbb{R}$ a continuous function. For every $x \in X$, assign a non-empty set $\Gamma(x) \subset Y$. We are interested in the function

$$h(x) \doteq \sup_{y \in \Gamma(x)} f(x,y).$$

We need the following definitions.

Definition 6

1. *We say Γ is* compact-valued *if for every x, the set $\Gamma(x)$ is compact.*
2. *We say Γ is* lower semi-continuous at x, *if for any $x \in X, y \in \Gamma(x)$, and every sequence $\{x_n\} \subset X$ such that $x_n \to x$, there exists $y_n \in \Gamma(x_n)$ such that $y_n \to y$.*

3. *We say* Γ *is* upper semi-continuous at x, *if for any* $x \in X, \{x_n\} \subset X$ *such that* $x_n \to x$, *and every sequence* $\{y_n\}$ *such that* $y_n \in \Gamma(x_n)$, *there exists a convergent subsequence of* $\{y_n\}$ *whose limit point belongs to* $\Gamma(x)$.

4. *We say* Γ *is* continuous at x, *if it is both lower semi-continuous and upper semi-continuous at* x.

We can state now the *Theorem of Maximum*.

Theorem 8 *Let* $f : X \times Y \to \mathbb{R}$ *be a continuous function. Assume that* Γ *is compact-valued and continuous, then the function*

$$h(x) \doteq \sup_{y \in \Gamma(x)} f(x, y)$$

is continuous. Furthermore, the set-valued function

$$L(x) \doteq \{y \in \Gamma(x) : f(x, y) = h(x)\}$$

is non-empty, compact-valued, and upper semi-continuous.

Proof. Since f is continuous and Γ is compact valued, the supremum is attained at every x, or $L(x)$ is non-empty for every x. The compactness is trivial since f is continuous, thus $L(x)$ is a closed subset of the compact set $\Gamma(x)$.

Now let $\{x_n\} \subset X$ be an arbitrary sequence such that $x_n \to x \in X$. Given any sequence $\{y_n\}$ such that $y_n \in L(x_n) \subset \Gamma(x_n)$, the upper semicontinuity of Γ implies that there exists a subsequence, still denoted by $\{y_n\}$, such that $y_n \to y \in \Gamma(x)$. We want to show $y \in L(x)$ (thus finish the upper semicontinuity of L), or equivalently, for any $z \in L(x)$, $f(x, y) \geq f(x, z)$. Indeed, due to the lower semicontinuity of Γ, there exists a sequence of $\{z_n\}$ with $z_n \in \Gamma(x_n)$, such that $z_n \to z$. But $f(x_n, y_n) \geq f(x_n, z_n)$. Letting $n \to \infty$, we arrive at $f(x, y) \geq f(x, z)$.

It remains to show that h is continuous. Let $x_n \to x$ be an arbitrary sequence. Without loss of generality, we assume $\lim h(x_n)$ exists. We want to show $\lim h(x_n) = h(x)$. Let $y_n \in L(x_n)$. Since L is upper semi-continuous, there exists a subsequence, still denoted by $\{y_n\}$, such that $y_n \to y \in L(x)$. Thus

$$\lim h(x_n) = f(x_n, y_n) \to f(x, y) = h(x).$$

This completes the proof. ■

Corollary 3 *Let $f : X \times Y \to \mathbb{R}$ be a lower-semicontinuous function. Assume that Γ is compact-valued and upper-semicontinuous, then the function*

$$h(x) \doteq \inf_{y \in \Gamma(x)} f(x, y)$$

is lower-semicontinuous.

Proof. For each x, the infimum is attained since f is lower-semicontinuous and Γ is compact-valued. Let $x_n \to x$ be an arbitrary sequence, and let $y_n \in \Gamma(x_n)$ be such that $h(x_n) = f(x_n, y_n)$. Due to upper-semicontinuity, there exists a subsequence, still indexed by n, such that $y_n \to y \in \Gamma(x)$. We have

$$\liminf h(x_n) = \liminf f(x_n, y_n) \geq f(x, y) \geq h(x).$$

This completes the proof. ■

Lemma 7 *Suppose $\{F_n : X \to \mathbb{R}\}$ is a sequence of non-negative, lower-semicontinuous functions, and $F_n(x) \uparrow F(x)$. Then for any compact set $\Gamma \subset X$, we have*

$$\lim_n \inf_{x \in \Gamma} F_n(x) = \inf_{x \in \Gamma} F(x).$$

Proof. Suppose $x_n \in \Gamma$ is the minimizer for $F_n(x)$ over Γ. Such x_n always exists, thanks to the lower-semicontinuity of F_n and the compactness of Γ. Without loss of generality, assume that $x_n \to x^* \in \Gamma$. It follows from assumption that

$$F_n(x_n) \geq F_m(x_n), \qquad \forall\ n \geq m,$$

and thus

$$\lim_n \inf_{x \in \Gamma} F_n(x) = \lim_n F_n(x_n) \geq \liminf_n F_m(x_n) \geq F_m(x^*).$$

Now let $m \to \infty$ we have

$$\lim_n \inf_{x \in \Gamma} F_n(x) \geq F(x^*) \geq \inf_{x \in \Gamma} F(x).$$

The reverse inequality is trivial. ■

Lemma 8 *Suppose $F : X \to \mathbb{R}$ is a non-negative, lower-semicontinuous function. Then there exists a sequence of non-negative, bounded continuous functions $\{F_n : X \to \mathbb{R}\}$, non-decreasing in n, such that*

$$F(x) = \lim_n \uparrow F_n(x).$$

Proof. Define for $n = 1, 2, \ldots,$

$$F_n(x) \doteq n \wedge \inf_{y \in X} \left[F(y) + n\rho(x, y) \right].$$

Clearly F_n is non-negative, non-decreasing in n, bounded, and $F_n(x) \leq F(x)$. Furthermore,

$$\left| \inf_{y \in X} \left[F(y) + n\rho(x, y) \right] - \inf_{y \in X} \left[F(y) + n\rho(z, y) \right] \right| \leq n\rho(x, z).$$

Thus F_n is continuous. It remains to show that $F_n(x) \to F(x)$ for an arbitrarily fixed $x \in X$. Suppose this is not true. Then for all $n \geq F(x)$, we have

$$F_n(x) = \inf_{y \in X} \left[F(y) + n\rho(x, y) \right] < F(x) - \varepsilon$$

for some $\varepsilon > 0$. This implies the existence of a sequence $\{x_n \in X\}$ such that

$$F(x_n) + n\rho(x, x_n) \leq F(x) - \varepsilon.$$

But since F is non-negative, we have $\rho(x, x_n) \to 0$, and thus $x_n \to x$, and

$$\liminf_n F(x_n) \leq F(x) - \varepsilon < F(x).$$

A contradiction. ■

References

[1] D. Bertsekas and S. Shreve. *Stochastic Optimal Control: The Discrete Time Case.* Academic Press, San Diego, California, 1978.

[2] D. Blackwell. Discounted dynamic programming. *Ann. Math. Statist*, 36:226–235., 1965.

[3] A. Granas and J. Dugundji. *Fixed Point Theory.* Springer-Verlag, NY, 2003.

[4] O. Hernández-Lerma and J. B. Lasserre. *Discrete-time Markov Control Processes: Basic Optimality Criteria.* Springer, NY, 1996.

[5] R.E. Strauch. Negative dynamic programming. *Ann. Math. Statist.*, 37:871–890, 1966.